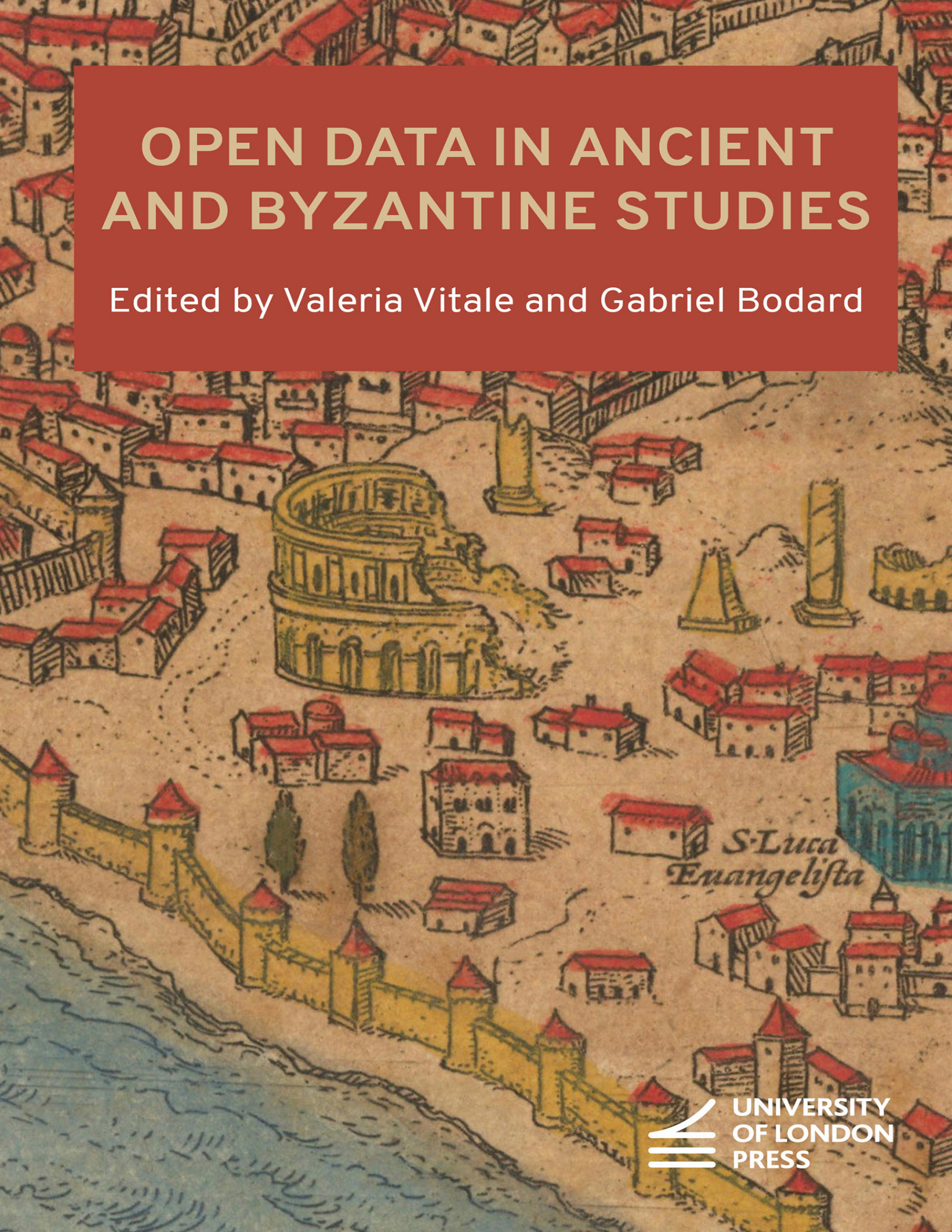


A detailed medieval manuscript map of Rome, showing the city's layout with numerous buildings, streets, and landmarks. The Colosseum is a central feature, surrounded by other structures and a fortified wall along the bottom. The map is drawn in a traditional style with red roofs and yellow walls.

OPEN DATA IN ANCIENT AND BYZANTINE STUDIES

Edited by Valeria Vitale and Gabriel Bodard

Open Data in Ancient and Byzantine Studies

Edited by
Valeria Vitale and Gabriel Bodard



Available to purchase in print or download for free at <https://uolpress.co.uk/>



First published 2026 by
University of London Press
Senate House, Malet St, London WC1E 7HU

© the Authors 2026

The right of the authors to be identified as authors of this Work has been asserted by them in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988. License

This book is published under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)) license.



Any third-party material reproduced in the book is not covered by the book's Creative Commons licence, unless indicated otherwise in the image or text's credit line. To reuse third-party material not published under the same licence as the book you will need to obtain permission from the copyright holder.

A CIP catalogue record for this book is available from The British Library.

ISBN 978-1-912250-82-0 (hardback)

ISBN 978-1-912250-83-7 (paperback)

ISBN 978-1-912250-84-4 (.epub)

ISBN 978-1-912250-85-1 (.pdf)

ISBN 978-1-912250-86-8 (.html)

DOI <https://doi.org/10.63674/shcx2882>

Cover image: 1572 Byzantium Nunc Constantinopolis – Georg Braun (1541-1622) and Frans Hogenberg (1535-90)

Cover design for University of London Press by Nicky Borowiec.

Book design by Nigel French.

Text set by Westchester Publishing Services UK

in Meta Serif and Meta, designed by Erik Spiekermann.

The editors and all authors would like to dedicate the work in this volume with gratitude, admiration and infinite affection to our mentor, teacher and friend, Professor Charlotte M. Roueché. Charlotte's work has for many years exemplified all of the virtues that this book explores: transparency, openness, rigour, collaborativeness, inquisitiveness, technological and methodological exploration, creativity and generosity. Her impact on the careers and lives of generations of students and scholars in classics, Byzantine studies and digital humanities has been, and continues to be, enlightening, transformative and revolutionary.

Table of contents

[List of figures](#)

[List of tables](#)

[Notes on contributors](#)

[Acknowledgements](#)

[Introduction](#)

[Valeria Vitale](#) and [Gabriel Bodard](#)

[Preamble](#)

[Digital scholarship](#)

[Open scholarship and open data](#)

[Weaving approaches and legacies](#)

[Final observations](#)

[Notes](#)

[References](#)

1. [Historical authority and the *Prosopography of the Byzantine World*: from factoids to assertions](#)

[Tara L. Andrews](#)

[The problem of medieval history as data](#)

[The factoid model as a solution](#)

[Linked Open Data approaches](#)

[The gaps that remain](#)

[Data as assertions and the STAR model](#)

[From factoids to assertions: the case of PBW](#)

[Says who? Authority for factoid statements](#)

[What makes you say that? Evidence basis and documentation source](#)

Who records the recorders? Authority and source for the assertion database itself

Assertions as a basis for historical reconstruction

Acknowledgements

Appendix

Notes

References

2. Connecting Late Antiquities: challenges and opportunities for prosopographical data

Charlotte Tupman and Richard Flower

Introduction

Background and current prosopographical initiatives

Undertaking the digital PLRE and CLA project

Creating an updated digital PLRE

Developing the CLA resource

Building and involving communities

The future of Connecting Late Antiquities

Sustainability and longer-term plans

Keeping other resources findable

Future work through a further funded project

Acknowledgements

Notes

Bibliography

Online resources

References

3. Venetian officials in sixteenth-century Cyprus: an online database and a pilgrimage account

Tassos Papacostas

Introduction

A register of Venetian officials

[The lure of the Holy Land and of holy relics](#)

[Across Cyprus in the summer of 1516](#)

[Dying far from home and the temptation of overseas posts](#)

[An appraisal](#)

[Acknowledgements](#)

[Notes](#)

[Bibliography](#)

[Online resources](#)

[Abbreviations](#)

[Archival sources](#)

[Primary sources](#)

[References](#)

[4. Connections through collections: linking archaeological and epigraphic archives through places and people](#)

[*Alessandra Giovenco and Nurdan Atalan Çayirezmez*](#)

[Introduction](#)

[BSR: 'Inscriptions of Roman Tripolitania' Photographic Collection project](#)

[BIAA: Connecting Archives, Connecting People \(CACP\) Project](#)

[Discussion and conclusion](#)

[Acknowledgements](#)

[Notes](#)

[Bibliography](#)

[Online resources](#)

[References](#)

[5. Engraving the soil: the challenges of defining spatial footprints in Cypriot cultural heritage](#)

[*Stuart Dunn*](#)

[Introduction](#)

[Digital gazetteers](#)

Defining place

H. H. Kitchener, and the first ‘datafication’ of Cyprus’s landscape

Challenging spatial normativity in Cyprus

Toponyms beyond text

The emerging map

Conclusion

Acknowledgements

Notes

References

6. You can’t step into the same dataset twice: iterative improvement through reuse of open Libyan place data

Valeria Vitale

Introduction

The case study: places in Libyan epigraphy

The Heritage Gazetteer of Libya

The digital catalogue of the BILNAS archive

Linked Open GeoData

Mapping Roman Libya with students

Back to the data

Closing the circle or, better, riding the spiral

Conclusions

Acknowledgements

Notes

References

7. The makeup of *Ajax*: reconstructing networks of interpretation through the commentary tradition

Charles Pletcher and Matteo Romanello

Introduction

Assembling the Multi-Commentary

[An anti-haphazard approach to the *Trugrede* \(Sophocles' *Ajax*, 646–92\).](#)

[Navigating the hazards of tralatitiousness](#)

[Conclusion](#)

[Acknowledgements](#)

[Notes](#)

[References](#)

8. [Traces of lost libraries in the works of Alexandrian and Byzantine scholars: bibliographic open data from antiquity to Wikimedia](#)
[Monica Berti](#)

[Introduction](#)

[Alexandria](#)

[Byzantium](#)

[Wikidata and the Wikimedia universe](#)

[Conclusions](#)

[Acknowledgements](#)

[Notes](#)

[References](#)

9. [Digital epigraphy at Dura-Europos: towards a FAIR-er inscriptional corpus](#)
[Anne Hunnell Chen](#), [Carl R. Rice](#), [Ilaria Bucci](#) and [J. A. Baird](#)

[Introduction](#)

[Dura-Europos: The site and its history](#)

[The legacies of early archaeology](#)

[Linked Open Data in practice at the Palmyrene Gate](#)

[Conclusion](#)

[Acknowledgements](#)

[Notes](#)

[References](#)

10. Reception of Sapphic language in Hadrianic verse epigraphy:

building on open philological data

Marja Vierros and *Gabriel Bodard*

Introduction

The Colossi of Memnon

Digitising Balbilla and open data

Research questions

Observations

Verbs

Personal pronouns

Grammatical case

Sentence complexity (visuality of trees and subordination)

Conclusions

Balbilla's imitation of Sappho

Methodological assessment

Future work

Acknowledgements

Notes

References

11. The next decade of Byzantine sigillography: the impact of open data as collaboration

Martina Filosa, *Jonathan Shea* and *Alessio Sopracasa*

Introduction

Edition of the sigillographic legend

Diplomatic transcription

Interpretive edition

Indexing and searching named entities

Towards an atomised search form

Cases requiring special attention

Personal names

Unique identifiers for seals and matrices

A sustainable publishing infrastructure

Teaching and outreach

Conclusions

Acknowledgements

Notes

References

Index

Figures

Figure 1.1 A typical Linked Open Data statement

Figure 1.2 A STAR assertion of the statement in **Figure 1.1**

***Figure 2.1** CLA People Search page (available from <https://connectinglateantiquities.org/texts.html>)

Figure 3.1 Notes in Latin recording the names of sixteenth-century Venetian officials, Jer. Saba 234, f.2v (Library of Congress Collection of Manuscripts in the Greek Orthodox Patriarchate of Jerusalem)

Figure 3.2 Map of Cyprus (Tassos Papacostas)

Figure 4.1 One of the boxes housing the IRT Photographic Collection in the BSR Research Collections

***Figure 4.2** Example of online prosopographical data: a cool URI is created for each person's data (<https://digitalrepository.biaa.ac.uk/biaaperson00085>)

***Figure 5.1** Ammochostos/Αμμόχωστος/Mağusa or 'Famagusta' in the Heritage Gazetteer of Cyprus: <https://cyprusgazetteer.org/hu/15>

***Figure 6.1** Central Church, Apollonia in the Heritage Gazetteer of Libya: https://slsgazetteer.org/irt_geo/recordview/?id=921

Figure 7.1 A screenshot of the *Ajax* Multi-Commentary, showing navigation on the left, the critical text in the centre and example glossae on the right. The critical text is highlighted on lines and lemmata, with darker shades corresponding to a greater density of glossae

Figure 7.2 Bar chart showing the number of glossae per lemma in the second episode of *Ajax*

***Figure 8.1** *Linked Ancient Greek and Latin* (LAGL) project: The Catalog of Authors and Works

***Figure 8.2** LAGL Catalog: References to the *Atthides* of Istros in the *Lexicon* of Harpocration

***Figure 8.3** LAGL Catalog: Experimental map with Wikidata location properties

Figure 9.1 Schematic mapping of the ecosystem of digital resources related to Dura-Europos and IDEA's use of the Wikidata/Wikimedia Commons platforms as a bridge among monolingual resources

Figure 9.2 Archival photograph (YUAG dura-b109~01; Q124841766) annotated by members of the IDEA team to clarify location of inscriptions on the walls of the city gate

***Figure 10.1** The trees of Balbilla 1, sentence 3 by Bodard et al. (left: <https://www.perseids.org/tools/arethusa/app/#/perseids?chunk=3&doc=61737>) and Vierros (right: <https://papygreek.com/text/466666/arethusa>)

***Figure 11.1** DOSeals 4, entry 11.19 (left) and Dumbarton Oaks Online Catalogue, DO BZS.1947.2.195 (right). Available at: <https://www.doaks.org/resources/seals/byzantine-seals/BZS.1947.2.195/view>

*Interactive versions of these figures are included in the open access edition of the book on the University of London Press instance of the Manifold platform: <https://read.uolpress.co.uk/projects/open-data-in-ancient-and-byzantine-studies>.

Tables

Table 1.1 CIDOC-CRM-based paths for conversion of PBW factoids to Linked Open Data assertions

Table 9.1 Example Statement 1

Table 9.2 Example Statement 2

Table 9.3 Example Statement 3

Table 10.1 Balbilla 1, sentence 3 with lemmata and morphological annotations

Table 11.1 Overview of the GitHub repository structure behind the SigiDoc portal (from Sopracasa et al. 2024: § 44)

Notes on contributors

Tara L. Andrews is university professor of digital humanities at the University of Vienna, specialising in the history of the medieval Christian Near East, digital philology and Armenian historiography. Most recently she led the ERC project RELEVEN, focusing on new models for digital representation of historical information about the eleventh century.

Nurdan Atalan Çayirezmez is a Near Eastern archaeologist with a PhD in information and records management. She specialises in digital cultural heritage, research data management, digital repositories and open-access scholarship. Since 2018, she has served as the British Institute at Ankara's digital repository manager and leads digital humanities projects aimed at improving access to and preservation of cultural heritage resources.

J. A. Baird is professor of archaeology and director of the Carena Institute of Sustainable Archaeologies at Birkbeck College, University of London. Her current work focuses on archaeological archives as sites of knowledge, engagement and repair.

Monica Berti is a classicist and digital humanist at Leipzig University. She researches Greco-Roman antiquity with computational text analysis and teaches ancient history, digital classics and digital philology. Founder of the teaching programme Sunoikisis Digital Classics, she leads international collaborations and digital-scholarship initiatives across global academic communities.

Gabriel Bodard is reader in digital classics at the Digital Humanities Research Hub and the Institute of Classical Studies at the University of London. He specialises in text encoding for epigraphy and papyrology, linked open data for prosopography and geography, imaging and modelling of cultural heritage, and ancient magic.

Ilaria Bucci is Leverhulme early career fellow at Durham University. An archaeologist of Southwest Asia, she specialises in graffiti, epigraphy and visual culture from Hellenism to late antiquity, and the study of archaeological archives and museum collections. She is the co-vice director of the Hatra Statuary Salvage Project (HaSSP).

Anne Hunnell Chen is assistant professor of art history and visual culture at Bard College. She specialises in Roman art and archaeology and in low-barrier Linked Open Data practices, focusing on equitable access and collaborative research. She directs the International Digital Dura-Europos Archive and is a co-chair of the Pelagios Network.

Stuart Dunn is professor of spatial humanities at King's College London. He has written extensively on the representation of place in the ancient world, ancient geography and the digital technologies that help to explore and understand it. With Charlotte Roueche and Tassos Papacostas, he was co-lead of the Heritage Gazetteer of Cyprus.

Martina Filosa is a researcher in Byzantine sigillography, prosopography and social history, with a particular interest in TEI-based encoding and the digital edition of text-bearing objects, including seals, coins and manuscripts.

Richard Flower is professor of classics and late antiquity at the University of Exeter. He co-directed the *Connecting Late Antiquities* digital prosopography project and has written or edited several books, including (with Marietta Horster, Frédéric Hurlet and Ralph W. Mathisen) *Brill's Companion to Roman Prosopography* (2025).

Alessandra Giovenco is the archivist at the British School at Rome, overseeing historical archives and records management. Holding an MA in medieval history and a PdG from the Archivio Segreto Vaticano, she has served as archivist since 2002. She specialises in conservation, collection description, digital curation and supporting research projects with colleagues and external partners.

Tassos Papacostas is reader in Byzantine and medieval culture at King's College London (Department of Classics/Centre for Hellenic Studies). His research focuses on the culture, art and architecture of Cyprus from late antiquity to the early modern period. With Charlotte Roueche and Stuart Dunn, he co-led the Heritage Gazetteer of Cyprus.

Charles Pletcher is a postdoctoral scholar at Tufts University. His current work sits at the intersection of ancient drama, reception studies and digital humanities, with a particular focus on how data are transmitted and manipulated as information and news.

Carl R. Rice is a visiting assistant professor of Greek and Roman studies at Vassar College. He completed his PhD in history and classics at Yale in 2023.

Matteo Romanello is senior data engineer at the Swiss Art Research Infrastructure in Zurich. Previously senior lecturer at the University of Lausanne with a project on the history of classical commentaries through computational approaches, he holds a PhD in digital humanities and works at the intersection between humanities and computer science.

Jonathan Shea is curator of coins and seals at Dumbarton Oaks. In this role he oversees the exhibition, study and publication of the collection of 17,000 Byzantine lead seals and 19,000 coins, including the online catalogue, the biennial Byzantine Coins and Seals Summer School, and the K-12 education programme.

Alessio Sopracasa is associate professor of Byzantine history at Sorbonne Université and deputy director of the 'Centre d'histoire et civilisation de Byzance' (Paris). His research is currently focused on digital Byzantine sigillography and on the edition of Venetian archival material related to Byzantium and the Eastern Mediterranean.

Charlotte Tupman (senior lecturer in classics and digital humanities, University of Exeter) is an epigraphist and digital humanities scholar whose interests include Roman funerary commemoration, standards for the digital

publication of ancient texts and machine learning approaches to the study of inscriptions and the processes involved in their creation.

Marja Vierros is a professor of classical philology at the University of Helsinki. She is a specialist in ancient Greek linguistics and papyrology. She led the project Digital Grammar of Greek Documentary Papyri (2018–23) and has worked as a researcher in several projects.

Valeria Vitale is lecturer in digital humanities at the University of Sheffield. She studies the representation and reception of historical places through a range of digital approaches including gazetteers, Linked Open Data, semantic annotation and 3D visualisation. She is currently focusing on digitised archives documenting archaeological excavations in Libya.

Acknowledgements

The editors would like to thank all the colleagues without whom this volume would not have been possible:

The University of London Press staff, and especially our expert commissioning editor Emma Gallon for her thoughtful support.

All the contributing authors for performing internal review with enthusiasm and kindness, providing feedback on and cross-referencing with one another's chapters.

All our generous external reviewers who helped improve the overall richness, clarity and completeness of this volume: Caroline Barron, Kasper Beelen, John Bodel, Chiara Cenati, James Chetwood, Maxime Guénette, Elizabeth Koch-Kölük, Francesco Mambrini, Sarah Middle, Elli Mylonas, Elisa Nury, Jun Ogawa, Christopher Ohge, Maria Parani, Mar A. Rodda, Rosario Rovira Guardiola and Polina Yordanova.

Introduction

Valeria Vitale and Gabriel Bodard

*‘But, Charlotte, aren’t you afraid that someone will **use** your data?’*

—anon.

Preamble

The chapters in this volume address different aspects of work that involves open data relating to the ancient and Byzantine worlds, from scholarship using accessible open access resources, to the development of methods, schemas and standards for the production and sharing of structured data with other researchers. We showcase this fundamental, infrastructural element of the digital turn in the humanities because it is undeniably revolutionising the study of the historical world in both quantitative and qualitative ways. The range of work and scope of methodological focus brought to the table by our contributors allows us to consider a broad spectrum of research influences and impacts of the use and sharing of open access, source, licences, standards and scholarship.

In this introduction we will begin with a brief discussion of digital scholarship in classics and Byzantine studies, tracing an unbroken thread from traditional academic and disciplinary standards to the ideals and impacts of open data. We then rehearse some of the key ways in which open scholarship and open data have been defined and prescribed in scholarly and cultural heritage discourse. Finally, we draw the web of links between approaches, methods and projects that appear within and among the chapters, more clearly to sketch the coherent themes across the volume as a whole.

We, the editors and authors of this collection of essays on open data, are not uncritical advocates of unlimited openness in all cases and contexts. There may be many reasons, including privacy, financial, legal and ethical issues, for some data to be embargoed, paywalled, protected by copyright or other intellectual property regulations, or otherwise have access to them limited. While making the data and tools behind digital research projects accessible for review, testing and further research is as much a scholarly imperative in ancient studies as it is in medicine or astrophysics, and indeed as using explicit bibliographies, primary sources and apparatus criticus are to traditional philology, there is nuance and discretion in all of these areas.

Digital scholarship

It is perhaps unsurprising that classics, together with neighbouring disciplines including ancient and medieval history, has been at the forefront of the adoption of academic digital methods since the early days of what is now called digital humanities. It is similarly to be expected, given the nature and history of the discipline, that this subject area was and continues to be influenced and innovated by digital tools and research methods ([Brunner 1993](#); [Crane 2004](#); [Terras 2010](#)). Classics has long depended on a common body of academic protocols and conventions. Epigraphic transcription standards, textual criticism and apparatus, and archaeological methods are all highly formalised, and now also documented digitally, and have in some cases spearheaded encoding standardisation across the humanities. Norms such as architectural vocabularies, classifications applied to artworks, or corpus linguistics methods, in contrast, are less explicitly formalised, and so benefit from the centring of interoperability, transparency and documentation in digital scholarship.

The arbitrary and arcane conventions for the philological transcription of inscriptions and papyri are neatly analogous to the equally oblique and technical codes of computer languages. It is but a step to appreciate the elective affinity between philology and digital encoding. Epigraphy is an especially appropriate example of such a field of study even within classics, as it both depends on established conventions and studies texts that themselves contain standardised codes, abbreviations, symbols and other

shorthand to deliver their messages more concisely and efficiently. The highly structured and largely adopted Leiden System of papyrological and epigraphic transcription features, therefore, is part of a conceptual continuity that includes the EpiDoc Guidelines, a specialisation of TEI XML for encoding ancient epigraphic and documentary editions ([Bodard and Stoyanova 2016](#): 54–5). The same thing could be said about the labyrinthine praxis of finding and citing primary sources or critical editions, in both traditional scholarship and digital platforms ([Smith 2009](#); [Romanello et al. 2009](#): 81–2).

It would be reductive to say that digital classics and Byzantine studies simply translated into a digital form of principles and practices that entirely already existed in long-standing scholarly practice. The ‘digital turn’ has also ignited a more profound transformation and opened new, exciting doors. Digital prosopographies were not simply a digital, online reproduction of printed ones: they enabled different methodological approaches to primary sources. Digital gazetteers were not simply digitised versions of the *Barrington Atlas* ([Talbert et al. 2000](#)), but became the de facto backbone of historical Linked Open Data (LOD) networks.

The focus of this volume is a particular aspect of this process that we think was the most radical in the evolution and transformation of the discipline: the openness of data, tools and standards.

If classics was one of the disciplines in the humanities best suited to dialogue with computational frameworks, it was also, by tradition, one of the most hostile to this kind of openness. Possibly due to its dealing with texts and objects that belong to a very distant past, researchers in classics have long held the ambition of creating complete and definitive publications that would become (to different extents) cornerstones for future scholars: the ‘seminal study’ that everyone interested in the topic will henceforth have to read and cite. This idealisation of the ‘definitive edition’ has sometimes encouraged classical scholars to sit on their findings for years, rather than publishing them, waiting for their corpus to be as complete and flawless as possible. These instincts have impeded collaboration and even sowed diffidence towards colleagues working on cognate topics.

In this context, the creation of open datasets, and all the related research practices driven by sharing and collaboration, reveal all their subversive

potential within the discipline. It took the genius and revolutionary ethos of some pioneers to start experimenting with datasets connected to the ancient and late antique world that were not only open, so that everyone could access and even reuse them, but also collaborative, in-progress and perpetually incomplete. This change in the idea of what constitutes a publication, and what constitutes a good publication, has probably been the most groundbreaking change that open data has brought to the discipline ([Crane et al. 2009](#); [Bodard and Garcés 2009](#)). The goal of (digital) classicists and Byzantinists is shifting from creating the most complete, accurate and immutable study or critical edition, to building the foundations for common resources that can be enriched, and even corrected, by many other researchers, contemporary and future.

Open datasets and standards are not only redefining centuries-old ideas and practices around publication and transmission of knowledge in these disciplines, but are also impacting the kind of questions that we can ask about the ancient world, and the ways in which we seek and communicate our results.

The most obvious consequence of creating and publishing open datasets is that more data about the historical world is readily available to researchers. This benefits the entire community, but perhaps especially scholars who do not have the means (in terms of time, money or institutional support) to visit the large and prestigious libraries. Publishing datasets online and making them available also means that they can be used as comparanda or the basis for further studies. Such opportunities broaden the scope of the discipline, and suggest cross-corpora comparisons that would have been much harder to achieve using traditional approaches. While the amount of text and tabular data in open datasets for classics and Byzantine studies will likely never amount to what is generally considered ‘big data’, once the information about the past has been modelled and structured as data, it also becomes accessible for different kinds of computational analysis, generating novel ways to interrogate centuries-old corpora.

But it is not simply the sheer quantity of available data that enables comparisons between projects; to dialogue with one another, datasets need to be compatible, built following interoperable scholarly, technical and computational standards. The importance of agreed practice and shared

vocabularies, ontologies and languages for data-rich projects cannot be overstated, as they play a fundamental role in enabling cross-dataset comparisons, facilitating computational approaches to analysis and visualisation, and supporting long-term sustainability and preservation. Creating, maintaining and updating such standards themselves has become one of the essential tasks of digital humanists, an invisible but necessary work on which much digital scholarship is built, comparable to the development of the Leiden System ([Wilcken 1932](#)).

Highly codified academic standards such as the Leiden System, canonical citations of authors and texts, and even the humble apparatus criticus are sometimes taken for granted as stable, reliable and appropriate tools of classical scholarship. All of these conventions have taken expertise and work, albeit undervalued, to establish and maintain. At the same time, these very standards are periodically re-examined, challenged, sometimes even wilfully ignored by scholars who imagine that incremental improvements outweigh the benefits of consistency and compatibility. The work of digital classicists and Byzantinists on the technical formats, formal standards and infrastructures, on the other hand, involves and is recognised for the creation, study or communication of historical data through computational methods, and the knowledge infrastructure to express, share and preserve such information. In other words, the design and maintenance of data models, schemas and ontologies, such as those behind gazetteers or prosopographical databases, are complex and valuable research activities in their own right, not supporting or ancillary disciplines as their analogues in traditional historical scholarship might have been.

Open scholarship and open data

There are several axes of openness along which scholarship, and especially digital scholarship, can be measured, and while not all are obviously analogous to one another and none are simple spectra of open/good to closed/bad, it is worth exploring the terminology and definitions of these values a little further. In this section we shall outline the parameters of open access, open standards, open licences, open data and open scholarship, and recap further fundamental concepts such as the five stars of Linked Open

Data, the FAIR principles (Findability, Accessibility, Interoperability and Reusability), and the CARE principles (Collective benefit, Authority, Responsibility and Ethics).

Open access (the most commonly understood, and increasingly expected, form of openness) is the freedom to read or otherwise consult a work unimpeded by cost, membership, technology, location or other arbitrary barriers. An open access work should not be paywalled, in a password-protected area (even if registration for access is free of financial charge) or in a format only accessible via proprietary software ([Swan 2012: 15–16](#)). There is an expectation that open access work should remain available, but otherwise no reason that the full protections of copyright should not apply (although some may of course be waived).

Open licensing is a recognised or legal basis on which to waive some or all of the normal protections of copyright; the best known are schemes such as Creative Commons for creative content or data, or licences such as GNU or Apache for open source software.¹ Licences may preserve the intellectual property features of copyright—the right to be identified as the author of a work is most commonly retained—but other rights may be waived and permissions given, so others may be granted rights including creating derivative works, remixes or new visualisations of data, redistribution or republication of new, combined (now effectively co-authored) contents. Most importantly, the open licence means that users of the material do not need to ask for permission to reuse content in this way.

Open standards, as discussed above, are the least alien feature to a classicist or Byzantinist working in these highly formalised and canonical disciplines. In addition to following well understood and shared subject matters, languages and vocabularies, digital standards allow us to structure data according to well documented criteria, and based on formats that do not depend on proprietary software solutions—usually readable as plain text. Standards or sets of guidelines such as TEI or EpiDoc use XML encoding; LOD is based on RDF and shared ontologies;² text may be structured as Markdown (plain text with simple formatting marks) or other easily interpreted languages; data may be represented as comma-separated values (CSV, a plain text representation of data) or other tabular formats. As well as technically and legally permitting collaborative use of data, shared datasets are likely to be encoded and published using common or

compatible formats and standards, enabling interoperability and the combination of data from multiple sources into new projects.

As related to classical antiquity and its medieval aftermath, **open data** can cover almost unlimited materials: from annotated texts or linguistic analyses to digitised archaeological or architectural objects; from geography, prosopography, bibliography or palaeography to scholarly or archival metadata, digital images, 3D models or modern creative receptions of ancient and Byzantine heritage. The surfacing of underlying digital materials, evidence, data and the code or methods that process it into information and conclusions, is a factor of its open accessibility, legal permissions and licences, and adherence to open formats and standards.

That such underlying data is as important to the reliability of **open scholarship** is only new inasmuch as we recognise our intellectual output as relying on hard evidence, actionable data and reproducible methods. The sciences—including of course archaeology—have long recognised the importance of sharing data and exposing methodologies to allow other scholars and reviewers to replicate our experiments and test our conclusions. The arts and humanities practise the same principles with the insistence on references, footnotes, bibliography and appeals to reason and criticism over dogma and authority.

In 2010, Berners-Lee enhanced his definition of linked data with the ‘five stars’ of LOD ([Berners-Lee 2010](#)). The first star simply required that the data be on the web, in any format: in the terms discussed above, this means (partly) open access and open licensed data, usable by others. The second star was earned by data being in a structured format, such as a database table rather than a screenshot, or OCRred text rather than page images. The third star added to this the recommendation to share data in non-proprietary formats—such as CSV rather than a commercial database or spreadsheet application—and thus completes the alignment with open access as defined above. The (cumulative) fourth star originally called for the use of the RDF open standard, although is often now expressed more broadly as ‘use URIs [web-addressable unique identifiers] to identify your data’; either way added open standards to the non-proprietary formats requirement. The final star, qualifying your data as full five star LOD, is granted to data that includes actionable links to other LOD datasets, and therefore forms part of the Semantic Web.

Public access to and usability of (especially publicly funded) research data has been an important concern for many years, and some of the key principles were outlined in an Organisation for Economic Co-operation and Development report ([OECD 2007](#)). A more recent formulation has gained traction for the term the 'FAIR principles' ([Wilkinson et al. 2016](#)), which groups the key concepts of open data under four headings. Findable data and metadata should be assigned referenceable identifiers, described with rich metadata, and deposited or indexed by a searchable resource. Accessible data is both encoded and made available through widely recognised and open standards and protocols, and ensures that the structural metadata remains available so that it can be properly interpreted and actioned by both people and software. Interoperable data is made compatible with other datasets and tools by being described in formal and open systems of knowledge representation, the use of standard and shared vocabularies, and links to other online data. Reusable research materials require accurate metadata and paradata, open licensing, a good account of data provenance, methodology and attribution, and adherence to academic and community standards; fuller project documentation is perhaps the greatest step towards making data reusable. All of these criteria depend on the quality of data citability, discoverability and documentation, but also require work to move from hypothetical findability to true interconnectedness with the wider ecosystem of collaborative research materials and Linked Open Data.

Where data relating to both tangible and intangible cultural heritage is concerned, particularly ancient material heritage whose creators are long past, this emphasis on openness and sharing runs the risk of being taken as a moral imperative. It is especially problematic when global north researchers insist that heritage materials collected (and profited from) over decades and centuries of colonial context must now be seen as the shared heritage of all humanity, and digital copies released into the public domain ([Pavis and Wallace 2019](#)). In the context especially of indigenous heritage and data sovereignty, the CARE principles were promulgated to provide further nuance to formulations that treat heritage only as scientific data that must be open in every regard (Carroll et al. 2020). Collective benefit is the idea that indigenous heritage materials and data should be treated and exploited for inclusive ends, including and especially those of the peoples whose

heritage it includes; that community engagement be embedded in data treatment and management; that equitable outcomes should be sought. Authority to control also recognises the rights and interests of communities in their own heritage data, the importance of data in supporting indigenous self-determination, including the governance of data itself. Responsibility is the need for respectful relationships with heritage communities, including in supporting capacity building, and in grounding the handling of data around local worldviews and languages. Ethics ties all this together with a concern for justice, especially around the assessment of our use of data for benefit (or harm) and potential future use.

Ancient and medieval historians or archaeologists may be tempted to imagine that such ethical niceties and cultural or curatorial sensitivities barely apply to their materials and data. That would be naïve if not wilfully shortsighted. Not only are many of the material objects of classical study held in global north museums, often from the colonial period, but significant parts of the former Roman and Byzantine spheres of influence are now part of the developing world, whose languages and populations are underrepresented in classical and Byzantine scholarship, and underserved by exhibition and publication practices. Responsible ancient historians already pay attention to dangers such as antiquities trafficking, lack of provenance, archaeology that emerges from warzones, and other connections between our objects of study and contemporary abuses, injustices and geopolitical crimes ([Nongbri 2022](#)). When thinking about open data and open scholarship, therefore, we should also be mindful of our colonial history, of the use of local languages and community engagement in the areas of our study, and of the good or harm to which our use of data is capable of contributing.

Open access, open licensing, open standards and open data are all key to academic transparency, freedom, integrity and independence in the digital humanities, as in all open scholarship since the Enlightenment. At the same time, open access is not an absolute; we recognise that there are plenty of good reasons to respect intellectual property protections, including when we work with modern translations, artistic representations, reception of antiquity, web archives and other large datasets, and when the protections afforded to minoritised or marginalised creators are what allow them to contribute to the study of ancient heritage at all. The several axes of data

openness discussed here are all means to the end of enabling responsible and transparent research; they are not an end in themselves.

Weaving approaches and legacies

The chapters in this volume bring a variety of methodological expertise, data types and projects to explore the impact of open data in classical and Byzantine studies. What they share is a broadly common discipline and overlapping set of subdisciplines with well-established canons and standards, and particular suitability to digital encoding and structured data. We have talked above about the long-standing use of standards and vocabularies in epigraphy; the other data-rich subdisciplines represented in this volume, including prosopography, geography, palaeography, linguistics and text editing (and intersections between them), have similarly been early adopters of digital humanities and data-driven approaches.

Individual chapters feature a range of perspectives, case studies and research questions, focusing on open data as a foundation of historical scholarship. They both exemplify and argue theoretically for approaches including: scholarly work based on shared open data; open licensing to enable transformative reuse; application of open standards and authorities; creation, enhancement and (re)distribution of new open data; publication of code alongside data and written argument; and curation, promotion and sustainability of the classical open data ecosystem. Each chapter discusses concrete instances of open data acting as methodology in its own right, and of the code with which data is collected, processed, interpreted and enriched serving as scholarly outputs of published research in classics and Byzantine studies.

Individual chapters range across the creation of data and data models, design and documentation of standards or methods, transformation of open licensed and standardised data, building on pre-existing materials, and the scholarly access to and citation of openly published data and code—often in rich combination. The volume as a whole therefore argues for and depends upon the necessity of open access delivery, mixed output formats, citation of datasets and crucially of code, and the ability to embed actionable code and data within the argument. In this respect, the volume has a

methodological affinity with its content, as it is only possible because of the affordances of its digital and interconnected format, and because it enables direct links to digital and computational components that, while integral parts of the research process, have traditionally been extremely hard to publish and cite. The other aspect that makes this publication consistent with its topic is, of course, that of being open. The editors and authors of these chapters are no longer afraid that ‘someone will *use* their data’, but actively encourage any reader and fellow researcher to do so: to apply the same standards to other corpora, to reuse and improve data models, to fork and customise code, to broaden comparisons, to comply with ethical guidelines and, in doing so, to join the most exciting transformation of the discipline.

In [Chapter 1](#), ‘Historical authority and the *Prosopography of the Byzantine World*’, Tara Andrews introduces structured statements in LOD format not only as the most functional representation of our fragmentary knowledge of Byzantine past lives, but perhaps also the most appropriate methodologically. The open data model for a Byzantine prosopography proposed in the chapter, STAR, enables one not only to link and disambiguate person-references, but also to distinguish transparently between the information held in historical sources and their interpretation, assigning a traceable provenance to each assertion, encouraging scientific soundness as well as sharing and reuse.

The volume proceeds with another analysis of digital approaches to ancient prosopography and the potential for open datasets and standards to reshape the discipline. Charlotte Tupman and Richard Flower offer valuable reflections about their *Connecting Late Antiquities* project in [Chapter 2](#), discussing in detail workflows and outcomes and providing the reader with replicable models.

The use of open digital resources to generate new research questions is engagingly explored by Tassos Papacostas in [Chapter 3](#), ‘Venetian officials in sixteenth-century Cyprus’, where he uses the *Rulers of Venice* database to discover connections among digitised historical documents and then follows these threads to reconstruct personal and public events and formulate new interpretative hypotheses.

[Chapter 4](#), ‘Connections through collections’, by Alessandra Giovenco and Nurdan Atalan Çayırezmez, brings to the conversation the perspective

of the cultural heritage sector, and discusses applications of LOD-oriented approaches in two archaeological archives—at the British School at Rome and the British Institute in Ankara respectively. The case studies analysed emphasise the crucial role that digital identifiers (URIs) for people and places are increasingly playing in libraries, museums and archives, not only enabling a more efficient cataloguing process that reduces redundancies and ambiguity, but also becoming a powerful tool that can drive exploration and discover connections outside the single collection or even institution.

After discussing in detail open standards, datasets and workflows in ancient prosopographies, the book moves towards digital historical gazetteers with [Chapter 5](#), ‘Engraving the soil’, by Stuart Dunn. He presents the sophisticated spatial knowledge modelling that informs the Heritage Gazetteer of Cyprus, discussing the rationale behind the open and collaborative online resource, and the cultural, political and ethical implications that such a project needs to address.

In [Chapter 6](#), ‘You can’t step into the same dataset twice’, Valeria Vitale retraces the evolution of a spatial dataset, from simple spreadsheet to an online open and reusable resource. The chapter discusses the changes and improvements that each new use, canonical or otherwise, has brought to the dataset, expanding not only the number of items recorded but also, or maybe mainly, its scope, methods and technologies, each new phase informing and shaping the next.

The concept of reuse is also pivotal in [Chapter 7](#), ‘The makeup of Ajax’, by Charles Pletcher and Matteo Romanello, albeit from a very different perspective. The chapter introduces a digital critical edition of Sophocles’ *Ajax* that, leveraging the availability of digitised copies of old textual criticism, offers new interpretative keys. Through the digital recovery of almost-forgotten nineteenth-century scholarship, the platform developed by the authors shows how the identification of most commented passages and the critical analysis of the evolving scholarship around them allows for a deeper understanding of the classical text, its tradition and its broad spectrum of possible interpretations. Besides proposing a compelling comparative analysis of one of the most iconic moments of the tragedy, this chapter also highlights nuanced issues around availability and copyright.

The volume expands on textual and philological approaches with Monica Berti’s contribution, [Chapter 8](#), ‘Traces of lost libraries in the works

of Alexandrian and Byzantine scholars'. It focuses on expressions used in historical documents to refer to the onomastics of ancient authors and the content of their works, including those who are less frequently cited and thus risk becoming, in time, almost completely invisible. Using the case study of two women poets, Cleobulina of Rhodes and Moero of Byzantium, Berti takes us on an exciting journey through ancient library records, leveraging automated digital methods to spot and recognise mentions of these two authors and their work. The research discussed here does not end with the identification and classification of these ancient records, but, truthful to the open data ethos, endeavours to contribute the data back to Wikidata, with the aim of giving those ancient authors a voice again.

In [Chapter 9](#), 'Digital epigraphy at Dura-Europos', Anne Hunnell Chen, Carl R. Rice, Ilaria Bucci and Jennifer A. Baird introduce the fascinating and eventful history of Dura-Europos as an ancient site first and archaeological one later. Through this reconstruction of fragmented, and heavily colonial, excavation and archival practices, the authors evidence the urgent need for a more conscious and ethical approach that not only accounts for present marginalised perspectives, but also paves the road for the ones that are bound to emerge in the future. The chapter proposes an original use of semantic web technologies and, in particular, explores the potential of the highly structured, multilingual and community-based Wikidata environment for a more diverse, and perhaps even subversive, reinterpretation of the archaeological documentation of the ancient site.

In [Chapter 10](#), 'Reception of Sapphic language in Hadrianic verse epigraphy', Marja Vierros and Gabriel Bodard experiment with treebanking on a very peculiar corpus: the verses inscribed by Julia Balbilla on the statue known as the Colossus of Memnon. The authors apply computational linguistics methods to compare these lines to those of other authors, and use open data and tools to highlight differences and similarities between Balbilla and her (alleged) literary models. The chapter challenges assumptions about the reception of Greek literary classics, but also creatively stretches quantitative approaches by testing them on very small and idiosyncratic text collections.

We conclude with [Chapter 11](#), 'The next decade of Byzantine sigillography', wherein Martina Filosa, Jonathan Shea and Alessio Sopracasa retrace how the availability of digital methods, tools and

standards adopted from sister disciplines such as epigraphy, papyrology and numismatics has enabled Byzantine sigillography to emerge and establish itself as an autonomous field. Looking forward, the authors argue for a new open and collaborative model not only for publishing sigillographic sources, but also to study and compare them, turning the discipline into a continuous interpretative dialogue.

While made up of independent chapters, this edited volume presents a strong and cohesive narrative, emphasised by copious intertextual connections. As is expected in this editorial format, all contributions offer their authors' perspectives on the topic of open data, standards and methods in research about the ancient world. In addition, the case studies discussed here are strongly related to each other, building on one another's outputs and improving each other's scholarship in a virtuous circle that enhances the richness, diversity and sustainability of the field and further supports the main thesis proposed by this volume. Among the several examples that we could cite, the volume offers a multilayered and multi-perspective analysis of how the data modelling of information about ancient people and places has shifted from relating authoritative (and opaque) interpretations by scholars to recording direct references to the sources. This attestation-based approach enhances the transparency of the research process, fronting the different provenance of the information and highlighting the potential for variant interpretations. The epistemological implications of this methodology can be seen in Andrews's work on the *Prosopography of the Byzantine World* analysed in [Chapter 1](#); but readers can also appreciate how this kind of approach has the potential to facilitate comparisons between different digital prosopographies, as explored by Tupman and Flower in their *Connecting Late Antiquities* project. Similarly, in [Chapter 5](#) Dunn explains how the data model underlying the Heritage Gazetteer of Cyprus only includes references to places and their names that can be directly attributed to a citable source, thus pursuing the double aim of increasing data transparency and navigating delicate political contexts. In all of the research processes described in these chapters, what may have started as a technological challenge has soon revealed its methodological potential: the need to expose the rationale behind the data and its provenance has blended into the need to 'make open' the scholarly process that inevitably interprets such information.

Another thread that runs through several chapters is the argument for the use of open data and technologies to improve the connectivity of information at all levels. LOD, in particular, is identified as a promising framework to reunite dispersed collections in Giovenco and Atalan Çayırezmez's case study in [Chapter 4](#), but also in [Chapter 9](#), by Chen, Rice, Bucci and Baird.

Supporting the discovery of new, non-obvious connections promises to enable new avenues for research. However, as the chapters by Papacostas, Pletcher and Romanello, and Berti evidence, easier access to a larger number of resources is only the first step into embedding open data and tools into humanities research. The wider comparison of corpora across time and space is one of the most recognisable opportunities that is enabled by the digitisation of primary sources and the publication in data-format of the secondary ones, and is supported by the use of common standards and models. However, it is only when we use these resources to ask new research questions that this increased access becomes also transformative of the discipline. As Pletcher and Romanello argue, for example, it is undoubtedly useful to have access to digitised versions of older commentaries of *Ajax*, but it is only when a new tool to compare and contrast them is built that these interpretations can be used to identify controversial passages in the text and to enrich our contemporary (and future) understanding and appreciation of the play's ambiguities.

Another key topic that echoes across the pages of the volume is our relationship with software. As Vierros and Bodard demonstrate by deconstructing lines of verses by Julia Balbilla, digital tools are not mere means to an end but can become integral parts of our thinking process. The digital tools that these authors have chosen are not convenient aids to the research, but they embody it, as also becomes apparent when Filosa, Shea and Sopracasa describe how the shape of a digital platform closely reflects the aims and objectives of Byzantine sigillography as a discipline. As several of our authors argue, it is exactly for this reason that digital tools need to be used critically and ethically, and, when necessary, they should be challenged and improved. The responsibility of researchers of the ancient world to advocate for an idea of 'openness' that does not become the inevitable replica of existing inequalities is acutely felt by Chen, Rice, Bucci and Baird as well as by Vitale.

Finally, all of these chapters are implicitly or explicitly connected through their intellectual debt to seminal open digital projects, tools and standards that came before them and laid the foundation on which we are building today. Far from being unoriginal derivatives, these projects take forward the scholarship of their predecessors, adding layers of complexity to the methodological debates and testing the robustness of the approach. The elaboration of the TEI guidelines, for example, was a prerequisite for the birth of the EpiDoc schema and guidelines, often mentioned in these chapters not only as a useful model to encode complex historical information but also as an example of an active and engaged community built around the use of an open digital tool. EpiDoc appears in this publication as a key reference for Vierros and Bodard, but also Tupman and Flower, Giovenco and Atalan Çayırezmez, Vitale, Filosa, Sopracasa and Shea, and others. The Pleiades gazetteer of the ancient world was a model and inspiration for several historical gazetteers employed by the digital ancient world community—including the Heritage Gazetteer of Cyprus discussed by Dunn that, in turn, acted as a model for the Heritage Gazetteer of Libya researched by Vitale—and played a fundamental role in the LOD-based Pelagios project. Similarly, the legacy of the latter, whose aim was to support a network of LOD connections among digitised resources based on common place references, is most directly apparent in the project to connect the archaeological archives discussed by Giovenco and Atalan Çayırezmez, but can also be clearly recognised in *Connecting Late Antiquities* despite the focus on person-data. Several chapters in the volume, Berti's and Chen, Rice, Bucci and Baird's most prominently, also bring to their research workflows one of the tools that over the years has become a flagship of the idea of openness: the Wikidata infrastructure.

We argue that the data literacy necessary for customisation of the available open tools is not simply a technological competence that would be useful to gain, but an empowering process that stimulates the researchers' creativity and problem solving, allowing for the use of existing and open/free digital resources that may have been born for other intended uses.

We envision reading this edited volume as navigating a dense web of thematic and methodological connections enabled by explicit and implicit internal cross references. These links allow the discovery of variant takes on the same issue, alternative applications of the same approaches, creative

reuses of the same datasets. In this respect, the authors represent not only themselves and their research, but the entire digital ancient world community, and a new idea of scholarly research as a collaborative endeavour.

Final observations

All of the contributors to this volume recognise the value of open data in at least some of its forms and definitions, but the purpose of this work is not to advocate for extreme open access in all conditions without restrictions, copyright abolition or a universal public domain for all research data. (Some may want to argue for some of these things in some contexts.) There may be diverse professional, commercial or legal reasons for limiting access to certain data, on top of which we should consider ethical concerns, the protection of heritage or people, and respect for indigenous heritage. The CARE principles ([Carroll et al. 2020](#)) may sometimes be in tension with FAIR and similar principles of openness, as discussed especially by Vitale in [Chapter 6](#) and Chen et al. in [Chapter 9](#).

There are other issues with open data and digital publication that require a nuanced approach to the celebration of open scholarship. An especially challenging problem is the preservation and sustainability of online open data, in particular the deployed or visualised versions of the data, which may depend on specific tools or platforms and therefore be less portable, harder to deposit in archives or repositories, and more vulnerable to obsolescence and security developments ([Ageh et al. 2024](#)). Many researchers struggle to find a stable home for their digital deliverables after the end of a period of funding or the support of a specific institution, and raw data may not be as accessible and visible to our colleagues in less-digital disciplines. Related to this last, there are still disciplines and localities where digital outputs are (or are perceived to be) of less value for research assessment or career advancement than traditional print monographs and peer reviewed journal articles; early career scholars in particular need to carefully consider the balance of benefit and cost ([Roueché 2009](#)).

Open scholarship and all the subcategories of openness are an essential part of research, in particular but not restricted to digital research. There are many issues still to address and discuss, and this work surely does not represent the end of the conversation. As you read the following chapters in this volume, please do so with a view to you joining and continuing the discussion.

Notes

1. Creative Commons licenses: <https://creativecommons.org/licenses/>; GNU and the General Public License (GPL): <https://www.gnu.org/licenses/licenses.html>; Apache licenses: <https://www.apache.org/licenses/>.
2. Text Encoding Initiative Guidelines: <https://tei-c.org/guidelines/>; EpiDoc Tools and Guidelines: <https://epidoc.stoa.org/>; Resource Description Framework (RDF): <https://www.w3.org/RDF/>.

References

- Ageh, T., Bayler, M., Durlacher, C. and Turner, J. (2024) *The Preservation of Knowledge in the Digital Age: An Analysis of Public and Research Libraries' Perpetual Access to Licensed Electronic Literature*. Arcadia Fund and Ageh Consulting. Available from: <https://arcadiahfund.org.uk/highlights/the-preservation-of-knowledge-in-the-digital-age>.
- Berners-Lee, T. (2010) 'Is your Linked Open Data 5 Star?' *World Wide Web Consortium: Design Issues*. Available from: <https://www.w3.org/DesignIssues/LinkedData.html#fivestar>.
- Bodard, G. and Garcés, J.L. (2009) 'Open Source Critical Editions: A Rationale', in M. Deegan and K. Sutherland (eds) *Text Editing, Print, and the Digital World*. Farnham: Ashgate Press, pp. 83–98. Available from: https://blog.stoa.org/files/2010/09/Bodard-Garces_2009_Open-source-digital-editions.pdf.
- Bodard, G. and Stoyanova, S. (2016) 'Epigraphers and Encoders: Strategies for Teaching and Learning Digital Epigraphy', in G. Bodard and M. Romanello (eds) *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange & Public Engagement*. London: Ubiquity Press, pp. 51–68. <https://doi.org/10.5334/bat.d>.
- Brunner, T.F. (1993) 'Classics and the Computer: The History of a Relationship', in J. Solomon (ed.) *Accessing Antiquity*, pp. 10–33.

Tucson: University of Arizona Press.

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020), ‘The CARE Principles for Indigenous Data Governance’, *Data Science Journal* 19(1), p. 43. <https://doi.org/10.5334/dsj-2020-043>.

Crane, G. (2004), ‘Classics and the Computer: An End of the History’, in S. Schreibman, R. Siemens and J. Unsworth (eds) *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 46–55. Available from: <https://companions.digitalhumanities.org/DH/>.

Crane, G., Seales, B. and Terras, M. (2009), ‘Cyberinfrastructure for Classical Philology’, *Digital Humanities Quarterly* 3.1. Available from: <https://dhq.digitalhumanities.org/vol/3/1/000023/000023.html>.

Nongbri, B. (2022), ‘The Ethics of Publication: Papyrology’, *Bryn Mawr Classical Review*, 2022.05.25. Available from: <https://bmcr.brynmawr.edu/2022/2022.05.25/>.

OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264034020-en-fr>.

Pavis, M. and Wallace, A. (2019), ‘Response to the 2018 Sarr-Savoy Report: Statement on Intellectual Property Rights and Open Access Relevant to the Digitization and Restitution of African Cultural Heritage and Associated Materials’, *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 10(2), pp. 115–29. <https://doi.org/10.2139/ssrn.3378200>.

Romanello, M., Boschetti, F. and Crane, G. (2009), ‘Citations in the Digital Library of Classics: Extracting Canonical References by Using Conditional Random Fields’, in *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*. Suntec City, Singapore: Association for Computational Linguistics, p. 80–87. Available from: <https://aclanthology.org/W09-3610.pdf>.

Roueché, C.M. (2009), ‘Digitizing Inscribed Texts’, in M. Deegan and K. Sutherland (eds) *Text Editing, Print and the Digital World*. Farnham: Ashgate Press, pp. 159–68.

Smith, N. (2009), ‘Citation in Classical Studies’, *Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure*.

- Digital Humanities Quarterly*, 3.1. Available from: <https://dhq.digitalhumanities.org/vol/3/1/000028/000028.html>.
- Swan, A. (2012) *Policy Guidelines for the Development and Promotion of Open Access*. Paris: UNESCO. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000215863>.
- Talbert, R.J.A. (2000) *Barrington Atlas of the Greek and Roman World*. Princeton, NJ: Princeton University Press.
- Terras, M. (2010) 'The Digital Classicist: Disciplinary Focus and Interdisciplinary Vision', in G. Bodard and S. Mahony (eds) *Digital Research in the Study of Classical Antiquity*. London: Routledge, pp. 171–89.
- Wilcken, U. (1932) 'Das Leydener Klammersystem', *Archiv für Papyrusforschung*, 10, 211–12.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-BeltranAlasdair, A., Gray, J.G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Mons, B., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., and Zhao, J. (2016). 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Chapter 1

Historical authority and the *Prosopography of the Byzantine World*: from factoids to assertions

Tara L. Andrews

The problem of medieval history as data

In his seminal article outlining origins and approaches to prosopography, Lawrence Stone opines that ‘For most social groups in most areas prosopography cannot usefully be employed before the explosion of record materials in the sixteenth century, caused by the invention of the printing press, the spread of literacy, and the growth of the bureaucratic, record-keeping nation state’ ([Stone 1971](#): 58). His point is that a prosopography based on such diverse and lacunose sources as we tend to have for the history of the Byzantine world, or the medieval eastern Christian world in general, will contain enough gaps to invalidate any statistical approach to historical phenomena. Even as Stone wrote his article, and as acknowledged within it, this limitation did not stop historians of classical Rome or the medieval empire from producing prosopographies. It could indeed be argued that, in the case of the Byzantine empire and its sphere of influence, it is exactly the diverse, diffuse and multilingual nature of our sources that makes a good prosopography so essential.

Where Stone saw the difference between older and newer prosopographical styles as a division between recording ‘the elites’ and ‘the masses’ ([Stone 1971](#): 47–48), Dion Smythe draws a distinction between ‘old’ and ‘new’ prosopography as a focus on the things (titles, property,

etc.) that were held by individuals, vs a focus on the connections between those individuals ([Smythe 2008](#): 177). Moreover, Smythe's definition of 'new prosopography', with its focus on relationships and connections between people, lends itself much more readily to the sort of network analysis that, arguably kicked off by [Padgett and Ansell \(1993\)](#), has grown increasingly popular in historical studies.

The factoid model as a solution

Enter the 'factoid' approach to prosopography. While Franz Georg Maier, who may have introduced the term to the historical field, defined factoids as 'speculations or guesses which have been repeated so often that they are eventually taken for hard facts' ([Maier 1985](#): 32), factoid prosopography defines a 'factoid' as something that may or may not be true, but has been claimed ([Smythe 2007](#): 130; [Pasin and Bradley 2013](#): 89). A factoid prosopography is a large and interconnected collection of these factoids, indexed by reference to the primary person to whom the factoid pertains. Here the work of the prosopographer is expressly limited to interpreting those sources and representing their claims in the data model, regardless of the prosopographer's opinion of their credibility. These claims may perfectly well conflict, and with the factoid approach, it is up to the scholar using the database to determine, given all the pieces of evidence, how best to interpret them. Smythe justifies the factoid approach by pointing out that 'Computer systems are binary systems ... [T]he information—all the information—must be one thing or the other; it is there or it is not there. There is no room for ambiguity, for the shades of grey' ([Smythe 2007](#): 128). Factoid-based systems adapt to this binary nature by reframing the prosopography as a set of statements that were certainly drawn by the prosopography from primary sources, rather than a set of ambiguous or contestable statements about 'what really happened'.

The *Prosopography of the Byzantine World* or PBW ([Jeffreys et al. 2017](#)), like many digital prosopographies conceived and implemented since the 1990s, uses modern database technologies and nigh-unlimited data capacity to transcend Stone's distinction between records of 'the elite' and 'the masses' and even Smythe's distinction between records based on personal holdings and those based on personal relationships. Rather, its

scope has been to record, as far as possible, information on certain sorts of biographical details, holdings and relationships for every person found to have been documented in some connection to Byzantium between 1025 and the late twelfth century. Even more than the recording of these factoids and publication of a novel form of prosopography on a bespoke website, a major contribution of PBW has been the openness of its editors to the free access and reuse of its data with the release, in MySQL format, of the data collection itself through a public GitHub repository. This openness has directly enabled the work that is described herein.

Linked Open Data approaches

Another approach to the collection and representation of historical information online is to collect the information directly as subject-predicate-object statements in Linked Open Data (LOD) format, using ontologies for the description of information about people such as Bio CRM ([Tuominen et al. 2017](#)), Symogih ([Beretta 2017](#)) or SDHSS ([Beretta 2024](#)) to structure it. One of the great advantages of the LOD models is that datasets can be interconnected—technically speaking, a subject or object of an LOD statement can easily be an entity that is defined in a different dataset. Another great advantage is the facility that ontologies provide to give semantic meaning to this data; this in turn enables certain sorts of reasoning that can be used to augment data collections, *as long as logical consistency is maintained*. In order to facilitate this reasoning, LOD datasets tend to present their information as fact, as gleaned by the builders of the resource from whatever sources they used. Moreover, the addition of contextual information to an LOD statement is generally not supported, as ‘statements about statements’ were not foreseen in the original technical implementation.¹ This has already been recognised as a serious problem especially in collections of cultural heritage data published by galleries, libraries, archives and museums ([Kahn and Simon 2020](#)). While we as medieval prosopographers would indeed like to have specific and uncontested facts about the life events of the people we record, we are very conscious of our lack of unambiguous and unimpeachable sources, and so are likely to look with suspicion and even disbelief at any piece of data asserting something so precise.

The gaps that remain

The main difficulty with the factoid model, and a reason that this author has frequently heard dissatisfaction expressed by scholars who use it, is precisely that it refrains from providing a coherent or authoritative narrative about its subject. Rather, the user is left to consider the sources and come to their own conclusions, quite possibly without much familiarity with the sources in question.² The more serious issue is that factoid prosopographies cannot quite escape the charge of historical positivism—the tendency to assume that, if something is not recorded in a source, it either did not happen or was not important enough to affect anything. ‘Sticking to the sources’ forestalls the possibility of including substantial reassessments of these sources—examples include the re-dating of certain Armenian histories and therefore of the lives of their authors ([Toumanoff 1961](#); [Thomson 1982](#); [Greenwood 2012](#)) or the identification of Michael Sikidites with the historian Michael Glykas ([Schreiner 1977](#); [Kazhdan 2005](#)) which is indeed not mentioned in the PBW entry for Michael 305. Conversely, the difficulty with the LOD model is that, as described above, the ontologies used are almost always written with the aim of presenting a single self-consistent set of historical facts.

Scholars who use prosopographical resources are naturally not keen to have to choose between arguments from authority that might quickly seem outdated on the one hand, or collections of factoids with no scholarly authority or even guidance provided on the other hand. How, then, can we find a way in the computational realm, where—as Smythe pointed out—shades of grey are not directly expressible, to express authoritative claims such that the prosopography as a whole remains relevant and useful for future generations of scholars? Even better, how can we represent statements such as ‘this might be the case’, which are poorly served in both the factoid model and the Linked Open Data model?

Since the digital medium tends rather to the quantitative, proposals for solutions that involve quantifying uncertainty frequently arise. Unfortunately, historians of the ancient and medieval periods do not tend to have guidelines for where the boundary between, for example, ‘high’ and ‘medium’ certainty should lie, much less a coherent measure for quantifying certainty numerically.

Data as assertions and the STAR model

A more profitable approach is to treat collections of prosopographical data not in terms of facts verified or accepted by the prosopographer(s) or even factoids positively arising from primary sources, but rather in terms of claims, or assertions, that have been made by any relevant person, whether historical or modern-day. The assertion- or claim-based approach to data collection has been at the root of several proposals for cultural heritage data modelling, including the LOUD data model used in the Historical Context Ontology ([Daquino and Tomasi 2015](#)), the Linked Art data model ([Newbury 2018](#)) and the ‘conjectures’ data model ([Vitali and Pasqual 2025](#)). While these models all provide a way for the provenance of assertions to be recorded, they each in their own way expect that conflicts in the data should be resolved in favour of some accepted set of facts—exactly the situation that is not always possible in the Byzantine context!

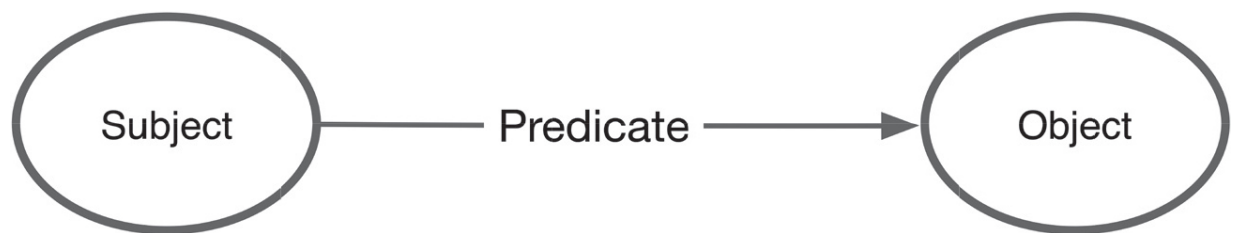


Figure 1.1: A typical Linked Open Data statement

This assertion-based collection of data is at the core of the RELEVEN project, which was funded by the European Research Council and ran from 2021 to 2026. In our work we have aimed to provide a clear counter-example to the common assumption that digital history projects cannot help but revert to an old-fashioned form of positivism. For these purposes we have developed the STAR model (short for STructured Assertion Record), a Linked Open Data strategy that is based on the well-known ontology provided by the CIDOC-CRM ([Doerr 2003](#)), which itself is used in an increasing number of cultural heritage collections and which provides the possibility to make formal inferences over large sets of data.

STAR is a reification model whose implementation is most similar to that of the LOUD model. ‘Reification’ means that the entity–property–

value statement that constitutes a data point becomes an entity in its own right to which more properties can be attached.

The schematic of a typical LOD statement is given in [Figure 1.1](#). It consists of an entity that is the subject of the statement, a relationship that denotes a property of that entity—one might think of this as the verb—and an object, which is the value (or content) of that property. An example of a simple statement in the CIDOC-CRM might have the subject ‘Porphyrogennetoi’ (an instance of an E74 Group, in this case the group of all those who were ‘born in the purple’ in the Byzantine era), the predicate P107 has current or former member and the object ‘Ioannes II Komnenos’ (an instance of an E21 Person).

The same statement, now a somewhat more complex object, is represented in [Figure 1.2](#) according to the STAR model. All three components of the original statement—the subject, predicate and object—are then connected to a core ‘assertion’ object, which also provides a connection to an entity representing the authority for the statement (that is, who made the claim) and one representing the evidence (that is, on what basis the claim was made), and is itself linked to the source (where the claim is found). To extend the example above, we might say that Anna Komnene (the authority), in writing of the birth of her brother Ioannes in the *Alexiad* (the source), has made the assertion that he was a porphyrogennetos. While Anna gives us no particular evidence for this assertion apart from her word, Zacos and Veglery (a different authority) have drawn the same conclusion on the basis (evidence) of the seal inscriptions they publish ([Zacos and Veglery 1972](#): n. 104a); the seal catalogue is then the source.

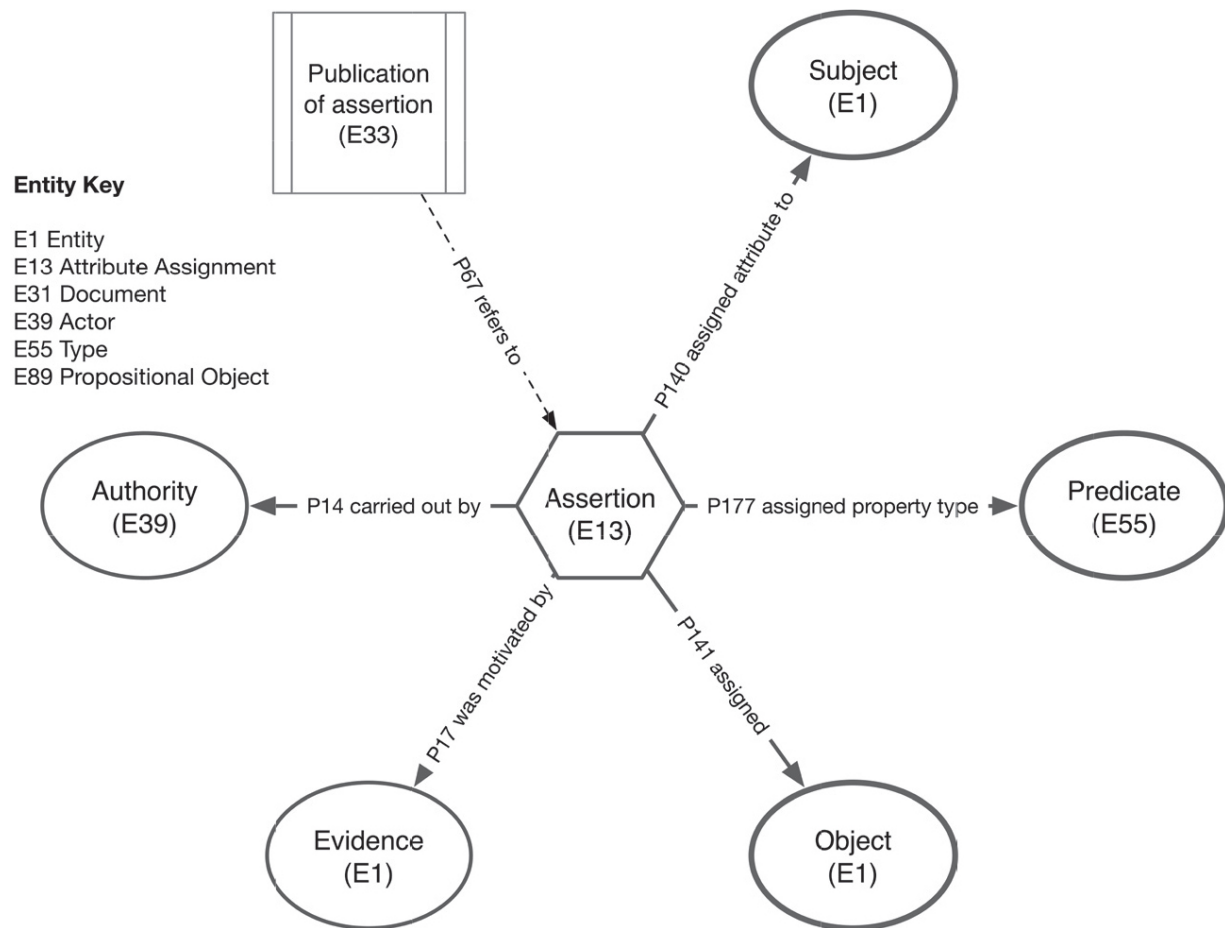


Figure 1.2: A STAR assertion of the statement in [Figure 1.1](#)

We have implemented this model using the CIDOC-CRM class E13 Attribute Assignment to represent an assertion, which takes the following properties according to the definition of the ontology:

- P140 assigned attribute to (the subject)
- P177 assigned property of type (the predicate)
- P141 assigned (the object).

As a subclass of E7 Activity, meant to represent any activity that occurred in time, the assertion may also take properties suitable for activities, such as:

- P14 carried out by (the person, group or other actor who made the assertion)
- P17 was motivated by (the evidence basis)

- any other properties compatible with an E7 Activity, such as when and where the assertion was made.

Finally, since the vast majority of our assertions are taken from secondary published sources, we can document their provenance by making the assertion itself an object of a CIDOC-CRM statement E33 Linguistic Object – P67 refers to – Assertion. Every historical statement in a STAR collection is thus linked not only to its content, but also to the authority—the person(s) from whom we have the information—and the provenance, which is the source for the claim and/or any evidence upon which they were basing their claim.

At first sight, the STAR model may seem unnecessarily complicated for the representation of data that no one disputes. However, in our view a wholesale switch from fact-based to assertion-based thinking simplifies our data collection rather than complicating it, and even has a key advantage: When we treat every statement as an assertion, we open the critically important possibility of re-examining old certainties in light of new evidence.

From factoids to assertions: the case of PBW

The *Prosopography of the Byzantine World* (PBW) has been an ideal dataset with which to test the recasting of factoids into assertions, for a variety of reasons. Firstly, the prosopography includes a large amount of data about nearly 10,000 people who have been documented during the eleventh century. Secondly, thanks to the foresight of Charlotte Roueché, who published on the PBW site a list of the scholars involved in the project and the sources they analysed, we have information not only about the primary sources that were used, but also about which project member was responsible for editing and reading these sources and converting them into data statements, for the vast majority of the database. Thirdly, as the source code for the entire project, including its data, has been made available on GitHub, we were able to find, understand and reuse the data in a transparent way. Fourthly, since PBW remains a published resource that is openly available, its use enabled us to develop a schema to link the reused data

back to its original source and so help to ensure cross-discovery of the respective resources. Fifthly and finally, thanks to the nature of PBW as a factoid database, there is quite a lot of divergent and even conflicting information to model and resolve.

The first step for the conversion was to map as many of the PBW factoid types as possible into an ontology compatible with CIDOC-CRM. This process also allowed us to understand the sorts of information that are not yet well-represented in existing ontologies, and thus give us a sense of the challenges we will face as we add other kinds of data.

Since CIDOC-CRM is an event-based ontology, its core concepts are expressed in terms of things that happened at some given place or time, which could involve people, objects or other entities. A good example is the expression of birth or death. It is easy to imagine, in Linked Open Data style, that a person may have properties such as ‘date of birth’, ‘place of birth’ and so on, which would allow us to model a natural-language description such as ‘Maria Komnene, daughter of Irene Doukaina and Alexios Komnenos, was born in the Porphyra of the Imperial Palace in 1085’ with a set of statements:

Maria Komnene – had birth date – 1085

Maria Komnene – had birth place – Porphyra

Maria Komnene – had mother – Irene Doukaina

Maria Komnene – had father – Alexios Komnenos

The disadvantage of this approach to modelling is firstly that every aspect of the birth will require a new property to be defined for each person, and secondly that these properties have no explicit connection to each other. It is only through semantic understanding of the property names that a person can infer that they are describing the same occurrence. An event-based model instead creates an entity called Birth, so that all the statements relevant to the birth of Maria are collected together. Maria’s birth would thus be represented as follows, using the vocabulary of CIDOC-CRM:

E67 Birth (of Maria Komnene) – P98 brought into life – E21 Person
(Maria Komnene)

E67 Birth (of Maria Komnene) – P96 by mother – E21 Person (Irene
Doukaina)

E67 Birth (of Maria Komnene) – P97 from father – E21 Person (Alexios Komnenos)

E67 Birth (of Maria Komnene) – P4 has time-span – E52 Time-Span (the year 1085)

E67 Birth (of Maria Komnene) – P7 took place at – E53 Place (Porphyrium)

The event model is thus a useful way to organise information about a person, insofar as their life can be treated as a series of events (birth, death, journeys undertaken, correspondence exchanged) or at least situations in time (kinship, holding of offices, possession of property) that have metadata about when, where, who was involved and so on.

Of the seventeen PBW factoid types, a few—‘Description’, ‘Second Name’, ‘Ethnic Label’, ‘Alternative Name’, ‘Possession’—were straightforward to model in CIDOC-CRM itself and did not require recasting as an event. Most others had to be recast into events; in all of these cases, the expression as events allows the addition of information that was not explicitly represented through these PBW factoid categories, such as the date range to which these factoids applied.³ It means that kinships, the holding of offices and so on can have a time and even a space element connected to them, if the source has provided one. Most of these event categories also required the use of supplemental ontologies (details are given in the appendix). Manual augmentation was performed by the RELEVEN team for the list of Death factoids, which are human-readable descriptions of the person’s death that appear in the sources, rather than machine-readable data. The dates of deaths, in particular, are very rarely explicitly recorded in the PBW database and the dates that are recorded are not published as such on the website.

Certain categories included too much semantic variance to model directly. The ‘Location’ factoid type, for example, includes items that associate a person with a location in any respect and are not directly associated with a date. Moreover, they are often implicitly paired with factoids from other categories—for example, Michael Psellos is recorded as having held the dignity of *krites* of Thrakesion, and is recorded separately with the same source basis as associated with the location Thrakesion.⁴ This made ontological modelling difficult, since the existence of a location

factoid did not necessarily imply the presence of that person in that place at any given time. Here the project team worked manually through the list of Location factoids and, for each one, marked whether it was a case of physical presence and, if so, any date that was attested by the source or by the editor. Only those cases of physical presence that are not implicitly associated with another factoid category are modelled independently.

The Narrative, Education and Authorship categories could not be programmatically converted. The Narrative factoids, which function as a sort of catch-all and thus include a wide range of semantics, cannot categorically be modelled as events. Like the Locations, the Authorship and Education categories are more specific but have been used in several different semantically incompatible ways—e.g. ‘Authorship’ as statements that a person authored a particular text, as statements that a person wrote some texts whose details no longer survive, or as statements that a person was known as an author. Where available, however, we did make use of certain Narrative or Authorship factoids in our modelling of the source texts used by PBW itself.

This leads to the topic of modelling the primary sources. Material sources such as lead seals were covered adequately by the CIDOC-CRM ontology, which was in any case initially conceived for material objects that might be found in museum collections ([Andrews 2025](#)). For text sources, which are not quite so comprehensively handled in the base CRM, we adopted the recently-released LRMoo ontology ([Bekiari et al. 2024b](#))—this is the successor to the more well-known FRBRoo ontology, for which development ended in 2017 ([Riva and Žumer 2018](#)). While FRBRoo and LRMoo were conceived for use by libraries where print publication is an assumed norm, the current version of the ontology is flexible enough to deal with unpublished documents such as charters or manuscript colophons.

Says who? Authority for factoid statements

The process of conversion of factoid data into an assertion model leads quickly to the question of what constitutes an authority, in the sense of the STAR model. We can illustrate the problem with a few statements that portray different sorts of authority, or different sorts of responsibility for the content of a statement:

- Matthew of Edessa writes that the emperor had the mercenary Frangopoulos executed in 1063.
- The editors of the Vałaršapat edition of Matthew's *Chronicle* claim that Matthew wrote a sentence that contains the information that the emperor had Frangopoulos executed in 1063 ([Matthew of Edessa 1898](#): 144).
- Gérard Dédéyan concludes from the text of Matthew's *Chronicle* that Frangopoulos was executed in 1063 ([Dédéyan 2003](#): I.172).
- Werner Seibt understands Matthew's *Chronicle* to contain the information that Frangopoulos was executed in 1063, although he disputes the factuality of the statement ([Seibt 2010](#): 91).
- A future scholar interprets the account in Matthew's *Chronicle* together with the evidence of Seibt's seal to propose that Frangopoulos was executed during the reign of Michael VII, and Matthew had been unclear about the dating and identity of the emperor.⁵

The variety of statements that can be said to have some authority attached to them, and the ways in which these different authorities can overlap, lead to several questions. Can the author of a primary source (for example, a contemporaneous annalist, about whom we may know nothing at all other than that a text exists and a passage therein names this person as the author) be treated as an authority in our sense? If a scholar today repeats a claim made in a primary source, does that scholar then become an additional authority for the statement? If the claim is based on a material source such as a coin legend or seal inscription, where does the authoritative reading of text end and authoritative historical interpretation begin? Finally, what do we do with statements that are usually considered so obvious that they are not typically asserted, such as the gender classification of an otherwise unattested person whose name is given as Maria? Should we assign an authority statement for these?

Each of these questions was the subject of deliberation among the project members. In the end we have adopted the following guidelines for assignment of authority for the primary source statements that underpin prosopographical factoids:

- In principle, authority belongs to the person who brought the information into the scholarly record. Narrative documents used as primary sources are always assigned an (anonymous if necessary) author.
- For certain documents, especially legal documents, the authority belongs both to the person who caused the document to be written (if known) and the person who actually wrote it (if known). Information from imperial chrysobulls, for example, arises from the emperor. Several documents in the various archives of Mt. Athos name both principal and scribe, in which case these hold joint authority for the contents.
- If the information arises from a non-narrative document for which authorship cannot be established, authority is currently assigned to the editor who published the text of the document.⁶
- If the author of a primary source is known, this person is the authority for a statement that appears in their text. To do otherwise would have been, ultimately, to take the viewpoint that the reader is the sole authority for all that they understand. Moreover, in the opinion of the project team, the rejection of the idea that medieval authors can be authorities would have had the unwelcome side effect of suggesting that only present-day scholars merit the respect implied by assignment of authority.
- The authority for a statement concerning the authorship of a primary source lies either with the author, if that author has identified themselves, or with the editor of the edition used by the PBW contributors. The statement of authorship might come from the title page or the preface of the edition, for example.
- The PBW factoids include many snippets of source text to underpin the information given in the respective factoid. Since a complete assertion-based modelling of a factoid must include the provenance of these supporting text snippets, the authority that the primary source actually contains the given text at the given reference point is assigned to the PBW contributor who was responsible for analysis of that source.
- Certain statements in the database, such as the gender or the descriptive identifier of a person, are represented outside of factoids.

These statements have been assigned the authority of the PBW project as a whole.

- Certain other statements—for example, the formal statement required by the CIDOC-CRM scheme that a given birth or death is the birth or death of a certain person—have not been assigned any authority or source basis, since the fact that every person who lived in the eleventh century did, at some point, die, is not in dispute. In these cases, the statements that concern the date, location or description of the death are given the appropriate authorities according to these guidelines.

Although PBW, by design, does not contain information from secondary sources, we have applied similar principles to the secondary source information that we have incorporated into the database. In the current phase of the project, we are building a proof-of-concept network of scholarship based on the CRMinf ontology ([Doerr et al. 2023](#)) in which those who accept, support, doubt or contradict certain assertions can be represented in full.

What makes you say that? Evidence basis and documentation source

The STAR model provides for two related but distinct concepts of ‘source’; these are the evidence base for a statement, and the documentation of the same. Evidence bases can be recorded in many instances of secondary scholarship, and in primary scholarship especially where the information arises from a material object. Examples include:

- An analysis of a lead seal that came into his possession led Werner Seibt to conclude that Frangopoulos was still alive in 1071, although the seal inscription did not directly say this.
- The identification of owners of many lead seals is based on comparison of the name and title in the seal’s inscription with the appearance of that name and title elsewhere in the historical record; dating can also be based on visual or iconographic similarity to other seals.

In these cases, where possible, the ‘evidence’ branch of a STAR assertion can be completed. As we do this, we are building an explicit picture of the ways in which scholars have used information to deduce other information about the past. This explicit record of provenance, which allows us to take a much more critical and informed approach to reconstructing the past, has always been a desideratum of historians; in late antique and Byzantine studies, the trouble has often been that such provenance information is simply not available. This absence of information is, however, also information that becomes visible in the assertion model and that is available to users of data collected in this style.

As the factoids in the PBW collection almost exclusively concern statements taken from primary sources, these sources themselves are usually considered a sufficient evidence base. For use in the assertion model, however, we must say that the evidence base is more often than not unknown—for example, we have seen that Matthew of Edessa claims that Frangopoulos was executed by the emperor, but Matthew gives no specific basis for his claim. Like many medieval writers of history, his only acknowledgement of the provenance of information comes in his statement that ‘we have put together [the *Chronicle*] having enquired for a long time, from a great number of those who saw and heard who were born in years past, and through the readers of earlier historians’ ([Matthew of Edessa 1898](#): 112; trans. [Andrews 2017](#): 212–13). For statements taken from the *Chronicle* as well as most other contemporary historical narratives, we thus cannot record any evidence basis.

Naturally, especially for the material evidence but also sometimes for the written sources, there can be conflict about some aspects of their production (for example, whether two lead seals were in fact produced from the same boulloterion, what an image on a coin actually depicts, or who in fact wrote the source text in question). Since we use the assertion model to represent information about our sources themselves, we are easily able to represent differing hypotheses of how these sources arose and what they signify.

Who records the recorders? Authority and source for the assertion database itself

It would hardly be proper to put so much effort into the modelling of provenance and authority for historical assertions without considering how we model our use of databases such as PBW. We have devised a system, based on the CRMdig ontology for provenance metadata ([Doerr et al. 2022](#)), to indicate responsibility for encoding the assertions themselves. The vast majority of conversion is handled by a Python script, which is an instance of D14 Software, according to this ontology; running it results in a D10 Software Execution event whose property L11 had output links this script to all assertions that it produced, and whose CRM property P14 carried out by links the execution to the person who invoked the script.

Equally important is to ensure the acknowledgement of our sources, in the form of credit and links to the other digital resources from which we get our information. PBW factoids and, in some cases, PBW person pages are themselves classed as instances of E31 Document, which bear a P70 documents relationship to the assertion(s) they gave rise to.

Assertions as a basis for historical reconstruction

Since assertions allow the expression of differing, and even conflicting, points of view concerning historical people and their life events, a database of these assertions will allow many operations useful to historians who wish to use it. In closing, we present a few examples of how an assertion database can enable the historian to make more effective use of the historical data we have.

First and foremost, assertions can be filtered based on criteria such as the identity of the authority or the type of source material from which the assertion arose. This allows the scholar to examine exclusively material evidence, for example, or to examine assertions made by certain clusters of secondary scholars. An interface for the RELEVEN assertion database is currently in development, which will allow users to create these filters and visualise the information they contain in a variety of ways.

Secondly, the assertion structure enables detection of logically conflicting information without compromising the ontological integrity of the database as a whole. A master's thesis is currently in preparation within

the project, for which a tool has been built to detect conflicting statements about kinship that occur in the PBW database. A prosopography that can automatically highlight the incompatibility of source statements with each other, rather than leaving the exercise to the mental energy of the user, will be a major advance that the digital medium can provide.

Thirdly, assertions can be complemented by inferences. An inference, in this context, is an assertion-like structure that has a defined set of rules, rather than a source or authority, which describes the basis for the inference. Baillie has pointed out the peculiarity that PBW and other similar prosopographical databases seem to have surprisingly few members of the majority ethnic group ([Baillie 2021](#): 126–7); indeed, PBW records only sixteen Byzantines and three Greeks in the whole of its dataset. This decision is justified by Smythe on the principle that ethnicity is a much more complex concept than simple labels, and that factoid prosopographies should therefore only use ethnicity fields in cases where a source explicitly mentioned it ([Smythe 2007](#): 129). Inferences, on the other hand, are much better suited to this problem. If, for example, a will drawn up in Greek mentions a person named Helene Diaxene, it is commonly assumed that this person was a Byzantine Greek based on her name and the fact that the source did not say otherwise. Making such assumptions explicit with a rule-set allows for more comprehensive statistical analysis of the whole dataset, even as it refrains from tainting the data with unsourced and arguable claims. Moreover, different scholars can apply different rule-sets to reach their conclusions.

In these ways, the assertion structure opens the possibility that users of digital prosopographies can finally ‘have it all’: a relatively exhaustive list of what is known about a set of people, the ability to learn the opinions of scholars whom they consider to be sufficiently authoritative, and a straightforward way of augmenting the data with their own deductions and conclusions that will render the economic and social analyses that Stone once considered impossible for pre-modern periods possible and even defensible.

Acknowledgements

This work arises from a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 101002357).

The author also expresses her deep gratitude to the honorand of this volume, whose early intellectual investment in a digitally minded scholar at the very beginning of her career has led directly to the work presented here and much more besides.

Appendix

This table gives the ontological paths that have been used to render the respective PBW factoid types into CIDOC-CRM compliant Linked Open Data. The following prefixes indicate the ontologies that were used for the PBW model:

- crm: CIDOC-CRM version 7.1.3 ([Bekiari et al. 2024a](#))
- lrmoo: LRMoo version 1.0 ([Bekiari et al. 2024b](#))
- sdh: Semantic Data for the Humanities and Social Sciences ([Beretta 2024](#))
- sdh-so: SDHSS Social, legal and economic life ([Beretta 2024](#))
- sim: Similarity Ontology ([Halpin et al. 2010](#)), as implemented internally
- xsd: XML Schema 1.0

Table 1.1: CIDOC-CRM-based paths for conversion of PBW factoids to Linked Open Data assertions

Narrative	Not transferred
Authorship	Not transferred
Death	crm:E69 Death – crm:P100 was death of – crm:E21 Person [PBW person] crm:E69 Death – crm:P3 has note – xsd:string [PBW factoid text] crm:E69 Death – <i>crm:P4 has time-span</i> – crm:E52 Time-Span [date range of death]

	<i>crm:E69 Death – crm:P7 took place at – crm:E53 Place [location of death]</i>
Description	crm:E21 Person [PBW person] – crm:P3 has note – xsd:string [PBW description]
Dignity/Office	sdh-so:C13 Social Role Embodiment – sdh-so:P26 is embodiment by – crm:E21 Person [PBW person] sdh-so:C13 Social Role Embodiment – sdh-so:P33 is embodiment of – sdh-so:C12 Actor's Social Role [PBW dignity / office] sdh-so:C13 Social Role Embodiment – <i>crm:P4 has time-span – crm:E52 Time-Span [date range]</i> sdh-so:C13 Social Role Embodiment – sdh-so:P70 is embodied in – sdh-so:C25 Political or Administrative Entity [political or administrative unit associated with dignity or office]
Education	Not transferred
Ethnic Label	crm:E74 Group [PBW ethnic label] – crm:P107 has current or former member – crm:E21 Person [PBW person]
Second Name / Alternative Name	crm:E21 Person [PBW person] – crm:P1 is identified by – crm:E41 Appellation [Second or alternative name]
Kinship	sdh-so:C3 Social Relationship – sdh-so:P16 has relationship type – sdh-so:C4 Social Relationship Type [PBW kinship] sdh-so:C3 Social Relationship – sdh-so:P17 has relationship source – crm:E21 Person [PBW person] sdh-so:C3 Social Relationship – sdh-so:P18 has relationship target – crm:E21 Person [PBW kinsperson] sdh-so:C3 Social Relationship – <i>crm:P4 has time-span – crm:E52 Time-Span [date range]</i>
Language Skill	crm:E21 Person [PBW person] – sdh:P38 has skill – sdh:C21 Skill – sdh:P37 concerns – sdh:C29 Know-

	How – crm:P129 is about – crm:E56 Language [PBW language]
Location	Not yet transferred; many of these factoids will be subsumed in other paths.
Occupation/Vocation	sdh-so:C1 Social Perception of an Actor – sdh- so:P13 pertains to – crm:E21 Person [PBW person] sdh-so:C1 Social Perception of an Actor – sdh- so:P14 has social quality – sdh-so:C2 Actor’s Social Quality [PBW dignity / office] sdh-so:C1 Social Perception of an Actor – <i>crm:P4 has time-span – crm:E52 Time-Span [date range]</i>
Possession	crm:E18 Physical Thing [PBW possession] – crm:P51 has former or current owner – crm:E21 Person [PBW person]
Religion	sdh-so:C23 Religious Identity – sdh-so:P36 pertains to – crm:E21 Person [PBW person] sdh-so:C23 Religious Identity – sdh-so:P35 is defined by – sdh-so:C24 Religion or Religious Denomination [PBW religion] sdh-so:C23 Religious Identity – <i>crm:P4 has time- span – crm:E52 Time-Span [date range]</i>
Eunuchs	Not transferred; these appear in the underlying database but not in the PBW2016 publication.
Uncertain Ident	crm:E21 Person [PBW person] – sim:ID8_identical – crm:E21 Person [PBW linked person]

For ease of reading, given here are the plain paths, before they were rendered into STAR assertions. In each case, the relevant assertions include the subjects, predicates and objects listed here as well as the authority and evidence base or documentation source. Those paths in upright type indicate direct conversions of PBW data; those in italics indicate where additional contextual information can be added, beyond what is represented explicitly in PBW.

Notes

1. This is not to say that these things are impossible. RDF-Star with its companion query language SPARQL-Star ([Arndt et al. 2021](#)) is an extension to the Linked Data standard that would allow a statement to carry context; named graphs ([Lassila and Swick 1999](#)) are perhaps a more widely used approach to similar problems.
2. That said, prosopographies have a long and storied tradition of their users complaining about what they do and do not provide; see [Barnish \(1994: 173\)](#) for the example of the *Prosopography of the Later Roman Empire*.
3. The details of how the different PBW factoid types were modelled using these ontologies are given in the Appendix.
4. <https://pbw2016.kdl.kcl.ac.uk/person/107822/> [accessed 9 October 2024].
5. As far as the author is aware, no one has made a case for or against this idea; it is included here as an example of the sort of deductive historical reasoning that is frequently employed in medieval studies in particular.
6. Another equally valid option would be to assign the authority to an anonymous person or group of people, on the principle that the document must have had some contemporary author; however, this seemed to us unnecessary, insofar as the editors of all such documents in PBW presented the documents as genuine.

References

- [Andrews, T.L. \(2017\)](#) *Matt'ēos Urhayec'i and His Chronicle: History as Apocalypse in a Crossroads of Cultures*. Leiden; Boston: Brill (Medieval Mediterranean, 108). Available at: <https://doi.org/10.1163/9789004330351>.
- [Andrews, T.L. \(2025\)](#) 'Byzantine Sigillography, Linked Open Data, and the Structured Assertion Record', *Digital Medievalist*, 18(1), pp. 1–24. Available at: <https://doi.org/10.16995/dm.16708>.
- [Arndt, D., Broekstra, J., DuCharme, B., Lassila, O., Patel-Schneider, P.F., Prud'hommeaux, E., Thibodeau, T. Jr. and Thompson, B. \(2021\)](#) *RDF-star and SPARQL-star*. Final Community Group Report. W3C. Available at: <https://www.w3.org/2021/12/rdf-star.html>.
- [Baillie, J. \(2021\)](#) 'Alternative Database Structures for Prosopographical Research', *International Journal of Humanities and Arts Computing*, 15(1–2), pp. 117–132. Available at: <https://doi.org/10.3366/ijhac.2021.0265>.
- [Barnish, S.J.B. \(1994\)](#) 'Late Roman Prosopography Reassessed', *The Journal of Roman Studies*. 84, pp. 171–177. Available at: <https://doi.org>

- [/10.2307/300875](#).
- [Bekiari, C., Bruseker, G., Canning, E., Doerr, M., Michon, P., Ore, C.-E., Stead, S. and Velios, A. \(eds\)](#) (2024a) ‘Definition of the CIDOC Conceptual Reference Model’. Available at: https://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_7.1.3.pdf.
- [Bekiari, C., Doerr, M., Le Bœuf, P. and Riva, P. \(2024b\)](#) ‘LRMoo: object-oriented definition and mapping from the IFLA Library Reference Model’. Available at: https://cidoc-crm.org/lrmoo/sites/default/files/LRMoo_V1.0.pdf.
- [Beretta, F. \(2017\)](#) ‘L’interopérabilité des données historiques et la question du modèle: l’ontologie du projet SyMoGIH’, in B. Juanals and J.-L. Minel (eds) *Enjeux numériques pour les médiations scientifiques et culturelles du passé*. Nanterre: Presses universitaires de Paris Nanterre (Notions et méthodes), pp. 87–127. Available at: <https://halshs.archives-ouvertes.fr/halshs-01559816>.
- [Beretta, F. \(2024\)](#) ‘Semantic Data for Humanities and Social Sciences (SDHSS): An Ecosystem of CIDOC CRM Extensions for Research Data Production and Reuse’, in T. Riechert and U. Gleixner (eds) *Professorale Karrieremuster Reloaded. Entwicklung einer wissenschaftlichen Methode zur Forschung auf online verfügbaren und verteilten Forschungsdatenbanken der Universitätsgeschichte*. Forthcoming, pp. 73–102. Available at: <https://doi.org/10.33968/9783966270502-05>.
- [Daquino, M. and Tomasi, F. \(2015\)](#) ‘Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects’, in E. Garoufallou, R.J. Hartley, and P. Gaitanou (eds) *Metadata and Semantics Research*. Cham: Springer International Publishing, pp. 424–436. Available at: https://doi.org/10.1007/978-3-319-24129-6_37.
- [Dédéyan, G. \(2003\)](#) *Les Arméniens entre Grecs, Musulmans et Croisés: étude sur les pouvoirs arméniens dans le Proche-Orient méditerranéen (1068–1150)*. Lisbon: Fundação Calouste Gulbenkian.
- [Doerr, M. \(2003\)](#) ‘The CIDOC CRM—An Ontological Approach to Semantic Interoperability of Metadata’, *AI Magazine*, 24, p. 2003.
- [Doerr, M., Ore, C.-E., Fafalios, P., Kritsotaki, A., Stead, S. et al. \(2023\)](#) ‘Definition of the CRMInf: An extension of CIDOC-CRM to support

- argumentation'. Available at: <https://cidoc-crm.org/crminf/sites/default/files/CRMinf%20v1.0%28site%29.pdf>.
- Doerr, M., Stead, S. and Theodoridou, M. (2022), 'Definition of CRMdig v 4.0'. Available at: <https://cidoc-crm.org/crmvig/sites/default/files/CRMdigg4.0.pdf>.
- Greenwood, T. (2012), 'A Reassessment of the History of Lewond', *Le Muséon*, 125(1–2), pp. 99–167. Available at: <https://doi.org/10.2143/MUS.125.1.2162441>.
- Halpin, H., Hayes, P.J., McCusker, J.P., McGuinness, D.L. and Thompson, H.S. (2010), 'When owl:sameAs Isn't the Same: An Analysis of Identity in Linked Data', in *The Semantic Web – ISWC 2010. International Semantic Web Conference*, Springer, Berlin, Heidelberg, pp. 305–320. Available at: https://doi.org/10.1007/978-3-642-17746-0_20.
- Jeffreys, M. et al. (2017), *Prosopography of the Byzantine World, 2016*. King's College London. Available at: <https://pbw2016.kdl.kcl.ac.uk/>.
- Kahn, R. and Simon, R. (2020), 'Feast and Famine: The Problem of Sources for Linked Data Creation', in T.L. Andrews et al. (eds) *Graph Technologies in the Humanities—Proceedings 2020. Graph Technologies in the Humanities 2020*, Vienna, Austria: CEUR (CEUR Workshop Proceedings), pp. 86–100. Available at: <https://ceur-ws.org/Vol-3110/#paper5>.
- Kazhdan, A. (2005), 'Glykas, Michael', in A. Kazhdan (ed.) *The Oxford Dictionary of Byzantium*. Oxford: Oxford University Press. Available at: <https://www.oxfordreference.com/view/10.1093/acref/9780195046526.001.0001/acref-9780195046526-e-2110>.
- Lassila, O. and Swick, R.R. (1999), 'Resource Description Framework (RDF) Model and Syntax Specification'. W3C. Available at: <https://www.w3.org/TR/PR-rdf-syntax/>.
- Maier, F.G. (1985), 'Factoids in Ancient History: The Case of Fifth-Century Cyprus', *The Journal of Hellenic Studies*, 105, pp. 32–39. Available at: <https://doi.org/10.2307/631520>.
- Matthew of Edessa (Matt'ēos Urhayec'i) (1898) *Žamanakagrut' iwn*. M. Mēlik'-Adamean and N. Tēr-Mik'ayēlean (eds). Vałaršapat: Tparan Mayr At'oroy Srboy Ėjmiacni.
- Newbury, D. (2018), 'LOUD: Linked Open Usable Data and linked.art', in *CIDOC Annual Conference 2018. CIDOC Annual Conference 2018*,

- Heraklion, Greece: International Council of Museums. Available at: <https://cidoc.mini.icom.museum/archive/past-conferences/2018-heraklion/>.
- Padgett, J.F. and Ansell, C.K. (1993), ‘Robust Action and the Rise of the Medici’, *American Journal of Sociology*, 98(6), pp. 1259–1319.
- Pasin, M. and Bradley, J. (2013), ‘Factoid-based Prosopography and Computer Ontologies: Towards an Integrated Approach’, *Literary and Linguistic Computing*, p. fqt037. Available at: <https://doi.org/10.1093/lc/fqt037>.
- Riva, P. and Žumer, M. (2018), ‘FRBRoo, the IFLA Library Reference Model, and now LRMoo: a circle of development’, in *World Library and Information Congress, 84th IFLA General Conference and Assembly, Kuala Lumpur, Malaysia, 24–30 August 2018*. Kuala Lumpur, Malaysia. Available at: <https://web.archive.org/web/20220309015738/http://library.ifla.org/id/eprint/2130/1/074-riva-en.pdf>.
- Schreiner, P. (1977), ‘Glykas, Michael, historian’, *Lexikon des Mittelalters*. Stuttgart: Verlag J.B. Metzler (Brepolis Medieval Encyclopaedias).
- Seibt, W. (2010), ‘Übernahm der französische Normanne Hervé (Erbebios Phrangopolos) nach der Katastrophe von Mantzikert das Kommando über die verbliebene Ostarmee?’, in J.-C. Cheynet and C. Sode (eds) *Studies in Byzantine Sigillography 10*. Berlin; New York: Walter de Gruyter, pp. 89–96. Available at: <https://www.degruyter.com/document/doi/10.1515/9783110227055.89/html>.
- Smythe, D. (2008), ‘Prosopography’, in R. Cormack, J.F. Haldon and E. Jeffreys (eds) *The Oxford Handbook of Byzantine Studies*. Oxford: Oxford University Press, pp. 176–181. Available at: <https://doi.org/10.1093/oxfordhb/9780199252466.013.0019>.
- Smythe, D.C. (2007), ‘“A Whiter Shade of Pale”: Issues and Possibilities in Prosopography’, in K.S.B. Keats-Rohan (ed.) *Prosopography, Approaches and Applications: A Handbook*. Oxford: Unit for Prosopographical Research (Linacre College) (Prosopographica et Genealogica / 13), pp. 127–137.
- Stone, L. (1971), ‘Prosopography’, *Daedalus*, 100(1), pp. 46–79.
- Thomson, R.W. (1982), ‘Introduction’, in Elishē, *History of Vardan and the Armenian War*. Translated by R.W. Thomson. Cambridge, MA: Harvard University Press, pp. 1–53.

Toumanoff, C. (1961), ‘On the Date of Pseudo-Moses of Chorene’, *Handes Amsorya*, 75, pp. 467–76.

Tuominen, J., Hyvönen, E. and Leskinen, P. (2017), ‘Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research’, in A. Fokkens et al. (eds) *Proceedings of the Second Conference on Biographical Data in a Digital World 2017. Biographical Data in a Digital World 2017*, Linz, Austria: CEUR (CEUR Workshop Proceedings), pp. 59–66. Available at: <https://ceur-ws.org/Vol-2119/#paper10>.

Vitali, F. and Pasqual, V. (2025), ‘The Representation of Historical Uncertainties as the Outcome of Competing and Incompatible Certainties’, in D. Raines (ed.) *Models of Data Extraction and Architecture in Relational Databases of Early Modern Private Political Archives*. Venice: Edizioni Ca’ Foscari, pp. 61–77. Available at: <https://edizionicafoscari.it/it/edizioni/libri/978-88-6969-919-1/the-representation-of-historical-uncertainties-as/>.

Zacos, G. and Veglery, A. (1972), *Byzantine Lead Seals*. Basel.

Chapter 2

Connecting Late Antiquities: challenges and opportunities for prosopographical data

Charlotte Tupman and Richard Flower

Introduction

In this chapter we discuss the capture, representation and dissemination of prosopographical data for late antiquity, examining how open data methods and standards can enhance this work and provide a basis for a more detailed understanding of the social and professional connections that existed during this period. We take as a case study the development of the *Connecting Late Antiquities* (CLA) digital resource, funded via an AHRC–DFG project led by Richard Flower (Exeter) and Julia Hillner (Bonn) with Co-Investigators Gabriel Bodard (London) and Charlotte Tupman (Exeter).¹ Following an exploration of previous and current approaches to the compilation and publication of large-scale prosopographies, we discuss how we implement open methods in the creation of a digital resource for the *Prosopography of the Later Roman Empire* (PLRE), examining the processes required to establish an accurate digital record, structure the data for publication and design an interface that enables users to query the prosopography in multiple ways. We discuss issues of sustainability, viewing the use of open standards as of key importance in addressing concerns about longer-term use of digital prosopographical data. Finally, we set out potential areas for future development of the CLA initiative, including plans to update and expand the PLRE and establish connections between entries in CLA and a

variety of sources and projects that contain data about known individuals in late antiquity.

Background and current prosopographical initiatives

The utility of prosopography has long been recognised by historians, particularly those wishing to understand the composition and functioning of particular groups and institutions within ancient societies. The discipline developed into its modern ‘scientific’ form during the nineteenth century, first in studies of classical Greece and then for the Roman Republic and Principate, drawing in particular on new critical editions of texts and monumental published corpora of inscriptions ([Destephen 2026](#)). These pioneering projects focused on office-holders and other elites, seeking in particular to understand the development and workings of the Roman state. For the late-antique period, the systematic creation of a comprehensive prosopography has been the subject of major scholarly endeavour for a significant period of time. It began in earnest at the start of the twentieth century, when a project was initiated in Germany by Theodor Mommsen and Adolf Harnack on the model of the *Prosopographia Imperii Romani* (PIR), which covers the preceding centuries ([Rebenich 1997](#): 247–326; see also [Destephen 2024](#)). The initiative came to a stop in the early 1930s, but was revived after the Second World War, with responsibility divided between groups of researchers in France and the United Kingdom. The former took charge of creating a Christian prosopography, the *Prosopographie chrétienne du Bas-Empire* (PCBE), primarily consisting of ecclesiastical office-holders as well as other individuals who made a significant contribution to the church, such as wealthy benefactors. The editorial work and publications are divided geographically according to the administrative divisions of the later Roman empire, with four volumes published to date, covering Africa (1982), Italy (1999–2000), Asia (2008) and Gaul (2013). The British team dealt with ‘secular’ prosopography, largely restricted to holders of high rank or major office in the imperial government and omitting many sub-elites, including local aristocrats and *curiales*. Funded by the British Academy, this produced the *Prosopography*

of the *Later Roman Empire* (PLRE), covering the period from 260 to 641 and arranged into three, chronologically structured volumes published between 1971 and 1992 ([Mathisen 2003](#): 23–40; [Martindale 2003](#): 2–10). A supplementary volume containing addenda and corrigenda was planned but never completed (see [Jones, Martindale and Morris 1971–92](#): III v). In the half-century since the appearance of the first volume, however, numerous articles have been published with the explicit purpose of proposing new entries for the PLRE (e.g. [Baldwin 1982](#); [Barnes 1973](#); [Handley 2005, 2010](#); [Martindale 1980](#); [Mathisen 1982, 1987](#); [Roda 1980](#)), while new discoveries and wider developments in scholarship mean that many of the existing ones are also in need of revision.

Scholars of late antiquity are therefore fortunate to have access to such rich and impressive resources, but nonetheless face limitations on their ability to use them to maximum effect. Beyond the ever-present need for updates in response to new research, the most obvious issue concerns their format, as both PLRE and PCBE exist only as print volumes. They are some of the last major projects of their type to be undertaken before the shift to digital approaches to the humanities began in earnest. For example, John Martindale, who spearheaded the editorial work on PLRE from the death of A. H. M. Jones in 1970 to its completion in 1992, then moved on to take charge of the *Prosopography of the Byzantine Empire*, which was published on CD-ROM in 2001, with a second edition online in 2014.² The separation of material between PCBE and PLRE also creates challenges for researchers who wish to approach late antiquity in a holistic manner. Although this division between the ‘religious’ and the ‘secular’ was established partly due to practical concerns, it runs the risk of imposing artificial boundaries onto late Roman society and its complex networks. On occasion this even affects individual lives: the fifth-century empress Galla Placidia appears in both PLRE II and PCBE 2, while PLRE I provides ample information on the career of Ambrose, *consularis Aemiliae et Liguria*, but stops abruptly at his ordination as bishop of Milan in 374, leaving the remaining twenty-three years of his life to the PCBE.³ Moreover, the scope of PLRE was also relatively circumscribed to only cover those figures perceived as most significant for the study of late-antique government and society.⁴ This was again partly due to the limitations of what could be achieved by the researchers, but it also

reflected prevailing scholarly interests at the time the project was first conceived. A substantial quantity of evidence survives, particularly from epigraphic and papyrological material, attesting to the lives of sub-elite and non-elite individuals from late antiquity, but it remains largely untapped in prosopographical research.

Moreover, a number of important developments have also taken place in approaches and formats for prosopographies. The PLRE largely followed the model adopted for the PIR (albeit written in English rather than Latin), with entries recounting lives of individuals, focusing on reconstructing their careers as far as possible and providing references to the relevant ancient sources. Although these frequently incorporate statements of uncertainty, especially concerning matters such as dating or the possibility that two attested homonymous figures are actually the same person, they nonetheless ultimately represent scholarly judgements made by the compilers and editors. These often affect the details of particular biographies but can also be at a more structural level, including assumptions regarding the order in which particular offices were usually held by members of the imperial bureaucracy. The advent of digital resources did, however, also bring with it new methodologies for prosopographical research. One of the most significant is the ‘factoid’ model, developed during the creation of the *Prosopography of the Byzantine Empire* (PBE) and implemented most fully in its successor, the *Prosopography of the Byzantine World* (PBW) ([Roueché, Cameron and Nelson 2023](#)).⁵ Pioneered by John Bradley and a team of researchers at King’s College London, this approach moves away from employing individual historical people as the essential building blocks of a prosopography and instead places individual statements about these people at the heart of its structure ([Bradley 2017](#)). These are assertions drawn from the sources, such as that a person was born in a specific town, held a particular office or died on a certain date. As Bradley puts it, however, ‘a factoid represents an interpretation of what a spot in a particular source says on its own, rather than what modern scholarship now believes to be true more generally about an historical person’ ([Bradley 2026](#): 138).

The database thus functions as an aggregation of all the extant assertions from the sources, enabling the user to view these and make their own judgements with the minimum of editorial intervention by the compilers. It is also therefore possible for mutually contradictory statements about an

individual to appear within a resource due to conflicting assertions in different pieces of evidence, but the prosopography does not have a core aim of declaring some of them right and some wrong, as is the case in more traditional approaches. Such ‘factoids’ are, of course, the raw material from which prosopographical research has always been constructed. They lie behind the biographical entries in the PLRE and PCBE, as is attested by their extensive references to their sources. Resources such as PBW do, however, reduce the interpretation of them to as low a level as possible. The result might therefore be regarded as a more unmediated guide to the evidence, although it does lack the sort of neat, easily accessible narrative account of an individual’s life that many users of a prosopography find useful.

Developments in digital prosopographical research over the past decade have expanded upon or suggested alternatives to the factoid approach, focusing particularly on the expression of editorial assertions and on linking the data that we publish about people in the ancient world, to aid disambiguation and enable researchers to analyse an individual’s network. James Baillie, Tara Andrews and others have advocated for approaches that allow more room for expressing editorial interpretations within prosopographical datasets, while resources such as the *Digital Prosopography of the Roman Republic* have taken a scholarship-based rather than solely source-based approach ([Baillie 2018](#)). Baillie recommends a move away from relational databases to graph databases (for his own database on the Kingdom of Georgia he uses neo4j),⁶ seeing practical advantages in using a node and edge model, with edge types being used to connect person nodes. Graph databases can also be provided with front-end capabilities for visualisation of the person records and their connections. Key to this approach is the ability to include what Baillie describes as ‘a significant interpretative element’ that enables the user to separate out and question for themselves the interpretations that the editor of a prosopography makes ([Baillie 2018](#): 4). Andrews’s approach in the RELEVEN project, meanwhile, as outlined in her chapter in this volume, recommends expanding the way we model data about ancient people, building upon the Linked Open Data model to include more information about both provenance and context ([Andrews 2023](#): 1). Andrews and her colleagues recommend the STAR model, which adds to the triple of subject-

predicate-object two further points: source and authority. Their project also uses an event-based model based on CIDOC-CRM that allows them to record changes in an editor's perspective over time rather than assigning a single viewpoint to a given scholar ([Andrews 2023](#): 3).

An important step in developing shared standards for creating Linked Open Data for ancient people was the Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names (SNAP) initiative, which set out recommendations for structuring person data as a step towards an eventual linked open data network of ancient people ([Bodard et al. 2017](#); [Bodard 2021](#)). According to the SNAP recommendations, each person is assigned a SNAP identifier that serves as both a canonical reference point—a unique, stable Uniform Resource Identifier (URI) that identifies that specific individual—and as the basis for enabling the creation of a link between a person record and a scholarly assertion made about that individual. The aim is to create a network of assertions about individuals and their relationships to permit research on larger and more interconnected datasets. Whilst SNAP was funded only in its first phase, until 2014, it worked with several major datasets including the Lexicon of Greek Personal Names (LGPN), a corpus of individuals recorded in Greek texts, Trismegistos, a database of names and people in Egyptian papyri, and the *Prosopographia Imperii Romani* (PIR), a corpus of Roman elites from the first three centuries C.E., and its ontology has been used by large-scale projects including the Epigraphische Datenbank Heidelberg.⁷ Its recommendations, ontology and visualisations remain available on the SNAP website.⁸ The SNAP recommendations are worth outlining here as they have been important in our approach to creating the digital PLRE. SNAP recommended that we 1) make our prosopographical data available online with a stable URI for each person; 2) publish our person data in the form of RDF (Resource Description Framework) triples, in addition to any other outputs; 3) identify any co-references for the individuals in our prosopography in other sources; 4) record assertions about the relationships between individuals who appear in different sources; and 5) annotate our published texts with SNAP person IDs. Its Cookbook establishes how to undertake the steps that are required for the first three recommendations.⁹ The Pelagios Network LOD People group is now building on this groundwork by developing further guidelines, ontologies

and materials for working with Linked Open Data for people, including ancient people.¹⁰ This collaborative, standards-based approach to making person data available allows for the establishment of a network of ancient individuals that can transcend individual project and disciplinary boundaries, without requiring any given project to submit their work to a larger, comprehensive database.

For CLA we follow the SNAP recommendations as closely as possible within this two-year pilot project, while acknowledging the importance of recording specific editorial assertions. We are creating a stable URI for the subject of each PLRE entry; creating RDF from our XML files as a project output; and recording editorial assertions about the relationships within the PLRE. We are not yet recording co-references or relationships with persons appearing in sources outside PLRE, as that falls outside the scope of the current project.

Undertaking the digital PLRE and CLA project

The technological possibilities for the construction and presentation of prosopographical information have therefore changed tremendously since the PLRE was compiled. Concurrently, scholars have also taken advantage of these developments to engage in new forms of research and ask different questions of the available data through methods such as Social Network Analysis ([Hillner and MacCarron 2026](#); [Hammond 2021](#)). An array of projects focusing on particular aspects of late-antique history have created their own datasets to enable this type of work, often drawing material from the PLRE and PCBE and retooling it into other formats to suit their own purposes. For example, *The Migration of Faith* explored the topic of clerical exile, tracing the physical movements of priests while also mapping their social connections and classifying these into different types such as ‘companion’, ‘correspondent’ or ‘responsible for exile’.¹¹ It would therefore be highly desirable to transform the PLRE so that it can be used for such analysis without having to be manually mined for relevant information by every new project in turn and then re-entered into bespoke databases. At the same time, however, the PLRE remains a core reference work, easily

comprehensible to scholars at all levels due to the traditional biographical, narrative format of its entries, as well as being cited in countless publications for more than half a century. For that reason, it would be undesirable to radically alter its appearance and structure for the online version. The PLRE might therefore be said to exemplify both the challenges and the possibilities of prosopographical resources in the present day: they need to be as flexible as possible to accommodate not only multiple audiences but also an ever-increasing array of research questions, methodologies and practices. For the initial phase of our project, we therefore decided to produce a digital edition of the existing PLRE marked up in TEI XML, employing person entries as the main building blocks for the structure of the resource in the traditional prosopographical manner. Creating individual, predictable and stable URIs for every person in the resource will make linking to existing and future resources much more straightforward. Moreover, the markup enables much greater functionality and the capability to filter and query the dataset for a range of variables, including dates, offices and relationships. The underlying data from the markup and searches can be exported by users for other purposes, such as the reconstruction of connections and networks. For these reasons, the two main areas of focus for the project are the creation of the marked up digital PLRE and the building of the *Connecting Late Antiquities* digital resource, which will enable the linking of data on late-antique people from multiple sources.

Creating an updated digital PLRE

To prepare the ground for the markup of the PLRE, the first step was to capture the text and create a checked, accurate digital version, a challenge common to any project dealing with data from large-scale printed sources. Beginning with a set of PDFs provided by Cambridge University Press, we used a combination of Optical Character Recognition (OCR) and manual correction to achieve an accurate text. The nature of the PLRE presented us with some particularly interesting challenges, not least how to achieve the best OCR results for the various languages and scripts in the PLRE entries. It was especially important for us to address this issue in order to make the checking process by our postdoctoral research assistants (PDRAs),

Aleksander Paradziński, Jessica van 't Westeinde and Jeroen Wijnendaele, as efficient as possible. This would then allow more time for the markup and linking tasks that are fundamental to the project's outputs.

The OCR process was twofold: initial experiments with ABBYY FineReader yielded mixed but in places promising results,¹² but the best outcomes were achieved through a collaboration between one of our Exeter Digital Humanities Lab interns, Heide Elton-Miller, and Prof. Bruce Robertson of Mount Allison University, who kindly gave some of his time to the project. We are very grateful for his assistance. Robertson and Elton-Miller's experiments using Robertson's XQuery-based in-browser LACE2 OCR tool enabled us to achieve a high degree of accuracy for the output of the ancient Greek, which had proved more problematic with the ABBYY software ([Robertson 2019](#)).¹³

Prior to the project's funded phase, a case study on correcting the text was carried out by Daisy Wakefield, a student intern funded by the University of Exeter. This preliminary work established the nature of common OCR errors and allowed us to make an estimate of the time required to perform this task for the entirety of the PLRE. Our PDRAs then worked to compare the OCR output with the text of the PDFs and to make manual corrections where needed in order to produce an accurate text ([Paradziński and van 't Westeinde 2025](#)). During this process they encountered several types of errors, some of which could be addressed collectively but most of which required individual correction, meaning that this proved to be a significant and time-consuming task for the PDRAs.¹⁴

Once accurate texts had been achieved for all three volumes, we moved to the main focus of the PLRE work: the TEI XML markup of the structure and key features of each of the entries in the prosopography according to the project's priorities. To enable the desired functionality, for each person entry we are encoding: unique identifier; personal name; relevant dates; rank, office, occupation and/or role; and other persons named in the entry. We are also recording our researchers' assertions about the relationships between the people noted in the entries, to enable research into personal networks during late antiquity. For a subset of the entries in Volume I we have also encoded each placename with its canonical Pleiades identifier, as a pilot for future expansion. We selected these elements of the entries for tagging based on the likely needs of different groups of users. As discussed,

many research projects that employ prosopographical data focus on the workings of government, but with an increasing interest in the tracing of connections and mobility, performing social network analysis on late-antique society more broadly rather than focusing exclusively on institutions such as the imperial bureaucracy or army. We have not yet encoded primary source references, partly due to time constraints but also because most relevant texts (in either the original language or translation) are not yet openly available to link to online. This reflects the fact that late-antique material has not yet benefitted to the same degree as ‘classical’ literature from digitisation initiatives. As the PLRE uses standardised abbreviations for all sources, this encoding work should be relatively straightforward to perform in future, when more texts will hopefully be available online. To make the task of encoding the prosopography as efficient as possible, we first worked on a batch markup process to add TEI XML that represents the structure of the entries along with aspects of the data relating to the subject of each entry. This process was developed in an iterative way, through feedback from the PDRAs on which parts of the batch markup had been successful and which parts contained errors.¹⁵ The PDRAs then undertook a second phase of corrections to the batch-applied TEI XML, while also adding further manual markup to match our requirements of unique identifier, name, dates, occupation and other individuals appearing in the entries.

The task of recording the relationships that we are asserting existed between the people in the entries necessitated a discussion in several stages as to which kinds of relationships needed to be recorded in the digital resource, and also of how best to structure the representation of our knowledge of these relationships. Again, we considered the needs of a range of potential users of the resource, as we want to ensure as far as possible that researchers can work directly with the data produced from our markup without having to read through the prosopography and recast the information into a new format, as has been the case with the printed volumes. We therefore built on the experience of our principal investigator, Prof. Julia Hillner, from her *Migration of Faith* project, which recorded and analysed a variety of types of relationship and connection. We categorised them into four groups: familial (such as nephewOf or sisterOf); professional (such as doctorOf or prosecutorOf); dependency (such as clientOf or

victimOf); and ‘other’ for relationship types that could not be classified in one of these ways (such as neighbourOf and enemyOf). There are, of course, many ways that we could have chosen to categorise the relationships between individuals in the PLRE, but these categories drew consensus across the team and enabled us to assign each relationship under one of these headings. Users can, however, choose to organise the individual relationship tags in other ways if required for their own research. Wherever possible, we used relationships already defined in the Friend Of A Friend (FOAF) ontology to assist with the mapping of relationships between projects.¹⁶ We recorded the relationship assertions using the TEI <relation> element with @active and @passive attributes linking the unique identifier of the subject of the entry to the person with whom they have the defined relationship. Each assertion also includes a @resp that records who is responsible for making the assertion, enabling us to distinguish between assertions made by the original PLRE editors and those made by the current editors of the digital resource. These relationship assertions will also be made available as RDF (though the current project is not hosting a triplestore itself). These and other methodological considerations relating to the markup of the prosopography and the encoding of its relationships are discussed in detail in a paper by two of the CLA PRDAs ([Paradziński and van ’t Westeinde 2025](#)).

Developing the CLA resource

The digital PLRE will be the first prosopography to be published in the *Connecting Late Antiquities* resource, with Volume I made available to users first. It will form the basis of our pilot digital resource, providing a searchable online version of the prosopography that allows users multiple entry points to, and through, the people, the entries, the relationships that we have identified, and the resources to which we are linking.

We are building the CLA resource in eXist-db, an XQuery-based open-source framework designed specifically for publishing and querying XML.¹⁷ eXist-db permits a high degree of customisation, enabling us to tailor the resource to the requirements of a prosopographical project. Within this framework we are building on a set of templates designed by Exeter’s Digital Humanities Lab (DH Lab).¹⁸ Working closely with a developer

assigned to the project within the DH Lab, Lucie Mingmei Hao, we are creating a resource that provides detailed faceted browsing functionality, enabling the user to narrow down the dataset by selecting their own research criteria.

The primary means of accessing information about the subjects of the PLRE entries will be through a faceted browse interface providing enhanced functionality (see [Figure 2.1](#)). This will include browse features familiar to users of the printed version, such as alphabetical lists of names, but also a map view for those entries that are geolocated with Pleiades IDs, enabling an entry point to the dataset via geographical location; a date slider for narrowing down the time period; and filters for occupation, role and gender, so that a researcher interested (for instance) in viewing entries for teachers in the fourth century C.E. can view these records in isolation. Text search will also be provided.

Each person featured in a volume entry will have a dedicated page in the CLA resource, which will contain: their unique identifier; details of their gender, attested period of activity and ranks, offices, occupations or roles; their full PLRE entry; links to any other entries in which they are mentioned; and details of relationships with other named persons in the prosopography. The entries themselves will have two views, selectable via a toggle button: the original PLRE entry, and the updated entry containing any addenda and corrigenda by the CLA editors. People, places and role names will comprise clickable links that enable the user to move through the data contained in multiple entries and to follow paths through the entries according to their particular interests. For example, users will be able to view other holders of the same role; entries that relate to the same individuals; and entries that mention the same places (initially for a subset of entries). The combination of the faceted browse functionality, text search and the links between categories of information contained on the person pages should allow for both finer-grained research and serendipitous discovery of the contents of the PLRE entries.

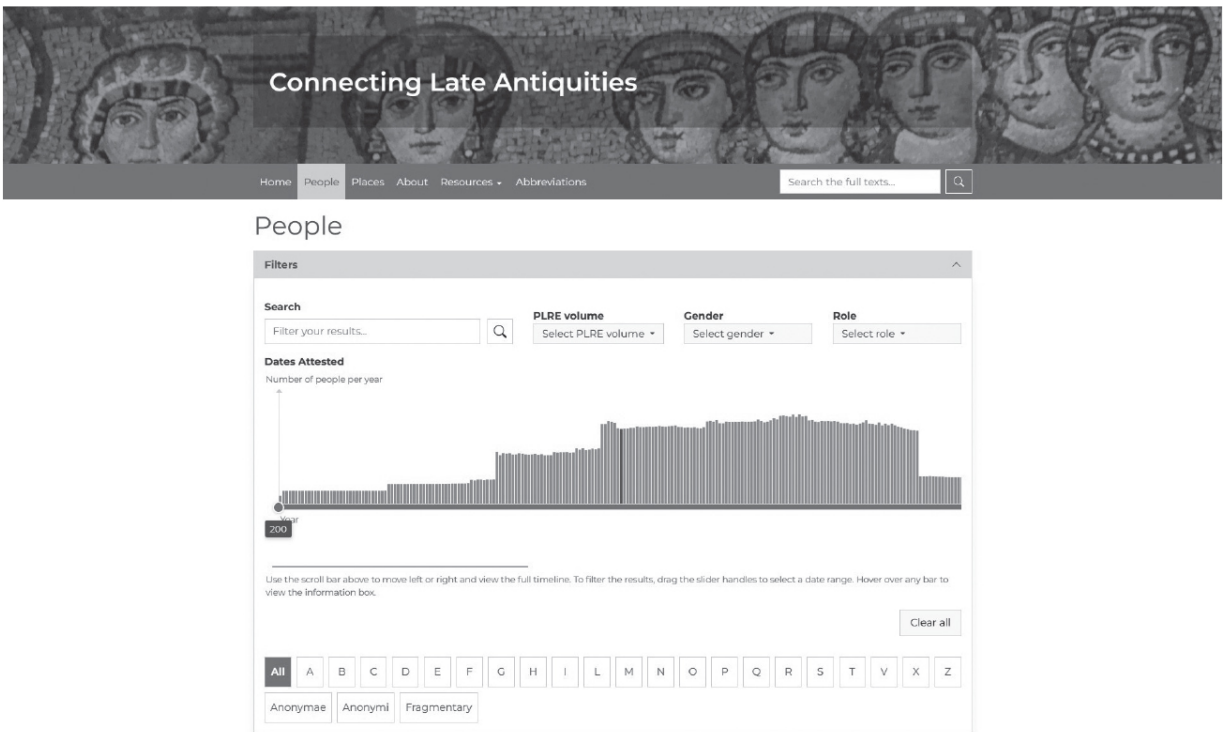


Figure 2.1: CLA People Search page (available from <https://connectinglateantiquities.org/texts.html>) (CC BY)

Although time limitations prevent us from including in this initial two-year project pointers to other resources that contain data about the same people, we anticipate that in due course this will include not only links to other project entries, but also assertions about the nature of the relationships between individuals in PLRE and those recorded in other resources, documented in the same way as our own internal PLRE relationships. As part of our project's outputs, we are providing guidance on how to construct URIs and how to record relationship assertions as RDF, drawing directly upon the earlier work carried out by Gabriel Bodard and colleagues for SNAP and the current work by the Pelagios Network LOD People working group ([Bodard et al. 2017](#)).

Building and involving communities

A key element of the *Connecting Late Antiquities* project is, of course, the development of meaningful links between projects, not only in a computational sense in terms of linking our data, but also in a broader sense that encompasses dialogue with colleagues and projects in our field, as well

as the accompanying iterative development of ideas and methods. The project is engaging with wider communities with prosopographical interests in three ways: first, through traditional dissemination methods including conference presentations, workshops and publications; second, by making a portion of our person dataset available before the project is launched to the Pelagios Network LOD People working group¹⁹ to stimulate discussion, experimentation and feedback on our approach; and third, by providing recommendations to other prosopographical projects about how to link their data to ours, provided they are happy to do so (and we make no assumptions about this), as well as how we will link ours to theirs. This will include examples and explanations of our URIs for the individuals who are present in CLA and details of the principles on which we have formed the URIs. We have already spoken to representatives from a number of projects that have agreed to incorporate our URIs into their own databases in order to enable this in the future.

Many fruitful conversations have already arisen through conferences and workshops that have helped us to refine our thinking and our approach, including gaining a deeper understanding of the challenges facing prosopographers of different periods in capturing and making available their datasets, as well as broadening our knowledge of potential user requirements and research questions, which in turn has helped to inform the design of the CLA interface. These workshops include the Edinburgh PROSOPON meeting organised by Niels Gaul and a workshop at St Andrews for the *Land and Loyalty in Late Antiquity* project. The sharing of a subset of our person data from PLRE I with the LOD People group is currently underway and we anticipate two potential outcomes: one, that feedback from that group should help us make better-informed choices about how we prioritise our markup to make it as useful as possible to the wider scholarly community; and two, that being able to engage publicly with our data ‘in the flesh’ before the launch of the project might encourage previously unanticipated forms of engagement. We expect that our current forms of participation in the wider communities of prosopographical scholarship are likely to broaden and develop further over time, particularly once a version of the CLA digital resource is available for users to experiment with for their own purposes. User testing will encompass two main phases and will include both scholars with strong familiarity with

prosopographical materials and those who work in broader but related research areas, including social and religious history, epigraphy and papyrology, as well as those interested in network analysis methods. Some users will go to the resource knowing precisely what they are looking for; others will want to be provided with pathways through the materials that enable them to discover the data in a more serendipitous way. Ultimately, we anticipate being able to demonstrate the utility and advantages of linking together datasets through CLA and to encourage new projects to consider making their datasets available to allow new connections to be made.

The future of *Connecting Late Antiquities*

Sustainability and longer-term plans

Questions of long-term sustainability are ever-present in any discussion about digital resources, not least because most research funding models do not allow for support that extends beyond the funded period of the project ([Cayless 2019](#); [Vannini 2022](#)). This means that the responsibility for sustaining a project beyond this period, where this is possible, tends to lie with the institution at which it has been created. In practice, this of course means that the responsibility lies with specific personnel, who are likely to have changing or growing responsibilities and who might move to other roles or institutions over time, making it harder to ensure that a project is supported after its completion. For this reason, institutions (including Exeter in this case) usually have a substantial but nevertheless time-limited period during which they agree to maintain a digital resource (including regular security updates) in the state in which it was published, in this case for five years. This is also the standard period for which the Arts and Humanities Research Council (AHRC) expects a project to be supported by an institution after funding has ended. Inevitably, this situation creates a potential problem for any project that is intended as a launchpad for further work, which the *Connecting Late Antiquities* pilot unquestionably is.

We are addressing this issue of sustainability in three ways: first, by making use of existing recommendations (in discussion with other prosopographical projects), standards (including TEI XML and SNAP) and ontologies (including FOAF) as the basis for recording and making

available our datasets; second, by documenting the main aspects of our decision-making, workflows and processes on the project website and in a GitHub repository so that others can choose to replicate our methods in their own projects; and third, by making available our encoded XML files and underlying eXist-db framework templates (including XQuery, XSLT, CSS and HTML). As well as being downloadable from the CLA digital resource and the project's GitHub repository, the files will be archived for the longer term in the free, publicly available Open Research Exeter (ORE) repository, which has a commitment to keeping its deposited data available to users of the repository indefinitely.²⁰ This will not only ensure that the encoded data remains available to researchers beyond the expected lifetime of any individual website, but should also help to encourage the reuse and repurposing of our materials by future scholars.

Keeping other resources findable

In addition to taking steps to address the sustainability of project data, CLA also seeks to contribute to the findability of not only its own datasets but also those of the projects to which it links. Any project that aims to bring together multiple datasets in a useful way for researchers should also be able to act as a means of disseminating information about those projects, and CLA is firmly committed to being a point of contact between varied, independent projects rather than an umbrella under which it gathers other projects to itself. In this we take inspiration from initiatives including Pelagios (and now the Pelagios Network), which enables access to multiple datasets based on shared geographical data ([Isaksen et al. 2014](#); [Simon et al. 2020](#)).²¹ CLA seeks to provide a means not only for connecting, but also for drawing attention to the highly varied datasets and areas of scholarship that form the basis for our understanding of ancient people.

Future work through a further funded project

This phase of the *Connecting Late Antiquities* project has been conceived as a pilot and testbed of methods, standards and sets of functionalities for a longer-term initiative to connect multiple prosopographies and related datasets on late antique people. We anticipate that its fruition will come when we have been able to expand the number of datasets that we can (a)

publish as part of the CLA resource and (b) link to from our materials. We are currently applying for further funding that would enable substantial expansion of the project and ensure its long-term stability. The first priority will be the expansion of the text markup of PLRE to enhance the usefulness of both the resource and its underlying data. This will involve completing the inclusion of Pleiades IDs for all locations mentioned in the text, as well as contributing new place data to the Pleiades Gazetteer²² where needed, especially for certain late Roman administrative divisions. We will also finish encoding all relationships recorded in PLRE, which will allow the data to be used for large-scale analysis of connections and social networks during late antiquity. This will provide us with the opportunity simultaneously to gain both an unprecedented level of detail and an overall view—including the possibility of network visualisations—of the nature of social and professional connections throughout this period, including changes over time or according to geographical location. For instance, it will be much easier to trace the relationships of and between the pupils of Libanius, exploring not only their origins and family ties but also their mobility, careers and developing associations across their lives. Where possible, we would also like to link references for literary, epigraphic and papyrological texts to online editions and translations of these works, although this will be dependent on their availability in an open format. In addition, we have recently acquired access to the (largely complete) archive of working papers from the original production of PLRE. We plan to digitise these and make it possible for users to view relevant material alongside individual entries in the online version.

Alongside completing this work on the original version of PLRE, we will also be embarking on a major campaign of updating and expanding the prosopography, with two main strands to this work. The first involves addenda and corrigenda, making changes to existing entries to reflect developments in scholarship since the publication of the volumes and also adding in new entries for individuals who ought to be included in accordance with the criteria used by the compilers, including a number of people already mentioned in passing in the work. As well as researching and writing these entries ourselves, we will also solicit suggestions from other scholars, many of whom often make relevant suggestions in passing or in footnotes within publications on a variety of aspects of late-antique

history, rather than just within specifically prosopographical studies. In all cases, any additions or corrections will be recorded in the XML files with standard TEI elements (<add>, <corr> and so forth) so that users will be able to choose whether to view the original PLRE entry or the updated version. For each revision the name of the contributor will be both acknowledged in the XML files and displayed with the revision in the resource. The second strand will expand PLRE beyond its original remit to encompass other groups and classes outside the imperial elite. In particular, we will focus our attention on *curiales*, the town councillors who constituted local elites and were often heavily connected to each other and to members of important institutions such as the bureaucracy, army and church. This research will involve a wide variety of different sources, with epigraphy, papyri and letter collections being particularly important for identifying these individuals and their relationships. At the same time, we will also be expanding significantly the work of linking our own resource to others, increasing its utility as a point of contact for the study of all aspects of late antiquity. We hope that it will prove valuable in the long term as both a repository of information and a model for scholarly collaboration.

Acknowledgements

All data from Jones, Martindale and Morris (1971–92) has been made available for us to use by kind permission of Cambridge University Press. The project has drawn upon eXist-db templates designed by Exeter Digital Humanities, based on eXist-db's open-source software. *Connecting Late Antiquities*' data is made available under a CC BY licence.

We would like to thank the Arts and Humanities Research Council and Deutsche Forschungsgemeinschaft; the CLA Advisory Board; our colleagues Julia Hillner, Gabriel Bodard, Aleksander Paradziński, Jeroen Wijnendaele, Jessica van 't Westeinde, Lucie Mingmei Hao, Gary Stringer, Gemma Poulton, Richard Holding, Julia Hopkin and Bruce Robertson; research assistants Dorka Kinti and Gregor Kirilov; interns Daisy Wakefield, Heide Elton-Miller and Vanessa Wong; and student assistants Azime Can and Josephine Probst. We thank the editors of this volume, Gabriel Bodard and Valeria Vitale, for their invitation to submit this piece

and for their suggestions for improvements; our reviewers for their thoughtful comments and suggestions; and last—but very definitely not least—we would like to thank Charlotte Roueché for her insight and inspiration, and for continuing to connect countless scholars around the world in pursuit of a deeper understanding of ancient societies.

Notes

1. *Connecting Late Antiquities* digital resource: <https://connectinglateantiquities.org/>.
2. J. Martindale et al., *Prosopography of the Byzantine Empire* (2014) available at: <https://pbe.kcl.ac.uk/>.
3. In fact, the PCBE volume for Italy does not contain an entry for Ambrose of Milan, since the editors concluded, as with Augustine of Hippo in the Africa volume, that such an undertaking would require a volume in its own right.
4. The remit was limited to holders of major military and civilian ranks or offices, as well as ‘lawyers, doctors, rhetors, grammarians and poets’, but excluded ‘private soldiers and NCOs ... and decurions of cities’: [Jones et al. 1971–92](#): I vi.
5. M. Jeffreys et al., *Prosopography of the Byzantine World, 2016* (King’s College London, 2017) available at <http://pbw2016.kdl.kcl.ac.uk>.
6. neo4j graph database and analytics tools: <https://neo4j.com/>.
7. *Epigraphische Datenbank Heidelberg*, download page including Linked Open Data (RDF): <https://edh.ub.uni-heidelberg.de/data/download>.
8. SNAP Cookbook <https://snapdrgn.net/cookbook.html>; SNAP Ontology <https://snapdrgn.net/ontology.html>.
9. SNAP Cookbook <https://snapdrgn.net/cookbook.html>.
10. LOD People GitHub repository: <https://github.com/DigiClass/LOD-People>
11. *Migration of Faith: Clerical Exile in Late Antiquity (325–600)* led by Julia Hillner: <https://www.clericalexile.org/>.
12. Our thanks are due to our student intern Daisy Wakefield and to Heide Elton-Miller for their work on the OCR process, and to Exeter’s Digital Humanities Lab Manager Gemma Poulton for facilitating this.
13. Bruce Robertson, LACE2 editing suite (including source code): <https://github.com/brobertson/Lace2>.
14. Some of the most common errors included Roman numerals appearing as numerals rather than as letters and sometimes even being converted into Greek letters; the apparently random capitalisation of some letters, especially E, S and V; diacritics not being converted correctly; and superscripts in some cases being converted to a capital P. Numerals presented several issues, including some numbers being italicised unnecessarily; 0 being converted to a lower-case o; the numeral 4 being converted to a + sign; and the numeral 8 being converted into a 6. These problems presented themselves inconsistently, making it harder to resolve the errors in the form of batch corrections. For instance, Isidore of Seville might appear correctly five times, but on the sixth occasion he would appear as Lidore of Seville.

15. Errors included false negatives, for instance where a name had not been recognised and therefore not encoded, and false positives, for instance where page numbers were incorrectly marked up as dates.
16. Friend Of A Friend (FOAF): D. Brickley and L. Miller: FOAF Vocabulary Specification, <https://xmlns.com/foaf/spec/>. See also Matsuo et al. 2001.
17. eXist-db environment for XML web application development: <https://exist-db.org/exist/apps/homepage/index.html>.
18. Exeter Digital Humanities Lab: <https://www.exeter.ac.uk/research/digitalhumanities/>.
19. Ancient People Google Group: <https://groups.google.com/g/ancient-people>.
20. A link to the data stored in our CLA GitHub repository will be provided from the main CLA project website. The data will also be findable via Exeter's institutional repository, currently Open Research Exeter, which is freely searchable and browsable. In addition, the data will be deposited in a Zenodo repository, which will provide another means of accessing the dataset. Links to all repositories will be provided from the project website. CLA GitHub repository: <https://github.com/ExeterDigitalHumanities/connecting-late-antiquities>. Open Research Exeter policy document: https://www.exeter.ac.uk/v8media/universityofexeter/research/openaccess/ORE_policy.pdf.
21. Pelagios Network: <https://pelagios.org/>.
22. Pleiades Gazetteer of Ancient Places: <https://pleiades.stoa.org/>.

Bibliography

Online resources

Ancient People Google Group: <https://groups.google.com/g/ancient-people>
Digital Prosopography of the Roman Republic (DPRR): <https://romanrepublic.ac.uk>
Epigraphische Datenbank Heidelberg, download page including Linked Open Data (RDF): <https://edh.ub.uni-heidelberg.de/data/download>
Exeter Digital Humanities Lab: <https://www.exeter.ac.uk/research/digitalhumanities/>
eXist-db environment for XML web application development: <https://exist-db.org/exist/apps/homepage/index.html>
Friend Of A Friend (FOAF): D. Brickley and L. Miller: FOAF Vocabulary Specification, <https://xmlns.com/foaf/spec/>
LACE2: Bruce Robertson, LACE2 editing suite (including source code repository): <https://github.com/brobertson/Lace2>
Neo4j: <https://neo4j.com/>
Pelagios Network: <https://pelagios.org/>
Pleiades Gazetteer of Ancient Places: <https://pleiades.stoa.org/>

Prosopography of the Byzantine Empire (PBE): J. Martindale et al.,
Prosopography of the Byzantine Empire (2014), available at: <https://pbe.kcl.ac.uk/>

Prosopography of the Byzantine World (PBW): M. Jeffreys et al.,
Prosopography of the Byzantine World, 2016 (King's College London, 2017), available at: <http://pbw2016.kdl.kcl.ac.uk>

Standards for Networking Ancient Prosopographies (SNAP): G. Bodard, H. Cayless, M. Depauw, L. Isaksen, K.F. Lawrence, S. Rahtz, *Standards for Networking Ancient Prosopographies*, available at: <https://github.com/SNAP-DRGN/Cookbook>

The Migration of Faith: Clerical Exile in Late Antiquity, 325–600: <https://www.clericalexile.org/>

References

- [Andrews, T. \(2023\)](#) ‘The STructured Assertion Record (STAR) Model for Event-based Representation of Historical Information’ (abstract for paper at *GraphHNR 2023: Graphs and Networks in the Fourth Dimension*, Mainz 18 July–21 July 2023), 1–3.
- [Baillie, J. \(2018\)](#) ‘Alternative Database Structures for Prosopographical Research’ (abstract for paper at *ALLC/EADH Conference 2018*, 1–4.
- Baillie, J. (2021) ‘Alternative Database Structures for Prosopographical Research’, *International Journal of Humanities Data*, 15(1–2), pp. 117–32.
- [Baldwin, B. \(1982\)](#) ‘Some Addenda to the Prosopography of the Later Roman Empire’, *Historia*, 31, pp. 97–111.
- [Barnes, T.D. \(1973\)](#) ‘More Missing Names (A.D. 260–395)’, *Phoenix*, 27, pp. 135–55.
- [Bodard, G. \(2021\)](#) ‘Linked Open Data for Ancient Names and People’, in S. Bond, P.C. Dilley and R. Horne (eds) *Linked Open Data for the Ancient Mediterranean: Structures, Practices, Prospects*, ISAW Papers 20. Available at: <http://dlib.nyu.edu/awdl/isaw/isaw-papers/20-4/>.
- [Bodard, G., Cayless, H., Depauw, M., Isaksen, L., Lawrence, K.F. and Rahtz, S. \(2017\)](#) ‘Standards for Networking Ancient Person Data: Digital Approaches to Problems in Prosopographical Space’, *Digital*

Classics Online, 3.2, pp. 28–43. <https://doi.org/10.11588/dco.2017.0.37975>.

[Bradley, J. \(2017\)](#), ‘What is Factoid Prosopography all about?’ Available at: <https://www.kcl.ac.uk/factoid-prosopography/about>.

[Bradley, J. \(2026\)](#), ‘Digital Models for Prosopographical Data’, in M. Horster, R. Flower, F. Hurlet and R.W. Mathisen (eds) *Brill’s Companion to Roman Prosopography*. Leiden: Brill, pp. 130–50.

[Cayless, H. \(2019\)](#), ‘Sustaining Linked Ancient World Data’, in M. Berti (ed.) *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter, pp. 35–50.

[Destephen, S. \(2024\)](#), *Du Christianisme et des hommes dans l’Antiquité Tardive. Essais de prosopographie*. Turnhout: Brepols.

[Destephen, S. \(2026\)](#), ‘The Development of the Science of Prosopography in the 19th Century’, in M. Horster, R. Flower, F. Hurlet and R.W. Mathisen (eds) *Brill’s Companion to Roman Prosopography*. Leiden: Brill, pp. 30–42.

[Hammond, M. \(2021\)](#), ‘From Digital Prosopography to Social Network Analysis’, *Medieval People: Social Bonds, Kinship, and Networks*, 36, pp. 235–62.

[Handley, M.A. \(2005\)](#), ‘One Hundred and Fifty-two Addenda to *PLRE* from Gaul, Spain and Britain’, *Historia*, 54, pp. 93–105.

[Handley, M.A. \(2010\)](#), ‘Two Hundred and Seventy-four Addenda and Corrigenda to the *Prosopography of the Later Roman Empire* from the Latin-speaking Balkans’, *Journal of Late Antiquity*, 3.1, pp. 113–57.

[Hillner, J. and MacCarron, M. \(2026\)](#), ‘Social Network Analysis’, in M. Horster, R. Flower, F. Hurlet and R.W. Mathisen (eds) *Brill’s Companion to Roman Prosopography*. Leiden: Brill, pp. 151–81.

[Isaksen, I., Simon, R., Barker, E. and de Soto Cañamares, P. \(2014\)](#), ‘Pelagios and the Emerging Graph of Ancient World Data’, in *Web Sci’14: Proceedings of the 2014 ACM Conference on Web Science*, pp. 197–201. Available at: <http://dl.acm.org/citation.cfm?doid=2615569.2615693>.

[Jones, A.H.M., Martindale, J.R. and Morris, J. \(1971–92\)](#), *The Prosopography of the Later Roman Empire*, 3 vols. Cambridge: Cambridge University Press.

- Martindale, J.R. (1980), 'Prosopography of the Later Roman Empire: Addenda et Corrigenda to Volume I', *Historia*, 29, pp. 474–97.
- Martindale, J.R. (2003), 'The Prosopography of the Later Roman Empire, Volume I: A Memoir of the Era of A. H. M. Jones', in A. Cameron (ed.) *Fifty Years of Prosopography: The Later Roman Empire, Byzantium and Beyond*. Oxford: Oxford University Press, pp. 2–10.
- Mathisen, R.W. (1982), 'PLRE II: Suggested Addenda and Corrigenda', *Historia*, 31, pp. 364–86.
- Mathisen, R.W. (1987), 'A Survey of the Significant Addenda to *P.L.R.E.*', *Medieval Prosopography*, 8, pp. 5–36.
- Mathisen, R.W. (2003), 'The Prosopography of the Later Roman Empire: Yesterday, Today and Tomorrow', in A. Cameron (ed.) *Fifty Years of Prosopography: The Later Roman Empire, Byzantium and Beyond*. Oxford: Oxford University Press, pp. 23–40.
- Matsuo, Y., Hamasaki, M., Mori, J., Takeda, H. and Hasida, K. (2001), 'Ontological Consideration on Human Relationship Vocabulary for FOAF'. Available at: https://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/ontological_consideration_on_human_relationship_vocabulary/.
- Paradziński, A. and van 't Westeinde, J. (2025), 'The Next Frontier: A Digital Edition of PLRE', in E. Gritti (ed.) *Prosopografie 2024. Atti del Convegno Internazionale di Bergamo, 16-18 Maggio 2024*. Bergamo: Università degli studi di Bergamo, pp. 21–42.
- Rebenich, S. (1997), *Theodor Mommsen und Adolf Harnack: Wissenschaft und Politik im Berlin des ausgehenden 19. Jahrhunderts*. Berlin: De Gruyter.
- Robertson, B. (2019), 'Optical Character Recognition for Classical Philology', in M. Berti (ed.) *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter, pp. 117–36.
- Roda, S. (1980), 'Supplementi e correzioni alla *PLRE*, Vol. I', *Historia*, 29, pp. 96–105.
- Roueché, C., Cameron, A. and Nelson, J.L. (2023), 'Prosopography Meets the Digital: *PBW* and *PASE*', in J. Nyhan, G. Rockwell, S. Sinclair and A. Ortolja-Baird (eds) *On Making in the Digital Humanities: The Scholarship of Digital Humanities Development in Honour of John Bradley*. London: UCL Press, pp. 51–65.

Simon, R., Vitale, V., Kahn, R., Barker, E. and Isaksen, L. (2020)

‘Revisiting Linking Early Geospatial Documents with Recogito’, *e-Perimtron*, 14(3). Available at: <http://oro.open.ac.uk/68009/>.

Vannini, L. (2022) ‘Online Availability, Impact and Sustainability of Digital Papyrological Resources’, *Digital Classics Online*, 8, pp. 23–45.

Available at: <https://doi.org/10.11588/dco.2022.8.87562>.

Chapter 3

Venetian officials in sixteenth-century Cyprus: an online database and a pilgrimage account

Tassos Papacostas

Introduction

This chapter uses a combination of published and unpublished sources in conjunction with the *Rulers of Venice* database in order to delve into the prosopography and activities of Venetian officials serving on Cyprus in the sixteenth century. The primary aim is to highlight certain trends, some of which have been well documented beyond the island's shores, while others have been little discussed and remain obscure. In essence this is an exercise in reconstructing elements of microhistory, yet the examination of these trends obviously affects any assessment of both Venetian-ruled Cyprus and Venice's administration of its overseas dominions. At the same time, it serves as a demonstration of the type of research questions that open data and prosopographical databases encourage us to pose and help us to answer, and of how such tools facilitate and enhance scholarly engagement with our sources.

A register of Venetian officials

The *Rulers of Venice* database (RoV) was compiled by Benjamin Kohl, Andrea Mozzato and Monique O'Connell and its first version was published online in 2009 (rulersofvenice.org). It contains *c.* 70,000 records

pertaining to almost 10,000 individuals elected to public office in Venice between 1332 and 1524 ([O'Connell 2009a](#)). Its source base is limited but fairly uniform. The records concern approximately 600 offices in the city itself as well as in the territories under its control, both on the Italian mainland and overseas, from nearby Dalmatia through the Ionian and Aegean seas to distant Cyprus ([O'Connell 2009b](#): 40–42; [Knapton 2013](#)). The structured information collected in RoV helps users to reconstruct more easily the career of an individual, the list of incumbents for a particular post, or the variety of offices to which members of the same family were elected within a specific time frame. As the case studies will show, this information can be combined with other historical sources to investigate facets of the administration, society and even indirectly for the culture of Venetian territories through the identification of officials involved with architectural, artistic or scholarly endeavours.

The RoV's search interface presents users with a page headed by a keyword field, followed by twelve fields with drop-down menus containing authority lists, grouped in four sections and dealing respectively with sources, offices, office holders (names) and date, plus a larger section with a number of clickable options for the display of additional details from each record ([Plakotos 2021](#): 158). Each record is given a unique identifier.¹

The records in RoV emanate from the aforementioned registers of elections but also from other relevant sources kept today in the State Archive of Venice (RoV: Guide; [Salmini 2014](#)). The manuscript codices can only be consulted in situ, often in far from ideal conditions that at different times may include fragile material not available for inspection, individual folios damaged and virtually illegible, and most importantly, accessibility restrictions (readers may order no more than three items per day during limited opening hours when, for scholars on a short-term visit to Venice, time is of the essence). Even if some series are available in microfilm, the same accessibility restrictions apply. The user is also confronted with a variety of problems ranging from entries difficult to decipher for a non-expert in Venetian palaeography, to the use of technical terminology and questions about the overall interpretation and reliability of what was actually recorded. The database solves most of these practical problems and in many cases offers the possibility to verify the information, through direct links to the primary sources.

In what follows I will discuss some peculiar evidence from Cyprus in the early sixteenth century that revolves around Venetian officials, their succession, social networks and spiritual concerns while serving overseas. At that time the island was the wealthiest, most populous, but also most distant territory controlled by the Republic ([Skoufari 2011](#), [Arbel 2017](#)). Having been acquired in the later fifteenth century through astute diplomacy from its erstwhile rulers of the Crusader Lusignan dynasty (King James II married the Venetian Caterina Corner who succeeded him on the throne in 1474), the island was governed until its Ottoman conquest in 1570/71 by a group of Venetian officials, usually serving two-year terms. At their head was the *luogotenente* (governor) with his two *consiglieri* (councillors), all residing in the capital Nicosia and known collectively as the *reggimento*. Military affairs were in the hands of the *capitano del regno*, based at the principal port, Famagusta. Further down the hierarchy, each of the island's districts was headed by magistrates who could be either Venetian or Cypriot and bore a variety of titles (*capitano*, *castellano*, *civitano*). The three most important posts among the dozen districts, however, were reserved for Venetian patricians only, namely the *castellano* of Cerines/Kyrenia and the *capitani* of Baffo/Paphos and of Saline/modern Larnaca ([Skoufari 2011](#): 57–8). The reason was the strategic significance of the harbour or anchorage at each of these locations, and their defensive infrastructure (Kyrenia fort) or the nearby resources (Saline salt pans).

The RoV database contains *c.* 460 records of election to a dozen Cypriot offices from the 1470s to the 1520s, that is roughly from the first half-century of Venetian rule. Published lists drawn from later registers and other sources, beyond the chronological scope of RoV, fill the gap for the remaining decades and for the most important offices down to the Ottoman conquest ([Birtachas 2019](#): 369–97). The relevant RoV records concern some 250 individuals belonging to around 90 families, that is about a third of those in the relevant RoV authority list and approximately 60 per cent of the patrician clans with members active in Venetian public life in the 1520s (Sanuto, *Diarii*, 45:569–72). This suggests familiarity with and at times intimate knowledge of the distant possession through direct contact for a large proportion of the Venetian elite during a relatively short period of time. As will be shown below, the data can be dissected in order to examine how particular families engaged consistently with office holding at an

individual location such as Cyprus across generations (Monique O’Connell has expertly demonstrated this for other territories), before delving further through other sources into the realities of this state of affairs that led to matrimonial alliances, commercial interests and even in some cases full integration into the local elite ([O’Connell 2009b](#): 51–56; [O’Connell 2013](#); [Papacostas 2022a](#): 192–95). In fact, the RoV data render feasible a number of fruitful investigations that shed welcome light on contemporary society and which would have otherwise been too onerous if not impossible to undertake without the aggregate information generated by RoV searches. What will be attempted below is by necessity limited in scope, employing a couple of case studies to illustrate certain developments and patterns within the specific Cypriot context and using a particular text as its starting point.

The lure of the Holy Land and of holy relics

On 24 July 1516, a group of Venetians set out from Kyrenia on the north coast of Cyprus for a journey that would take them across the island and then over the sea to Christianity’s holiest shrine, the Holy Sepulchre in Jerusalem. At the head of the travelling party was Alvise Contarini di Galeazzo, the outgoing *castellano* of Kyrenia, accompanied by his wife Maria Moro di Alvise (Barbaro, *Arbori*, 18:487). The unpublished Italian account that the *castellano* compiled constitutes the source of most of what we know about their journey and will serve as the connecting thread for the discussion that follows (Contarini, *Account*, ff.1r–30r).² Before dissecting this account, however, a few words about pilgrimage within and from Cyprus in this period are in order.

Such pious endeavours, beyond the scope of professional duties, were not uncommon and Venetians serving and other Italians living temporarily on the island often took advantage of their proximity to the holy sites on the Muslim-ruled mainland to undertake the ultimate pilgrimage ([Skoufari 2011](#), 126–27). Among a number of documented examples there is the *consigliere* Marcantonio Calbo di Girolamo, who sought permission in February 1534 from the authorities in the metropolis to go away on pilgrimage, only two months after he arrived in Cyprus (usually such trips were undertaken at the end of service); he must have completed his journey

in the late summer of 1534 before returning to his duties in Nicosia ([Birtachas 2019](#): 383 no.45; Aristeidou, *Documents*, 4:168–9 no.82; Barbaro, *Arbori*, 18: 202). The *capitano* of Saline, Marino Gradenigo di Zacharia, crossed to Tripoli at the end of April 1553, visited Aleppo and Antioch before heading south to Jerusalem, and returned to Tripoli for the journey back to Cyprus after a three-month absence ([Skoufari 2011](#): 127; Gradenigo, *Itinerarium*).³ The *cancelliere* of the Latin archbishopric of Nicosia Giovanni Battista Perotto undertook the journey in 1561 to fulfil a vow and compiled a detailed account of his peregrinations. He arrived in Jerusalem during the Easter celebrations and returned to Nicosia shortly before embarking for Venice ([Skoufari 2011](#): 127; Perotto, *Memoria*; [Harkas and Schabel 2026](#): 121-9). As he notes in his account, he was joined on his pilgrimage by Luca Foscarini and Luigi (Alvise) Bragadin. While Bragadin may be the illegitimate son of the late Girolamo Bragadin, procurator of San Marco, Foscarini is almost certainly to be identified with yet another official serving on Cyprus, namely the homonymous *camerlengo*, whose tenure was coming to an end in April 1561 (Elezioni MC, reg.3, ff.168v–169r; [Harkas and Schabel 2026](#): 85-6).⁴ The erudite *cancelliere* of the Venetian administration Vincenzo Fedeli (1551–57), who also completed his pilgrimage at the end of his tenure in Nicosia, stands out for good reason: while in Jerusalem he obtained a fragment of what was believed to be the Column of the Flagellation, which he donated to the Republic upon his return. It was kept in the relic collection of San Marco until it was eventually placed in the chapel of Saint Isidore, attached to the north cross arm of the ducal church ([Targhetta 1995](#); Ongania, *Documenti*, 216 no.876, 288).⁵

Fedeli was of course not alone among officials returning to the metropolis with a precious relic, replicating time and again the event that was at the origin and heart of Venice's Christian identity, namely the translation of Saint Mark's body from Alexandria back in the ninth century. The *consigliere* Nicolò Michiel di Francesco, whose tenure in Nicosia overlapped with that of Alvise Contarini in Kyrenia, returned home with parts of the relic of Saint Hilarion and a fragment of the True Cross (rulersofvenice.org, record 45,647; [Birtachas 2019](#): 381 no.27; Sanuto, *Diarii*, 17:372; [Giuriato 1871](#): 2:37). These were presumably acquired on the island, however, rather than in the Holy Land. Although the fourth-

century Palestinian hermit's body was returned to the mainland from Cyprus soon after his death near Paphos, late medieval and later sources imply that it was still venerated on Cyprus (Donà, *Memorie*, f.154r). This was the result of confusion with a much later local holy man of the same name, whose cult was associated with a mountaintop site overlooking Kyrenia. Known as Saint Hilarion and crowned by an imposing fortress since at least the Comnenian era (late eleventh/twelfth century), the site may have initially hosted the younger Hilarion's abode and subsequently a monastic community (*iBCC* 57; [Bakirtzis 2022](#): 111–16). What's more, we know that the *consigliere* did in fact inspect the fortress during his tour of defensive structures on the island in early 1516, by which time most of its buildings lay in ruins (Aristeidou, *Documents*, 2:323 no.166; [Arbel 2023](#)). His relic acquisition may relate to this very visit and to likely contacts with not only Latin but also Greek monasteries and churches where such relics were kept, showing that the Holy Land did not hold a monopoly on their provision.

Evidence for the *consigliere*'s familiarity with Cypriot shrines comes from two almost identical manuscript notes in Latin, among miscellaneous Greek notes at the beginning of a thirteenth-century codex that contains various works by Maximos the Confessor (Jerusalem Saba 234 (*Diktyon* 34,490), f.2v; [Papadopoulos-Kerameus 1891](#)–1915: 2:365 no.234; [Darrouzès 1957](#): 147). Wherever the Greek manuscript was copied, it was certainly on Cyprus by the mid-fifteenth century, when it was donated to the church of the Holy Cross Mesokēpou in the hinterland of Limassol. Coincidentally, this church was a Venetian foundation from a much earlier period, erected by a merchant from the lagoon resident on the island in the late twelfth-century (*iBCC* 63). Whether in the sixteenth century there was still any residual memory or visible reminder of this early Venetian link we cannot know. The undated manuscript notes give the names of all three members of the *reggimento*, indicating that presumably one of them recorded their visit to the shrine where the manuscript was kept. They are the *luogotenente* Donato Marcello di Antonio and the *consiglieri* Giovanni Dolfin di Daniele and Nicolò Michiel di Francesco ('In nomine sancte et individue trinitatis amen. Nos Donatus Marcellus pro illustrissimo et excellentissimo duchallis dominio Ioannes Dolphinus et Nicholaus Michael consilliarrii Regni Cypri': [Figure 3.1](#)).⁶ The relevant records in RoV show that they overlapped in

office between late 1514 and early 1516 (rulersofvenice.org, records 45,610, 52,018, 45,648, 45,647; [Birtachas 2019](#): 373 no.13, 381 no.27 and 28; Barbaro, *Arbori*, 20:483, 19:272, 21:130). Nicolò Michiel is of course the aforementioned donor of relics of Hilarion and of the True Cross. He may be the one who showed particular interest in the Greek manuscript, for we know that he was an erudite man, possibly with a knowledge of Greek.⁷ Where exactly the three officials inspected this manuscript and added their names, however, remains uncertain. Considering that Venetian magistrates had to seek permission to spend the night outside their seat ([Szépe 2018](#): 158), it seems unlikely that all three would have been absent from Nicosia simultaneously and, moreover, at such a relatively distant location as the Holy Cross Mesokēpou shrine ([Figure 3.2](#)). They may have seen the manuscript in a Greek church or monastery within or near Nicosia instead (other undated Greek notes mention monks and a hegumen).

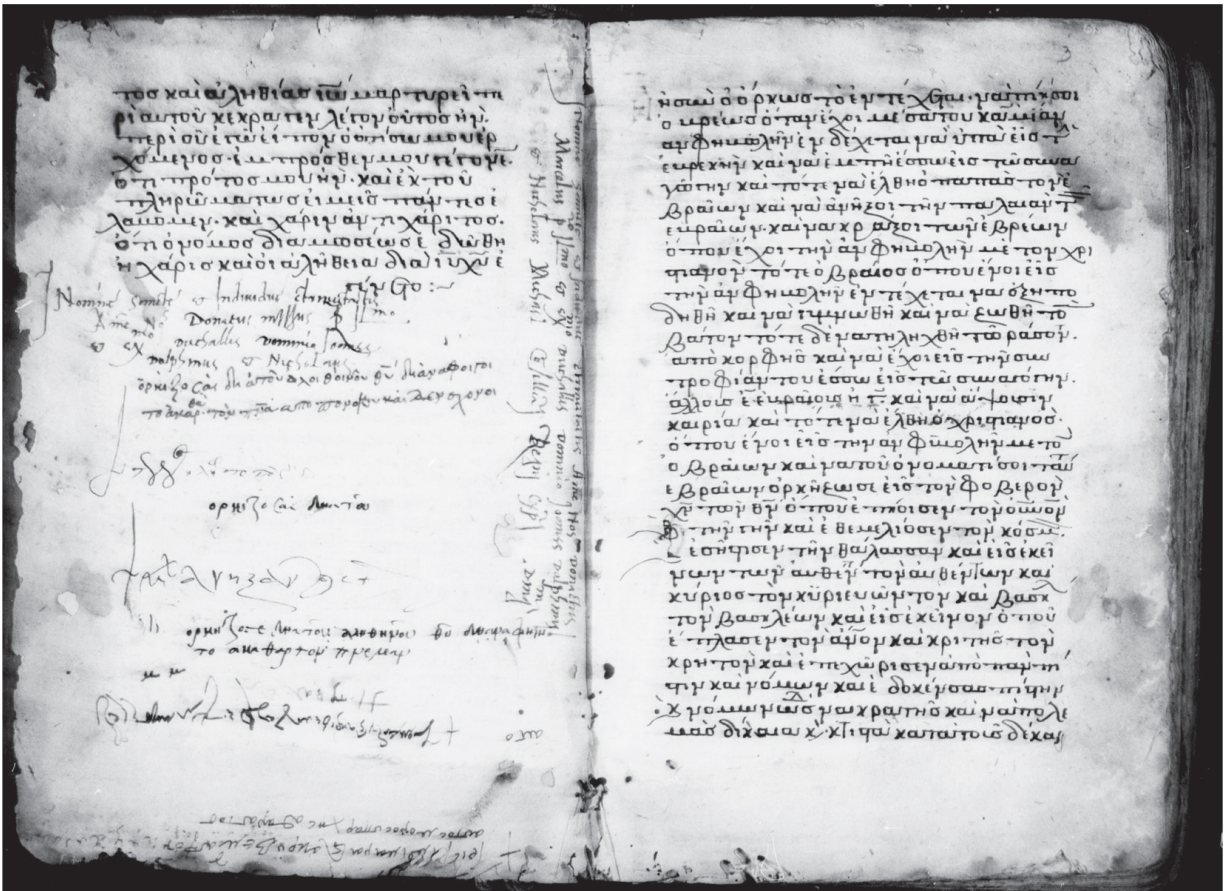


Figure 3.1: Notes in Latin recording the names of sixteenth-century Venetian officials, Jer. Saba 234, f.2v (Library of Congress Collection of Manuscripts in the Greek Orthodox Patriarchate of Jerusalem. Public domain)

Be that as it may, what matters for the present discussion is that Nicolò Michiel's documented relic collecting is matched by evidence of his engagement with local shrines that may indicate channels through which he acquired such prized articles. Although attestations for this type of engagement remain scarce, there are other examples. Another note in a twelfth-century manuscript, this time added in Greek by the hegumen of an unknown monastery (in Nicosia?), appears to record a visit by two Venetians ('λωβυζου ταρμεριου και φραζεσκου μαρτζελλου ... ιερομοναχος και καθηγουμενος της μεγαλης μονης (?): Par.gr. 632 (*Diktyon* 50,213), f.63r; [Darrouzès 1950](#): 178).⁸ A quick search through RoV reveals that they were the *luogotenente* Alvise d'Armer di Simone and the *consigliere* Francesco Marcello di Andrea. Their visit surely took place during their overlapping tenures in Nicosia that the RoV records allow us to place

between mid-1518 and mid-1520 (rulersofvenice.org, records 52,020, 45,613, 49,641, 45,652; [Birtachas 2019](#): 373 no.15, 382 no.33). Alvise d'Armer was no stranger to Nicosia, for he had served there as *consigliere* one decade earlier (rulersofvenice.org, record 45,640; [Birtachas 2019](#): 381 no.22). Francesco Marcello was even better acquainted with Cyprus for he may have grown up on the island, if his father is Andrea Marcello di Antonio. Andrea was among the eighty-eight Venetians selected in 1478 to settle there with their families (Navagero, *Historia*, 1150; [Skoufari 2011](#): 55) and although very few actually made the move from Venice, he was one of them (Barbaro, *Arbori*, 20:466).⁹ He subsequently served as *camerlengo* (treasurer, elected in 1494: rulersofvenice.org, records 45,719, 32,869) and his son Girolamo, presumably the 1518 *consigliere*'s brother, married a Cypriot and was later on elected himself *consigliere*, in 1526 ([Birtachas 2019](#): 382 no.42).¹⁰ Through these multiple links, partly illustrated through the RoV records, Francesco's family was particularly well acquainted with Cyprus and, one assumes, with its holy sites and monasteries as well, including the one he visited in the company of the *luogotenente*. Such visits of course did not necessarily entail spiritual motives, for very often Venetian officials were called to resolve disputes, control revenues and claim tax arrears, or were on regular tours updating tax registers and inspecting military installations and other elements of infrastructure. If the opportunity to acquire relics arose during such visitations, they may well have taken it up.

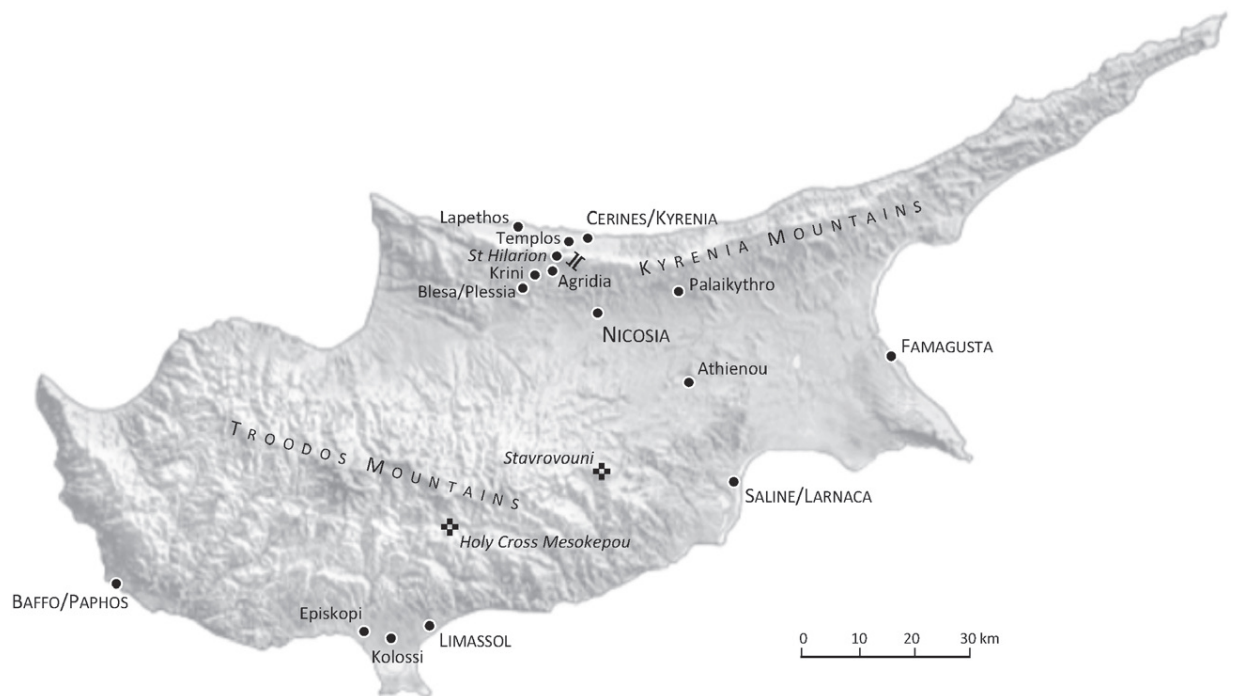


Figure 3.2: Map of Cyprus (Tassos Papacostas)

At the end of his service in late 1516, Nicolò Michiel loaded his belongings on the *Nicolosa* in order to send them back to Venice. His fellow *consigliere* Giovanni Dolfin did the same but the two men must have boarded a different vessel for their return journey. While sailing through the Aegean, the *Nicolosa* was attacked in the waters of Milos, the cargo was looted and both officials lost their possessions (Sanuto, *Diarii*, 24:11–12; [Gluzman 2023](#): 103). Yet the precious relics were clearly not part of the lost cargo, their owner probably having wisely kept them with him. Upon his return to Venice, he duly donated them to San Giorgio Maggiore ([Giuriato 1871](#): 2:37). No doubt there must have been many others whose relic acquisition and pious travels have not been recorded or are documented in archival sources awaiting scrutiny. The pilgrimage of the *castellano* of Kyrenia in 1516 with which I began this investigation is not only particularly well documented through his aforementioned account, but it is exceptionally valuable for one specific reason: its author describes in some detail the trajectory through Cyprus, during both legs of the journey. This, in turn, sheds much welcome and exceptionally rare light on the links between a Venetian magistrate in post far away from the metropolis and members of the local elite.

Across Cyprus in the summer of 1516

Alvise Contarini had taken up his post at Kyrenia two years earlier, having been elected in July 1514 (rulersofvenice.org, record 46,271).¹¹ Around forty years of age at the time, he had already served in a variety of offices in Venice and its mainland territories. Indeed, the records in RoV allow us to follow his career for about twenty years, from 1499 when he was first elected to public office (in the Mint of Venice) down to 1519 (RoV coverage stops thereafter: rulersofvenice.org, records 41,008, 35,397, 43,829, 26,870, 38,688, 37,549, 39,806).¹² Cyprus appears to have been his first appointment beyond the Gulf of Venice and at Kyrenia he oversaw the continuation of work on the town's crumbling fortifications, for which he was commended warmly by his superiors (Sanuto, *Diarii*, 22:41; Aristeidou, *Documents*, 2:322–23 no.166). Early in 1516, having already decided to undertake a pilgrimage to Jerusalem in order to fulfil a vow before returning to Venice at the end of his magistracy in the summer, he requested the necessary permission from the authorities in the metropolis. It was granted on 6 April, when it was also duly noted that since it was unlikely that his successor would arrive in time for the handover, another Venetian noble would need to fulfil temporarily his duties. The successor in question was Antonio Malipiero di Domenico, elected on 30 March, one week before the permission was granted to Contarini (rulersofvenice.org, record 46,272; Sanuto, *Diarii*, 22:91, 22:112); he was not expected to reach Cyprus until well into the summer. In his account Contarini reveals that the temporary substitute was Daniele Dolfín di Giovanni, without elaborating further (Contarini, *Account*, f.1r). The latter is almost certainly to be identified with the son of one of the two *consiglieri* serving at that time in Nicosia, Giovanni Dolfín di Daniele, who would no doubt keep an eye on his inexperienced son, probably in his early twenties at that time (rulersofvenice.org, record 45,648; [Birtachas 2019](#): 381 no.28; Barbaro, *Arbori*, 19:272).

Alvise and his entourage travelled across the mountain pass below Saint Hilarion into the central plain of Cyprus, heading towards the south coast whence they intended to sail away ([Figure 3.2](#)). By nightfall on the first day (24 July 1516) they reached the outskirts of Nicosia and spent the next few

days at the house of Zaco Contostefano, three Venetian miles (*c.* 5km) from the city. Contostefano was a wealthy member of the local elite, of noble status, who also owned since 1511 an estate (casal Blesa/Plessia) near the south approaches to the aforementioned mountain pass and was soon to secure the lease of the village of Lapethos near the north coast, not far from Kyrenia ([Arbel 1989](#): 187; BMCVe, ms. Donà dalle Rose 215, f.131v; ASVe, *Capi del Consiglio di dieci*, Dispacci (lettere) dei rettori e pubblici rappresentanti, 288 no.195). That the *castellano* maintained such friendly relations with a local aristocrat is noteworthy, as we shall see below, and may have played a role in Contostefano's successful Lapethos bid.

Our pilgrims' next stop was the mountaintop shrine of the Holy Cross where, as the *castellano* notes in his account that demonstrates his awareness of local traditions, Saint Helena had left a particle of the True Cross on her way back from Jerusalem (Contarini, *Account*, f.1v). This refers to Stavrovouni, the most prominent goal of pilgrimage on Cyprus in this period, not least because of its proximity to and visibility from Saline, one of the busiest anchorages of the south coast (Grivaud, *Excerpta*; [Papacostas 2022b](#)). It was popular well beyond the local population, among foreign pilgrims sailing back and forth from the Holy Land and of course among Venetians. Back in 1474, Pietro Bembo, a Venetian in the service of Queen Caterina Corner who had settled and married on Cyprus, was similarly attracted to the shrine but had to cut short his journey from Nicosia due to the volatile political situation in the wake of the death of King James II (Boustronios, *Diegesis*, 167; Barbaro, *Arbori*, 17:341). In 1516, a few months after the Kyrenia *castellano*'s visit, it was utter misfortune that brought to Stavrovouni another group of Venetians: survivors of a shipwreck in the waters between Alexandria and Cyprus just before Christmas of that year, they witnessed their fellow passengers drown or die of hunger and thirst during the week they subsequently spent on a boat drifting in Levantine waters until it reached the southeast coast of Cyprus; they made a vow to the True Cross and duly presented the boat that saved their lives to the shrine (Sanuto, *Diarii*, 24:23–25). The pious *castellano*'s visit was surely intended to secure divine protection against precisely this kind of misfortune.

The next station in the party's progress was Limassol, where they arrived on 30 July and lodged with a certain Giacomo Corner. He is

probably to be identified with the homonymous administrator of the Grand Commandery of the Hospitaller Order at nearby Kolossi (*governator*, attested in office between 1507 and 1517: [Phillips 2021](#): 84, [Coureas 2016](#): 119).¹³ Alvise Contarini may have known Giacomo Corner through his contacts with other officials of the Order in the region of Kyrenia, as we shall see below, and he took advantage of his protracted stopover in Limassol to visit Kolossi, no doubt accompanied by his host. He also toured the adjacent village of Episkopi with its sugar cane plantations, where the valuable commodity had been produced and exported since the fourteenth century by another branch of his host's family, the Corner Piscopia ([Ouerfelli 2008](#): 114–21).¹⁴ When the long awaited Venetian pilgrim galley finally dropped anchor in the bay of Limassol, the *castellano* and his travelling companions embarked for the seaborne leg of their journey on 15 August and headed for Jaffa, the most common point of arrival for pilgrims along the Syro-Palestinian coast in this period.

After a month spent touring and worshipping in the shrines of the Holy Land, including Bethlehem and Bethany as well as Jerusalem, Alvise Contarini and his fellow pilgrims returned to Cyprus on 18 September 1516. This time they disembarked at Saline and, just as at the start of their journey, they were offered hospitality for their first night by a local aristocrat, Giovanni Costanzo (Contarini, *Account*, f.29v), who belonged to a well-known family of south Italian origin, established on the island since the 1460s ([Papacostas 2022a](#): 175–6, 189–92, 196–7). The *castellano* most probably had met his host through the latter's brother Bruto Muzio Costanzo, who was commander of Templos on the foothills very near Kyrenia, the smallest of the three Hospitaller commanderies on Cyprus ([Coureas 2016](#): 123).¹⁵ Tuzio, the Costanzo brothers' father, owned estates at Krini and Agridia, adjacent to the Contostefano properties by the southern entrance to the mountain pass above Kyrenia (and below Saint Hilarion), also within the area of jurisdiction of the *castellano* (Bustron, *Historia*, 418). Although Tuzio lived in Italy, his Cypriot estates were managed by his son Giovanni and this may have furnished additional opportunities for contacts with Alvise Contarini in his official capacity, perhaps explaining the choice of temporary lodgings upon the pilgrims' return to Cyprus. The friendly relations intimated by the *castellano*'s overnight stay with Giovanni Costanzo, just as with Zaco Contostefano

earlier on, confirm that such ties between Venetian officials and the local aristocracy were common, whether based on familial, professional or other links. Although this is hardly surprising, rather unusually here we have the details of the individuals involved and precious concrete evidence of a pattern that troubled the authorities, as we shall see below.

The *castellano* does not name the location of the Costanzo property where the pilgrims spent their first night on Cypriot soil. Giovanni and his family owned numerous estates and villages, most of them along a west-east axis running halfway between Saline and Nicosia and including Athienou (Boustronios, *Diegesis*, 109; Bustron, *Historia*, 408, 418; [Patapiou 2013/14](#): 157–62). The latter is the most likely candidate for the travelling party's stopover, for not only was it known in later centuries as an important station along the road from Larnaca (Saline) to the capital (Mariti, *Viaggi*, 85; Cobham, *Excerpta*, 436, 446) but it also preserves one of the very few material testimonies of the Costanzo presence on Cyprus, namely a painted panel from c. 1500 with the family coat-of-arms ([Patapiou 2013/14](#): 167–9). The following day the pilgrims left their host and reached Nicosia. In the capital they visited the *consigliere* Giovanni Dolfìn, likely father of Contarini's temporary replacement at Kyrenia, who no doubt will have briefed the *castellano* about any developments during his absence. Later on the same day (19 September), the travelling group entered Kyrenia, almost two months after their departure.

Dying far from home and the temptation of overseas posts

By this time Contarini's successor Antonio Malipiero, elected some six months earlier, must have arrived on the island, allowing for a smooth handover. Yet the biennial succession of officials was not always straightforward. The new incumbent died halfway through his tenure at Kyrenia in November 1517 (Sanuto, *Diarii*, 25:210). Both earlier and later *castellani* are known to have met the same fate, including Antonio Badoer di Marino in 1512 and Antonio Gritti di Francesco in 1526 (rulersofvenice.org, record 46,269; Sanuto, *Diarii*, 15:258, 41:470–71; Barbaro, *Arbori*, 20:171). Mortality while in office was of course not

restricted to those serving at Kyrenia, who resided after all in one of the island's most salubrious districts. According to my rough estimate based on information from Marino Sanuto's *Diaries* and circumstantial evidence from other sources, as many as one in five of all officials sent from the metropolis to Cyprus died in the course of their tenure on the island, succumbing to disease, epidemics or even the tribulations of the long journey and the island's frequently unforgiving climate. The standard procedure in such cases was to appoint as temporary replacement one of the Venetian officials already in post in the territory until the arrival of the new appointee. This could entail a substantial wait, considering that it usually took on average between six and eight weeks for news from Cyprus to reach Venice, and even longer for the election to take place, the official to embark on the journey and to finally take up his post.

In the case of Malipiero there is some discrepancy in the sources concerning what happened after his untimely death. While Marino Sanuto notes that the *castellano* of Famagusta Nicolò Bragadin di Vittore was swiftly dispatched to ensure a degree of continuity in the affairs of Kyrenia (November 1517) with half the salary of the deceased (the other half going to the latter's grieving family), the election records in RoV imply that as soon as news reached Venice in January 1518, an appointment was made in the person of Pietro Malipiero di Antonio, almost certainly the late incumbent's own son who must have been present on the island with his father (rulersofvenice.org, record 46,277; Sanuto, *Diarii*, 25:210, 40:708).¹⁶ Then, in April of the same year, the election of Girolamo Barbaro di Leonardo is recorded, although presumably he did not reach Kyrenia until well into the summer (rulersofvenice.org, record 46,273).¹⁷ It is likely that the late *castellano*'s son was formally acknowledged as his short-term successor in order to allow Bragadin to return to his duties in Famagusta while awaiting the election and arrival from Venice of a new *castellano* for a full two-year term.

Despite the trauma of his father's death in Cyprus, Pietro Malipiero was subsequently keen to return to the island. He was clearly following in his late father's footsteps, whose career can be partly reconstructed once more through the RoV records: he had served twice at Famagusta well before his Kyrenia appointment (elected *collaterale* in 1497 and then *castellano* in 1502: rulersofvenice.org, records 45,684, 36,253, 45,772, 35,479). Pietro's

first attempt was thwarted when in 1526 he bid unsuccessfully for the position of *capitano* of Paphos, but a few days later he managed to secure his father's last magistrature, at Kyrenia, where he held office from September 1528 for two full years (Sanuto, *Diarii*, 42:427, 42:429, 42:588, 42:589; Elezioni MC, reg.1, ff.171v–172r; Barbaro, *Arbori*, 20:396).

The vicissitudes of the succession in a minor office far away from Venice, glimpsed through the examples discussed above, are in no way atypical of what was happening all over the Venetian maritime state. The only difference here is that the distance slowed down even further the requisite processes and forced the Republic to rely even more on the preparedness and shrewdness of its local representatives to act swiftly and implement the necessary temporary measures. It also illustrates most eloquently the already well documented trends I alluded to above. Venetian magistrates took their families with them, forcing the authorities later on to vote through legislation in an attempt to restrict the abuse of this practice that often involved more than just the close relatives of the elected magistrate ([O'Connell 2009b](#): 61; [Fortini Brown 2017](#): 62). The officials, as well as their family members, would often develop links with the territory where they served and attempted, often successfully, to return whenever the opportunity arose, having already established useful social networks, lucrative financial interests and even beneficial familial connections. These were not necessarily detrimental to good governance, a constant concern of the Republic, for the collective experience gathered within a family, as in the cases of both Francesco Marcello and Pietro Malipiero, must have served well its members in the performance of their duties ([O'Connell 2009b](#); [O'Connell 2013](#)). Yet, as we also saw, in many instances the bond with the territory turned out to be less advantageous and more permanent than anticipated for those who died in office.

An appraisal

The unique evidence for Alvise Contarini's social circle, illustrated through his close relations with members of the Cypriot elite, brings into sharp relief yet another unavoidable characteristic of office holding. Once more the central government attempted in vain to ensure that its magistrates

remained objective in their exercise of power by restricting to the official sphere their contacts with the local population and especially with members of the elite, prohibiting for example officials from entertaining locals at home ([Fortini Brown 2017](#): 62; [Szépe 2018](#): 158). As usual, such measures were ineffective and overridden by the realities on the ground. The Republic did after all seek the support of these very same elites to buttress and consolidate its rule over their territory, and Cyprus was certainly no exception to this. It was no exception either to the common pattern of office holding by several members of the same family in the same location, as noted above, although never simultaneously. By transcribing and making available the historical documents, and identifying key information (such as names, places, dates and relationships), the RoV database allows us to trace trends very clearly, to follow their development across time and to compare data about offices at different levels, across territories throughout the Venetian empire, and families across the patrician class.

As we have also seen, holding office so close to the Holy Land prompted many Venetians to go on pilgrimage to Jerusalem. Relics were acquired both there and on Cyprus itself. They enriched the already overflowing treasuries and churches of the metropolis, demonstrating the continuity and vibrancy of a pattern established centuries earlier. Those acquired on the island suggest that Venetian officials were well acquainted with local monasteries and churches through the carrying out of their various duties and perhaps after serving repeatedly on the island. Once more RoV affords a good picture of the latter trend. Its search tools allow users to gather information about the careers of both individuals and members of the same family and to trace patterns across time for the same location. Such an investigation would have required more than one expedition to the Venetian archives to gather relevant information; RoV makes the latter infinitely more accessible in terms of availability and legibility. Beyond the improved access, RoV's search tools enable other types of investigation in any combination of location, office, family or individual, illustrating most vividly the benefits of open digital resources.

Unlike other digital prosopographical projects like *Open Seals*, *Connecting Late Antiquities* or *Prosopography of the Byzantine World*,¹⁸ RoV was not conceived from the outset with the FAIR principles for open data in mind ([Wilkinson et al. 2016](#)). Data collection had started already in

2000 using FileMaker. As a result, the RoV data is not available for analysis and querying in any environment other than the current website, which leads to limitations in search and findability as well as interoperability and reusability. An online search for the name of an official will not lead to any records in RoV; only a user who is aware of its existence and queries its own search page will be able to locate relevant records. The provenance of the data, that is the specific archival source, is clear and can be verified often through the ‘Transcriptions’ tab (top and bottom of the RoV interface) that gives access to scans of election registers and their transcription, although these are not linked to the records (RoV: Transcriptions). Despite these limitations, the case studies in the chapter illustrate how a scholarly, historical database can not only facilitate research but also encourage users to ask new and consequential questions of their primary material.

Acknowledgements

I am grateful to the editors for their invitation to contribute to this volume and to Stuart Dunn, Maria Parani and Alessio Sopracasa for their useful suggestions. This essay is dedicated to Charlotte Roueché, an enthusiastic creator and promoter of online tools, as a small token of gratitude and appreciation.

Notes

1. Despite its fundamental role for referencing purposes, a search by identifier is not currently available in RoV.

2. Short extract published in [Cicogna 1824](#)–61: 6.2:651.

3. Gradenigo was in office from 6 May 1552 to 5 January 1555: Elezioni MC, reg.2, ff.188v–189r (not to be confused with Marino Gradenigo di Pietro, who held the same office in 1561–64: Elezioni MC, reg.4, ff.178v–179r). His journey was therefore undertaken halfway through his tenure. He served again briefly as *luogotenente* on Cyprus, where he died in 1565: Elezioni MC, reg.4, ff.177v–178r, and BNM, It. VII, 153 (8143), f.103R.

4. Alvisè Bragadin, who is known to have owned a particle of the True Cross and is referred to as ‘magnifico’ by Perotto (i.e. of noble status), is more difficult to identify with certainty but he may be the homonymous son (also called Zuan Alvisè in some sources) of Girolamo Bragadin: ASVe, *Notarile*, Testamenti (Marcantonio Cavanis), 195 no.596; ASVe, *S. Francesco della Vigna*, Atti 3.2, ff.37r–41r; [Fortini Brown 2004](#): 92. This identification is strengthened by the fact that the travelling party on their way back to Nicosia made a stopover at Palaikythro, a property of the illegitimate Alvisè’s stepbrothers (through their Cypriot mother Caterina Negron).

5. A fragment of the same relic may have been kept in Nicosia as well: Grivaud, *Excerpta*, 86; [Bacci 2014](#): 240.
6. 'In the name of the holy and indivisible Trinity, amen. We, Donatus Marcelus for the most illustrious and most excellent ducal dominion, Ioannes Dolphinus and Nicholaus Michael councillors of the kingdom of Cyprus.'
7. He is mentioned as a 'dotor/doctor' in the sources, customarily reserved for men of letters (Sanuto, *Diarii*, 24:12, 24:24; Aristeidou, *Documents*, 2:295 no.157, 300 no.158, 311 no.162) and may also be identical with the 'scrittore' Nicolò Michiel di Francesco di Angelo di Francesco (Barbaro, *Arbori*, 21:130).
8. 'Of Lobyzos Tarmerios and Phrazeskos Martzellos ... hieromonk [ordained monk] and hegumen [abbot] of the Great Monastery (?).'
9. Another Andrea Marcello, di Fantino, also appears in the sources with Cypriot links (elected *castellano* of Famagusta in 1492: [rulersofvenice.org](#), record 32,890) but he seems less likely to be Francesco's father, as he is not known to have had male issue: Barbaro, *Arbori*, 20:463.
10. The bride was Caterina Crespi, daughter of 'Domino Nicolò cipriotto' (1496: Barbaro, *Arbori*, 20:466), a Florentine established on Cyprus (Loredan, *Libro di nozze*, f.111r).
11. The year is mistakenly given as 1515 in RoV; see the relevant register folio: https://www.fulcrum.org/concern/file_sets/rj430701d
12. For a reconstruction of his career down to his death in 1557, see [Baiocchi \(1983\)](#).
13. Several homonymi are recorded in the sources in this period, including a Giacomo Corner active on Cyprus: [Papacostas \(2022a\)](#): 192).
14. Despite spending a fortnight in Limassol, Contarini makes no reference to his colleague in charge of local affairs, the *capitano* of Limassol. The incumbent at that time must have been Jason Lusignan, from a branch of the former royal family and father of the historian Etienne de Lusignan (Lusignano, *Chorografia*, 79a).
15. Bruto Muzio was appointed in 1494 ([Anderson 1973](#): 296), confirmed in possession of Templos in 1514 ([Phillips 2021](#): 80), and died in 1547 (ASVe, *Senato*, Deliberazioni, Mar, filza 4, f.146r); see also [Birtachas \(2019\)](#): 93, 112).
16. Bragadin's family maintained close links with Cyprus: [Papacostas \(2022a\)](#): 26).
17. He may be identical with the *capitano* of Paphos Girolamo Barbaro di Leonardo, elected in 1504: [rulersofvenice.org](#), record 46,461; Barbaro, *Arbori*, 17:209.
18. *Open Seals*: see Filosa, Shea and Sopracasa 2026 ([Chapter 11](#) in this volume); *Connecting Late Antiquities*: see Tupman and Flower 2026 ([Chapter 2](#) in this volume); *Prosopography of the Byzantine World*: <https://pbw2016.kdl.kcl.ac.uk>, also discussed in Andrews 2026 ([Chapter 1](#) in this volume).

Bibliography

Online resources

DBI: Dizionario biografico degli Italiani, ed. A.M. Ghisalberti and M. Pavan, 100 vols. Rome: Istituto dell'Enciclopedia Italiana, 1960–2020.
Available at: <https://www.treccani.it/biografico/>

Diktyon: Pinakes / Πίνακες. Textes et manuscrits grecs. Available at: <https://pinakes.irht.cnrs.fr>

iBCC: T. Papacostas, *Inventory of Byzantine Churches on Cyprus*. London: King's College London, 2015. Available at: <http://ibcc.digium.kcl.ac.uk>

PBW: M. Jeffreys et al., *Prosopography of the Byzantine World*, 2016.

London: King's College London, 2017. Available at: <https://pbw2016.kdl.kcl.ac.uk>

RoV: rulersofvenice.org: B.G. Kohl, A. Mozzato and M. O'Connell, 'The Rulers of Venice, 1332–1524'. Available at: <http://rulersofvenice.org>

Abbreviations

ASVe: Archivio di stato di Venezia

BMCVe: Biblioteca del Museo Correr, Venice

BNM: Biblioteca nazionale marciana, Venice

Archival sources

ASVe, *Notarile*, Testamenti (Marcantonio Cavanis), 195.

ASVe, *Capi del Consiglio di dieci*, Dispacci (lettere) dei rettori e pubblici rappresentanti, 288.

ASVe, *S. Francesco della Vigna*, Atti.

ASVe, *Senato*, Deliberazioni, Mar, filza 4.

BMCVe, ms. Donà dalle Rose 215.

BNM, It. VII, 153 (8143).

Barbaro, *Arbori*: ASVe, *Miscellanea codici*, s.I: storia veneta, nn17–23

[Marcantonio Barbaro, *Arbori de' patritii veneti*]. Available at: <https://asve.arianna4.cloud/ricerca?query=barbaro%20genealogie&orderby=score&orderdirection=DESC&page=1&limit=10>

Contarini, *Account*: BNM, It. VI, 179 (6350) [untitled pilgrimage account by Alvise Contarini].

Donà, *Memorie*: BMCVe, ms. Donà dalle Rose 45 [ff.143r–175r: Leonardo Donà, *Memorie per le cose di Cipro*].

Elezioni MC: ASVe, *Segretario alle voci*, Elezioni in maggior consiglio, registri 1–4.

Gradenigo, *Itinerarium*: BMCVe, ms. Cicogna 796 [*Itinerarium clarissimi domini Marini Gradonici praefecti et capitani Salinarum in Insula*

- Cypro].
- Loredan, *Libro di nozze*: BNM, It. VII, 538 (7734) [Girolamo Loredan, *Libro di nozze e origine delle famiglie nobili*].
- Perotto, *Memoria*: BMCVe, ms. Correr 1208/2657 (= Misc. Correr LXXIX 2657) [*Memoria del viaggio fatto per me pre Giambattista Perotto padovano da Cipro in Gierusalemme, et del rittorno alla patria l'anno 1561*].

Primary sources

- Aristeidou, *Documents*: Aristeidou, A. (1990–2003) *Ανέκδοτα έγγραφα της κυπριακής ιστορίας από το κρατικό αρχείο της Βενετίας*, 4 vols. Nicosia: Cyprus Research Centre.
- Boustronios, *Diegesis*: Kechagioglou, G. (1997) *Τζώρτζης Μπουστρός, Διήγησις Κρονίκας Κύπρου*. Nicosia: Cyprus Research Centre.
- Bustron, *Historia*: de Mas Latrie, R. (1886) *Florio Bustron, Chronique de l'île de Chypre*. Paris: Imprimerie nationale (repr. as *Historia over Commentarii di Cipro*, Nicosia: Archbishop Makarios III Foundation, 1998).
- Cobham, *Excerpta*: Cobham, C.D. (1908) *Excerpta Cypria. Materials for a History of Cyprus*. Cambridge: Cambridge University Press.
- Grivaud, *Excerpta*: Grivaud, G. (1990) *Excerpta Cypria Nova, I. Voyageurs occidentaux à Chypre au XV^{ème} siècle*. Nicosia: Cyprus Research Centre.
- Lusignano, *Chorograffia*: Lusignano, S. (1573) *Chorograffia e breve historia universale dell'isola de Cipro principiando al tempo di Noè per il fino al 1572*. Bologna: Alessandro Benaccio.
- Mariti, *Viaggi*: Mariti, G. (1769) *Viaggi per l'isola di Cipro e per la Soria e la Palestina*. Lucca: Jacopo Giusti.
- Navagero, *Historia*: Navagero, A. (1733) *Historia Veneta*, in Muratori, L.A. (ed.), *Rerum italicarum scriptores*, vol.23. Milan: ex typographia Societatis Palatinae.
- Ongania, *Documenti*: Ongania, F. (1886) *Documenti per la storia dell' augusta ducale basilica di San Marco dal nono secolo sino alla fine del decimo ottavo*. Venice: Ferdinando Ongania.
- Sanuto, *Diarii*: Sanuto, M. (1879–1903) *I Diarii*, 58 vols. Venice: F. Visentini (repr. Bologna: Forni Editore, 1969–70).

References

- Anderson, J. (1973) ‘Some Documents Relating to Giorgione’s “Castelfranco Altarpiece” and His Patron Tuzio Costanzo’, *Arte Veneta*, 27, pp. 290–99.
- Arbel, B. (1989) ‘The Cypriot Nobility from the Fourteenth to the Sixteenth Century: A New Interpretation’, in B. Arbel, B. Hamilton and D. Jacoby (eds) *Latins and Greeks in the Eastern Mediterranean After 1204*. London: Cass, pp. 175–97.
- Arbel, B. (2017) *Studies on Venetian Cyprus*. Nicosia: Cyprus Research Centre.
- Arbel, B. (2023) ‘The Cypriot Mountain Fortresses and the Protection of the Rural Inhabitants of Venice’s Sixteenth-Century Colonies’, *Frankokratia*, 4.2, pp. 166–94. Available at: https://brill.com/view/journals/fra/4/2/article-p166_2.xml.
- Bacci, M. (2014) ‘Patterns of Church Decoration in Famagusta (Fourteenth to Sixteenth Centuries)’, in A. Weyl Carr (ed.) *Famagusta: Art and Architecture*. Turnhout: Brepols, pp. 203–276.
- Baiocchi, A. (1983) ‘Contarini, Alvise’, *DBI*, 28, pp. 76–8. Available at: [https://www.treccani.it/enciclopedia/alvise-contarini_res-20594dd4-87eb-11dc-8e9d-0016357eee51_\(Dizionario-Biografico\)/](https://www.treccani.it/enciclopedia/alvise-contarini_res-20594dd4-87eb-11dc-8e9d-0016357eee51_(Dizionario-Biografico)/).
- Bakirtzis, N. (2022) ‘Tradition and Transition on the Slopes of the Pentadaktylos Mountain in Thirteenth-Century Cyprus’, in J.P. Albani and I. Christoforaki (eds) *Πάντα ῥεῖ. Change in Thirteenth-Century Byzantine Architecture, Art, and Material Culture*. Turnhout: Brepols, pp. 105–125.
- Birtachas, S. (2019) *Βενετική Κύπρος (1489–1571): Οι Εκθέσεις των αξιωματούχων του ανώτατου διοικητικού σχήματος της κτήσης*. Thessaloniki: Epikentro.
- Cicogna, E.A. (1824–61) *Delle iscrizioni veneziane*, 6 vols. Venice: Giuseppe Orlandelli editore.
- Coureas, N. (2016) ‘A Process of Secularization? Venetian Hospitallers and Hospitaller Estates on Cyprus After 1474’, *Ordines Militares. Colloquia Torunensia Historica*, 21, pp. 111–27.

- Darrouzès, J. (1950) ‘Les manuscrits originares de Chypre à la Bibliothèque Nationale de Paris’, *Revue des Études Byzantines*, 8, pp. 162–96.
- Darrouzès, J. (1957) ‘Autres manuscrits originares de Chypre’, *Revue des Études Byzantines*, 15, pp. 131–68.
- Fortini Brown, P. (2004) *Private Lives in Renaissance Venice: Art, Architecture and the Family*. New Haven; London: Yale University Press.
- Fortini Brown, P. (2017) ‘Vain Legislation Against *vana ostentazione*: Sumptuary Laws in the Venetian Dominion’, *Artibus et Historiae*, 76, pp. 53–76.
- Giuriato, G. (1871) ‘Lepanto (MDLXXI-MDCCCLXXI)’, *Archivio Veneto*, 1, pp. 247–97 and 2, pp. 17–49.
- Gluzman, R. (2023) ‘Ship Biographies as a Form of Microhistory’, in M. Aymard and E. Orlando (eds) *Venezia e il senso del mare. Percezioni e rappresentazioni*. Venice: Istituto Veneto di Scienze, Lettere ed Arti, pp. 93–110.
- Harkas, I., Schabel, C. (2026) ‘Between the *Chronicle of Amadi* and the *Historia* of Florio Bustron: Luca Foscarini’s Genova, Biblioteca Universitaria, ms. E.V.15’, *Frankokratia*, 7, pp. 77–129.
- Knapton, M. (2013) ‘The *terraferma* State’, in E. R. Dursteler (ed.) *A Companion to Venetian History, 1400–1797*. Leiden; Boston: Brill, pp. 85–124.
- O’Connell, M. (ed.) (2009a) *Rulers of Venice, 1332–1524*. New York: ACLS Humanities E-book. Available at: <https://www.fulcrum.org/concern/monographs/2514nm30s>.
- O’Connell, M. (2009b) *Men of Empire. Power and Negotiation in Venice’s Maritime State*. Baltimore: The Johns Hopkins University Press.
- O’Connell, M. (2013) ‘Individuals, Families, and the State in Early Modern Empires: The Case of the Venetian *Stato da Mar*’, *Zgodovinski casopis/Historical Journal of Slovenia*, 147.1–2, pp. 8–27.
- Ouerfelli, M. (2008) *Le sucre. Production, commercialisation et usages dans la Méditerranée médiévale*. Leiden; Boston: Brill.
- Papacostas, T. (2022a) ‘Suspended in Time and Space: A Carved Heraldic Panel from Nicosia. Family Histories and Marital Strategies in Venetian

- Cyprus', *Frankokratia*, 3.2, pp. 169–206. Available at: https://brill.com/view/journals/fra/3/2/article-p169_3.xml.
- Papacostas, T. (2022b), 'The Architectural Setting of Pilgrimage in the Byzantine World: Learning from Cyprus', in R.A. Etlin (ed.) *The Cambridge Guide to the Architecture of Christianity*, 2 vols. Cambridge: Cambridge University Press, vol. 1, pp. 192–201.
- Papadopoulos-Kerameus, A. (1891–1915), *Ιεροσολυμιτική βιβλιοθήκη ἥτοι κατάλογος τῶν ἐν ταῖς βιβλιοθήκαις τοῦ ἀγιοτάτου ἀποστολικοῦ τε καὶ καθολικοῦ ὀρθοδόξου πατριαρχικοῦ θρόνου τῶν Ἱεροσολύμων καὶ πάσης Παλαιστίνης ἀποκειμένων ἐλληνικῶν κωδίκων*, 5 vols. St Petersburg: B. Kirschbaum.
- Patapiou, N. (2013/14), 'Σκιπίωνας Costanzo: Ο τελευταίος φεουδάρχης της Αθηνών', *Στασίνος*, 14, pp. 153–82.
- Phillips, S.D. (2021), *Rhodes in the Final Hospitaller Years, 1502–1521: Latin and Greek Interaction in the Aegean and Eastern Mediterranean*. Unpublished doctoral dissertation, University of Athens.
- Plakotos, G. (2021), 'Review of *The Rulers of Venice, 1332–1524*. Database by Benjamin G. Kohl, Andrea Mozzato and Monique O'Connell', *Renaissance and Reformation/Renaissance et Réforme*, 44.1, pp. 157–60.
- Salmini, C. (2014), 'Il Segretario alle voci: un primo contributo sulle origini dell'incarico e la formazione dell'archivio', in M. Knapton, J.E. Law and A.A. Smith (eds) *Venice and the Veneto During the Renaissance. The Legacy of Benjamin Kohl*. Florence: Firenze University Press, pp. 47–70.
- Skoufari, E. (2011), *Cipro veneziana (1473–1571). Istituzioni e culture nel regno della Serenissima*. Rome: Viella.
- Szépe, H.K. (2018), *Venice Illuminated. Power and Painting in Renaissance Manuscripts*. New Haven; London: Yale University Press.
- Targhetta, R. (1995), 'Fedeli, Vincenzo', *DBI*, 45, pp. 611–15. Available at: [https://www.treccani.it/enciclopedia/vincenzo-fedeli_\(Dizionario-Biografico\)/](https://www.treccani.it/enciclopedia/vincenzo-fedeli_(Dizionario-Biografico)/).
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016), 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', *Sci Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Chapter 4

Connections through collections: linking archaeological and epigraphic archives through places and people

Alessandra Giovenco and Nurdan Atalan Çayirezmez

Introduction

Digital innovation in information has enabled exponential growth in the digitalisation of galleries, libraries, archives and museums (GLAM). This expansion of digital offerings aims to broaden public access and foster new forms of engagement with preserved cultural heritage, particularly in cultural organisations ([Goudarouli et al. 2019](#): 1–2). Institutions such as the National Archives of the UK, the British Museum and the Smithsonian in the USA are engaged in digitising their collections, creating databases and digital repositories, and promoting reuse for various purposes.¹ On a smaller scale, organisations such as the Society of Antiquaries, Sir John Soane’s Museum and the Warburg Institute are proactive in publicising their collections online in various capacities.² While institutions digitise and preserve their collections, new digital technologies and infrastructures have opened new ways to share and engage researchers and visitors, such as Europeana.³ GLAMs (cultural and memory institutions) also faced access challenges during and after the COVID-19 pandemic, and subsequently have become more digital, more open and more accessible ([Avdikos et al. 2024](#); [Kennedy 2020](#)). Additionally, collaboration between institutions and individuals has become increasingly important, leading to a rise in the establishment of GLAM communities and the sharing of knowledge,

particularly regarding data curation, data sharing and public engagement activities ([Kennedy 2020](#)).

The British Academy's British International Research Institutes (BIRI), worldwide research organisations fostering international engagement, cultural heritage preservation, and new knowledge discovery in the humanities, arts and social sciences, are similarly participating in this wave of innovative digitisation.⁴ The nine institutes that make up the BIRI, all part-funded by the British Academy, span the regions of the Black Sea and Türkiye, Southern Europe, Africa, the Middle East, Iran and Central Asia. Archaeological research across these regions has created deep-rooted connections between BIRI,⁵ and the places and the people involved in such epic endeavours. The relationship that has been formed between the various BIRI archives and their special collections, and the degree to which they complement and enhance each other, represents considerable added value. As Charlotte Roueché puts it,⁶ 'the whole is greater than the parts' (DCDC21 2021: BIRI). Working in collaboration with each other, the BIRI have so far unearthed a multitude of connections between people and places, across time as well as collections:

- Eugénie Sellers Strong (1860–1943), for example, was the first female student at the British School at Athens (BSA) in 1890–1891 but then became Librarian and Assistant Director at the British School at Rome (BSR) (1909–1925);
- Mortimer Wheeler (1890–1976), President of the British Academy and co-founder of the British Institute in Eastern Africa (BIEA), worked closely with John Bryan Ward-Perkins (1912–1981), Director of the BSR, to protect Libyan antiquities during the Second World War;
- Olwen Brogan (1900–1989) served as the BSR's Faculty of Archaeology, History and Letters Secretary between 1931 and 1961 and then went on to co-found the Society of Libyan Studies, now known as BILNAS, in 1969;
- Donald E. Strong (1927–1973), Rome Scholar in Classical Studies in the early 1950s and BSR Faculty member, was also key in the founding of BILNAS;

- Another BSR fellow, David Oates (1927–2004), after working in Tripolitania went on to become the Director of the British Institute for the Study of Iraq (BISI) in the 1960s.

As of 2022, the BIRI have been granted funding from the British Academy to support the BIRI Digital Coordination and Strategy Project,² which focuses on the many connections between their archival and special collections through people and places. A key aim is that discussions of the nature and interdependence of physical and born-digital resources amongst the BIRI will lead to the adoption of Linked Open Data (LOD)-oriented data models to share the essential properties in each institution's catalogue (BIAA Digital Repository 2025: Projects). For example, a paper by Eleni Gkadolou (BSA) and Nurdan Atalan Çayırezmez (BIAA), presented at the CAA-Greece Conference in April 2024, identifies in the 'Places' and 'People' common datasets the foundations on which to build interactive digital storytelling drawn from archival records ([Gkadolou and Atalan Çayırezmez 2025](#)). A concrete output of such collaborative endeavour is the 'Women Transcending Boundaries' StoryMap published online in May 2025 (BIRI Digital Coordination and Strategy Project 2025: Women Transcending Boundaries). A StoryMap is essentially a narrative that leverages the power of maps to visualise people's activities across spaces. This visualisation, if supported by LOD principles, helps uncover hidden connections within the data to create narratives that were previously unexplored.

As a BIRI community initiative for open digital collections, the BIRI Project also supports the creation of a space to discuss open access and copyright-related issues and create Findable, Accessible, Interoperable, and Reusable (FAIR) archives ([Ariadne Plus 2025](#)). Another important part of the project is to create standards for metadata and terminology while structuring the data for sharing purposes. LOD provides a solution for generating machine-readable data and enhancing findability and interoperability during the curation of legacy data processes ([Schmidt et al. 2022](#); [Daquino 2024](#); [Hawkins 2022](#)). 'People' and 'Places' are easily identifiable through stable Uniform Resource Identifiers (URIs) and georeferenced datasets and are ideal for highlighting such connections through collections. Moreover, the adoption of appropriate LOD

infrastructures appreciably increases opportunities for engagement and better collaboration by increasing the accessibility, interconnectedness and usability of data ([Kansa and Kansa 2022](#)).

To illustrate such interconnections and demonstrate the potential of this archival material, this paper will explore two distinct digital projects carried out by two of the BIRI: the British School at Rome (BSR) and the British Institute at Ankara (BIAA).

The BSR project, the ‘Inscriptions of Roman Tripolitania (IRT) Photographic Collection’, makes use of place URIs to preserve the history and context of the collaborative archaeological research that it documents. A particular focus of the work carried out on the IRT Photographic Collection is its use of the Heritage Gazetteer of Libya (Heritage Gazetteer of Libya/SLS Archives 2015) and the IRT 2021 Digital Edition ([Reynolds et al. 2021](#)). The BIAA project—entitled ‘Connecting Archives, Connecting People (CACP)’ (Imagining Futures 2022)—is a collaboration between BIAA and BILNAS concerning data relating to people based on the content derived from their archival collections. The aim of CACP is to create a set of multilingual guidelines for specialists in the study of cultural heritage to enable them to extract the information relating to individuals from these archives. The eventual goal is to build an LOD infrastructure for all BIRI. This pilot project will lead to the development of English, Turkish and Arabic guidelines and facilitate the creation of workflows so that the basic structured information for ‘person data creation’ can be entered. This, in turn, will encourage BIRI digital archivists to compile sets of ‘person data’ to adopt on a shared platform.

BSR: ‘Inscriptions of Roman Tripolitania’ Photographic Collection project

Ostensibly, these two distinct projects arising from two of the BIRI—the BSR and the BIAA, respectively—seem to point in different directions due to their diverse historical questions and intellectual content. In essence, however, both institutes are dedicated to promoting the use of reliable and stable digital resources, such as online gazetteers, authority lists and Wikidata entry points, to enhance the accessibility of their cultural

collections under the umbrella of the BIRI Digital Coordination and Strategy Project.

A representative example of enhancing archival accessibility in the case of the BSR is offered by the digitised IRT Photographic Collection.⁸ This collection comprises a corpus of more than a thousand photographs assembled by John Bryan Ward-Perkins (1912–1981), classical architectural historian and archaeologist, and Director of the BSR (1945–1974)⁹ and classicist Joyce Reynolds¹⁰ (1918–2022) in the aftermath of the Second World War. A voluminous correspondence with Italian academics (preserved in the BSR Institutional Archive), as well as several expeditions to Libya from 1946 to the early 1950s, prepared the ground for a pioneering book, *Inscriptions of Roman Tripolitania*, which was published in 1952 ([Reynolds et al. 1952](#); BSR Research Collections, Boxes 755–7). The IRT Photographic Collection stands as the visual counterpart to the book, which draws on these images but reproduces very few of them. It has been at the centre of various digital activities for fifteen years, from 2009 to the present. These have included two digital editions of the IRT publication (*IRT2009* and *IRT2021*) as well as the online publication of the IRT Photographic Collection (BSR Digital Collections 2020).

In the 1940s and 1950s, Ward-Perkins and Reynolds could not have imagined that their work would one day be the subject of successive digital collaboration projects. Nonetheless, the spirit of collaboration and common endeavour underpinning their research informed the whole enterprise. Conversations with Italian scholars began immediately after the end of the war, driven by the urgent need to publish their excavations and make their discoveries available to the international community of archaeologists, thereby supporting the advancement of the discipline ([Ward-Perkins, Toynbee and Fraser 1949](#): 165). In a brief report addressed to Major D.E.L. Haynes,¹¹ dated April 1946, John Bryan Ward-Perkins, for example, remarked that ‘every opportunity and encouragement should be given to the present Superintendent Professor Caputo, and to his assistant Doctor Pesce¹² to publish the results of their excavations conducted under their direction before the war’ ([Gandolfo 2014](#); BSR Research Collections, Boxes 756–7). Giacomo Caputo¹³ had been the last Superintendent of Antiquities for Tripolitania before the outbreak of the Second World War. In November 1945, Caputo was invited back by the British Administration to act again as

superintendent and to assist with surveying, protecting and preserving the archaeological sites and antiquities in Tripolitania ([Rizzo and Turjman 2022](#): 37). Under the British Administration, Richard Goodchild took over as Antiquities Officer from 1946 to 1948 ([Brogan 1968](#): ix–x). In Goodchild's words, 'the British School at Rome [...] can serve and is indeed already serving as a link with those Italian Scholars who are already doing good work in this field and with the appropriate Italian learned journals' (BSR Research Collections, Boxes 756–7). Research and intellectual exchanges such as these played an important role in the postwar regeneration of a community of students, researchers and academics in Italy and the Mediterranean in general. Maintaining this community remains at the heart of the work of all the BIRI.¹⁴

The IRT Photographic Collection is essentially a distinctive presentation in its own right of the history and context of the archaeological research it documents. Like an excavation, the collection effectively comprises various layers that the digitisation process makes more visible.

The first layer concerns *the study of inscriptions as artefacts* illustrating historical events, people's lives and activities. These objects, indeed, were seen and examined in situ in a particular sequence, and their relevant findspots were identified to provide an accurate picture of their distribution. For instance, the inscription numbered IRT 918 (BSR Digital Collections 2021: WP-INSCRIPTIONS:2106; IRT2021 918)¹⁵ is a limestone stele located in the Lepcis Museum but found in the Baths at Bu Njem, a Roman outpost and oasis in the desert. In it, centurion *Q. Avidius Quintilianus* poetically describes the lives of soldiers fighting an everyday battle against the fierce climate in the southernmost fortress of the Roman frontier. Photographers were hired or photographic equipment was brought directly from the UK to create an up-to-date record of their status quo at a given time.

The second layer is *the finding and recording process*, partially illustrated by the filing system conceived for the 'Tripoli Inscriptions'.¹⁶ The inscriptions were divided into four main series: Lepcis (Leptis Magna) Serial L, Sabratha Serial S, Tripoli (Oea) Serial O and other sites, Serial T. Each of these series were subdivided into two parts: Non Christian and Christian. This filing system formalises Ward-Perkins's and Reynolds's archaeological research methods.

Finally, *assembling, annotating and editing* so many photographs entailed activities relating to ‘putting a collection together’, and this reveals much of the entire research process. The photographic record in its entirety was available for consultation only at the BSR, for very few of the images used by Ward-Perkins and Reynolds to support the exposition in their book were actually included in the volume ([Reynolds, Ward-Perkins and Aurigemma 1952](#); cf. [Bodard 2008](#)).¹⁷ To maintain the spirit of the earlier collaboration, informed by an extensive exchange of research data, the purpose of the first digital edition (2009) was ‘to make the original material available again, and to provide the full photographic record, together with geographical data linking the inscriptions to maps and gazetteers, and so to other resources’ ([Reynolds et al. 2009](#)). The first online edition was the first attempt to provide online access to the numerous photographs that had been deposited at the BSR, even though their online presentation was ancillary to the browsing and understanding of texts. The same can be said of the second edition issued in 2021 ([Reynolds et al. 2021](#)). In fact, very little attention was given to the photographic objects themselves or to the cards on which they had been glued. For obvious reasons, content prevailed over context. The second digital edition of the IRT nevertheless offered further enhancements and provided links to a set of resources with stable URIs, the Heritage Gazetteer of Libya, as well as the Pleiades Gazetteer (*Pleiades*). As we shall see, this newer edition proved extremely important for the IRT photographs and cards metadata infrastructure.

The most significant advancement offered by the IRT Photographic Collection is its attention to the rich materiality of these materials and their original contexts. It was deemed appropriate to not only provide a description at the item level—the single photograph—but also include the description and the image of the card to which each photograph had been attached.¹⁸ By presenting the archival context and photograph layout, our aim was to ensure that information about the original contexts of creation was also available. It was important to allow online users to get a grasp of the work done by the original researchers in assembling the cards for the printed publication. It is no surprise that the way the IRT photographs are arranged, and can be physically browsed in the archive, reflect an organisational practice that mirrors other series in the same archival collection. The majority of the photographs from the J.B. Ward-Perkins

Series were printouts of the negatives, glued on a purpose-sized light-blue card and organised thematically by place or site. The photographs of the Libyan campaigns and those of another pioneering survey conducted in Italy, in the South Etruria region, between the 1950s and the early 1970s, follow this pattern too and are definitely a distinctive feature of the way Ward-Perkins and his team organised materials. His wife, Margaret, was also involved in such curatorial activities. This family/community-driven archival practice, as [Figure 4.1](#) shows, is one of the attributes that constitute the visual identity of the BSR collections, at least for the period concerned with Ward-Perkins's directorship (1945–1974).

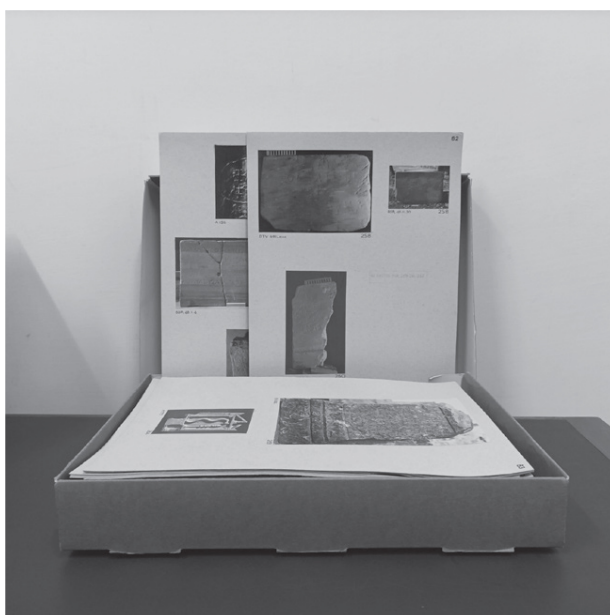


Figure 4.1: One of the boxes housing the IRT Photographic Collection in the BSR Research Collections (CC BY-SA)

Before starting the cataloguing process for the IRT Photographic Collection, the BSR Research Collections cataloguing team analysed the structure of the archival collection and its association with the printed publication and immediately realised that it would be essential for users to have access to both each individual item and its relevant card. The process was not without its challenges, as various issues concerning the call numbers and the sorting of the photographs, as well as the cards, emerged. One of these was the adoption of distinct numbering systems for the photographs owned or commissioned by the BSR and those belonging to

the *Soprintendenza alle Antichità della Libia*—the Italian structure operating under the British Military Administration in Tripoli at that time—the majority of which were undated. Then, the team had to find a suitable way to correctly associate the card with the photographs attached and the numbers assigned to each inscription, highlighting the complexity of the archival context. It goes without saying that looking at these resources from another perspective—that of archivists and librarians—may lead to other considerations that help to improve the contextualisation of the research patterns of those who originated and assembled such materials.

Some examples from this collection may serve to demonstrate the range of questions these photographs may prompt. The first is the extent to which the archiving of photographs provides information about the original research itinerary in Tripolitania. In the IRT card 233 (BSR Digital Collections 2022: WP-INSCRIPTIONS:1956), for example, users may see the same subject commissioned by two different organisations, the BSR and the *Soprintendenza alle Antichità della Libia*.¹⁹ The juxtaposition of both photographs, and the context in which they have been placed, can give us some hint about different archival practices relating to the same artefacts. A possible explanation for the difference in the Tripoli instance is the further information yielded in documents in the BSR archives that enable us to identify the photographer as Flaminio De Liberali,²⁰ who worked closely with Gennaro Pesce and Giacomo Caputo. De Liberali, who produced outstanding images both for the BSR and the archive of the *Soprintendenza alle Antichità della Libia*, was hired by Ward-Perkins on many occasions, especially when it was urgent to get additional shots to complete research in his home country.

As far as the chronological sequence of the photographs belonging to the BSR is concerned, what we understand is that the 35mm strips were numbered with an identifier comprising the following: the year of the photographic campaign, the number of the roll, usually expressed in Roman numerals, and the number of the shot, in Arabic numerals. So, through this example, we can deduce that two shots were taken one after another, perhaps on the same day. With such information, we can map chronologically the research itineraries undertaken in Libyan territory, provided that the year is accurately recorded on film negatives.

The availability of the geolocation data, as well as the variants of placenames provided by the Heritage Gazetteer of Libya (HGL), and the enhanced metadata from the second digital edition of IRT, offered the opportunity to envisage an LOD framework for the BSR's digital records. Although the BSR's digital platform is not yet compliant with LOD, geolocation data and associated URLs are embedded into each record and can be manipulated later. Work on 'Places' and 'People' according to an LOD framework will also be crucial in the near future, in light of the recently established collaboration with the other BIRI, which is laying the foundations for a common digital archive.²¹ The collaboration initiated with the BIRI additionally prompted another partnership with the Manar al-Athar Digital Archive, based at the University of Oxford (Manar al-Athar Digital Archive 2025). This partnership involves the extension of links to the HGL through descriptive BSR records relating to Libyan cultural heritage sites. These links are now embedded into the relevant metadata of all IRT photographs taken between 1946 and the 1960s.

BIAA: Connecting Archives, Connecting People (CACP) Project

Retrieving and sharing information related to cultural heritage collections requires significant effort for collection management and the creation of open access data standards. A lot of work is ongoing behind the scenes, such as digitisation, cataloguing, curation, data sharing standards and controlled vocabularies, and communication with actors and stakeholders ([Belteki et al. 2025](#): 2). One of the key challenges with standardised data is using controlled vocabulary and linking datasets with international authority files and verification ([Hawkins 2022](#): 331). People and places are two important elements of the archives that help enrich metadata, provide additional information about the material and link the different archives ([Woodley 2016](#)).

To add people's names into archival name authority records using the standards ([Stevenson 2020](#)) in the archives and library field, however, is not always straightforward. For example, the same name might be presented inconsistently across various collections within the same archive, different

people can have the same name, and also the same person can have different roles in the archive and have different titles. Therefore, additional time and effort are needed to identify and match the names with the correct authority records.

Several projects and studies have been undertaken regarding names in the fields of cultural heritage and prosopographies created. The examples given here were chosen because one is related to ancient people and the other two are related to modern people in the archives. The final example shows the potential of Wikidata to provide more information about people data in the archives, which the CACP project aimed to do by using Wikidata for enriching and linking person information.²²

- ‘Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names (SNAP:DRGN or SNAP)’ focused on ancient people data and produced a cookbook named ‘SNAP:DRGN Cookbook v. 1.0’ specifically for ancient people prosopographies ([Bodard 2021](#)).
- The ‘Beyond Notability’ Project (Beyond Notability: Re-evaluating Women’s Work in Archaeology, History and Heritage, 1870–1950) explores the histories of women active in archaeology, history and heritage as revealed in the archives of the Society of Antiquaries of London and the Royal Archaeological Institute and works on modern people data ([Harloe et al. 2024](#)).
- ‘People are Essential to Linking Biodiversity Data’, ‘Connecting Te Papa Natural History Collectors and Identifiers’ and ‘Guidelines for Using Wikidata to Mobilise Information about People in Collections: A Paleontology Perspective’, which have assisted in the development of ideas such as ‘LOD’ and ‘Wiki tools’ ([Groom et al. 2020](#)).²³

Another project, by the University of Reading, titled ‘Navigating Networks and Developing Practice: People and Biography in the British Museum Collection Database’, aims to document the biographies and networks of people associated with the British Museum collections. Understanding the people and the connections between them within a collection is essential for museum cataloguing. Building on the biographical authority work of the British Museum²⁴ ([GTR UKRI 2025](#)), this project investigates the implications of creating a network for biographical data that

can enhance future research and cataloguing practices across the wider sector.

Adding creator information to archival records is crucial for the interoperability and discoverability of digital collections. The examples of people involved in more than one BIRI across time and space, as mentioned in the introduction, offer an excellent opportunity here. Similarly to many other GLAM institutions, members of BIRI began developing a standard vocabulary for person data across their archives in the course of the BIRI Digital Coordination and Strategy Project. The first step was to identify people's names in different BIRI archives, and spreadsheets were created to match similar names. For example, Eugénie Sellers Strong worked both at the BSA and the BSR; James G. Crow appears both in the BIAA and BILNAS archives. Similar to many other archives, the BIRI archives lack a standardised method for recording people's names, resulting in variations of the same name appearing differently. The fact that the BIRI work in different geographic regions and use different languages adds another layer of complexity. To address these difficulties, in 2022, two of the BIRI, the BIAA and BILNAS, launched the 'Connecting Archives, Connecting People (CACP)' project²⁵ supported by the Arts and Humanities Research Council, GCRF Network, and Imagining Futures.

The CACP project specifically addressed two key concerns: the presence of unidentified or 'invisible' individuals in archival records/archives and the criteria for determining a person's notability. Linking identified individuals to established controlled vocabularies remains a complex process and requires reconciliation with international authority files. To effectively address these challenges, the collaborative effort extended to multilingual access and training programmes. The CACP project brought together multiple stakeholders, including the BIAA, BILNAS and Wikimedia Community User Group Turkey (WMTR),²⁶ to create a shared approach to person-data management. Collectively, these institutions align their efforts with international best practices and enable access to archival resources across various platforms.

The structured workflow of the CACP project included key stages, both mandatory and optional: compiling an initial list of individuals (including both identified and unidentified/anonymous people), normalising the data, assigning identifiers to person data, linking person data with archival

records, integrating identifiers from name authority records (such as VIAF, LOC and Wikidata), publishing linked data and publishing data using Wikidata.

Compiling an initial person list for archives in tabular format is a simple step to record the initials and/or visible names written on the archival materials. The next step is to normalise the data by deleting the duplicate records and identifying similarities, e.g. James G. Crow, Jim Crow, James Crow and JGC are the same person. This step is mostly manual, and human effort is needed to understand the similar records. It is possible to use some Excel or Google Sheets functions in data normalisation. Next, controlled vocabularies information (URIs) from sources such as VIAF,²⁷ Library of Congress Authorities²⁸ and Wikidata²⁹ is incorporated into the list. Afterwards, the OpenRefine³⁰ tool was utilised in order to verify the data, followed by the creation of a person database—such as the BIAA person database—to share the structured data.³¹ While tools like OpenRefine assist in cleaning and verifying the information, human judgement remains crucial for recognising subtle similarities and resolving ambiguities. Despite these challenges, normalising and structuring the data is a necessary step to ensure the digital repository is accurate, consistent and interoperable. The entry for James G. Crow serves as an example to illustrate the format used for entries (person terms can be entered into the taxonomy fields in the BIAA Digital Repository (Islandora/Drupal, [Figure 4.2](#)). As [Figure 4.2](#) shows, the Authority Sources area provides links to the Library of Congress and Virtual International Authority Files, and to Wikidata and the BILNAS Archive. James Crow's record in the BILNAS Archive links back to the BIAA record, thus connecting the different BIRI archives.³²

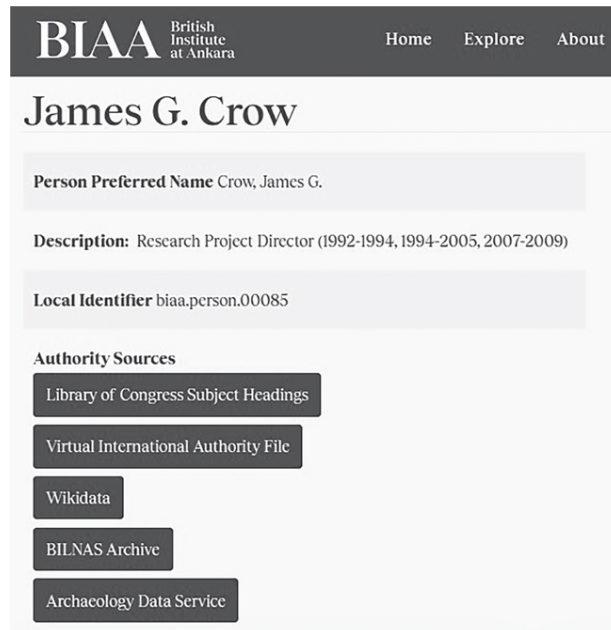


Figure 4.2: Example of online prosopographical data: a cool URI is created for each person's data (<https://digitalrepository.biaa.ac.uk/biaaperson00085>, courtesy of the British Institute at Ankara, CC BY-SA 4.0)

Consequently, Wikidata linking processes were conducted manually and semi-manually to synchronise the BIAA Digital Repository System with Wikidata. After the initial processes were complete, pilot studies for person data were conducted for the automatic processes by utilising the Application Programming Interface (API), which was set up with the support of the BIAA Digital Repository system developers, in order not to take Wikidata item information directly from the Wikidata service. Within the BIAA Digital Repository system, the API service can automatically provide and send the required data, as well as allow archivists to control the data transfer process between the BIAA Digital Repository System and Wikidata.

In order to ensure the synchronisation of the data between an institute's unique archive and Wikidata runs smoothly, the property defined for the institute—'British Institute at Ankara person ID (P11940)', as an example—will be added to the identifiers section within Wikidata. Adding additional data to platforms like Wikidata can be a valuable solution to avoid the need for institutions to store detailed information in their systems, such as birth and death dates, as well as other biographical details,

including sexual orientation and educational history. Smaller GLAMs often lack sufficient staff to update information related to individuals, unlike large platforms like Wikidata. Wikidata is thus a reliable alternative for storing this potentially sensitive information, due to its robust infrastructure and clear policies on data retention and management, which many smaller institutions may not have the capacity to address.

CACP project outputs and guidelines have now been developed in English, Arabic and Turkish.³³ Likewise, three workshops have been held to train professionals in different geographic regions and using different languages, for example a Turkish workshop organised for Türkiye and an Arabic workshop organised online for Arabic-speaking countries.³⁴ Corresponding Wikidata GLAM pages³⁵ were also created in these three languages. This process ensures that a broader network of GLAM professionals can engage with and apply the developed workflows. Workshops conducted in these three languages provided training on best practices in structuring archival data, demonstrating the benefits of linked data and digital interoperability.

By the conclusion of the project, the BIAA had created a person-data template and opened their data to access through its BIAA Digital Repository Page.³⁶ With the assistance of this model, BILNAS also began work on the creation of its person database and ultimately created a unique list for the BILNAS archives.

Certainly, the utilisation of Wikidata for the creation and verification of person data in archives, as well as the contribution of data to Wikidata from GLAM institutions, presents both advantages and disadvantages ([Hawkins 2022](#)). Wikidata is not an entirely reliable data provider in itself; it nonetheless provides statements from other providers—such as VIAF, Library of Congress Authorities, etc.—and other controlled vocabularies' IDs can also be found easily. Those IDs then can be compared using OpenRefine. It would be appropriate to point out that in relation to IDs in particular, having property IDs as a GLAM organisation would also be helpful to share open person data, increase the visibility of the archives and share the person vocabulary with other GLAM institutions. For the BIRI in particular, such an arrangement would mean the creation of a common vocabulary that would save a great deal of time and avoid duplication of effort. During the project workshops, two examples were created; the first is

the ‘James Crow’ example (see [Figure 4.2](#)). From the person data entry made for James Crow by the BIAA Digital Repository, a user can access the James Crow entry made by the BILNAS archive using the relevant links provided. The second example is the entry for ‘Michael Ballance’, by which one can similarly access the BSR Archive and Special Collections entry through the BIAA Digital Repository.³⁷ And the Wikidata record of Michael Ballance includes in its property information a ‘British Institute at Ankara person ID’ for the BIAA Digital Repository and a statement ‘described at URL’ for the BSR Digital Collections.³⁸ In addition to person data, this model will also be applicable to settlement data/places to create common shared place information for all the BIRI. For GLAM organisations, using these common vocabularies and datasets would be useful, especially since archiving unidentified people and place information has proven difficult.

Ultimately, all the methods exemplified above demonstrate that creating LOD infrastructures will prove to be valuable for other GLAM organisations. Creating LOD infrastructures ensures that person data is not only accessible within institutional repositories but also linked to broader digital resources, improving discoverability and interoperability.

Discussion and conclusion

How to reflect such potential in a digital environment supported by these technologies poses challenges that can be addressed only by cultivating and reinforcing a sense of community. Such is the pioneering impact of the work of archaeologists such as Joyce Reynolds, Richard Goodchild or J.B. Ward-Perkins, and so meticulous their observation and assembly of a documentary and photographic record, that these resources are of international importance.

There is no doubt that technologies and tools developed in the digital humanities domain offer the opportunity to profile intentions, contexts and content better. As this chapter has shown, the level to which communities can be engaged effectively depends on individuals, their relationships, the research materials they produced, assembled and left, and the spaces and places in which they operated. Although each project has different scopes and aims, both are fundamentally built on the assumption that ‘people’ and

‘places’ are the common denominator through which institutions can link data to complement one another. More than subjects or topics, which are by nature subjective and questionable, places and people may be anchored to tools or resources that offer some level of disambiguation, as is the case of the Heritage Gazetteer of Libya or Wikidata, and are ideal for the implementation of an LOD/Semantic Web infrastructure.³⁹ Moreover, thanks to conversations on these matters started in 2022, BIRI have been acquiring a much stronger sense of community, through the endeavours of their librarians, archivists, digital repository managers and information management specialists.

Distributed amongst the archives and libraries of the BIRI is a remarkable concentration of documentary, cartographic and photographic material, which both relates to and underpins UK research in Southern Europe, the Black Sea and Central Asia across Iran, Türkiye and the Middle East, to North and East Africa. Although each institution’s holdings are unique, and the identity and diversity of each collection should be maintained, our resources nevertheless complement one another and together constitute a powerful body of knowledge for the history of British archaeological research in the Mediterranean and the Levant.

It is to be hoped that building a community where both knowledge of the collections and technical expertise are shared, and creating a virtual union of our collections, will generate an unrivalled digital gateway.⁴⁰ For the BSR, it would be advisable to change the paradigm according to which its information is organised, by translating the rich corpus of metadata into a knowledge system compatible with an LOD/Semantic Web environment. Such technologies based on semantically structured data that can be easily queried and interpreted through sets of triples or graphs, as opposed to ‘search engines that use keyword matching algorithms’ ([Miller 2011](#): 322), will enable the BIRI to discover even more connectivity than before. The use of Wikidata can help highlight information on minor figures or hidden voices that are embedded in collections and can become a powerful means of engagement for broader audiences. The adoption of an LOD infrastructure primarily founded on ‘People’ and ‘Places’ therefore seems to be the most fruitful way through which the BIRI can harmonise and integrate diversity in a unified environment.

Acknowledgements

This chapter draws on openly accessible digital resources (including IRT digital editions, the Heritage Gazetteer of Libya, the CACP Project [all CC BY-NC-SA], Pleiades [CC BY], and Wikidata [CC0]) as well as archival materials from the British School at Rome and the British Institute at Ankara, accessed under institutional permission. Data included in this chapter is licensed under the Creative Commons Attribution-ShareAlike 4.0 International, except where otherwise stated.

The authors would like to express their gratitude to Professor Charlotte Roueché for her guidance and advice regarding the BIRI Digital Coordination and Strategy Project and the CACP Project. Thanks are also due to the British Academy and BILNAS, through the encouragement of Professor Roueché, for the financial support that enabled the online publication of the IRT Photograph Collection in 2021–2022. Thanks to the University of Exeter, ‘Imagining Futures through Un/Archived Pasts’, for funding the CACP Project. We would also like to thank Dr Valeria Vitale and Dr Gabriel Bodard for their support and assistance during the CACP Project, their contributions to linking data discussions, and their help in the writing of this chapter; Rosario Rovira Guardiola and Alessio Sopracasa for generous comments; Alara Demir for the editing and Dr Eleanor Careless for the proofreading of the chapter. Finally, we would like to thank our colleagues from the BIRI and our institutes, the BIAA and the BSR, for their support.

The BIRI Digital Coordination and Strategy Project team: Board of BIRI Directors: Rudaina Al Momani, CBRL; Nurdan Atalan Çayırezmez, BIAA; Eleanor Careless, BSR; Felicity Crowe, BILNAS; Silvia Ferreri, BIPS; Konstantina Georgiadou, BSA; Alessandra Giovenco, BSR; Eleni Gkadolou, BSA; Jessica Holland, BSA; Freja Howat, CBRL; Amalia Kakissis, BSA; Alice Lubbock, CBRL; David Maina, BIEA; Raphaële Mouren, BSR; Shatha Mubaideen, CBRL; Gonca Özger, BIAA; Charlotte Roueché, KCL; Eman Shahin, CBRL; Rainer Simon; Valeria Vitale, University of Sheffield; Anne Marie Williamson, BILNAS.

Notes

1. The National Archives: <https://www.nationalarchives.gov.uk/explore-thecollection/>; British Museum: <https://www.britishmuseum.org/collection/>; Smithsonian: <https://library.si.edu/collections>.
2. Society of Antiquaries: <https://stories.sal.org.uk/>; Soane Museum: <https://collections.soane.org/home>; Warburg Institute: <https://warburg.sas.ac.uk/photographic-collection>.
3. Europeana: <https://www.europeana.eu/en>.
4. The British Academy (n.d.) 'British International Research Institutes (BIRI)'. Available at: <https://www.thebritishacademy.ac.uk/international/research-institutes/>.
5. These connections were already being made in the early 1960s through the publication of reports in the journal *Antiquity* (Megaw et al. 1964). The uniqueness and interdependence of the BIRI collections were further illustrated at the Discovering Collections, Discovering Communities Conference in 2021 (DCDC21) in a collective presentation chaired by Charles Tripp, then BIRI Vice President, with contributions from: John Bennet (BSA), Alessandra Giovenco (BSR), Nurdan Atalan Çayirezmez (BIAA) and Charlotte Roueché (KCL).
6. Charlotte Roueché (1946–) is a British academic of archaeology and classical studies at King's College London and former British Institute for Libyan and Northern African Studies (BILNAS) Honorary Archivist: <https://www.wikidata.org/wiki/Q28925071>. Professor Roueché used the expression 'the whole is greater than the parts' to highlight the value and potential of the BIRI collections as a unique body of knowledge throughout various exploratory meetings at the DCDC21 Conference in 2021.
7. The Digital Coordination and Strategy Project started in 2022 and is still ongoing. The initial application was written by Charlotte Roueché, Lutgarde Vandeput, Valeria Vitale and Rainer Simon.
8. See <https://ipervisions.digitalcollections.bsr.ac.uk/irt/>.
9. See <https://www.wikidata.org/wiki/Q827842>.
10. See <https://www.wikidata.org/wiki/Q18160815>.
11. Denys Eyre Lankester Haynes (1913–1994) was an English classical scholar, archaeologist and museum curator: <https://www.wikidata.org/wiki/Q1190035>.
- 12.. Gennaro Pesce, Italian archaeologist (1902–1984): <https://www.wikidata.org/wiki/Q15432772>.
13. See <https://www.wikidata.org/wiki/Q20644955>.
14. The role of the BSR is explained well in a report titled 'British School at Rome. Expedition to Tripolitania' outlining the main points of the extensive research programme undertaken in Tripolitania between 1946 and 1951: BSR Research Collections, Box 756. Published references to that can be found in the British School at Rome Annual Reports, particularly those relating to the years 1973–74 and 1974–75: 3–4. For further reading, see also: [Reynolds, Ward-Perkins and Aurigemma 1952](#): v–vi.
15. See <https://digitalcollections.bsr.ac.uk/islandora/object/WP-INSCRIPTIONS:2106>; <https://irt2021.inslib.kcl.ac.uk/en/inscriptions/IRT0918.html>.
16. Many references to photographic equipment, film negatives, chemicals, etc. are in the copious correspondence that J.B. Ward-Perkins maintained with Giacomo Caputo and Richard Goodchild: BSR Research Collections, Box 755, folder 'Inscriptions of Roman Tripolitania. Correspondence, 1947–1950; 1950–1951'; BSR Research Collections, Boxes 756–7.

17. Eleven (XI) plates in total depicting some of the inscriptions are published at the end of the volume.
18. See <https://digitalcollections.bsr.ac.uk/islandora/object/WP-INSCRIPTIONS:1582>.
19. For further information on the *Soprintendenza alle Antichità della Libia*, see [Brogan \(1968\)](#) and [British Archaeology in Libya, 1943–70 \(1970\)](#).
20. Flaminio De Liberali, known as ‘Nino’, worked extensively for the *Soprintendenza alle Antichità della Libia* as a photographer. Ward-Perkins commissioned him to take numerous photographs of the archaeological sites and inscriptions in the Tripolitania region. Caputo speaks very highly of him when referring to the continuity and consistency of the archive: BSR Research Collections, Box 756, folder ‘Tripoli 1949. Expedition funds, vouchers, etc.’.
21. See Vitale 2026 ([Chapter 6](#) in this volume) on the reuse of Heritage of Gazetteer IDs in other datasets and as a possible model for displaying place and person connections across the BIRI.
22. See Chen et al. 2026 ([Chapter 9](#) in this volume) for the use of Wikidata as shared research infrastructure.
23. Wikidata helps to improve access to, linking of and reuse of data and helps enrich metadata for botanists and institutions. For more information see [von Mering et al. 2025](#).
24. [Szrajber \(2008\)](#). Please see the example of a biographical record from the British Museum: <https://www.britishmuseum.org/collection/term/BIOG30463>.
25. See <https://imaginingfutures.world/projects/connecting-archives-connecting-people/>.
26. See https://meta.wikimedia.org/wiki/Wikimedia_Community_User_Group_Turkey.
27. VIAF: <https://viaf.org/>.
28. Library of Congress Authorities: <https://authorities.loc.gov/>.
29. Wikidata: <https://www.wikidata.org/>.
30. OpenRefine: <https://openrefine.org/>.
31. More information about the project process can be found in the relevant Guidelines: <https://hdl.handle.net/20.500.12842/5e371c45-1bfc-4374-b2a6-d5c28f2484e3>.
32. See <https://slsgazetteer.org/person/18>.
33. More information about these processes can be found in the ‘Connecting Archives, Connecting People Project Guidelines for Creating a Person Database’; see: for English <https://hdl.handle.net/20.500.12842/5e371c45-1bfc-4374-b2a6-d5c28f2484e3>, for Turkish <https://hdl.handle.net/20.500.12842/5e206d75-e40e-4a5e-b4a5-6deec8cb5878> and for Arabic <https://hdl.handle.net/20.500.12842/ae4de444-cfd3-4d76-9ac8-dd2c386d20be>.
34. For the workshop details, see: for English <https://hdl.handle.net/20.500.12842/acb99160-ed1a-4fce-9885-f38dad66fb9c>, for Arabic <https://hdl.handle.net/20.500.12842/79ecb306-391f-464c-934e-34ac234d0c2e> and for Turkish <https://hdl.handle.net/20.500.12842/0c0f13cc-c63c-4047-8c27-fcd0d88f208e>.
35. For the Wikidata page, see https://m.wikidata.org/wiki/Wikidata:GLAM/Connecting_Archives_Connecting_People_Project.
36. All data related to CACP was published through the BIAA Digital Repository. More information about the project can be found in the BIAA Digital Repository: BIAA Project No: PRJ240.
37. See <https://digitalrepository.biaa.ac.uk/biaaperson00089>.
38. See <https://www.wikidata.org/wiki/Q123347237>.

39. The adoption of the HGL to geolocate the main building sites in Libya is essential. Searching and browsing through geolocations would hugely improve the experience of users. Resources such as the World Historical Gazetteer (2019–2025) might open up new ways of engaging with archives and special collections for specialised as well as general audiences.

40. See Filosa, Shea and Sopracasa 2026 ([Chapter 11](#) in this volume) on the importance of sharing and consolidating linked datasets between projects via community standards.

Bibliography

Online resources

BIAA Digital Repository (2025) ‘Projects’. Available at: <https://digitalrepository.biaa.ac.uk/projects> (Accessed 18 September 2025).

BIRI Digital Coordination and Strategy Project (2025) ‘Women Transcending Boundaries Over a Century of Scholarship, Fieldwork and Friendship Across the British International Research Institutes’. Available at: <https://storymaps.arcgis.com/stories/47f520f8a244442cb13fc605ee8d95ac> (Accessed 27 May 2025).

The British Academy (2020) ‘British International Research Institutes (BIRI)’. Available at: <https://www.thebritishacademy.ac.uk/international/research-institutes/> (Accessed 25 May 2025).

BSR Digital Collections (2020) ‘The Inscriptions of Roman Tripolitania’. Available at: <https://ipervisions.digitalcollections.bsr.ac.uk/irt/> (Accessed 25 May 2025).

BSR Digital Collections (2021) ‘WP-INSCRIPTIONS:1582’. Available at: <https://digitalcollections.bsr.ac.uk/islandora/object/WP-INSCRIPTIONS:1582> (Accessed 25 May 2025).

BSR Digital Collections (2021) ‘WP-INSCRIPTIONS:2106’. Available at: <https://digitalcollections.bsr.ac.uk/islandora/object/WP-INSCRIPTIONS:2106> (Accessed 25 May 2025).

BSR Digital Collections (2022) ‘WP-INSCRIPTIONS:1956’. Available at: <https://digitalcollections.bsr.ac.uk/islandora/object/WP-INSCRIPTIONS:1956> (Accessed 25 May 2025).

DCDC21 (2021) ‘British International Research Institutes (BIRI)’. Available at: <https://www.youtube.com/watch?v=BB7ptzokGX0>. (Accessed 26 May 2025).

Heritage Gazetteer of Libya/SLS Archives (2015). Available at: <https://slsgazetteer.org/> (Accessed 26 May 2025).

Imagining Futures (2022) ‘Connecting Archives, Connecting People (CACP)’. Available at: <https://imaginingfutures.world/projects/connecting-archives-connecting-people/> (Accessed 18 September 2025).

IRT 2021 918 (2021). Available at: <https://irt2021.inslib.kcl.ac.uk/en/inscriptions/IRT0918.html> (Accessed 26 September 2025).

Manar al-Athar Digital Archive, University of Oxford (2025). Available at: <https://www.manar-al-athar.ox.ac.uk/welcome.html> (Accessed 27 May 2025).

Pleiades (n.d.) Available at: <http://pleiades.stoa.org/> (Accessed 18 September 2025).

References

Ariadne Plus (2025), ‘The Fair Principles and Open Data.’ Available at: <https://ariadne-infrastructure.eu/the-fair-principles-and-open-data/> (Accessed 18 September 2025).

Avdikos, V., Dragouni, M., Michailidou, M. and Pettas, D. (2024) ‘Rethinking GLAMs as Commons: A Conceptual Framework’, *Open Research Europe*, 3. <https://doi.org/10.12688/openreseurope.16473.2>.

Belteki, D., Rees, A.J. and Sichani, A.M. (2025) ‘Datafication and Cultural Heritage Collections Data Infrastructures: Critical Perspectives on Documentation, Cataloguing and Data-sharing in Cultural Heritage Institutions’, *Journal of Open Humanities Data*, 11(1). <https://doi.org/10.5334/johd.277>.

Bodard, G. (2008) ‘The *Inscriptions of Aphrodisias* as Electronic Publication: A User’s Perspective and a Proposed Paradigm’, *Digital Medievalist*, 4. <https://doi.org/10.16995/dm.19>.

Bodard, G. (2021) ‘Linked Open Data for Ancient Names and People’, in S.E. Bond, P. Dilley and R. Horne (eds) *Linked Open Data for the Ancient Mediterranean: Structures, Practices, Prospects*, ISAW Papers 20. Available at: <http://dlib.nyu.edu/awdl/isaw/isaw-papers/20-4> (Accessed 18 September 2025).

British Archaeology in Libya, 1943–70 (1970). *Annual Report—Society for Libyan Studies* (1), pp. 6–11. <https://doi.org/10.1017/S0263718900009560>.

British School at Rome Faculty of Archaeology, History and Letters (1911–1985) *Annual Report to the Subscribers*. London: The British School at Rome.

Brogan, O. (1968). ‘Richard Goodchild’, *Azania: Archaeological Research in Africa*, 3(1), ix–x. <https://doi.org/10.1080/00672706809511483>.

BSR Research Collections, Boxes 754–69; Boxes 777–8.

Daquino, M. (2024). ‘Photo Archives and Linked Open Data. The Added Value’, *Open Library of Humanities*, 10(2), pp. 1–18. <https://doi.org/10.16995/olh.15232>.

Gandolfo, F. (2014). *Il Museo Coloniale di Roma (1904–1971): Fra le zebre nel paese dell’olio di ricino*. Roma: Gangemi editore.

Gkadolou, E. and Atalan Çayırezmez, N. (2025). ‘Digital Stories for the BIRI Collections: Methods, Limitations and Challenges’, *Transforming Heritage Research in a Transforming World: 5th CAA-GR Conference 2024*, pp. 367–75. https://doi.org/10.1007/978-3-032-06389-2_35.

Goudarouli, E., Sexton, A. and Sheridan, J. (2019). ‘The Challenge of the Digital and the Future Archive: Through the Lens of The National Archives UK’, *Philosophy & Technology*, 32, pp. 173–83. <https://doi.org/10.1007/s13347-018-0333-3>.

Groom, Q., Güntsch, A., Huybrechts, P., Kearney, N., Leachman, S., Nicolson, N., Page, R.D.M., Shorthouse, D.P., Thessen, A.E. and Haston, E. (2020). ‘People Are Essential to Linking Biodiversity Data’, *Database*, 2020. <https://doi.org/10.1093/database/baaa072>.

GTR UKRI (2025). ‘Navigating Networks and Developing Practice: People and Biography in the British Museum Collection Database’. Available at: <https://gtr.ukri.org/projects?ref=studentship-2927010> (Accessed 27 September 2025).

Harloe, K., Thornton, A., Baker, J., Mahal, A.K. and Howard, S. (2024). ‘Beyond “Notability”: Re-evaluating Women’s Work in Archaeology, History and Heritage in Britain, 1870–1950’. Available at: <https://beyond-notability.wikibase.cloud> (Accessed 18 September 2025).

Hawkins, A. (2022). ‘Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-digital Archives Via the

- Semantic Web', *Archival Science*, 22, pp. 319–44. <https://doi.org/10.1007/s10502-021-09381-0>.
- [Kansa, E.C. and Kansa, S.W. \(2022\)](#) 'Promoting Data Quality and Reuse in Archaeology Through Collaborative Identifier Practices', *Proceedings of the National Academy of Sciences*, 119(43). <https://doi.org/10.1073/pnas.2109313118>.
- [Kennedy, M.L. \(2020\)](#) 'GLAM Collaborations Under COVID-19 Conditions and Beyond', *Research Library Issues*, 300, pp. 32–8. <https://doi.org/10.29242/rli.300.4>.
- [Megaw, A.H.S., Ward Perkins, J.B., Kenyon, K., Wiseman, D.J., Gough, M., Chittick, N., Stomach, D. and Emery, W.B. \(1964\)](#) 'British Archaeology Abroad, 1963', *Antiquity*, 38(149), pp. 7–17. <https://doi.org/10.1017/S0003598X0006871X>.
- [von Mering, S., Leachman, S., Santos, J. and Meudt, H.M. \(2025\)](#) 'Wikidata for Botanists: Benefits of Collaborating and Sharing Linked Open Data', *Annals of Botany*, 136, 3, pp. 491–511. <https://doi.org/10.1093/aob/mcaf062>.
- [Miller, S.J. \(2011\)](#) *Metadata for Digital Collections: A How-to-do-it Manual*. London: Facet Publishing.
- [Reynolds, J.M., Ward-Perkins, J.B. and Aurigemma, S. \(1952\)](#) *The Inscriptions of Roman Tripolitania*. London: British School at Rome.
- [Reynolds, J.M., Ward-Perkins, J.B., Roueché, C.M. and Bodard, G. \(2009\)](#) *The Inscriptions of Roman Tripolitania*. London: King's College London. Available at: <http://inslib.kcl.ac.uk/irt2009/> (Accessed 27 May 2025).
- [Reynolds, J.M., Ward-Perkins, J.B., Roueché, C.M., Bodard, G. and Vagionakis, I. \(2021\)](#) *The Inscriptions of Roman Tripolitania*. London: King's College London. Available at: <http://irt2021.inslib.kcl.ac.uk/> (Accessed 27 May 2025).
- [Rizzo, M.A. and Turjman, M.A. \(2022\)](#) 'Le Missioni archeologiche italiane in Libia', in M.A. Turjman and L. Musso (eds.) *Libia-Italia: Un'archeologia condivisa: Castello Rosso, Tripoli, 22 settembre –22 dicembre 2021*. Roma: L'Erma di Bretschneider, pp. 37–40.
- [Schmidt, S., Thiery, F. and Trognitz, M. \(2022\)](#) 'Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata', *Digital*, 2(3), pp. 333–64. <https://doi.org/10.3390/digital2030019>.

Stevenson, J. (2020) ‘Names Project (1): Creation of Name Records’, *Archives Hub Blog*. Available at: <https://blog.archiveshub.jisc.ac.uk/2020/06/09/names-project-1-creation-of-name-records/> (Accessed 18 September 2025).

Szrajber, T. (2008) ‘Public Access to Collection Databases: The British Museum Collection Online (Col): A Case Study’, *The Digital Curation of Cultural Heritage, Annual Conference of CIDOC, 2008*. Athens, 15–18 September 2008.

Ward-Perkins, J.B., Toynbee, J.M.C. and Fraser, R. (1949) ‘The Hunting Baths at Lepcis Magna’, *Archaeologia: Or Miscellaneous Tracts Relating to Antiquity*, 93, pp. 165–95.

Woodley, J.S. (2016) ‘Metadata Matters: Connecting People and Information’, in M. Baca (ed.) *Metadata Matters: Connecting People and Information*. Los Angeles: Getty Research Institute. Available at: <https://www.getty.edu/publications/intrometadata/metadata-matters/> (Accessed 18 September 2025).

Chapter 5

Engraving the soil: the challenges of defining spatial footprints in Cypriot cultural heritage

Stuart Dunn

‘Archaeologists discover various inscriptions of the past in the earth which they excavate. These inscriptions are published in corpuses ... The toponyms constitute a kind of inscription engraved on the soil.’

—Andreas Christofidis, Minister of Education of the Republic of Cyprus; preface to [Christodolou and Konstandinidis \(1987\)](#).

Introduction

During its long history, many ethnic groups have settled in Cyprus, inhabiting and naming its places. Lusignan, Ottoman, Turkish and British communities among others have lived there, and so its placenames, or toponyms, have always represented a multiplicity of cultures and traditions. The toponyms of Cyprus thus echo its geographical location at the intersection of the Greek, Turkish, Christian, Muslim, Near Eastern and African worlds—a story of trade, connection, collaboration and exchange, not merely one of conflict or occupation. As with any region, toponyms are generated by people and woven into the physical and cultural landscape. Because of their multicultural context, Cypriot toponyms are complex, multilingual and, in many cases, culturally charged, embodying and encoding the linguistic diversity of many cultures. Many toponyms in use today derive from the Greek and Turkish communities and have existed on

the island since the second half of the sixteenth century, although many earlier influences and direct examples also exist. At the same time, the study of Cypriot toponyms is also a study of the imperial project, most obviously manifested by the map of Cyprus produced by Herbert Horatio Kitchener between 1878 and 1885 on behalf of the British colonial authorities, primarily as a means of taxation and administration ([Chalkias et al. 2023](#)). A clear understanding of the toponyms' heritage and context, which takes into account their contested contexts and colonial histories, is essential background for any research into the island's history, culture and archaeology ([Giraut 2020](#)).

The contested and politicised nature of Cypriot toponyms makes them ideal case material for the open scholarship of historical topography, and for the creation of open (geo)data. By augmenting and broadening the single-authorial model of toponym production with agreed community standards, it is possible to demonstrate how the structures of open digital scholarship can reflect and represent the full diversity of the island's toponymic history. The Heritage Gazetteer of Cyprus (HGC), which was led by King's College London between 2013 and 2015 and funded by the A.G. Leventis Foundation, provides such a case study, enabling the community-led creation of a gazetteer documenting the historical geography of Cyprus as expressed over time in the form of toponyms recorded in written sources up to 1918, the date of the publication of G. Jeffery's *Description of the Historic Monuments of Cyprus* ([Jeffery 1918](#)). Before this, attested toponyms may be considered 'historic', as they precede the contentions of the present political landscape. This is not to assert that Jeffery's own recording of the toponyms is neutral: he was himself a colonial administrator with administrative responsibility for the state and preservation of Cyprus's 'ancient monuments'. Indeed, he used Kitchener's military-led map to coordinate and plan his own survey ([Anon. 1919](#): 385). The HGC therefore does not accord Jeffery's toponyms any special status, but rather uses the publication date as a suitable cut-off point. As well as recording the toponyms, variants, spellings and attestations, the HGC links different attestations to individual places, whether that be a point on the map or an area, and to provide each place, as well as all its attestations, with Uniform Resource Identifiers (URIs), which distinguish them from any others. This follows well-established practice in the design of historical and

ancient world scholarly gazetteers, notably that of the open Pleiades Gazetteer, from which the HGC owes substantial inspiration. This chapter discusses this approach to open gazetteer scholarship, and reviews how a such a process—where multiple users contribute toponyms based on their own readings and expertise, and operating within an agreed framework of standards—can create a deeper, richer gazetteer; a dataset that is ‘open’ in its creation, as well as in the more conventional sense of its publication and availability.

Open gazetteers enable the understanding of the political and cultural significance of toponyms, which is in turn crucial to the scholarly understanding of a region’s historical geography. Myriad historical and cultural research questions depend, whether directly or indirectly, on a full understanding of these. The toponym Famagusta, for example, simultaneously echoes the Greek past of the town (as a corruption of Ammochostos) and a fascination for the Roman past with the medieval fabricated etymology ‘Fama Augusta’, and its modern Northern Turkish Cypriot moniker of *Gazimağusa*, which is not internationally recognised. The modern western town of Paphos incorporates the archaeological park of Roman Nea Paphos, whereas the Bronze Age archaeological site ten miles to the east, Palea Paphos, is contained within the modern village of Kouklia. An open gazetteer to which multiple authors contribute can capture all of this.

The given name of a place in use at any one time has different textual witnesses across different textual sources. A toponym is never just a single name; it is a network of different versions of multiple names. Modelling this network as a series of connections between different toponyms (whether they are homophonic or not) and places, based potentially on multiple authorial interpretations, means taking into account the toponym’s different variants, the spellings of those variants, and the original documents in which those attestations were recorded. This network is dynamic in both time and place. The HGC, and any gazetteer that seeks to represent the textual history of placenames, offers a formal framework by which the multivocality of toponyms can be framed. As well as enabling historical scholarship, such scholarly gazetteers offer a general critique of the way in which the World Wide Web (WWW) represents names of places, historic or modern, and how this mode of representation has gained a

normative status as a result of the WWW's ubiquity. There is much literature exploring this, pointing out that bias on contemporary mapping platforms such as Google Maps (which are, themselves, fundamentally gigantic web gazetteers) is frequently invisible, and stressing the normative nature of digital place, and how it is appropriated by corporate entities ([Leszczynski 2014](#): 75). The case of Cypriot places in particular illustrates how scholarly digital gazetteers can *challenge* toponymic normativity and ontological stability, rather than reinforce it. A key contention of this chapter is that this should be a core mission for all scholarly digital gazetteers.

Given this challenge, it may seem paradoxical that, when written down, the toponyms of any region assume what might be termed 'ontological stability' ([Rossetto 2010](#)). That is, a written, as opposed to an orally transmitted or memorialised toponym, gives an impression of being an immutable 'thing', if not in a physical, then in a metaphysical sense. This is evident both in the cultural record and in everyday usage. For example, in Cyprus, as in many other parts of the world, it is common for street signs to acknowledge alternative variant names of the same place by giving the name in two (or more) languages. The authority and stability of the names themselves do not appear to be questioned in such a context. Toponyms in this context are an articulation of the capacity of governments and communities to give names to places; essentially to decide what forms their names should take. The public setting and authoritative purpose of street signs imbues them with a sense of ontological stability, which is reinforced by any official process which documents and records them. As will be shown, however, the toponyms of Cyprus are ontologically *unstable*, despite appearances to the contrary.

Structuring digitised toponyms into a formal gazetteer and defining their interrelationships depends, to an extent, on the toponyms themselves being ontologically stable. However, in the context of places with historically documented names (which includes physical features and settlements in existence long before their names were documented in written records), gazetteers in fact highlight toponyms' *lack* of such stability. Representing placenames as data, whether on paper or in digital form, highlights the ontological stability of the written word, in contrast to that of the place. In order to capture the nuance and uncertainty inherent in placenames, the

capacity to link to further context is needed. Placenames always extend beyond their textual form (whether or not they take such a form at all). This can be illustrated by considering the parallels between the names of places and the names of people. For example, personal names function as a formal proxy for identity, supportable by official documentation. A name properly verified allows its owner to pass through airports; it establishes that they have the necessary qualifications and authorisation to drive cars, or to carry out legitimate financial transactions, among myriad other functions. A person's legal identity depends on their name being 'ontologically stable'. At the same time, a personal name forms a fundamental part of a person's identity. Article 7 of the UN's Convention on the Rights of the Child, for example, specifies 'a name and nationality' as a fundamental right,¹ highlighting the importance of a name as part of one's persona. Finally, a name can establish implied kinship with different heritages and social and cultural groupings. In the European Anglophone world, the prefixes 'O' and 'Mc' or 'Mac' on a surname indicate association with the particular ethnic groupings of, respectively, Irish and Scots; although 'Irish' and 'Scots' are ethnonyms in the pure sense of the word. These factors all find direct parallels in 'ontologically stable' toponyms and the role they play at the centre of spatial and cultural identity; and yet at the same time they are both explicit/formal and implicit/informal.

Many of the *administrative* functions of a name (whether a personal name or a toponym) are directly related to and enabled by its textual form—by its *attestation*. This has different implications for different kinds of situations. There are immediate potential consequences, for example, if one's name is misspelled on one's passport or bank account; but not if it is misspelled on an invitation to an event (unless formal identification is required for admittance) or on a contact list. Toponyms are much the same: they serve many purposes of varying degrees of formality, ranging from the administrative to the cultural to the political. In governed, regulated societies, authorised or official toponyms share certain characteristics with authorised or official boundaries. They represent and encode a particular view of the world and enforce it, for example by delimiting access. In agrarian or pastoral societies—such as rural Cyprus before and under British administration—such attested/written positions, when held by an authoritative body or individual, can conflict with informal local

knowledge, inherited traditions and unofficial usage of land. Michael Given, for example, describes the case of a goatherd from Lanzia, in the eastern Troodos Mountains, called Michael Yiakoumi, who in 1935 ‘trespassed’ over boundary cairns placed by the British administration limiting access to grazing land in the Makheras Forest and was arrested. Yiakoumi would have been well aware that his transgression was forbidden, yet Given describes how nonetheless his action occupied an ambiguous place between deliberate disobedience to imperial rule and the exercise of de facto access rights that the pastoral farmers of Lanzia had occupied since time immemorial ([Given 2002](#): 1–3).

Local understanding is just one form in which the ambiguities behind the written toponym manifest themselves, but they can take many other forms as well. Informal understandings of toponyms can reflect natural characteristics of the landscape, cultural associations between place and community or identity, and cultural memory ([Giraut 2020](#)). They can also reflect historic geographical distinctions that emerge over time. For example, Nea Paphos and Paleo Paphos in western Cyprus emerged as separate entities following the foundation of the former in the Classical period. Finally, and especially in the case of Cyprus, toponyms represent the view of an ‘other’, such as an occupier or colonial authority.

Digital gazetteers

In the past ten to fifteen years, one of the principal responses of scholarship in digital classics to the complexity of historic toponyms, and to their perceived ontological stability, has been the development of the digital gazetteer ([Elliott 2021](#): 43–4). Gazetteers are lists of placenames, often accompanied by a map, providing spelling form, sometimes variants, etymologies, references and other information, adding the capacity to connect place references with one another, and with other sources ([Elliott 2021](#); [Mostern and Johnson 2008](#)), and thus form part of the so-called Semantic Web ([Berners-Lee et al. 1992](#)).

Digital gazetteers emerged from a period of rapid growth in digital scholarly infrastructures in the 1990s and early 2000s, which reflected an emerging recognition of the importance of (free) resource sharing and

collaboration, driven in turn by a general shift in the WWW away from its origin as a system for publishing and linking electronic documents, towards one of participation and exchange—so-called Web 2.0 ([O'Reilly 2009: 226](#)). Contemporary accounts noted that this environment ‘provide[d] a much-improved technological basis for sharing resources of all sorts from all sources’ ([Buckland and Lancaster 2004](#)). The key role of place as a concept was further noted:

Place, along with time, topic, and creator, is one of the fundamental components in how we define things and search for them. In addition, place is pivotal for interdisciplinary inquiry and latitude and longitude constitute a *lingua franca* across all disciplines and all countries for specifying places. ([Buckland and Lancaster 2004](#))

Place was thus recognised as a concept that links dispersed and disparate data from many sources. However, such ‘fundamental components’ typically focused on ontologically stable vocabularies—physical, topographic and quantitative geographic terms—which are the natural counterparts of timestamps, formal topics and the names of creators—rather than the ‘fuzzy’ and intuitive geographies of history and cultural heritage ([Smith and Frew 1995: 61](#)). The need to deal with the lack of certainty in place motivated the development of digital gazetteers to serve research in the humanities, and especially history ([Southall, Mostern and Berman 2011](#)). In this context, the toponyms of Cyprus and their ontological (in)stability present a complex set of challenges for the design and implementation of formal placename gazetteers, both to represent placenames accurately, and to support research into their past. These challenges are of a linguistic, textual, ethical and political nature.

The HGC is a collaborative project, a guiding principle of which is that any scholar working with written materials containing Cypriot placenames can contribute to the gazetteer, as well as making use of material that is already there. Newly submitted entries are subject to a light touch peer-review process to ensure that a set of basic guidelines are adhered to. These are that a new record must either be a toponym for a location that is not already in the database, or a new variant spelling of a place that is already there. For each entry, a distinction is made between specific monuments (described as Archaeological Entities, AE) for which a geographical point is

defined by the user as accurately as they can; and larger and/or composite groupings of features, such as villages, or excavation areas (described as Historical Units, HU). In these cases, deliberately schematic polygon outlines are created to represent the place, since the precise dimensions of such an area at every point in its history cannot be known. Broadly speaking, this reflects the distinction that commonly appears in heritage management contexts as ‘sites’ (HUs) versus ‘monuments’ (AEs). A more granular distinction can be made in temporal terms: an AE is generally considered to be an object, such as a building or a monument, which was constructed within a definable period of time, however broad that period might be; whereas an Historic Unit is a site that developed over a less definable period. In this, there are some conceptual similarities between the HGC and the Harris Matrix, a fundamental diagrammatic structure describing the sequences of events visible in archaeological stratigraphy ([Harris 2014](#)). This also illustrates that the evolution of toponyms over time, both in terms of their usage on the ground and their recording in the literature, is a chronological as well as a spatial process. This distinction reflects the project’s choices on how abstraction is best dealt with, a key point that will be returned to later in the chapter.

The HGC is thus driven by toponyms recorded in text. This in turn raises fundamental questions of how digital infrastructures, such as those described above, related text content and geographic data. This grounds the creation of HGC places solidly in the hermeneutic approaches of reading, parsing and textual analysis, which is significant because the digital representation and analysis of ancient and/or historic place is often linked with the formal analysis of textual and literary evidence. Existing schemata for structuring the semantics of text, such as the Text Encoding Initiative (TEI), provide mechanisms formalising these representations. The TEI’s coordinate declaration function closely follows the formal stylistics of textual structure, enabling spatial references to be embedded as ‘a particular coordinate notation, or a particular datum, has been employed in a text’.² Other common spatial methods in the humanities take advantage of the stylistic, semantic and grammatical structure of text. These methods include geoparsing sources to identify spatial references ([Gregory et al. 2015](#)), and semantic annotation for identifying narrative structure and chronotopes in ancient travel writers ([Barker et al. 2023](#)). It can certainly be argued

convincingly that textual analysis linked with geocomputational methods is a major element of the spatial humanities, if not its dominant paradigm. This is partly because, as pointed out by Matt Kirschenbaum, text is by far the form of media expression that lends itself most readily to conversion into computational form, and thus to computational analysis ([Kirschenbaum 2013](#): 8).

Defining place

Existing within such a formal geo-textual structure, the HGC is one of many digital infrastructures designed to deal with the constraints of quantified Cartesian space in the humanities (and social sciences), in what has been the ‘spatial turn’ in those disciplines (especially history and archaeology), where scholars have sought to use the characteristics of place as a means of interrogating the human record ([Arias 2010](#): 30). These methodological approaches include ‘deep mapping’, which explicitly rejects the normativity and authority of the ‘traditional’ map and, by implication, of the formal data structures behind it, focusing instead on the processes and argument construction that leads to the map ([Bodenhamer, Corrigan and Harris 2015](#)). The creation, evolution and contestation of toponyms is part of this process. While the technological imperatives that drive them (or rather have helped to drive them) are relatively recent phenomena, both the spatial turn and deep mapping build on long theoretical histories. Yi-Fu Tuan defined place as a subjective product of human experience, where the nation state is defined by the meaning it has to its citizens, the home is defined by its meaning to its family, the smallest political unit, and place in the day to day is defined by the ‘passive experiences’ of passing through the world ([Tuan 1975](#): 161). Similarly, Henri Lefebvre saw place as the cumulative product of an endless series of human interactions and transactions (1991: 15). In the same way, a toponym agreed on by a community represents the outcome of a cumulative series of such transactions. Such Marxist inspired approaches became popular in geographical scholarship in the second half of the twentieth century, a period when counter-culturalists such as Guy Debord were challenging both the epistemology and politics of the Cartesian map, with projects such as

the *Naked City*, which subverted and deconstructed the normative physical map and re-imagined what public space might look like if it were not so constrained ([Sadler 2017](#): 1–3). This was also a period of great political and cultural upheaval in Cyprus, with its independence from Britain in 1960, and the Turkish invasion and partition of the North in 1974. The development of the HGC provides an opportunity to reflect on both the advantages and the limitations of the Cartesian model, and thus the gazetteer, as a means of representing ancient and historic places, and indeed place in general, digitally. To record a place in a digital gazetteer such as the HGC is to define it. In the context of the creation of user-generated open data, it is essential to provide a simple and straightforward way for the user to do this, despite the definition of a ‘place’ being ‘multivocal’ in any context.

H. H. Kitchener, and the first ‘datafication’ of Cyprus’s landscape

Cartographic normativity is a state facilitated by authoritative entities that produce maps, such as governments, national mapping agencies and, in some cases, private corporations. Projecting a particular way of mapping as ‘natural’ or ‘normal’ reinforces that status on the entity doing the mapping—especially when that entity is an occupying power. Herbert Horatio Kitchener was born in 1850 in Ballylongford, Ireland. Three years after being commissioned into the Royal Engineers, at the age of twenty-four, he was commissioned to undertake a mapping survey of Palestine in 1874. Between 1878 and 1882, he was seconded to the Foreign Office and tasked to undertake a major trigonometrical survey of the island of Cyprus, following the establishment of Britain’s protectorate over the island under the aegis of the Ottoman empire. Kitchener’s scientific nature led him to clash with the new governor of Cyprus, Sir Garnet Wolseley, who wanted nothing more than a high-level sketch-map for the purposes of taxation. Kitchener’s insistence that he be given the time and resources to conduct a full scientific survey led to a brief interlude in Kastamonu, Turkey, before he was recalled to the island in 1880 following the appointment of a more amenable governor, Sir Henry Biddulph. Kitchener continued his survey for

the next two years. The fifteen-sheet Trigonometrical Survey of Cyprus was finally published by Stanfords of London in 1885.

The result of his efforts was a milestone in the history of cartography. It made use of the hill shading technique in use at the time by the British Ordnance Survey, which led the *Geographical Magazine* to note that ‘[t]he clearness of the map is no doubt to a great extent due to the hills being printed separately in a purple tint, which at the same time gives more effective prominence to the physical relief’ ([Anon. 1885](#): 590–91), although the same author went on to complain that the cost of such a survey, despite being ‘useful from a strategic point of view’ would have been better spent on further exploration of ‘little-known parts of the world’, for the purposes of trade and economic expansion ([Anon. 1885](#): 591).

Starting from a baseline that was constructed and surveyed using a 100-foot chain supplied by the Ordnance Survey, the survey established a grid of intervisible trigonometrical points, parts of which were still in use until the advent of digital mapping ([Chalkias et al. 2023](#)). The map itself was scaled at one inch to one mile, again adopting contemporary practice from the Ordnance Survey, or 1:63,360. Contemporary analysis of the Kitchener Map has shown that it is remarkably accurate in its representation of the Cypriot terrain, with an overall variation of just 68.76 metres, although there is significant variation in accuracy across the island, with its mountainous interior being, generally, mapped at a lower level of accuracy ([Hadjikyriacou et al. 2021](#)).

In order to populate the map with toponyms, surveyors and translators were dispatched to interview local residents to establish, as far as possible, the correct names of towns and villages, although the names of localities, regions, rivers and natural features were also recorded. Toponymic labels are transliterated from Greek and Turkish into English, with a schema for the transliteration provided on sheet fifteen of the map ([Kitchener 1885](#): nf). Each settlement is marked according to whether it was predominantly Christian with a cross or predominantly ‘Moslem’ with a crescent; those with a ‘mixed population’ are not accorded a symbol. The exact criteria that Kitchener used to categorise the settlements is not recorded; however, this activity illustrates that, as with subsequent maps (and gazetteers), Kitchener engaged in a process of both compromise and abstraction in representing complex information in ‘datafied’ visual form. This reflects many of the

challenges faced by gazetteers in the digital age. Like all maps, Kitchener's *Trigonometrical Survey* represents a datafication of observable reality, simultaneously containing and encoding all the ambiguities that human observation entails. In doing so, it negates ambiguity, presenting a univocal view of the world.

Challenging spatial normativity in Cyprus

By contrast, in the HGC, each attested place, and each attestation of each attested place, is assigned a Uniform Resource Identifier (URI), which provides a single and unique virtual proxy for that place or attestation. A bibliographical, or carto-bibliographical, source must be given for each entry, stressing the link with textual sources and textual interpretation. The URI structure, which is based on the 'cyprugazetteer.org' domain, differs slightly between HUs and AEs. Each HU is indicated by the inclusion of '/hu' directly after the domain, which is followed by a unique number string, indicating that place as a fundamental concept. 'Famagusta', for example, is represented as <https://cyprugazetteer.org/hu/15>. The attestation spelling provided at this top level is given as a 'default toponym' in its Greek, English and Turkish form, drawn from the *Complete Gazetteer of Cyprus* ([Christodolou and Konstandinidis 1987](#)). This provides an authoritative form of each name, standardised according to internationally recognised protocols. This ensures that the HGC operates within a legal and ethical framework that is acceptable within local legal frameworks. For attestations that are not the same as the default toponyms (as noted above, only historic toponyms are included, up to 1918), a further randomised number string is added to denote that variation. In the case of *S. Andronico*, for example, a variant of Famagusta attested in 1542 by Leonidas Attar, the URI is <https://cyprugazetteer.org/hu/472/3190>.

equates to the GeoNames record of <https://www.geonames.org/146737>. A nested association is then made between the HU and the AE, whose latitude/longitude point will be enclosed within the schematic HU polygon.

It is recognised that this is a manually designed process that hard-codes attributes straight into URIs, and that is not an extensible or scalable model, nor does it make use of semantic web standards such as Resource Description Framework or of purposefully designed protocols for linking historical gazetteers such as the Linked Pasts format ([Grossner and Mostern 2021](#)). This was a limitation of the project, the bulk of whose resources and effort was directed towards investigating the conceptual and historical nuances of Cypriot toponyms, and employed the manual format described to do this. The application of such standards would certainly be a focus of future work. The fundamental approach of the HGC as it is, is to formalise distinctions between different variants of toponyms, and to establish links between them; and at the same time to establish a schematic relationship between their spatial footprints.

The representation of the spatial footprint itself is deliberately schematic—in this sense it can be construed to be ‘post-representational’ ([Rossetto 2014](#)). The abstraction of place is a key premise of the digital gazetteer concept: when toponyms are transposed into digital form, then as well as continuing to occupy their role as linguistic, cultural or ethnic markers (or some combination of all three), they become *data*, and take on the properties of data. This is one example of what might be termed ‘the digitisation of place’ ([Dunn 2017](#)). A full discussion of the various philosophical and epistemological meanings of the word ‘data’ is beyond the scope of this chapter, but for the purposes of this discussion, it is sufficient to define data in its computational sense—that is the state of information, such as toponyms, after it has been digitised. Extending the (computational) definition, ‘digitisation’ is derived from Latin *digitus*, literally ‘finger’, and refers to a number less than ten (i.e. counting up to ten on the fingers). Thus, again for the purposes of this discussion, digitisation refers to the ontological representation of the world and anything it contains in digital format. The digitisation of Cypriot placenames from analogue text into digital text renders them computable and, as we will see, crucially, linkable. What it does not solve—indeed what it complicates further—is the *cartographic* representation of the toponyms.

The HGC's toponym data is visualised on a 'standard' Mercator web map. This allows AEs and HUs to be visualised in abstract and functional terms, using a point for AEs and an enclosing polygon defined by the user for HUs. This is a basic form of standardisation that recognises the limitations of the Mercator web map, an issue addressed by many other web mapping services. Google Maps, for example, does not make any claim to be a sophisticated or aesthetically attractive web map; it is driven by imperatives of function, usability and commercialisation that masks, or at least renders implicit, many of the non-formal attributes of toponyms described above. Indeed, recent research has highlighted the way in which such masking can entrench and amplify discrimination and other problematic practices. In her survey of OpenStreetMap and Google MapMaker, for example, Monica Stephens established that the features represented in many 'open' mapping platforms replicate and represent the concerns of the majority of their user base—in these cases white, older and male users, biases that are completely invisible in the visual interface ([Stephens 2013](#); see also [Bhandari and Noone 2023](#)). As with most web gazetteers with map visualisations, the HGC base map employs the familiar web Mercator format, derived in this case from the OpenStreetMap. The compromises made when creating geographical representations to appear on the map—the 'spatial footprint'—are justified by the need to avoid presenting 'false positives'. This is an important part of the process of datafication; however, it also illustrates the limitations of the Web Mercator format for representing the informal characteristics of the place associated with a toponym beyond the purely formal/linguistic. It begs a much deeper question about formal web gazetteers: is their main purpose (or should it be) to generate maps from datasets, or is it (or should it be) to represent the datasets that underlie the map?

Toponyms beyond text

A major question raised by the HGC project in general is how a digital online gazetteer can and should relate to, and represent, the informal, extra-linguistic contexts of placenames. A digital online gazetteer is, fundamentally, a set of data objects linked and organised according to the

model of the World Wide Web (WWW). The fact that the WWW is an infrastructure whose primary purpose is to link digital documents or, in the case of the Semantic Web, data objects and is *not*, by design, a tool to facilitate mapping or visualisation, is a fundamental limitation that all online mapping platforms exist within—hence the visual compromises described above. Although not (yet) formally mapped into a Linked Data ontology, the HGC, like all such gazetteers, reflects the binary nature of the data it contains, and supports links between them which are ‘value-free’—that is, they connect two or more data objects, encoding only the fact that such a connection exists, including none of the connection’s properties. The Linked Data paradigm is an immensely powerful vehicle for connecting dispersed online gazetteers (or any other forms of data source) together, and aligning semantic and ontological meaning across collections; but it is worth emphasising the fact that the URI-based structure and design of an online gazetteer enable purely URI-based links. Such a link would connect (again, for example) *Saline* with *Άλικάις*, given that they are different forms of names of the same place—the salt lake at Larnaca, in the south of the island, attested in, respectively, 1788 and 1542—but without expressing any evidential or interpretive evidence as to why the link exists. The user must click through to the original sources for any further explanation of the toponyms’ etymology or cultural significance. They must also recognise the limitations of mapping, both digital mapping on the WWW and the conceptual limitations of mapping in two-dimensional Cartesian space. On the map, the nature of the feature means that there is a relatively clear topographic delineation (the boundary of the lake), which the contributor has approximately followed, but there is no more sophisticated or formal way of verifying it, apart from visually. These limitations are rarely acknowledged in scholarly web mapping projects.

The emerging map

To an extent, the emergence of the HGC (and certainly the main argument of this chapter) may be seen as an example of ‘critical cartography’, which critiques two-dimensional Cartesian map representations, and stresses that they represent merely one among many ways in which place can be

visualised. Critical cartography draws heavily on scholars such as Fu Tuan, Lefebvre and Debord, who have already been mentioned, but it has rarely been applied in any context of ancient or historic place. Most online gazetteers (historic and modern), including the HGC, present their data in Cartesian map form, for the only reason that there is no viable alternative—if not accompanied by critical interventions and frameworks, the web imposes and normalises a certain way of representing place, and most spatial humanities projects are obliged to follow this. In contrast, most such representations on the generic WWW provide these highly normative and conventional web visualisations with little or no critical context. Conscious exceptions to this are projects such as Native Land Digital, which seeks to visually represent indigenous territories using overlaid schematic polygons⁴—an approach not dissimilar to that of the HGC, which represents HUs with schematic, non-normative polygons, and AEs as points. Rossetto has described this form of non-normative mapping as ‘post-representational’ cartography—visual mapping that seeks not to reflect spatial reality on the ground, but rather the abstract reality of the object being mapped ([Rossetto 2014](#)). However, most of these exceptions, like Native Land Digital, are explicitly concerned with indigenous mapping, which highlights the lack of such critical context in spatial infrastructures not explicitly engaged with that topic. As Rossetto points out, ‘conventional’ maps, printed on paper, are always static devices and, crucially, are ‘only an archetypical form of geovisualisation’ ([Rossetto 2010](#)). There are many ways in which a map can be crafted and constructed; the symbology, colour schemes and cylindrical Mercator projection are simply the most familiar. Just as Kitchener chose to adhere to these cartographic norms of his time, so, in the HGC, a number of decisions were made about how the information therein was presented, and the structure it was presented in. A key difference is that whereas Kitchener and his surveyors observed the physical reality of the Cypriot landscape, in creating an HGC record, the act of ‘observation’ is, in practice, limited to the act of reading and interpreting toponyms in text, identifying their appropriate place in the existing structure and, in the case of a HU, providing a schematic polygon to approximate the extent of the record on the map. This process follows the idea of ‘post-representational cartography’, the idea that maps can, and indeed should, be considered distinctly from the direct geographical ‘ground truth’ that they represent,

and rather be seen as ‘emerging’ as significant objects in the social and cultural context in which they are used. The Kitchener map was, as described above, a product of colonial administration, produced in the context of military convention and the British cartographic norms of the late nineteenth century. One could surely go further and argue that it also fitted into the broader context of Victorian intellectualism, which placed a strong emphasis on accurate observation, classification and taxonomy. This begs the question as to what context the HGC might have in the twenty-first century. It provides an ‘accurate’ representation of *purely textual* space on Cyprus but, as described above, this relies on the fact that text is inherently ‘datafiable’.

The act of creating an HGC record is simultaneously a direct and textual act of datafication (reading/translating a toponym), and an abstract and visual one (creating a schematic polygon in the case of an HU, or assigning a latitude/longitude coordinate in the case of an AE). This reflects the distinction that has been drawn elsewhere, in fields such as art history, between ‘digitisation’ and ‘digitalisation’ ([Drucker 2013](#)): the former is a flat and value-neutral representation of an analogue artefact in digital form, and the latter a complex transformation of semantic and epistemological knowledge. The tension between these two processes of interpretation—which, to one extent or another, lie at the heart of most digital gazetteers—is significant in the context of ‘the digital’. Rossetto goes on to develop her thesis on post-representational cartography by pointing out that in the digital age, ‘maps’ now include ‘map-like spatial images and digital media aimed at providing a visual understanding of space by employing different visual vocabularies’ ([Rossetto 2016](#): 258).

But to what extent do digital gazetteers constructed explicitly to support and enable open scholarship (like the HGC) embrace ‘different visual vocabularies’? Like many such gazetteers, the HGC is deliberately limited in this regard, displaying the HU and AE visualisations on a standard, web-Mercator projected map, which is draggable and zoomable in a way which is familiar to any user of a major contemporary web mapping platform, such as Google Maps. As explicitly described in the project’s online documentation,⁵ this is precisely to avoid the impression that the visualisation is imparting any kind of ‘ground truth’; rather it is imparting only the—perforce—partial and subjective interpretation of the record

creator. This undoubted limitation in the visual representation of semantic textual information highlights both the main strength and the main weakness of the web gazetteer format: that they are first and foremost data structures, not maps. Just as with the Kitchener map, there is an inherent set of compromises involved in visualising the information meaningfully; however, it is a challenge of data visualisation rather than cartographic visualisation.

Conclusion

The foregoing discussion of the HGC highlights its inherent limitations as a means of structuring, interrogating and visualising historic and ancient world geodata. At the same time, it has sought to highlight the necessity of contextualising these limitations with some significant recent literature on neogeography and post-processual mapping. The World Wide Web is fundamentally *not* an analytical or a representational medium; it is a medium designed first and foremost for *connecting* documents and information (which is marked up in schemata designed to enable web-like connectivity between separately defined segments of content), and enabling it to be retrieved and structured. It is not, fundamentally, a medium that was ever designed to convey, describe, impart or visualise spatial or geographic information.

This is not meant to be an attack on the concept of Linked Open Data. Quite the reverse: the aim is to situate Linked Open Data as a key mechanism for understanding historic geography, provided that such context is given. The concept and theory of post-representational mapping indeed allow users and developers of historic gazetteers to move beyond debates, which are now well-rehearsed, about accuracy and (/versus) precision, and to focus on what value gazetteers based on the Linked Open Data concept might bring to nuanced historical and archaeological enquiry. The schematic visualisation of HUs in the HGC, and the point-based representation of AEs, firstly acknowledges the conceptual distinction between the two types of spatial entity, one that is backed up by extensive literatures on heritage management and documentation. It also allows the user to navigate a collated corpus of Cypriot toponyms, without claiming

that in doing so they are also navigating a connected understanding of those sites. The creation of HGC entries is a contingent expert process of reading and interpretation, the use of the HGC as a scholarly resource is no less so.

Acknowledgements

All data in the HGC is available under a CC BY-NC-SA 4.0 licence. As this paper describes, it is created as part of a broad community effort, and I conclude by thanking and acknowledging all those who have contributed to it over the years. As any who know her will agree, the inspirational mentorship and collaboration of Charlotte Roueché is ever delivered with sparkle, humour and generosity. Both have touched me, as so many others, over many years at King's College London, and in the world far beyond, and I will always be grateful for them. In addition, I am very grateful to Ilaria Bucci, Tom Elliott and Tassos Papacostas for their invaluable feedback on an earlier draft of this chapter. It is greatly improved as a result, although all errors and misinterpretations remain mine alone.

Notes

1. <https://web.archive.org/web/20240413121943/https://www.unicef.org.uk/wp-content/uploads/2016/08/unicef-convention-rights-child-uncrc.pdf>. Accessed 26 July 2024.
2. <https://web.archive.org/web/20240409130113/https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-geoDecl.html>. Accessed 26 July 2024.
3. The HGC was nonetheless ingested into the now-defunct Peripleo viewer of the Pelagios project.
4. <https://web.archive.org/web/20240427201847/https://native-land.ca/>. Accessed 26 July 2024.
5. <https://cyprusgazetteer.org/about>.

References

- Anon. (1919) 'Review of A Description of the Historic Monuments of Cyprus', *Geographical Journal*, 54, pp. 385–6. Available at: <https://www.jstor.org/stable/1779421>.
- Anon. (1885) 'New Maps', *Scottish Geographical Magazine*, 1(11), pp. 589–93. Available at: <https://doi.org/10.1080/00369228508732317>.

- Arias, S. (2010), ‘Rethinking Space: An Outsider’s View of the Spatial Turn’, *GeoJournal*, 75(1), pp. 29–41.
- Barker, E., Konstantinidou, K., Kiesling, B. and Foka, A. (2023), ‘Journeying through Space and Time with Pausanias’s Description of Greece’, *Literary Geographies*, 9(1), pp. 124–60.
- Berners-Lee, T., Cailliau, R., Groff, J.-F. and Pollermann, B. (1992), ‘World-Wide Web: The Information Universe’, *Internet Research*, 2(1), pp. 52–58. Available at: <https://doi.org/10.1108/eb047254>.
- Bhandari, A. and Noone, R. (2023), ‘Support Local: Google Maps’ Local Guides Platform, Spatial Power and Constructions of “the Local”’, *Communication, Culture and Critique*, 16(3), pp. 198–207. Available at: <https://doi.org/10.1093/ccc/tcad018>.
- Bodenhamer, T.M., Corrigan, J. and Harris, T.M. (2015), *Deep Maps and Spatial Narratives*. Bloomington: Indiana University Press. Available at: https://muse.jhu.edu/pub/3/edited_volume/book/49063.
- Buckland, M. and Lancaster, L. (2004), ‘Combining Place, Time, and Topic’, *D-Lib Magazine*, 10(5). Available at: <https://www.dlib.org/dlib/may04/buckland/05buckland.html>.
- Chalkias, C., Papadias, E., Vradis, C., Polykretis, C., Kalogeropoulos, K., Psarogiannis, A. and Chalkias, G. (2023), ‘Developing and Disseminating a New Historical Geospatial Database from Kitchener’s 19th Century Map of Cyprus’, *ISPRS International Journal of Geo-Information*, 12(2), p. 74. Available at: <https://doi.org/10.3390/ijgi12020074>.
- Christodolou, M. and Konstandinidis, K. (1987), *A Complete Gazetteer of Cyprus*. Nicosia: United Nations. Available at: <https://www.geonoma.gov.cy/myfiles/ekdoseis/cygazetteer/>.
- Drucker, J. (2013), ‘Is There a “Digital” Art History?’ *Visual Resources*, 29(1–2), pp. 5–13. Available at: <https://doi.org/10.1080/01973762.2013.761106>.
- Dunn, S. (2017), ‘Praxes of “The Human” and “The Digital”’: Spatial Humanities and the Digitization of Place’, *GeoHumanities*, 3(1), pp. 88–107. Available at: <https://doi.org/10.1080/2373566X.2016.1245107>.
- Elliott, T. (2021), ‘The Pleiadic Gaze: Looking at Archaeology from the Perspective of a Digital Gazetteer’, in *Archaeology and Economy in the Ancient World*. Cologne; Bonn: Propylaeum, pp. 43–51.

- Giraut, F. (2020), 'Plural Toponyms: When Place Names Coexist', *EchoGéo*, no. 53 (October). Available at: <https://doi.org/10.4000/echogeo.20760>.
- Given, M. (2002), 'Maps, Fields, and Boundary Cairns: Demarcation and Resistance in Colonial Cyprus', *International Journal of Historical Archaeology*, 6(1), pp. 1–22. Available at: <https://doi.org/10.1023/A:1014862125523>.
- Gregory, I., Donaldson, C., Murrieta-Flores, P. and Rayson, P. (2015), 'Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research', *International Journal of Humanities and Arts Computing*, 9(1), pp. 1–14. Available at: <https://doi.org/10.3366/ijhac.2015.0135>.
- Grossner, K. and Mostern, R. (2021), 'Linked Places in World Historical Gazetteer', in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Geospatial Humanities*, pp. 40–43. Available at: <https://doi.org/10.1145/3486187.3490203>.
- Hadjikyriacou, A., Papadias, E., Vradis, C. and Chalkias, C. (2021), 'Combining Historical Maps and Censuses of Cyprus from the Sixteenth to the Twentieth Century: A Geospatial Approach', in *Proceedings of the ICA 3* (August), pp. 1–13. Available at: <https://doi.org/10.5194/ica-proc-3-7-2021>.
- Harris, E.C. (2014), *Practices of Archaeological Stratigraphy*. London: Academic Press Ltd.
- Jeffery, G. (1918), *A Description of the Historic Monuments of Cyprus*. London: Zeno Booksellers.
- Kirschenbaum, M.G. (2013), 'What Is Digital Humanities and What's It Doing in English Departments?' in M. Terras, J. Nyhan and E. Vanhoutte (eds) *Defining Digital Humanities: A Reader*. London: Routledge, pp. 55–61. Available at: <https://doi.org/10.1632/ade.150.55>.
- Kitchener, H.H. (1885), *A Trigonometrical Survey of the Island of Cyprus, Executed and Published by Command of H.E. Major General Sir R. Biddulph K.C.M.G. C.B. R.A., High Commissioner Under the Direction of H.H. Kitchener, R.E., Director of Survey*. London: Stanford.
- Lefebvre, H. (1991) *The Production of Space*. Oxford: Blackwell.
- Leszczynski, A. (2014), 'On the Neo in Neogeography', *Annals of the Association of American Geographers*, 104(1), pp. 60–79. Available at:

- <https://doi.org/10.1080/00045608.2013.846159>.
- Mostern, R. and Johnson, I. (2008) ‘From Named Place to Naming Event: Creating Gazetteers for History’, *International Journal of Geographical Information Science*, 22(10), pp. 1091–1108. Available at: <https://doi.org/10.1080/13658810701851438>.
- O’Reilly, T. (2009) *What is Web 2.0?* Sebastopol, CA: O’Reilly Media, Inc.
- Rossetto, T. (2010) *Object-Oriented Cartography: Maps as Things*. Abingdon: Routledge.
- Rossetto, T. (2014) ‘Theorizing Maps with Literature’, *Progress in Human Geography*, 38.4, pp. 513–30. Available at: <https://doi.org/10.1177/0309132513510587>.
- Rossetto, T. (2016) ‘Geovisuality’, in D. Cooper, C. Donaldson and P. Murrieta-Flores (eds) *Literary Mapping the Digital Age*. London: Routledge.
- Sadler, S. (2017) ‘The Naked City: Guy Debord and Asger Jorn’, in H.F. Mallgrave (ed.) *Companion to the History of Architecture*, 1st ed. Hoboken, NJ: Wiley, pp. 1–12. Available at: <https://doi.org/10.1002/9781118887226.wbcha142>.
- Smith, T.R. and Frew, J. (1995) ‘Alexandria Digital Library’, *Communications of the Association for Computing Machinery*, 38.4, pp. 61–2. Available at: <https://doi.org/10.1145/205323.205340>.
- Southall, H., Mostern, R. and Berman, M. (2011). ‘On Historical Gazetteers’, *International Journal of Humanities and Arts Computing*, 5(2), pp. 127–45.
- Stephens, M. (2013) ‘Gender and the GeoWeb: Divisions in the Production of User-Generated Cartographic Information’, *GeoJournal*, 78(6), pp. 981–96. Available at: <https://doi.org/10.1007/s10708-013-9492-z>.
- Tuan, Y.-F. (1975) ‘Place: An Experiential Perspective’, *Geographical Review*, 65(2), pp. 151–65. Available at: <https://doi.org/10.2307/213970>.

Chapter 6

You can't step into the same dataset twice: iterative improvement through reuse of open Libyan place data

Valeria Vitale

Introduction

This chapter uses the case of historical Libyan placenames to illustrate the increasingly prominent role that open datasets are playing in current research about the ancient world, and how they are changing not only the quantity and quality of information available to scholars but the research methodologies themselves. Following the evolution of a spatial dataset from a simple spreadsheet to an open authority list with stable unique identifiers, this contribution argues that each reuse participates in the transformation of the original dataset, expanding its content as well as the horizon of its possible applications. This process steers, at the same time, technological development and research enquiries while illuminating their mutual relationships.

The case study: places in Libyan epigraphy

The dataset that will be discussed in this chapter was created in 2007 to support two epigraphic digital publications: the corpus of *Inscriptions of Roman Cyrenaica* (IRCyr)¹ and subsequently that of the *Inscriptions of Roman Tripolitania* (IRT),² both based on the work of Joyce Reynolds. These pioneering digital editions were powered by the use of XML and, in

particular, of a domain-specific set of encoding rules called EpiDoc ([Bodard 2010](#); [Cayless et al. 2009](#)). The EpiDoc Guidelines³ support the enrichment of the historical text with contextual information such as disambiguation of named places and people. The disambiguation is usually enabled via links to external knowledge bases and authority lists such as gazetteers (for places) and prosopographies (for people).⁴

To enrich IRCyr and IRT from a spatial perspective, the authors,⁵ and in particular Libyan archaeologist Hafeed Walda, started compiling a list of the places that were relevant to the two epigraphic corpora, like findspots, places mentioned in the texts or current locations of the ancient artefacts. The work was performed manually and entailed the identification of the named places and the creation of a catalogue in the form of a simple spreadsheet. The task was not entirely straightforward as the mix of ancient and modern places, as well as the misalignment between the original names and their transliteration in European languages, required a certain degree of normalisation.⁶ In addition, interpretative choices were occasionally needed to establish if two places with similar names may, or may not, have been the same one. Both the normalisation and interpretation processes required disambiguation of the place references. This goal was achieved by matching the spatial entity to Uniform Resource Identifiers (URIs)⁷ ([Berners-Lee et al. 2005](#)) minted by existing authorities such as Pleiades or GeoNames ([Elliott and Gillies 2008](#); [Asciutti and Dunn 2011](#)). Both these digital resources are currently well established ([Horne 2020](#); [Barker et al. 2016](#)) and considerably rich, albeit at different scales.⁸ However, at the time of the creation of the IRCyr (and then IRT) geo-dataset, the gazetteers, like many digital collaborative endeavours, presented several lacunae in their coverage ([Elliott 2021](#); [Ahlers 2013](#)). It is not surprising that, even in the small dataset of Libyan epigraphic places, there were a number of references that did not already have an entry (and, therefore, a URI) in either of the two gazetteers. The IRCyr and IRT teams were keen to keep feeding the spatial dataset, and Walda contributed further data about hand-recorded Libyan archaeological locations. However, such information was becoming arduous to obtain, due to the lack of available digital surveys and the increasing political instability in the country. The authors sought collaborations with archaeologists and heritage professionals that might have already recorded relevant locations during their previous work in Libya. The call was kindly

answered by British archaeologist Philip Kenrick, who had published two archaeological guides to Libya, to Tripolitania ([Kenrick 2009](#)) and Cyrenaica ([Kenrick 2013](#)), and agreed to contribute information about the Libyan places and locations featured in his publications. What was initially a simple spreadsheet turned into a more complex dataset, with three main sources (i.e. Reynolds's epigraphic references, Kenrick's locations and placenames, and Walda's hand measurements) and, when possible, alignments with Pleiades or GeoNames.

[The Heritage Gazetteer of Libya](#)

To avoid duplication of effort and maximise the benefits for the community, the IRT and IRCyr teams contributed the newly added Libyan places back to Pleiades, to enrich its coverage of classical Libya. However, the granularity of the locations that were relevant for the two digital publications started to exceed the scope of Pleiades—or at least its scope at the time. It became increasingly clear that the resource created by Walda and his colleagues had a value beyond epigraphy, and that the information could be reused in other contexts. In 2016, Charlotte Roueché assembled a dedicated research team⁹ to create the Heritage Gazetteer of Libya (HGL) ([Roueché et al. 2016](#)). The HGL is an open gazetteer that combines and reinterprets, in both its content and structure, two pre-existing open resources: the IRT and IRCyr epigraphic digital publications and the Heritage Gazetteer of Cyprus (HGC) ([Dunn 2016](#); [Dunn et al. 2015](#)).¹⁰ The latter, explicitly recalled in the name, was not only a conceptual reference but, more concretely, provided the HGL with a reusable model for a heritage gazetteer that was tweaked and customised to accommodate the unique needs of the Libyan dataset. The connection was strengthened by the substantial overlap in the development teams that worked on both the digital projects, maximising the reuse not only of the data but also of the workflows and know-how.

If the initial content of the geo-dataset was a spreadsheet, then translated to EpiDoc, the same information now had to be reshaped to fit a new data format, GeoJSON ([Butler et al. 2016](#)), and a whole new conceptual model ([Vitale, forthcoming](#)). While the HGL was meant from the beginning to

expand beyond the scope of the epigraphic corpora, its authors endeavoured to preserve information about the composite provenance of the data in the initial dataset, to guarantee transparency and to give due credit to contributors. A data model that would allow connecting a place entity to multiple locations, like the one already used by Pleiades, was then chosen, not only because it was methodologically more suited to the representation of our ever-changing knowledge of ancient places, but also because it enabled the HGL to display the heterogeneous origin of its geographic information.

Although it was relatively simple to represent the association of a single place to multiple locations, there were other characteristics of the original open spatial dataset that were less straightforward to address. The epigraphic digital publications mostly delegated the definition of the referenced places to the external knowledge bases (i.e. Pleiades and GeoNames). In the absence of an existing URI, the authors associated geo-coordinates to a place reference without necessarily including it in a web of spatial relations and definitions. But expressing such geographical, historical and administrative relationships (hierarchical or otherwise) in structured data is one of the main roles of a gazetteer. This goal becomes especially challenging when dealing with ancient places and descriptions of space, often based on fragmentary, incomplete and partly unreliable sources.

In the HGL a further layer of complexity was added by the multiplicity of data sources and by the fact that in several cases they already displayed a certain level of scholarly interpretation and classification. It was not uncommon, for example, that Reynolds and Kenrick would attest different groups of Libyan places as comprising the same ancient regions. Therefore, the HGL features, for example, two separate but related entities for ‘Cyrenaica’: one as defined by Reynolds,^{[11](#)} and one as defined by Kenrick.^{[12](#)} The lists of settlements that are modelled as being ‘within’ each area are only partly overlapping.^{[13](#)} It is apparent that the shape of the initial, heterogeneous dataset influenced the HGL’s data model, as well as its visualisation approach. This choice is not only the most transparent with respect to data provenance, but it is also an important statement about the purpose of historical gazetteers and, more generally, of Linked Open Data (LOD) resources. Pushing forward the HGC attestation-based approach and

echoing the use of factoids in digital prosopographies,¹⁴ what the HGL data modelling suggests is that it is not the role of the author of the gazetteer to decide which interpretation is the most accurate, but to record that there are multiple ones, and that they all collectively contribute to shape our knowledge and understanding of a historical place ([Grossner and Mostern 2021](#)).¹⁵

The HGL also inherited the original relationship between the Libyan places relevant to the epigraphic corpora and the Pleiades gazetteer, and made it prominent by displaying an active link to the corresponding Pleiades URI just underneath the map view. Maintaining this connection not only contributed to the overall transparency of the new gazetteer by emphasising its ‘genealogy’, but it also facilitated the process of aligning the HGL URIs with those of Pleiades and GeoNames. Far from being a duplication of information, alignment is a crucial process for a historical gazetteer ([Berman et al. 2012](#)) and it greatly enhances its connectivity to other resources and its usefulness in a Linked Open Data ecosystem, an ecosystem that the HGL was aiming to enter.

Digital corpora like IRT and IRCyr demonstrated the relevance of connecting epigraphic texts to one or more related places. But digital classics projects with a stronger focus on the ‘biographies’ of material objects, like iSicily ([Prag and Chartrand 2018](#)) or more recently IDEA ([Chen 2022](#)), increasingly highlighted the need for higher granularity in historical gazetteers. For larger archaeological sites, such as Syracuse in Sicily or Sabratha in Libya, simply associating all findings with the settlement would be of limited use: querying for objects found in a prominent archaeological site would return such an overwhelming amount of matches that it would lose informative value. Much more interesting and useful relationships could be discovered by limiting the search to a single building, such as all the artefacts found in the same necropolis or a single temple. The Pleiades gazetteer featured a few isolated examples of urban-level URIs,¹⁶ but lacked the resources to ensure the same coverage for all areas of the Ancient Mediterranean.

Combining the interest in building-specific IDs with a manageable geographic scope, the HGL team decided to invest, experimentally, in enhancing the granularity of the gazetteer’s data by introducing a small number of urban-level URIs. In addition to making the HGL a richer and

more useful digital infrastructure, this process quickly revealed its potential for preservation and documentation. The locations of the selected buildings were derived by comparing published excavation reports with current georeferenced satellite images on Google Maps. What the process showed, though, was that a number of heritage buildings in Libya were becoming less and less visible, to the point of being now hard to detect on the photographs ([Rayne et al. 2017](#); [Bennett and Barker 2011](#)). This discovery gave a new impulse to the inclusion of building-level information in the gazetteer, and the initial workflow evolved to include retracing the mass-geometry of the ground level of the ancient buildings that were still visible on Google Maps. The coordinates were then recorded in GeoJSON as georeferenced polygons. The effort to preserve this fragile information was supported by the implementation of further visualisation options in the HGL: the introduction of Google satellite maps as one of the available layers, and the enabling of a higher level of magnification. In this way entries for individual buildings were visually associated with their current context, enabling a better understanding of the edifice's extent, preservation status and relationships with other natural and built features.

The digital catalogue of the BILNAS archive

The first use case for the HGL outside of epigraphy presented itself when the British Institute for Libyan and Northern African Studies (BILNAS)¹⁷ started planning the digitisation of their archaeological archives. Born in 1969, BILNAS sponsors research in Libya and Northern Africa, with a particular focus on archaeology. The bulk of the collections in the BILNAS archive comprises records generated by British-led archaeological expeditions in Libya from 1910 to the early 2000s. BILNAS's remit made it a perfect match for the HGL, especially when the idea of an online searchable interface for its digital catalogue started taking shape. Ideally, the catalogue was meant to replicate some of the features already successfully developed for the epigraphic publications, including spatial search. The places that were relevant to the BILNAS archive, though, were not only exceeding Pleiades' granularity, but fell sometimes outside of the scope of the classical world. The archive, for example, features detailed

documentation of Christian basilicas, and occasionally references Ottoman buildings and even Italian Futurist architecture from the time of colonial occupation. Current locations of historical artefacts such as museums or antiquaria are also not uncommon. The collaboration with the BILNAS archive drove the HGL to further expand its chronological scope to accommodate references to the country's diverse heritage spanning from prehistoric art to contemporary architecture. On the one hand, the HGL's URIs could strengthen the metadata of the documents in the BILNAS digital catalogue; on the other, the archival resources offered new attested place information to enrich the gazetteer. In the growing landscape of historical gazetteers, the new, stronger identity of the HGL contributed to the necessary diversification of these digital resources.

While the technological connections between the HGL and the BILNAS digital catalogue never moved beyond the stage of proof-of-concept, the obvious potential for collaboration remains very promising. BILNAS's digitised collections, and their metadata, are now officially deposited with the Archaeology Data Service,¹⁸ but several connections to BILNAS resources can still be browsed in the HGL.

Linked Open GeoData

From the beginning, the HGL had the ambition to become a reusable resource and to serve the needs of different audiences ([Roueché 2018](#)). The HGL team was especially keen on the idea of making the gazetteer an open tool for Libyans to contribute information about their own historical places and cultural heritage. This view of the gazetteer as a community platform aimed to generate engagement around Libyan heritage and, at the same time, improve the quality and diversity of the information in the HGL. Contributions from Libyan communities seemed especially crucial in disambiguating unidentified places that were common in Joyce Reynolds's records as well as in the BILNAS archive's metadata. Unfortunately, this potentially impactful aim was never fulfilled. The failure is probably due to the lack of investment in building, strengthening and maintaining a community of interest ([Kahn et al. 2021](#)) that involved Libyan stakeholders, but other concerns and barriers are also worth mentioning. The first was of

an ethical nature. Information about the location of cultural heritage sites and artefacts is potentially problematic as it can involuntarily facilitate looting and destruction, especially in times of political unrest. While the HGL team felt strongly that recording heritage sites could help protect them ([Nebbia et al. 2016](#)), it was impossible to ignore the possibility that this information could be misused. The issue, only partially addressed by the use of approximate locations, also anticipated discussions now deepened and explored by the CARE (Collective benefit, Authority to control, Responsibility and Ethics) set of Indigenous data governance principles.¹⁹ These community-developed guidelines question the right of Western researchers to manage access to information about the cultural heritage of other countries and peoples ([Carroll et al. 2020](#)). The second issue was mainly technical. Creating an account to contribute to the HGL was supposed to be a relatively easy task. Such a low threshold was a prerequisite to involve members of the community outside of academia. However, the HGL was developed and hosted at King's College London (KCL), and had to comply with institutional security requirements. This scenario translated into a fairly slow and cumbersome process for new users to obtain an account and limited editorial privileges. The complexity of the procedure made the HGL overall too difficult and discouraging to access. This impasse also highlighted the amount of invisible (and often unpaid) labour that is necessary to develop and maintain open resources, especially after the end of funding.

Although the attempt to involve more Libyan stakeholders in the curation of the gazetteer was not successful, the HGL actually managed to cross the boundaries of IRT and IRCyr, and even of BILNAS internal use, in another way. By providing unique and stable identifiers for places in Libya, the HGL was able to become a component of the growing ecosystem for historical LOD ([Davis and Heravi 2021](#)). The interest in semantic approaches to the study of the ancient world was championed by institutions such as the Getty Research Institute ([Cobb 2015](#)) and the German Archaeological Institute in the GLAM sector, and by projects such as Linked Ancient World Data ([Elliott et al. 2012](#)) and Pelagios in more academic circles ([Vitale et al. 2021](#)). Started in 2011, Pelagios played a key role in promoting the sharing and reuse of resources about the ancient world, leveraging semantic web technologies. As a proof-of-concept, the

Pelagios team developed a map-based visualisation Application Programming Interface (API), called Peripleo, that displayed dynamically generated place-based connections between resources held in partner open datasets in GLAM and academia ([Simon et al. 2016a](#)). The opportunity to explore these non-obvious spatial connections made a convincing case for the potential of LOD approaches for research. Among the open datasets that powered Peripleo, a mix of object metadata and historical gazetteers, there were three that we have already discussed: Pleiades, the de facto backbone of Pelagios, GeoNames and the HGL itself.

To become part of Peripleo, the HGL data had to be exported as a single GeoJSON dump, and then converted into what was called the Pelagios Gazetteer Interconnection Format: a lightweight standard based on the idea of connecting information through their ‘minimum common denominator’ ([Simon et al. 2016b](#)). The alignment with Pleiades and GeoNames, which was already present in the HGL structured data, facilitated this process, strengthening and enriching Pelagios’s shared infrastructure. Peripleo’s composite dataset library helped to further the academic reflection on the nature and role of gazetteers, and on how each of them expresses only a fragment or a perspective of our knowledge of a place. This intellectual framework was handled technologically in Peripleo through the use of simple semantic statements to define the relationships between different gazetteer URIs for the same place, stating, for example, if the match was complete (i.e. `skos:exactMatch`) or partial (i.e. `skos:closeMatch`) ([Simon et al. 2016a](#)). To better represent this approach, in the last release of Peripleo it was possible to visualise a little graph showing how many sources of information (i.e. how many different historical gazetteers) had a relationship to the queried place, and how they related to each other. The pioneering thinking that supported the Gazetteer Interchange Format has been further developed and expanded by the World Historical Gazetteer, an independent project aiming to create an open resource for interconnected historical gazetteers.²⁰

Until more recent years,²¹ Pleiades focused explicitly on the Ancient Mediterranean.²² As the implementation of Pleiades was (and is) driven by the research interests of the contributors, areas that are more traditionally at the centre of classical studies, such as Greece and Italy, were privileged and likely over represented. Not only was the coverage for ancient Libyan

places much less complete and systematic, but all the Libyan historical locations not falling within the chronological remit of Pleiades were outside of its scope. These circumstances made room for the HGL to become the official and most reliable authority list for Libyan places in a historical and heritage-oriented LOD landscape. Due to the exposure gained within the Pelagios community, HGL's identifiers started being used by other projects and archives that dealt with or managed digital resources related to Libya and wanted to join the growing LOD ecosystem. The first external archive to reuse the HGL identifiers was that of another British Research Institute: the British School at Rome (BSR). The BSR has a strong connection with BILNAS rooted in both national and institutional relationships.²³ When Libyan places appeared in the BSR digital catalogue, like, for example, in the collection of epigraphic photographs by John Bryan Ward-Perkins, HGL URIs started being used in the metadata, often as a complement to other, larger-scope gazetteers such as Pleiades, GeoNames or the Getty Thesaurus of Geographic Names (TGN).²⁴ While the addition of HGL URIs to existing ones in BSR metadata play an important role in strengthening and consolidating the LOD knowledge infrastructure, the most exciting potential of the HGL becomes apparent when less common place-references appear in the BSR archival documents. Only the HGL, for example, could provide a specific URI for a place such as the Central Church in Apollonia²⁵ (see [Figure 6.1](#)) or a reference for the Arabic name of the ancient settlement of Thenadassa, Ain Wif,²⁶ both occurring in Ward-Perkins's collection of Libyan photographs. Likewise, HGL identifiers have been used by Manar Libya, a spin-off of the seminal Oxford digital archaeological library Manar al Athar,²⁷ focusing on Libyan heritage across time periods. While, obviously, the simple use of place URIs does not translate immediately into becoming an active component of an LOD network, it surely contributes to make the metadata one step closer to this final, ambitious goal and could be exploited to create different internal exploration pathways.

Mapping Roman Libya with students

If it is true that urban gazetteers may enable more detailed and specific connections, the modelling of single buildings as places poses a few

additional challenges. Compared to regions or settlements, edifices are more ephemeral and are frequently subject to phenomena such as renovations, partial or complete destructions, additions and spolia reuse. Researchers introducing urban-level IDs also seem to agree that the function served by the building is an important aspect of its ‘being a place’ and, as such, should be recorded in an historical gazetteer alongside its more tangible components ([Jenstad 2016](#)). Due to the experimental nature of this knowledge-modelling approach in 2016, the HGL team decided to set aside both the information about the buildings’ functions as well as the complex recording of their transformations through time. Given the heritage focus of the dataset, buildings in the HGL gazetteers were modelled based on their status as recorded at the time of the excavation reports. The real limitation to the increase of the HGL’s granularity, however, was not the complexity of the data model but the lack of resources, as the funding for the HGL no longer covered data collection and enhancement.

Heritage Gazetteer of Libya / SLS Archives

Homepage Search Gazetteer Search Archive Search People About the projects Contact

Central Church, Apollonia

URI ↗

Coordinates

Latitude: 32.902028, Longitude: 21.9685 Provenance: GPS Walda
 Latitude: 32.901982, Longitude: 21.968323 Provenance: Google Earth

Variant names

- الكنيسة المركزية (Arabic)
- Basilica Centrale (Italian)
- Central Basilica (English)
- Eglise centrale (French)

Feature type(s)

Church/monastery/Christian building(s)

Relationships with other locations ?

Parent features ?

- Central Church, Apollonia forms part of Susah

Child features ?

Notes

VI century AD church in Apollonia, excavated and partially restored in 1940 (and 1950); Kenrick, site plan 4

Related catalogue entries

- Plans of Various Libyan Sites, Including Apollonia, Ghirza and Tocra., SLS Archive
- Central Church., SLS Archive
- Apollonia., SLS Archive
- "Cyrenaica - 1971", SLS Archive

External URIs

http://www.manar-al-athar.ox.ac.uk/dams/?c=1749&k=e7f80751f1

Figure 6.1: Central Church, Apollonia in the Heritage Gazetteer of Libya: https://slsgazetteer.org/irt_geo/recordview/?id=921 (CC BY-SA-NC 4.0)

An opportunity to improve the quantity and quality of the data in the gazetteer, as well as to promote its use to a different audience, presented itself during unexpected circumstances. In 2020, due to the COVID-19 pandemic and its related restrictions, the internship offered to undergraduate students of archaeology at University College London (UCL) had to be radically rethought in order to be delivered fully online. Although online teaching was then much more of an uncharted territory, the module convenor, Prof. Corisande Fenwick, decided to transform the internship into a ‘virtual excavation into archaeological archives’. Being at the time also BILNAS’s director, Fenwick’s choice to work on the institute’s archive was a natural one. The internship module was then renamed ‘Mapping Roman Libya’, and featured a selection of digital approaches that aimed at

developing in students a more critical approach to archival and archaeological records, leveraging spatial perspectives.²⁸

As part of this internship, the students were introduced to spatial humanities tools and methodologies ([Dunn and Vitale 2021](#)) that included semantic geo-annotation, principles of GIS, introduction to geodata formats and, of course, historical gazetteers. The potential for HGL identifiers to become aggregators of archival resources as well as to enable spatial-oriented analysis took a prominent role during the internship. Among the practical tasks the students were asked to complete was that of creating new entries for the HGL. In particular, the students were involved in the ongoing experiment with the enhanced granularity of HGL records and were invited to create individual identifiers, and associated entries, for the extant buildings in the ancient city of Sabratha.

An ad hoc workflow was devised to harmonise the module's learning objectives, the HGL growth, COVID restrictions and the students' digital skills. The first step in this workflow involved using a digitised version of the archaeological map of Sabratha published in Kenrick's archaeological guide as a compass to identify the major monuments and divide them among the students. Using Kenrick's map as a reference, the students had to locate the building on Google Maps and Google Earth georeferenced satellite images. In order to find the clearest available photograph, students were invited to take advantage of the Timelapse feature, which enables users to look at images of the same area taken at different points in time. This process also highlighted the rapid changes in the archaeological landscape, and made students aware of the urgency of recording such information. Once the chosen building had been located in Google Earth, the students could trace its mass base geometry and derive its coordinates. The simplified polygon was then turned into a GeoJSON file with the help of the online widget [geojson.io](#), which was also employed to check the file's validity. The simple GeoJSON code was then added by the students as 'location' in the entry they had created in the HGL for the building. The building's geometry would appear in the map visualisation associated with the HGL place. The students were responsible for populating all the fields related to the new place entry, including relationships with other places and alternative names (when appropriate). Within a few weeks, all major extant buildings in Sabratha appeared as individual entries in the HGL. The

enthusiastic contribution of this cohort of archaeology undergraduate students made it possible to showcase the potential of building-level identifiers for an entire ancient city, and stimulated reflections on new ways to leverage these increasingly granular connections. On the other hand, the students' dedication evidenced an important flaw in the HGL interface: names of contributors were not displayed in the public, browsable pages. Such absence not only felt unfair towards the new contributors but also revealed that much work had to be done to achieve a higher level of transparency.²⁹

The same group of students helped the HGL team to test another possible development of the gazetteer that involved using a customised version of [Hypothes.is](#)³⁰ to turn web annotations directly into active links in the HGL entries ([Puig et al. 2018](#); [Hanson Shrout 2016](#)). Students were invited to use this new widget in the pages they had created for Sabratha's buildings, to generate links to any related resource they could find online. The latter could be Wikipedia or Wikimedia entries, museum pages for artefacts associated with the building, or even Flickr photos (as long as they were published under a suitable licence). The experimental addition of the [Hypothes.is](#) widget positioned the HGL in a liminal space, in between being a curated authority list for places and a community-oriented aggregator of resources related to Libyan heritage. The students' contributions were very useful in steering the identity of the gazetteer and shaping its scope. The number of connections generated in just a few weeks confirmed the enormous potential for engagement and the added value of place-based connections. However, driven by the students' idiosyncratic preferences and inclinations, the slightly chaotic addition of these external connections overloaded the gazetteer from a visual and informational point of view. Produced in high number, connections were sometimes duplicated,³¹ generating redundancy and, possibly, even a false sense of completeness. The general ephemerality of web resources would also make it very challenging for the HGL team to check and guarantee validity of all the external links through time. More than revealing the practical barriers to this approach, the more valuable outcome of this test was, perhaps, that of confirming that gazetteers are most successful when they perform a single, well-defined function: collect and display reliable information about places in structured data format, under a given perspective. The rich potential of

gazetteers' URIs for the enhancement of archival resources, research or community engagement is probably better left to be explored by separate and dedicated projects.

Back to the data

As an open gazetteer, the HGL exists as a digital infrastructure (with its web URL, front and back-end, and user interface) but also as a downloadable and reusable dataset in GeoJSON format. As such, the HGL can become part of other gazetteer-based projects and resources. For example, HGL identifiers are now featured as external authorities in Wikidata entries for Libyan places. Recently, the HGL dataset has also been ingested by the World Historical Gazetteer (WHG) project, becoming part of an open, global historical resource that also relies substantially on Wikidata for reconciling data coming from different sources.³²

The HGL dataset, now fairly grown in scope and complexity compared to the initial epigraphic spreadsheet, also allows for more computationally oriented reuses. While it is impossible to know when and how researchers decide to use this resource, we can document at least one of these applications. At the Digital Humanities 2022 conference, the HGL GeoJSON dataset was used as a case study in the workshop³³ showcasing the application of DeezyMatch ([Hosseini et al. 2020](#): 62–69), an open-source software library for fuzzy string matching and candidate ranking that was developed within the Living with Machines project.³⁴ Combining spatial coordinates with information on Arabic placenames and their many variants and transliterations in several languages, mostly English, French and Italian, the HGL presented itself as a valuable training dataset to improve prediction of less obvious matches between geographic entities with multiple names, recording patterns of linguistic transformation from Arabic to other languages, or even recurring transliteration mistakes (learnt from the attestations in BILNAS's archaeological sources). Besides its pedagogical aim within the workshop, this experiment with DeezyMatch ultimately wanted to explore possible ways to improve Wikidata's own automatic matching mechanism, to support the pairing of HGL IDs with

those of Wikidata, thus strengthening the consistency and coherence of both resources, and help spot gaps and errors.

Closing the circle or, better, riding the spiral

After discussing how the initial dataset created for IRT and IRCyr has expanded and transformed, it is perhaps natural to guess that, after a few years, the paths of the digital publications of Libyan epigraphy and the HGL data have crossed again. In the new updated and enriched edition of IRT, places relevant to the epigraphic corpus appear associated with HGL identifiers. Compared to the first digital edition, which mostly leveraged the few available Pleiades IDs and manually encoded geocoordinates, the new one could benefit from the improved consistency between places with alternative names, including names that have changed through time, variant transliterations and different languages. The HGL identifiers could also, indirectly, offer information about relationships between the places associated with the epigraphic sources, such as places being part of a certain area or sub-area. The enhanced granularity of the data in the HGL made it, overall, a more useful tool to discover and highlight connections between the artefacts, and between the artefacts and the places.

The several additions and transformations of this open geo-dataset, along with the fast-paced growth of digital applications for the study of the ancient world, have opened up new scenarios for reuse. Each, potentially, brings new developments, in content or methodology, to the dataset itself.

The BSR reuse of HGL IDs in their metadata for artefacts connected to Libya is one of the inspirations for a more ambitious project, funded by the British Academy,³⁵ that aims to leverage common place- and person-references to discover and display connections not only between BSR and BILNAS but across all British International Research Institutes (BIRIs).³⁶ Considering the different digital literacies of each BIRI, the project has started with an investment in capacity-building and has looked at the HGL as a possible model to be efficiently replicated in other institutes that want to create their own gazetteers. The encouragement of an LOD-oriented perspective has also facilitated dialogue on best practice around the modelling of person-data, originating a dedicated spin-off project. As a

consequence, the HGL has now expanded its scope beyond places, and is hosting a small sample of person-data, in support of this proof-of-concept.

While the end of the pandemic does not make it necessary for archaeology undergraduate students to attend virtual internship modules such as Mapping Roman Libya anymore, the HGL has been used in teaching again more recently, as part of the Sunoikisis Digital Cultural Heritage programme.³⁷ The 2022 seminar on computational approaches to digitised map collections combined the HGL framework with the new ‘search by word’ function available on the David Rumsey online map collection ([Vitale 2023](#)).³⁸ The interface on Rumsey’s digital library enables users to search a large number of georeferenced, digitised historical maps as if they were digitised books. The search allows for basic regular expressions as well as variations within one character. In the SunoikisisDC seminar, this latter function is creatively exploited to search for Libyan places in Rumsey’s collection and find new and less expected attestations on the maps that are currently not present in the HGL nor Wikidata.³⁹ The exercise had a twofold aim: showcase the potential of computer vision for digitised map collections, but also demonstrate how these automated methods can become a powerful means to expand digital gazetteers through historical attestations that come directly from maps ([Larsen et al. 2024](#)).

We would like to conclude this journey by mentioning one more possible reuse/application of the geo-dataset that is not yet implemented but could become the next step in the HGL spiral-shaped life trajectory. This application was supposed to be one of the HGL’s main goals from the beginning, that is the opportunity to involve Libyan stakeholders in the enrichment of the dataset. Learning from past mistakes, in 2023 BILNAS started a collaboration with Scene,⁴⁰ a Libyan group of heritage experts and activists based in Tripoli. Thanks to seed funding from The National Archives,⁴¹ BILNAS and Scene organised a small number of workshops, in English and Arabic, to record views, memories and interpretations about selected photos in the archive. Despite its small scale, the project was successful in creating, for the first time, a group of interested and active Libyans, which opened up communication channels with other local groups and associations, including Libyan archives themselves. These conversations also hold promising potential for future inclusion of Libyan contributions to the HGL, especially around unofficial names or local

variations. Ideally, the HGL could even play a key role in the attempt to connect dispersed archival resources about Libya, including those hosted in Libya itself.

Conclusions

In this chapter we have looked at a geo-historical dataset created around Libyan epigraphy to discuss the role of open data in research, and to argue that open data sharing is a crucial component of innovation. The emphasis on reproducibility and reusability that often accompanies the discourse around open data ([Middle 2023](#)) does not necessarily imply that data and processes have to be replicated in a prescriptive way. On the contrary, open and documented datasets offer the opportunity to understand the way in which data have been created, so that they can be tweaked, changed and creatively reused (and even misused) beyond the intention of the original creators. It is a transparent history of changes and transformation that helps guarantee the integrity of the documents and resources.⁴² Retracing the eventful life of this dataset, we wanted to show not only the usefulness of data reuse but also, or perhaps in particular, how each reuse usually brings transformations and improvements, not only in the quantity and quality of the data, but also in the methodologies necessary to create them, substantially contributing to both praxis and theory.

Acknowledgements

The author would like to acknowledge the work of the team who developed the Heritage Gazetteer of Libya, under the direction of Charlotte Roueché: Hafeed Walda, Gabriel Bodard, Stuart Dunn, Neil Jakeman and Ginestra Ferraro. The gazetteer is currently hosted and maintained by King's Digital Lab. The author would also like to thank the editor, Gabriel Bodard, for his friendly advice and all the reviewers who provided useful and insightful feedback: Sarah Middle, J.A. Baird, Maxime Guénette and Elizabeth Koch Kölük. Last, the author wishes to express her profound gratitude to Charlotte Roueché for being an extraordinary model, mentor and friend.

The structure and data of the HGL are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Notes

1. IRCyr ([Reynolds et al. 2020](#)) was edited by J.M. Reynolds, C. Roueché and G. Bodard, in collaboration with C. Dobias-Lalou. The digital edition was designed and prepared by G. Bodard, S. Stoyanova and I. Vagionakis and published in 2020 by the Society for Libyan Studies (now British Institute for Libyan and Northern African Studies).
2. IRT ([Reynolds et al. 1952](#)) by J.M. Reynolds and J.B. Ward-Perkins with contributions from S. Aurigemma, R. Bartocchini, G. Caputo, R. Goodchild and P. Romanelli was first published for the British School at Rome in 1952. An ‘enhanced electronic reissue’ ([Reynolds et al. 2009](#)) edited by G. Bodard and C. Roueché, with new translations by J. Reynolds, maps by H. Walda and illustrations from the Ward-Perkins photographic archive of the British School at Rome, was published online in 2009. An expanded edition ([Reynolds et al. 2021](#)) incorporates contributions from C. Barron, F. Bigi, C. Dobias-Lalou, U. Gad, P. Kenrick, R. Kerr, M. Mackensen and I. Tantillo.
3. The EpiDoc Guidelines ([Elliott et al. 2017–2025](#)) are themselves an open resource developed, updated and maintained by a community of users.
4. The obvious advantages of text encoding and disambiguation against authority lists are exemplified by the enhanced search capabilities within single epigraphic corpora, such as the one produced by the IOSPE project (available at: <https://iospe.kcl.ac.uk/>), or even across different corpora as explored by the Crossreads project (<https://crossreads.web.ox.ac.uk/>).
5. The IRCyr and IRT teams were largely overlapping and, in both projects, the initial work on locations was undertaken by Walda.
6. An easy example is offered by one of the best known ancient Libyan cities, قناتريس. The current Arabic name appears in different documents transliterated as ‘Sabrata’, ‘Sabratha’, ‘Siburata’ and ‘Ṣabrātah’.
7. In its most abstract terms, a URI is defined by [Berners-Lee et al. \(2005\)](#) as ‘a compact sequence of characters that identifies an abstract or physical resource’.
8. As of June 2025, Pleiades features 41,620 places and 39,220 names while GeoNames counts 25 million geographical names and over 11.8 million unique features.
9. Charlotte Roueché, Hafed Walda, Valeria Vitale and Neil Jakeman.
10. See Dunn 2026 ([Chapter 5](#) in this volume).
11. <https://slsgazetteer.org/1680>.
12. <https://slsgazetteer.org/1679>.
13. For example, Zawiyat Asqafah (<https://slsgazetteer.org/1178>) can be found in both Reynolds’s and Kenrick’s Cyrenaica, while Haqfat at-Tayrah (<https://slsgazetteer.org/1605/>) is only recorded by Kenrick as belonging to Cyrenaica.
14. On the factoid model and its application in prosopographies, see Andrews 2026 ([Chapter 1](#) in this volume).

- 15.** This approach is also the methodological framework of the gazetteer of time periods PeriodO, discussed in [Golden and Shaw \(2016\)](#).
- 16.** Most notably within the ancient city of Pompeii, but also for example Rome, Athens and Aphrodisias.
- 17.** Formerly the Society for Libyan Studies.
- 18.** See <https://archaeologydataservice.ac.uk/archives/collections/view/1007105/index.cfm>.
- 19.** Born from the especially complex relationship between indigenous tribes and their heritage as digitised by colonial and imperialist institutions, these principles of ethical access and management of digital archives have subverted the concept of access and ‘openness’, inviting researchers and heritage practitioners to reflect on who has the right to decide what information is made available online and to what degrees.
- 20.** The new gazetteer interconnection format, Linked Places, was developed within the World Historical Gazetteer project. Its specifications can be found at: https://github.com/LinkedPasts/linked_places-format.
- 21.** The latest concept overview of the Pleiades gazetteer (published in 2020) now defines as its scope ‘any spatial feature that is connected to the pre-modern past and that a human being has noticed and discussed as such between the past and the present’. See <https://pleiades.stoa.org/help/conceptual-overview>.
- 22.** In [Elliott’s 2019 Description of Pleiades](#) (<https://pleiades.stoa.org/docs/papers-and-presentations/description-of-pleiades-2019>), the scope of the gazetteer is still officially described as ‘the ancient Mediterranean world’.
- 23.** The Libya Committee (later Society for Libyan Studies) was created in 1968 by members of the British School at Rome and Athens to honour the memory of Richard Goodchild.
- 24.** Available at: <https://www.getty.edu/research/tools/vocabularies/tgn/>.
- 25.** See the record for Apollonia’s Central Church in the Ward Perkins Libya Collection: <https://digitalcollections.bsr.ac.uk/islandora/object/WP-LIBYA%3A3232>.
- 26.** https://slsgazetteer.org/irt_geo/recordview/?id=677.
- 27.** Manar al Athar community sources photographic archive is available at: <https://www.manar-al-athar.ox.ac.uk/>.
- 28.** The digital component of the module was designed and delivered by the author, as part of BILNAS’s Kathleen Kenyon Fellowship.
- 29.** The lack of recognition of voluntary work, and especially student work, in collaborative projects is a key issue in the digital humanities, and was already pointed out by [Nowviskie \(2011\)](#). While not a perfect solution, gazetteers such as Pleiades that make every contribution traceable back to the originating user account (i.e. treating every contribution as a ‘nanopublication’) represent the current best practice. While HGL development is currently frozen due to lack of funding, redesigning the public interface in order to display contributors’ usernames is already flagged as a top priority.
- 30.** An annotation layer that was, at the time, free to use and modify. The customised version was developed by Neil Jakeman.
- 31.** For example, when an ancient artefact had both a Wikipedia entry and a page on the British Museum website.
- 32.** See <https://reviewsindh.pubpub.org/pub/world-historical-gazetteer/release/1>.
- 33.** The programme of the workshop can still be accessed on the conference website at <https://dh2022.adho.org/workshops-and-tutorials/wt-13>.

34. See <https://livingwithmachines.ac.uk/>.
35. The Digital Collaboration and Strategy project, which started in 2022 and is still ongoing. The initial application was written by Charlotte Roueché, Lutgarde Vanderput, Valeria Vitale and Rainer Simon.
36. Further discussed in Giovenco and Atalan Çayırezmez, 2026 ([Chapter 4](#) in this volume).
37. The SunoikisisDC programme produces and makes available open educational resources for the study of digital classics and digital cultural heritage: <https://sunoikisisdc.github.io/>.
38. <https://www.davidrumsey.com/>.
39. The open online seminar is available from <https://github.com/SunoikisisDC/SunoikisisDC-2023-2024/wiki/3.-Gazetteers>.
40. <https://web.archive.org/web/20260412185355/https://www.scene.org.ly/>.
41. The project was written and delivered by Felicity Crowe (then BILNAS archivist), Reem Furjani (Scene) and Valeria Vitale.
42. The epistemological and ethical dangers connected to unrecorded transformation of the data in archival collections are discussed in [Baird \(2023\)](#).

References

- [Ahlers, D. \(2013, November\)](#) ‘Assessment of the Accuracy of GeoNames Gazetteer Data’, in *Proceedings of the 7th Workshop on Geographic Information Retrieval*, pp. 74–81.
- [Asciutti, V. and Dunn, S. \(2011\)](#) ‘Connecting the Classics: A Case Study of Collective Intelligence in Classical Studies’, *The Digital Classicist 2013*, Bulletin of the Institute of Classical Studies, supplement 122, pp. 147–60. <https://doi.org/10.14296/917.9781905670802>.
- [Baird, J.A. \(2023\)](#) ‘Unclassified: Structured Silences in the Archaeological Archive’, in R. Raja (ed.) *Shaping Archaeological Archives: Dialogues Between Fieldwork, Museum Collections, and Private Archives*. Turnhout, Belgium: Brepols, pp. 19–32.
- [Barker, E., Simon, R., Isaksen, L. and de Soto Canameres, P. \(2016\)](#) ‘The Pleiades Gazetteer and the Pelagios Project’, in M.L. Berman, R. Mostern and H. Southall (eds) *Placing Names: Enriching and Integrating Gazetteers*. Spatial Humanities. Bloomington: Indiana University Press, pp. 97–109.
- [Bennett, P. and Barker, G. \(2011\)](#) ‘Protecting Libya’s Archaeological Heritage’, *African Archaeological Review*, 28, pp. 5–25.
- [Berman, M.L., Ahlfeldt, J. and Wick, M. \(2012\)](#) *Historical Gazetteer System Integration: Chgis, Regnum Francorum, and Geonames. Placing*

- Names: Enriching and Integrating Gazetteers*. Bloomington and Indianapolis: Indiana University Press.
- Berners-Lee, T., Fielding, R.T. and Masinter, L. (January 2005), *RFC 3986 Uniform Resource Identifier (URI): Generic Syntax*. Internet Engineering Task Force.
- Bodard, G. (2010), ‘EpiDoc: Epigraphic Documents in XML for Publication and Interchange’, in F. Feraudi-Gruénais (ed.) *Latin On Stone: Epigraphic Research and Electronic Archives*. Lanham, MD: Lexington Books, pp. 101–18.
- Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S. and Schaub, T. (2016), *The GeoJSON Format* (No. rfc7946).
- Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020), ‘The CARE Principles for Indigenous Data Governance’, *Data Science Journal*, 19(1), p. 43. <https://doi.org/10.5334/dsj-2020-043>.
- Cayless, H., Roueché, C.M., Elliott, T. and Bodard, G. (2009), ‘Epigraphy in 2017’, *Digital Humanities Quarterly*, 3.1(2009).
- Chen, A. (2022), ‘Digital Archaeological Archives 2.0: Answering Past Critiques with Linked Open Data (LOD)’, in O. Bobou, A.C. Miranda and R. Raja (eds) *Archival Historiographies*. Brepols Online, pp. 41–53.
- Cobb, J. (2015), ‘The Journey to Linked Open Data: The Getty Vocabularies’, *Journal of Library Metadata*, 15(3–4), pp. 142–56.
- Davis, E. and Heravi, B. (2021), ‘Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation’, *Journal on Computing and Cultural Heritage*, 14(2), pp. 1–18.
- Dunn, S. (2016), ‘Praxes of “The Human” and “The Digital”: Spatial Humanities and the Digitization of Place’, *GeoHumanities*, 3(1), pp. 88–107.
- Dunn, S. and Vitale, V. (2021), ‘Linking Text and Maps: Annotation as a Critical Tool for Teaching in the Spatial Humanities’, *Literary Geographies*, 7(2), pp. 292–310.
- Dunn, S., Papacostas, T. and Roueche, C.M. (2015), ‘Heritage Gazetteer of Cyprus’. Available at: <https://cyprusgazetteer.org/>.
- Elliott, T. (2021), ‘The Pleiadic Gaze: Looking at Archaeology from the Perspective of a Digital Gazetteer’, in *Archaeology and Economy in the*

Ancient World: Proceedings of the 19th International Congress of Classical Archaeology, Cologne/Bonn, 2018, pp. 43–51.

Elliott, T. (2019), ‘Description of Pleiades’. Pleiades. Available at: <https://pleiades.stoa.org/docs/papers-and-presentations/description-of-pleiades-2019>.

Elliott, T. and Gillies, S. (2008), ‘Pleiades: The un-GIS for Ancient Geography’, *International Journal of Geographical Information Science*, 22, pp. 1,091–1,108.

Elliott, T., Bodard, G., Mylonas, E., Stoyanova, S., Tupman, C. and Vanderbilt, S. (2017–2025), *EpiDoc Guidelines: Ancient documents in TEI XML* (Version 9). Available at: <https://epidoc.stoa.org/gl/latest/>.

Elliott, T., Heath, S. and Muccigrosso, J. (2012), ‘Report on the Linked Ancient World Data Institute’, *Information Standards Quarterly*, 24(2/3), pp. 43–5.

Golden, P. and Shaw, R. (2016), ‘Nanopublication Beyond the Sciences: The PeriodO Period Gazetteer’, *PeerJ Computer Science*, 2, e44.

Grossner, K. and Mostern, R. (2021, November), ‘Linked Places in World Historical Gazetteer’, in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Geospatial Humanities*, pp. 40–43.

Hanson ShROUT, A. (2016), ‘Hypothes.is’, *Journal of American History*, 103(3), pp. 870–71.

Horne, R. (2020), ‘Beyond Lists: Digital Gazetteers and Digital History’, *The Historian*, 82(1), pp. 37–50.

Hosseini, K., Nanni, F. and Ardanuy, M.C. (2020, October), ‘DeezyMatch: A Flexible Deep Learning Approach to Fuzzy String Matching’, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 62–9.

Jenstad, J. (2016), ‘Building a Gazetteer for Early Modern London, 1550–1650’, in M.L. Berman, R. Mostern and H. Southall (eds) *Placing Names: Enriching and Integrating Gazetteers*. Bloomington: Indiana UP, pp. 129–45.

Kahn, R., Isaksen, L., Barker, E., Simon, R., de Soto, P. and Vitale, V. (2021), ‘Pelagios—Connecting Histories of Place. Part II: From Community to Association’, *International Journal of Humanities and Arts Computing*, 15(1–2), pp. 85–100.

- Kenrick, P. (2009), *Libya Archaeological Guides: Tripolitania*. London: Society for Libyan Studies.
- Kenrick, P. (2013), *Libya Archaeological Guides: Cyrenaica*. London: Society for Libyan Studies.
- Larsen, K.L., Thornberry, E. and Vitale, V. (2024), 'Teaching and Learning with Text on Maps: A David Rumsey Collection Case Study', *Imago Mundi*, 76(2), pp. 290–96.
- Middle, S. (2023), 'A Documentation Checklist for (Linked) Humanities Data', *International Journal of Digital Humanities*, 5(2), pp. 353–71.
- Nebbia, M., Leone, A., Bockmann, R., Hddad, M., Abdouli, H., Masoud, A.M., Elkendi, N.M., Hamoud, H.M., Adam, S.S. and Khatab, M.N. (2016), 'Developing a Collaborative Strategy to Manage and Preserve Cultural Heritage During the Libyan Conflict. The Case of the Gebel Nāfusa', *Journal of Archaeological Method and Theory*, 23(4), pp. 971–88. <https://doi.org/10.1007/s10816-016-9299-6>.
- Nowviskie, B. (2011), 'Where Credit is Due: Preconditions for the Evaluation of Collaborative Digital Scholarship', *Profession*, pp. 169–81.
- Prag, J.R.W. and Chartrand, J. (2018), 'I. Sicily: Building a Digital Corpus of the Inscriptions of Ancient Sicily', in A. De Santis and I. Rossi (eds) *Crossing Experiences in Digital Epigraphy. From Practice to Discipline*. Warsaw: De Gruyter, pp. 240–52. <https://doi.org/10.1515/9783110607208-020>.
- Puig, V., Gilmozzi, G. and Haussonne, Y.M. (2018, October), 'Contributive Research: Hypothes.is Implementation for Academic Research Purpose', in *Proceedings of the 2nd International Conference on Web Studies*, pp. 19–23.
- Rayne, L., Sheldrick, N. and Nikolaus, J. (2017), 'Endangered Archaeology in Libya: Recording Damage and Destruction', *Libyan Studies*, 48, pp. 23–49.
- Reynolds, J.M., Roueché, C.M. and Bodard, G. (2020), *Inscriptions of Roman Cyrenaica*. London: King's College London. Available at: <http://ircyr2020.inslib.kcl.ac.uk/>.
- Reynolds, J.M., Roueché, C.M., Bodard, G. and Barron, C. (2021), *Inscriptions of Roman Tripolitania 2021*. London: King's College London. Available at: <http://irt2021.inslib.kcl.ac.uk/>.

Reynolds, J.M., Ward-Perkins, J.B., Aurigemma, S., Bartoccini, R., Caputo, G., Goodchild, R. and Romanelli, P. (1952) *The Inscriptions of Roman Tripolitania*. Rome: British School at Rome.

Reynolds, J.M., Ward-Perkins, J.B., Roueché, C.M. and Bodard, G. (2009) *Inscriptions of Roman Tripolitania*. London: King's College London.
Available at: <http://inslib.kcl.ac.uk/irt2009/>.

Roueché, C.M. (2018) 'Libya Matters: London 2016', *Quaderni di Archeologia della Libya*, 21(nsI), pp. 359–62.

Roueché, C.M., Walda, H.M., Vitale, V. and Jakeman, N.A. (2016) *Heritage Gazetteer of Libya*. Available at: <https://slsgazetteer.org/>.

Simon, R., Isaksen, L., Barker, E. and de Soto Cañamares, P. (2016a) 'Peripleo: A Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time', *The Code4Lib Journal*, 31. Available from: <https://journal.code4lib.org/articles/11144>.

Simon, R., Isaksen, L., Barker, E. and de Soto Cañamares, P. (2016b) 'The Pleiades Gazetteer and the Pelagios Project', in M.L. Berman, R. Mostern and H. Southall (eds) *Placing Names: Enriching and Integrating Gazetteers*. Bloomington: Indiana University Press, pp. 97–109.

Vitale, V. (2023) 'Searching Maps by Words: How Machine Learning Changes the Way We Explore Map Collections', *Journal of Cultural Analytics*, 8(1). <https://doi.org/10.22148/001c.74293>.

Vitale, V. (Forthcoming) 'Modelling Ancient Places in Historical Gazetteers', in J. Corrigan and A. Gardner (eds) *Oxford Handbook of Spatial Humanities*. Oxford: Oxford University Press.

Vitale, V., de Soto Cañamares, P., Simon, R., Barker, E., Isaksen, L. and Kahn, R. (2021) 'Pelagios—Connecting Histories of Place. Part I: Methods and Tools', *International Journal of Humanities and Arts Computing*, 15(1–2), pp. 5–32.

Chapter 7

The makeup of *Ajax*: reconstructing networks of interpretation through the commentary tradition

Charles Pletcher and Matteo Romanello

Introduction

In a note from 2018, John Davidson divides readers of the so-called *Trugrede* ('Deception Speech') of Sophocles' *Ajax* (ll. 646–92) into two camps: the proponents of modern orthodoxy, who maintain that Ajax already intends to die by suicide during this speech; and the unfashionable—which is not to say heterodox—view, championed by the likes of T.B.L. Webster and Maurice Bowra, that 'in the *Trugrede* Ajax had genuinely changed his mind and was now not intending to kill himself but to seek reconciliation, only for Athena's anger to catch up with him when he got to the shore, and abruptly and totally return him to his former suicidal state of mind' ([Davidson 2018](#): 472). As Davidson's own, at times lengthy, qualifications attest, the line between these two camps does not always appear neatly—as he writes, 'It would be unhelpful to rehearse in detail all the different orthodox interpretations of the speech, especially as found in the scholarship of much earlier generations, but an indication of the range that they have covered in more recent years is instructive' ([Davidson 2018](#): 475). With so many variations on the side of orthodoxy, two camps can hardly suffice to describe the whole range of interpretations on the *Trugrede*—to say nothing of the other disagreements that the play has provoked, such as whether Ajax's shield ever appears onstage, or how Sophocles

intended to stage Ajax's suicide. We need to return to a foundation of sorts, and we can do so by examining the commentary tradition of the past two centuries. These commentaries, more even than the articles that engage in the real mud-wrestling of interpretation, lay the groundwork for understanding how we might view the *Trugrede* (to take a single example from the disagreements above) and where we ought to draw interpretive lines in the sand.

To contribute to this foundation, we have built the *Ajax* Multi-Commentary (AjMC).¹ This platform collects seventeen modern commentaries on *Ajax*—although the public-facing version can only show the seven commentaries that are in the public domain²—and makes them simultaneously accessible, linked to the modern Oxford Classical Text edition, without the apparatus criticus, which remains under copyright ([Lloyd-Jones and Wilson 1990](#)). By 'simultaneously accessible' we mean that the glossae—interpretative notes that explicate and interpret words or passages—of the commentaries are presented in a single column, with each glossa linked to the line(s) or lemma(ta) to which it refers.³ The glossae themselves are initially collapsed, but opening one shows an automatic transcription of the text, using optical character recognition (OCR), as well as the option to display a facsimile of the page from which the glossa came, with the relevant portion highlighted. The OCR process can transcribe multilingual commentaries with state-of-the-art accuracy, but it is not perfect. Uncorrected errors might still appear in the transcriptions, which is one of the reasons why we also provide a facsimile of the corresponding page.

In the critical text of Sophocles, each glossed lemma and line is highlighted in blue, with darker shades indicating a greater density of glossae from different commentaries. We provide further detail on the construction and implementation of the Multi-Commentary below, but it is perhaps most helpful at this point to see the outcome of this feature, which creates a heat map of sorts ([Figure 7.1](#)).

In this chapter, we use the AjMC to investigate not only which commentaries say what when it comes to the *Trugrede*, but also *where* they say it—that is, to what lines and lemmata they anchor their analyses ([Pletcher, Najem-Meyer and Romanello 2024a](#)). This kind of analysis, while technically possible using paper commentaries, does not scale well in

analogue form: one needs a large desk in order to hold seventeen commentaries open at once, but when we consider the additional material cited by these commentaries and the information that we can add to them, such as links to authoritative entries on named entities, the value of the born-digital multi-commentary comes into clearer focus. The *Ajax* Multi-Commentary allows for reading multiple commentaries in a relatively compact virtual space, visualising the extent of their overlap through the citation heat map, and quickly accessing additional references through our own annotations on the commentaries. Further, it supports exporting the number of glossae per lemma as tabular data for more in-depth statistical analysis.

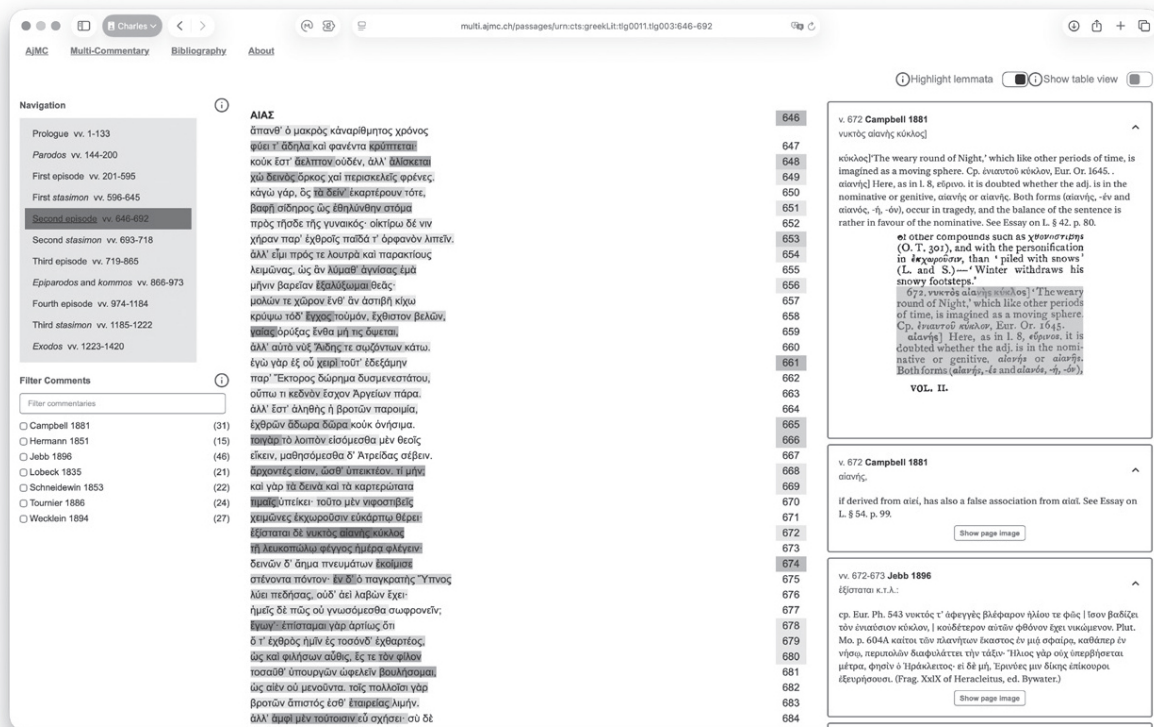


Figure 7.1: A screenshot of the *Ajax* Multi-Commentary, showing navigation on the left, the critical text in the centre and example glossae on the right. The critical text is highlighted on lines and lemmata, with darker shades corresponding to a greater density of glossae (CC BY)

By engaging with the commentaries on the *Trugrede* in this way, we aim to show first of all that the commentary tradition demands, to a measurable

extent, that we return to texts that have been considered obsolete not because of their antiquated methods but rather because of the passage of time.⁴ Further, we aim to provide access to these older commentaries in a way that takes advantage of the hypertextual ability to link references and display multiple commentaries at once, demonstrating the value of approaching a text through a digital multi-commentary. Finally, the heat map of citations visualises commentators' emphases and thus offers a new kind of guide to their interpretations: at a glance, we can see the words and phrases that attracted the most scholarly attention by looking for darker shades of blue. The heat map brings the commentators' focus into focus, as it were, and clarifies how their interpretations might drive our own.

Assembling the Multi-Commentary

Seventeen commentaries on Sophocles' *Ajax* were digitised and parsed using an OCR, natural language processing (NLP) and named entity recognition (NER) pipeline.⁵ Particularly important for the purposes of this chapter, the lemmata and citations in each commentary were linked to their corresponding lines in a common edition of the play, the Oxford Classical Texts version by Lloyd-Jones and Wilson. This linking process allows us to generate a 'heat map' based on how densely commented a given lemma is, providing a visualisation of what areas of the text have attracted commentators' attention. This visualisation, as we discuss below, guides our metatextual analysis of the commentaries by identifying areas of concentration and dispersion in their own analyses of the text.⁶

The lemma-linking process was performed by several annotators, including the authors, using the INCEpTION annotation tool alongside a bespoke implementation of CTS URN—Canonical Text Services Uniform Resource Names—citations in the Lloyd-Jones and Wilson text of *Ajax*.⁷ CTS URNs grew out of the Homer Multitext project's need for citations that can address multiple editions of a text. CTS URNs provide markers that can point both to specific lemmata within a given edition of a classical text and to glossae in a commentary.⁸ Importantly, CTS URNs can also point to an abstract work, rather than a specific edition—this feature is particularly helpful when dealing, for example, with Plato's citations of Homer, where

we may not want to resolve them by pointing to a Homer based on a different tradition. For each glossa in a commentary, we can assign a CTS URN that points to the specific line(s) or lemma(ta) that the glossa references in Sophocles' *Ajax*. Simultaneously, we can assign URNs to each glossa, so that readers can navigate the text via the URNs of glossae—that is, by reading the commentaries in a nonlinear fashion—or via the URNs of the text of Sophocles—that is, focusing on the critical text.⁹

We diverge from the official CTS URN specification in one important respect, although this divergence might best be understood as an expedience during the proof-of-concept phase of the project. Rather than tying the CTS URNs to specific tokens during the annotation process, we link them to character offsets in each line, allowing us to approximate citations for other editions of Sophocles' text.¹⁰ We have thus created a multi-commentary not only in terms of the number of commentaries included but also in terms of the number of critical texts. However, these citations are not always accurate, as they are linked not to the underlying word tokens but to the integer offsets of Unicode glyphs in a line, meaning in editions other than Lloyd-Jones and Wilson, citations sometimes include partial or incorrect word tokens. Nevertheless, as a secondary point of comparison—the main focus remains on Lloyd-Jones and Wilson—the linking in other editions is particularly useful when commentators disagree with the OCT text.

Notably, we have not included ancient scholia in the collection of commentaries. In part, we have excluded them because the project focuses on the modern commentary tradition, but a further reason for their exclusion arises from the difficulty of aligning both multiple editions of the critical text of *Ajax* and multiple editions of the texts of the scholia. As each scholion assumes a particular edition of the text, whether or not that edition is reproduced alongside the glossae, it quickly becomes intractable to display multiple versions of both commentary and text. Future work could well address this problem, but for the moment it poses issues both for designing the data infrastructure for encoding and storing these multiple layers and, perhaps more importantly, for displaying the multiple layers of text and commentary to the user.

Even without the scholia, each commentary comes with its preferred version of the text of *Ajax*. Although the critical text of AjMC does not account for the discrepancies that arise among the commentaries' editions

of the text, the multi-commentary nevertheless renders differences easier to spot by virtue of collocating the commentators' different voices. In turn, these differences often give way to more interesting—and not always overlapping—troves of other resources. For example, Hermann's comment on αἰανῆς at l. 672 offers a lengthy list of parallels and discussions on this easy-to-overlook adjective. [Lobeck \(1835\)](#), *ad* 672 likewise provides insights into the history of this line's interpretation that we might otherwise miss. These old commentaries do not necessarily provide the most convincing interpretations—they do not need to. As Peter Heslin puts it in 'The Dream of a Universal Variorum', 'Much, and indeed ... most of the value of a classical commentary lies in its role not as an arbiter of interpretation but as a collection of pointers to other resources' ([Heslin 2015](#): 498). Heslin continues, offering the following examples: 'In addition to the parallel passages with similar language or ideas, ... these data may also consist of manuscript readings, works of art, material artefacts, historical or mythical persons or episodes and so on' ([Heslin 2015](#): 498). The multi-commentary platform means to present these materials in one easy-to-find place and to provide methods, like CTS URNs, for avoiding what Heslin calls the 'completely haphazard and unsuitable manner' in which material artefacts, to take the chief example, are cited.

An anti-haphazard approach to the *Trugrede* (Sophocles' *Ajax*, 646–92)

As mentioned above, in Davidson's view, recent work on the *Trugrede* has tended to fall in line with twentieth-century orthodoxy, which understands the speech as Ajax's genuine attempt to mislead Tecmessa and the chorus about his intentions ([Davidson 2018](#)). With ample, if dated, support on his side, Davidson counter-proposes that Ajax had a genuine change of heart during the so-called 'Deception Speech', only to return to his self-destructive plans when he reaches the shore.¹¹ This argument rests on a false dichotomy, however: it is too simple to say that either Ajax meant to deceive or he didn't. We must also ask *whom* he meant to deceive and how. The ambiguity of the speech at 646–92 seems, in our view, to leave much to directorial discretion, and discussions that hinge on Ajax's intentions as

they are *written* all seem to flounder on the rocks of the inscrutability of those intentions as they are or would have been *performed*.¹²

Indeed, one might argue that most of the recent work on the so-called *Trugrede* merely acknowledges its difficulties and moves on.¹³ Even the ‘so-called’ epithet for the speech has become a cliché. Our aim here is thus less to resolve the ambiguity and more to lean into it and use it productively in an interpretation that marshals the immense bibliography on this pivotal speech to suggest that Sophocles, if not Ajax, aims for ambiguity and exploits the polysemous possibilities of staging to build up a bulwark around any definitive interpretation of Ajax’s sincerity or lack thereof.

In other words, by turning to commentaries that have been neglected by modern scholarship for over a century, we do not mean to resolve the interpretive difficulties presented by this speech but to show how they have evolved. As P.J. Finglass writes in his chapter, ‘Jebb’s Sophocles’, ‘[Modern] authors appear to assume (probably subconsciously) that whatever value that pre-twentieth-century criticism has for our understanding of Sophocles, is to be found mainly or exclusively in Jebb. This seems to me a widespread assumption, if an erroneous one’ ([Finglass 2015](#): 24). The structure of the AjMC—and, to be sure, the limitations of copyright—offer tools for applying Finglass’s consideration to the text, making nineteenth-century commentaries more accessible to modern readers by applying state-of-the-art OCR techniques and collocating all of the comments on the passage at a single URL.¹⁴

We should note that modern interpreters of Sophocles are limited by constraints of the primarily print publications that are still their aim—we should not expect their citations of commentaries to be exhaustive, and perhaps it suffices for the most part if they stop at Jebb. Nevertheless, as Finglass drives home, it can be instructive to look to older commentaries for what they might tell us about the text and its history and for how they can challenge the readings that we have received by shedding a new and different light on the text itself.

[Navigating the hazards of tralatitiousness](#)

One of the virtues of nineteenth-century commentaries, which we often take for granted or learn to approach too late in graduate school, is their handling of the nitty-gritty matters of text. A cursory scroll through the AjMC shows that many, if not most, of the glossae from pre-Jebb commentaries focus on matters of textual criticism: what authorities print which variants, what accent and thus agreement seems best for a given adjective, and so on. To our modern eyes, it can sometimes seem that these glossae miss the forest for the trees, focusing on minutiae to the detriment of their interpretations. We would rather trust the authority of the received text and its more recent interpreters than challenge our notion of the text's fixity and thus its meaning. This penchant for tralatitiousness in commentaries has been noted by [Heslin \(2015: 496\)](#). By taking an extended example of one of these 'old-school' glossae alongside parallel glossae from other commentators, we hope to draw attention to the usefulness of confronting the text's sources and, perhaps, contradicting certain authorities.

The most heavily glossed token in the *Trugrede* is, perhaps surprisingly, αἰανῆς (so Lobeck, Jebb, Lloyd-Jones and others) or αἰανής (thus Hermann, Schneidewin and so on), 'wearisome', with a long second α, or 'never-ending (usu. in a bad sense)', with a short second α, in l. 672. Paul Demont has a brief and instructive parenthetical: '[L]e sens et la métrique sont incertains: αἰανής signifie-t-il "lugubre" (α long) ou "éternel" (α bref)? κύκλος semble signifier ici "cycle"' ([Demont 2022: 181](#); 'The meaning and the meter are uncertain: does αἰανής mean "gloomy" (long α) or "eternal" (short α)? κύκλος seems to mean "cycle" here'). Even if we resolve the case of the lemma, we are left with a metrical and interpretive ambiguity. Where we might have expected glossae to revolve around lemmata that address Ajax's intentions, we instead have a long history of glossae dealing with the accentuation of a word that does very little to change the ultimate meaning of the passage. This word has attracted the attention of a wide array of commentators, including Hermann, Campbell, Jebb, Lobeck, Wecklein and Finglass. By looking at the glossae on this word, we immediately detect some patterns that conform to the general tenor of the commentaries. To wit, earlier commentaries tend to focus on transmission of this difficult form and its grammatical place in the sentence—Hermann, Lobeck and Campbell all spend time debating whether the word is nominative (from αἰανής, -εν) or genitive (from αἰανός, -η, -ον)—while later commentators

foreground the interpretive value of this difficult word. These functions are not mutually exclusive but rather characterise the focus of each commentator's gaze on this passage of text. The multi-commentary view shifts *our* focus, in turn, towards the interactions of these commentaries. The apparent concentration on αἰανῆς, and the dispersion of much of their other criticism, draws our attention to a central but seemingly tacit difficulty in Ajax's speech: what is the νυκτὸς [αἰανῆς | αἰανῆς] κύκλος, and how does it fit in with the rest of his reflections?

Hermann's commentary provides a useful case study:

Vulgo αἰανῆς. Suidas in hac voce et hanc et alteram scripturam αἰανῆς affert. Tertiam etiam exstitisse ex Favorino conicias, qui sic scribit, αἰανῆς κύκλος καὶ αἰανός, ὁ σκοτεινός. καὶ αἰανόν, τὸ δεινὸν καὶ χαλεπόν. Zonaras p. 64. αἰανός, σκοτεινός. Σοφοκλῆς· νυκτὸς αἰανός (ita cod. K.) κύκλος. ἢ θρηνητής, ἀπὸ τοῦ αἰάζω, τὸ θρηνῶ.

Quamquam dubito, an hic ex Suida corrigendus sit, apud quem haec ita leguntur: αἰανῆς κύκλος, σκοτεινός· ἀδιάλειπτος. Σοφοκλῆς· νυκτὸς αἰανῆς κύκλος· ἢ αἰανῆς, θρηνητικῆς, παρὰ τὸ αἰάζω, τὸ θρηνῶ ...

Exempla attulit Blomfieldius ad Aeschyli Persas 935., qui alteram formam αἰανός, cuius unum, quod sciam, satis certum exemplum in Sophoclis Electra v. 506. exstat, corrigendam arbitratur. Non ausim accedere, quum etiam Hesychius scribat: αἰανόν, χαλεπόν, αἰνόν. Ac re vera αἰανῆς et αἰανός idem esse atque ex eadem origine natum videtur, unde αἰνός est: quod, nisi fallor, ab αἰεὶ deductum primo diuturnum, deinde diuturnitate molestum et grave, ideoque taedii plenum significabat. ([Hermann 1851](#) a.d. 672)¹⁵

Usually transmitted as αἰανῆς. The Suda transmits both this spelling and the alternative αἰανῆς under the entry for this word.¹⁶ You might infer that a third variant also existed from Favorinus, who writes, αἰανῆς κύκλος καὶ αἰανός, ὁ σκοτεινός. καὶ αἰανόν, τὸ δεινὸν καὶ χαλεπόν. Zonaras, p. 64 [of his lexicon, *sub* αἰανός], [writes,] αἰανός, σκοτεινός. Σοφοκλῆς· νυκτὸς αἰανός (thus cod. K.)¹⁷ κύκλος. ἢ θρηνητής, ἀπὸ τοῦ αἰάζω, τὸ θρηνῶ.

Though I doubt whether this should be corrected based on the Suda, in which these are read thus: αἰανῆς κύκλος, σκοτεινός·

ἀδιάλειπτος. Σοφοκλῆς· νυκτὸς αἰανῆς κύκλος· ἡ αἰανῆς, θρηνητικῆς, παρὰ τὸ αἰάζω, τὸ θρηνῶ ...

Blomfield cited examples in his commentary at Aeschylus, *Persians* 935, where he supposes that the alternate form αἰανός—of which one sufficiently sure example exists, as far as I’m aware, at Sophocles, *Electra* 605—ought to be corrected. I wouldn’t dare agree, since Hesychius also writes: αἰανόν, χαλεπόν, αἰνόν. Indeed, αἰανῆς and αἰανός seem equivalent and to have come from the same root as αἰνός—which, unless I’m mistaken, was first derived from αἰεῖ meaning ‘long-lasting’, and then (by extension) ‘bothersome’ and ‘grievous’ due to its long duration, and by the same token came to signify ‘full of weariness’.

By invoking Hermann, we do not mean merely to resuscitate the dense philology at the core of his and other commentators’ arguments. Rather, we additionally want to provide entry into the repositories not only of interpretation but also of literary material per se that various commentators cite in various ways. Hermann dwells on the conflicting evidence in the *Suda* before offering his own etymological conjecture and questioning the parallel readings of contemporaries like Charles James Blomfield. Hermann clearly believes that this is not merely an argument over a diacritical marking, which Sophocles wouldn’t have used anyway, but is rather an argument about the meaning and import of the passage. Hermann’s edition prints αἰανῆς, which modifies νυκτὸς; but his commentary seems to make a case for considering the alternative αἰανῆς, modifying κύκλος.

Hermann’s gloss illuminates the difficulty of making sense not only of this word but also of the textual history that transmits it. After the third sentence, he stops discussing *Ajax* and turns instead to a discussion of the meaning of the word and its variants.¹⁸ Like Lobeck on the same phrase, Hermann homes in on the meaning of αἰανῆς / αἰανῆς to justify not only the variant that he chooses to print but also his understanding of the line ([Lobeck 1835](#): 308 a.d. 672).¹⁹ Together, Lobeck and Hermann might demonstrate some of the predilections of nineteenth-century Sophoclean criticism, but they certainly reveal much of the citational network driving their understanding of the text. From their readings of the ‘endless circle of night’ / ‘circle of endless night’, along with modern comparative linguistic

approaches like that of Chantraine, we can see that the distinction between the two forms of αἰανής is itself lost in the web of references that scholars have summoned in support of their readings. Like Ajax's intentions, this central word in the so-called *Trugrede* is rendered inscrutable.²⁰ Finglass, by contrast, focuses explicitly on the meaning of the sentence as a whole.²¹ His gloss on αἰανής is relatively short: 'αἰανής can mean either "grim" or "long-lasting" (Cf. Braswell on Pind. *P.* 4.236(a), Degan (1962) 45 = (2004) II 825); the former is appropriate' ([Finglass 2011](#): 336).

The preponderance of glossae on this word, however, militates against such a terse justification of one reading over another, let alone one meaning over another. Indeed, by examining the distribution of glossae by lemma, we can see that αἰανής in 672 is particularly unusual ([Figure 7.2](#)). This figure draws both from the public-domain commentaries and from the commentaries that are still under copyright. The lemma αἰανής with nine glossae is clearly an outlier; 75 per cent of the lemmata have up to six glossae, and only 1 per cent of them have seven or more glossae, including αἰανής.

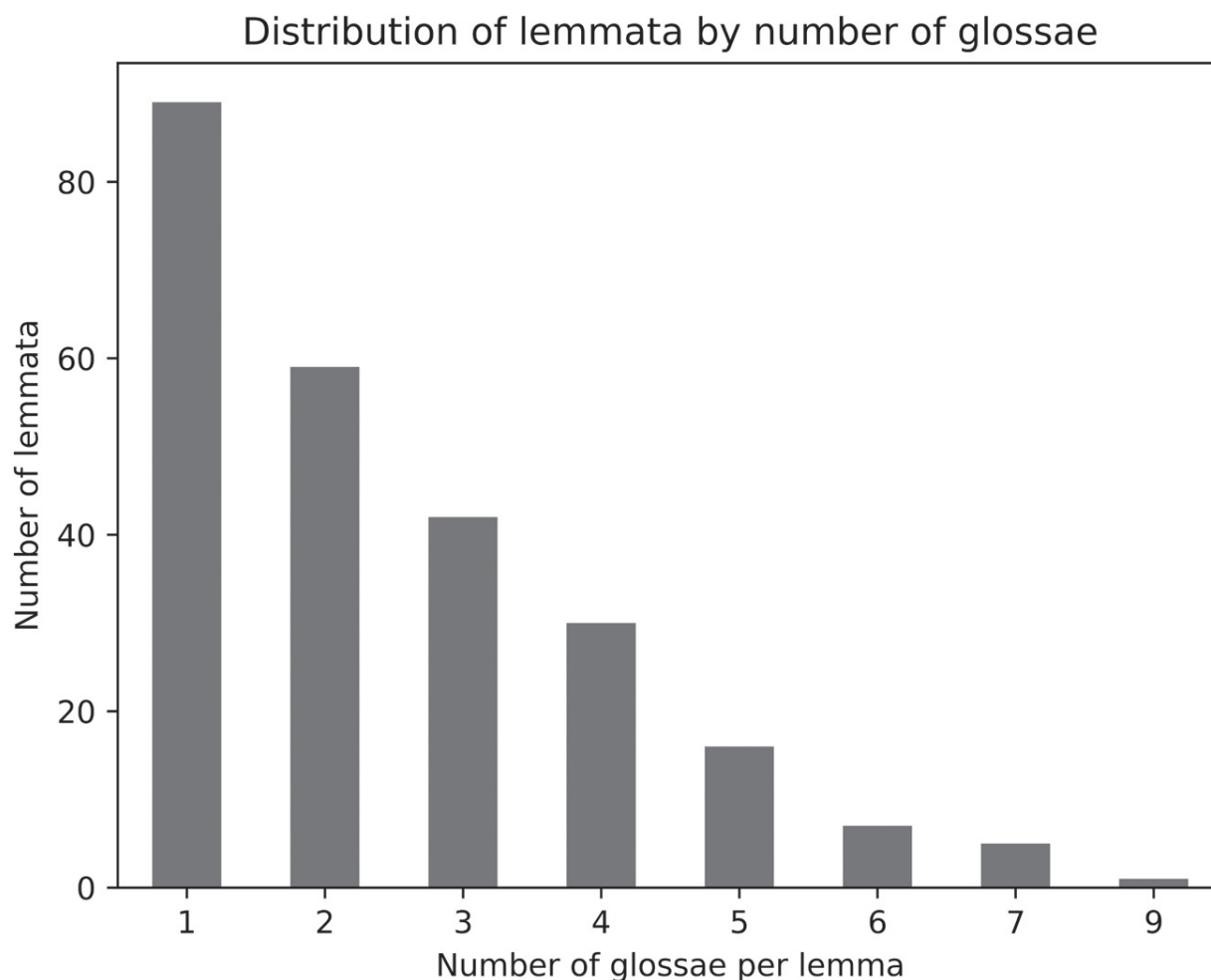


Figure 7.2: Bar chart showing the number of glossae per lemma in the second episode of *Ajax* (CC BY)

The density of glossae that include this lemma—not all of them mention it specifically, but all of those that overlap, even in the modern commentaries, address its problems—might suggest that there is little need for further comment. It bears repeating, however, that this same density of explication reflects the difficulty of interpreting this passage with any authority. Sophoclean textual scholarship has advanced considerably since the first edition of Hermann’s commentary appeared in 1817, such that Finglass can confidently cite the αἰανῆς κύκλος of Laurentianus 32.9 (ms. L) and leave νύκτος αἰανῆς for the apparatus. (See [Finglass 2011](#): 60–65.) As Demont notes in his more recent commentary, however, ambiguity over the meaning of the word still remains, as metrically the second α could still be long or short ([Demont 2022](#): 181, also cited *supra*, p. 10). While we do

not have space to delve much into the textual history here, it is worth drawing attention to the double ambiguity of this much-cited lemma as emblematic of the passage and its commentary tradition as a whole. By paying attention to commentaries that are usually ignored as ‘pre-Jebb’, we see how the broader debate about the passage has often veered into inconsequential territory. This misdirection does not mean that we can just cite the most reliable manuscript, however. Rather, it urges us to account for the minutiae of our own interpretations while at the same time recognising when a passage demands a more synoptic view. Put succinctly: the debate about the meaning and accentuation αἰανῆς might be (mostly) inconsequential for the passage as a whole, but the debate itself can show us new avenues of inquiry by attuning us to the limits of close-reading and encouraging attentiveness to the broader contours of the text.

Conclusion

We are not saying that scholars must accept that there is nothing more to be said on the *Trugrede*—far from it. Rather, we want to suggest that by thinking through the ambiguity of this passage and its lemmata, as shown by the evidence in these commentaries, we might arrive at a richer understanding of the speech than we will by vying to define heterodoxy and orthodoxy. Relatedly, we can also identify places where sufficient ink has been spilt. It is not simply a matter, as John Davidson suggests, of taking the heterodox view seriously ([Davidson 2018](#): 480). Rather, as Ajax himself suggests through his emphasis on the cycles of concealment and revelation (646–47), sleep and release (675–76), and misfortune and salvation (692), it is instead a matter of rejecting dichotomous interpretation entirely and embracing the ambiguous and endless cycles of creating meaning. As Hubbard observes, ‘Ajax seems to change course about suicide, but comes back to his original position, just like the cycles of summer and winter, day and night, calm and storm, waking and sleep that he makes emblematic for the working world (vv. 670–76)’ ([Hubbard 2003](#): 168).

In this paper, we have demonstrated what it is like to read with a multi-commentary through the seemingly insignificant case study of one heavily-glossed lemma in Sophocles’ *Ajax*. Starting from the textual problems that

commentators have encountered with αἰανής since antiquity, our discussion looked both quantitatively and qualitatively at how glossae elaborate on the subsequent etymological and interpretive issues that arise from the change of a diacritical marking. Working with the nineteenth-century commentaries forces us to take a historical view of the development of Sophoclean scholarship over the past 200 years, rather than merely having that view filtered through the lenses of modern commentators.

We also hope to have shown that a digital multi-commentary has the potential to provide a heretofore unattainable view of a text and its variant readings, even at this early stage before a digital critical apparatus has been implemented. By way of conclusion, we want to mention briefly how the *Ajax* Multi-Commentary platform has support for this kind of collaboration already built in, and how it might be used for future work.

As discussed above, the main multi-commentary application works in part by translating the XML → JSON commentaries for the OCR and NLP pipeline to standard Markdown with one customisation: each glossa or entry in each Markdown file is linked via CTS URN to its lemmata in the critical edition. This feature, along with the lightweight and adaptable nature of the Markdown parser that we use ([Jeffrey 2024](#)), means that anyone can write a commentary if they have citable references to valid TEI XML files. Moreover, thanks to tools like Pandoc ([MacFarlane, Krewinkel and Rosenthal 2024](#)), researchers can continue to use the word processing environments with which they are most comfortable: as part of the multi-commentary project, we are packaging Pandoc filters that can transform DOCX and ODT files to multi-commentary-compatible Markdown.

We have begun working on a small demonstration of this capability at <https://purl.org/ajmc/meta>.²² The flexibility of this platform means that it can be of use for elementary and intermediate language students—not just of Greek and Latin but of any language with texts available in TEI—as well as scholars. Peer review can be implemented on top of a source control system like GitHub, GitLab or Codeberg, or it can be handled manually by a project’s maintainers with minimal support from technical staff.

As the exploration of glossae on αἰανής showed, the power of this kind of platform lies in its ability to collocate a plurality of interpreters of a text around a shared source and follow the evolution of their conversations interactively. Its digital roots mean that access to supplementa, such as the

Suda, the works of Zonaras or encyclopaedic entries on named entities in the text and comments, is readily accessible. Further, the multi-commentary includes an extended bibliography of works cited by each of the commentators, providing easy access to second-level bibliography that is not usually collocated with the primary commentaries. While one view of the density of glossae might suggest that commentary on certain lemmata has been exhausted, a more optimistic take is that a multi-commentary's perspectives can sustain an interpretive conversation for far longer, and with far fewer resources, than multiple print publications spread out over centuries can accomplish. In so doing, we can continue to renew the relevance of texts like *Ajax* based on the needs of an ever-evolving readership.

Acknowledgements

The *Ajax* Multi-Commentary (AjMC) benefited from seven commentaries in the public domain; nine commentaries that, while still under copyright, provided essential training data; and three TEI XML editions of Sophocles' *Ajax*. Digitised copies for the vast majority of the commentaries were acquired from open digital libraries such as the Internet Archive, the Bayerische Staatsbibliothek and the Dresden University Library. The AjMC produced TEI XML editions of the public domain commentaries and shared them under the AGPLv3 licence on GitHub.

We would like to thank the book editors, Gabriel Bodard and Valeria Vitale, as well as the reviewers, Kaspar Beelen, Anne H. Chen and Elisa Nury, for their insightful comments and suggestions, which helped to strengthen the arguments presented here. We would also like to thank Ombretta Cesca for her careful scrutiny and philologist's attention to detail. This research was funded by the Swiss National Science Foundation under the Ambizione scheme (grant number 186033) and was carried out while both authors worked at the University of Lausanne.

Notes

1. <https://purl.archive.org/ajmc/multi>

2. For a complete list of public domain commentaries, please see below, [n. 5](#).

3. Internally, the project refers to ‘glossa(e)’ and ‘lemma(ta)’ as opposed to ‘comment(s)’ and ‘reference(s)’ both to highlight the technical nature of these notes and to connect them to the millennia-old (analogue) commentary tradition, despite their presentation in digital form.
4. On the continued usefulness of ‘outmoded’ commentaries, see especially [Finglass \(2015\)](#) (discussed further below).
5. The following commentaries are available in the public domain: [Lobeck \(1835\)](#); [Hermann \(1851\)](#); Friedrich Wilhelm [Schneidewin \(1853\)](#); [Tournier \(1886\)](#); [Wecklein \(1894\)](#); [Jebb \(1896\)](#); F.W. [Schneidewin, Nauck and Radermacher \(1913\)](#). For in-depth discussions of specific parts of the text-processing pipeline, see [Romanello, Najem-Meyer and Robertson \(2021\)](#); [Najem-Meyer and Romanello \(2022\)](#); [Romanello and Najem-Meyer \(2024\)](#); and [Najem-Meyer \(2025\)](#).
6. For a more technical discussion of the AjMC’s text processing pipeline, see especially section 2 of [Pletcher, Najem-Meyer and Romanello \(2024b\)](#).
7. [Klie et al. \(2018\)](#). See <https://inception-project.github.io/>.
8. See [Smith and Weaver \(2009\)](#). For some updates that have also influenced our approach, see [Blackwell and Smith \(2019\)](#). For the usage of CTS URNs to refer to fragmentary authors and works, see Berti, 2026 ([Chapter 8](#) in this volume).
9. A CTS URN for citing the main text might look something like `urn:cts:greekLit:tlg0011.tlg003.ajmc-lj:672`; for citing a commentary, `urn:cts:greekLit:tlg0011.tlg003.ajmc-cam:672@αἰανῆς`. Note that the orthography of the token ‘αἰανῆς’ in the URN matches its appearance in the text; it is not normalised here.
10. At present, our tools can display annotations on Lobeck and Pearson’s editions as well as Lloyd-Jones and Wilson’s, although the glossae are manually aligned only to Lloyd-Jones and Wilson.
11. See, for example, Maurice [Bowra \(1944: 39\)](#), whom Davidson also cites: ‘The great speech which he makes at 646–92 shows that he has recovered himself, got rid of his evil passions and illusions, learned the lesson which the gods have taught him.’
12. From a slightly different angle, Paul Demont highlights the friction created by, on the one hand, the characters’ ‘intrascenic’ responses to Ajax and, on the other, the audience’s ‘extrascenic’ knowledge of the mythic tradition surrounding Ajax and their own affective responses to the play’s internal displays of pity: ‘On peut penser qu’une bonne part de l’intérêt dramatique vient du jeu entre les deux perspectives de l’énonciation théâtrale, la perspective intrascénique (les personnages sont dupés par Aïas) et la perspective extrascénique (le public sait qu’Aïas ne renoncera pas au suicide, prescrit par le mythe comme par l’iconographie, et doit chercher dans ses paroles les indices de sa duplicité), car ce jeu, qui permet au spectateur de comprendre la situation mieux que les personnages, lui procure et plaisir et angoisse’ ([Demont 2022: 177](#); ‘One might suppose that a good part of the dramatic interest comes from the interplay between the two perspectives of the theatrical enunciation, the intrascenic perspective (the characters are deceived by Ajax) and the extrascenic (the audience knows that Ajax will not renounce his suicide, stipulated by the myth as well as by the iconography, and must search for indications of his duplicity within his lines), since this interplay, which enables the spectator to understand the situation better than the characters, gives one both pleasure and anguish’). In some ways, then, we are searching for a method to use the extrascenic elements of the performance to interpret and recontextualise the scene.
13. See Demont’s comments, cited just above.
14. <https://purl.archive.org/ajmc/multi/passages/urn:cts:greekLit:tlg0011.tlg003:646-692>
15. <https://purl.org/ajmc/multi/passages/urn:cts:greekLit:tlg0011.tlg003:646-692?gloss=urn:cts:greekLit:tlg0011.tlg003.ajmc-her:672>

16. ‘αἰανῆς κύκλος.’ Suda On Line. Tr. Anne Mahoney. 21 May 1999. 24 September 2024. <https://www.cs.uky.edu/~raphael/sol/sol-entries/alphaiota/7>.
17. This is Hermann’s note in brackets.
18. We should note that [Chantraine \(1983\)](#) regards the association with αἰεῖ as stemming from a folk etymology. [Beekes \(2016\)](#) likewise lists the word’s primary meaning as ‘terrible’, with a secondary association with ‘eternal’.
19. Lobeck cites the first printing of Hermann’s commentary, from 1817, as an opposing view.
20. In terms of a tally, most commentators under discussion here favour rejecting the majority manuscript reading, genitive αἰανῆς, which is retained by Lobeck (1835), [Schneidewin \(1853\)](#) and [Tournier \(1886\)](#), with more or less vigorous defences; and printed in the edition of [Hermann \(1851\)](#), with caveats noted in his commentary. [Campbell \(1881\)](#), [Wecklein \(1894\)](#) and [Jebb \(1896\)](#) all defend the minor emendation to nominative αἰανῆς. [Finglass \(2011\)](#) follows suit, as does [Demont \(2022\)](#).
21. [Finglass \(2011\)](#): 336 a.d. 672–3) prints, ἐξίσταται δὲ νυκτὸς αἰανῆς κύκλος | τῇ λευκοπόλῳ φέγγος ἡμέρᾳ φλέγειν·, which he translates, ‘Night’s dread sphere makes way for white-horsed day to blaze forth its light’.
22. Source code is available at <https://github.com/AjaxMultiCommentary/kodon-meta>. Moreover, the sources, code and a statically compiled version of the AjMC platform are archived at <https://doi.org/10.5281/zenodo.1555111>.

References

- [Beekes, R.S.P. \(2016\)](#) *Etymological dictionary of Greek*. Leiden Indo-European etymological dictionary series. Leiden: Brill.
- [Blackwell, C.W. and Smith, N. \(2019\)](#) ‘The CITE Architecture: A Conceptual and Practical Overview’, in M. Berti (ed.) *Digital Classical Philology*. Berlin: De Gruyter, pp. 73–94. <https://doi.org/10.1515/9783110599572-006>.
- [Bowra, C.M. \(1944\)](#) *Sophoclean Tragedy*. Oxford: Clarendon Press.
- [Campbell, L. \(1881\)](#) *Sophocles*. Oxford: Clarendon Press.
- [Chantraine, P. \(1983\)](#). *Dictionnaire Étymologique de La Langue Grecque: Histoire Des Mots*. Paris: Éditions Klincksieck.
- [Davidson, J. \(2018\)](#) ‘Challenging Orthodoxy: The “Deception Speech” in Sophocles’ *Ajax*’, *Classical Philology*, 113(4), pp. 472–82. <https://doi.org/10.1086/699552>.
- [Demont, P. \(2022\)](#) *Sophocle: Ajax*. Paris: Les Belles Lettres.
- [Finglass, P.J. \(2011\)](#) *Sophocles: Ajax*. Cambridge: Cambridge University Press.
- [Finglass, P.J. \(2015\)](#) ‘Jebb’s Sophocles’, in C.S. Kraus and C. Stray (eds) *Classical Commentaries: Explorations in a Scholarly Genre*. Oxford:

- Oxford University Press, pp. 21–38. <https://doi.org/10.1093/acprof:oso/9780199688982.001.0001>.
- Hermann, G. (1851), *Sophoclis Ajax*. 4th ed. Leipzig: Ernsto Fleischer.
- Heslin, P. (2015), ‘The Dream of a Universal Variorum: Digitizing the Commentary’, in C.S. Kraus and C. Stray (eds) *Classical Commentaries: Explorations in a Scholarly Genre*. Oxford: Oxford University Press, pp. 494–511.
- Hubbard, T.K. (2003), ‘The Architecture of Sophocles’ *Ajax*’, *Hermes*, 131(2), pp. 158–71.
- Jebb, R.C. (1896), *Sophocles: The Plays and Fragments*. London: Cambridge University Press. Available at: <http://archive.org/details/sophoclesplaysa05campgoog>.
- Jeffrey, C. (2024), ‘Markedjs/Marked’. JavaScript. Marked. Available at: <https://github.com/markedjs/marked>.
- Klie, J.-C., Bugert, M., Boullosa, B., Eckart de Castilho, R. and Gurevych, I. (2018), ‘The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation’, in *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Santa Fe, USA: Association for Computational Linguistics, pp. 5–9. Available at: <http://tubiblio.ulb.tu-darmstadt.de/106270/>.
- Lloyd-Jones, H. and Wilson, N.G. (1990), *Sophoclis Fabulae*. Oxford: Clarendon Press.
- Lobeck, C.A. (1835), *Sophoclis Ajax*. 2nd ed. Leipzig: Weidmann.
- MacFarlane, J., Krewinkel, A. and Rosenthal, J. (2024), ‘Pandoc’. Haskell. Available at: <https://github.com/jgm/pandoc>.
- Najem-Meyer, S. (2025), *Document Processing in Data-Scarce, Domain-Specific Environments: The Case of Multilingual Classical Commentaries*. PhD thesis, École polytechnique fédérale de Lausanne (EPFL). <https://doi.org/10.5075/epfl-thesis-11208>.
- Najem-Meyer, S. and Romanello, M. (2022), ‘Page Layout Analysis of Text-Heavy Historical Documents: A Comparison of Textual and Visual Approaches’, in *Proceedings of the Conference on Computational Humanities Research 2022*. Antwerp: CEUR-WS, pp. 36–54. Available at: https://ceur-ws.org/Vol-3290/long_paper8670.pdf.

- Pletcher, C., Najem-Meyer, S. and Romanello, M. (2024a) ‘Ajax Multi-Commentary Reading Environment’. Elixir. Lausanne. Available at: <https://github.com/AjaxMultiCommentary/ajmc-elixir>.
- Pletcher, C., Najem-Meyer, S. and Romanello, M. (2024b) ‘The Ajax Multi-Commentary: A Digital Platform for the Comparative Analysis of Classical Commentaries’, in J. Karajgikar, A. Janco and J. Otis (eds) *Digital Humanities 2024: Book of Abstracts*. <https://doi.org/10.5281/zenodo.13828231>.
- Romanello, M. and Najem-Meyer, S. (2024) ‘A Named Entity-Annotated Corpus of 19th Century Classical Commentaries’, *Journal of Open Humanities Data*, 10(January), p. 1. <https://doi.org/10.5334/johd.150>.
- Romanello, M., Najem-Meyer, S. and Robertson, B. (2021) ‘Optical Character Recognition of 19th Century Classical Commentaries: The Current State of Affairs’, in *The 6th International Workshop on Historical Document Imaging and Processing (HIP '21)*. Lausanne: Association for Computing Machinery, pp. 1–6. <https://doi.org/10.1145/3476887.3476911>.
- Schneidewin, F.W. (1853), *Sophokles*. Vol. 1. Leipzig: Weidmann. Available at: <https://archive.org/details/sophokle1v3soph>.
- Schneidewin, F.W., Nauck, A. and Radermacher, L. (1913), *Sophokles*. Berlin: Weidmann.
- Smith, D.N. and Weaver, G.A. (2009) ‘Applying Domain Knowledge from Structured Citation Formats to Text and Data Mining: Examples Using the CITE Architecture’, in G. Heyer (ed.) *Text Mining Services: Building and Applying Text Mining Based Service Infrastructures in Research and Industry*, pp. 129–39.
- Tournier, É. (1886), *Les tragédies de Sophocle*. 3rd ed. Paris: Librairie Hachette et Cie.
- Wecklein, N. (1894), *Die Tragödien des Sophokles / zum Schulgebrauch mit erklärenden Anmerkungen versehen von N. Wecklein*. Vol. 4. München: Lindauer.

Chapter 8

Traces of lost libraries in the works of Alexandrian and Byzantine scholars: bibliographic open data from antiquity to Wikimedia

Monica Berti

Introduction

At the time of writing this chapter, Artificial Intelligence (AI) and Large Language Models (LLM) present us with new challenges that affect our responsibilities as scholars working on the heritage from the past and as educators training students to ensure continued access to it and its preservation.¹ The rapid spread of these technologies is generating new questions for research and teaching, but it also requires us to constantly transfer our sources from analogue to digital environments, paying careful attention to their digital quality, especially when it concerns OCR errors, character encoding problems and inconsistencies in crowdsourced distributed annotations. These issues are not new to scholars, as the transfer of data from one medium to another has been a longstanding phenomenon in previous eras. What characterises the present is the vast amount of data produced so far, both in paper and digital formats, and the emergence of new types of technologies.

Considering these premises, this chapter focuses on the language of bibliographic citations in ancient sources and modern library systems that collect information about Greek and Latin authors. By ‘language of

bibliographic citations', I mean expressions used in historical documents to refer to the onomastics of authors and the content of their works. This analysis is different from the traditional study of textual quotations and fragmentary literature, although it is closely related to them ([Berti 2021](#); [Berti 2024](#); [Berti 2025a](#)).

Two examples of bibliographic language are the references to the female poets Cleobulina of Rhodes and Moero of Byzantium. In ancient sources, Cleobulina is said to be from Lindus (Κλεοβουλίνη ἡ Λινδία), the mother of Thales (Θαλῆς ... μητρὸς δὲ Κλεοβουλίνης), the daughter of Cleubulus the son of Euagoras from Lindus (Κλεόβουλος Εὐαγόρου Λίνδιος ... αὐτῷ θυγατέρα Κλεοβουλίνην), named after him and nicknamed Eumetis (τὴν δὲ Ῥοδίαν Εὐμητιν ἄχρι νῦν Κλεοβουλίνην πατρόθεν). She is attributed riddles (αἰνίγματα), and the playwrights Alexis and Cratinus entitled comedies after her (Κλεοβουλίνας) (Ath. *Deipn.* 171b = 4.70, 448b = 10.69, 586a = 13.50; Diog. Laert. *Vitae philosoph.* 1.22 and 1.89; Plut. *Mor.* 148d and 401b; *Suda*, κ 1718).

Moero is said to be from Byzantium (ἡ Βυζαντία), the daughter of the tragic poet Homer (Ὁμήρου τοῦ τραγικοῦ θυγάτηρ), the wife of the famous philologist Andromachus (γυνὴ δὲ Ἀνδρομάχου τοῦ ἐπικληθέντος φιλόλογου), the mother of the grammarian and tragic poet Homer (Ὁμηρος· Ἀνδρομάχου καὶ Μυροῦς Βυζαντίας, γραμματικὸς καὶ τραγωδιῶν ποιητής) and her name is transmitted in two forms (Μοιρώ and Μυρώ). She is said to be an epic, elegiac and lyric female poet, and she is attributed a poem entitled *Mnemosyne* (Μνημοσύνη), a curse poem *Curses* (Ἀραί) and a hymn to Poseidon (εἰς Ποσειδῶνα ὕμνον) (Parthen. *Narr. amat.* 27; Paus. 9.5.8; Ath., *Deipn.* 490e = 11.80; *Suda*, μ 1464 and ο 253; Eust. *Comm. Hom. Il.* ad B 711, 247).

Leaving aside the debated questions of identity and family ties, such as, for example, the exact relation between Moero and the tragic poet Homer, the language of the ancient references to Cleobulina and Moero is rich in linguistic data and proper names related to people and places. Generally, this language is hidden in modern philological resources, where proper names are translated or paraphrased into Latin and bibliographic references are covered by the label *fragmenta* or citations of modern editions. This also happens in digital libraries that ingest metadata from library systems, like the *TLG Canon* and the *Perseus Catalog*, where the identifiers *tlg0244*

and urn:cite:perseus:author.380 are used for Cleobulina, while tlg0220 and urn:cite:perseus:author.945 refer to Moero. These metadata also enrich the Wikidata knowledge base, where Q2576040 and Q541746 uniquely identify the two women respectively. By exploiting computational data, we can extract the language of bibliographic citations from ancient sources to give ancient authors a voice again, access their language directly without the filter of modern conventions and evaluate it critically in context.

If we look for other examples, instead of simply registering Hom., *Ilias* 2.493–760 or Theop. *FGrHist* (BNJ) 115 F 249, we can extract and annotate the forms ἐν νεῶν καταλόγῳ (‘in the catalogue of ships’) and ἐν δὲ τῷ ἐπιγραφομένῳ τοῦ Θεοπόμπου συγγράμματι περὶ τῶν ἐκ Δελφῶν συληθέντων Χρημάτων (‘in the treatise by Theopompus entitled *On the Money Plundered from Delphi*’). These forms were used by Thucydides, Plutarch and Athenaeus to refer to the famous catalogue of ships preserved in the *Iliad* of Homer and to a lost work of Theopompus of Chius (Thuc. 1.10.4; Plut. *Thes.* 25.3; Ath. *Deipn.* 532d = 12.43). These expressions can be collected to enrich a digital dynamic resource that is not constituted by modern citations and abbreviations, but is based on the linguistic variety of ancient sources to express bibliographic information. This variety enables us, for example, to reveal forms that in the past were used to refer not simply to the content of a portion of a work, but to defined and partly independent sections, like the catalogue of ships of the *Iliad*, the nature of which is still a matter of debate in modern scholarship ([Castelli 2020](#): 53–5). In the case of fragmentary authors like Theopompus, this linguistic variety enables us to transcend the concept of *fragmentum* and examine the range of topics addressed in independent works, thereby revealing the limitations of modern categories applied to ancient sources.

Given the complexity and extension of ancient corpora, in the following paragraphs I describe research on open historical data, arranging them geographically and chronologically. The focus is on literary sources that preserve bibliographic information about Greek and Latin authors. Alexandria and Byzantium are the scholarly places taken under analysis by selecting authors who were active in those cities between the second and tenth centuries C.E.: Athenaeus of Naucratis, Valerius Harpocration, Clement of Alexandria and the author of the *Suda*. Both places were fundamental for collecting and accessing ancient sources and transmitting

them to us. The reasons for choosing these authors are explained below, but also depend on the current availability of their works in digital formats. Additional sources will be incorporated into this research, as current limitations of digital data are expected to be partially overcome in the coming years with increased digitisation. In addition to ancient sources and to show the continuous work from past to future libraries, I devote the last section of this chapter to Wikidata, which is proving to be a powerful collector of metadata for accessing structured information on classical literature according to the Linked Open Data (LOD) principles ([Cayless 2019](#); [Middle 2024](#)).

Alexandria

Past accounts of the libraries of Alexandria in Egypt always confront us with the issue of access. Applying this term to the ancient world is difficult and might even be misleading. However, even though the sources are scarce, we still have significant evidence of Alexandrian efforts to provide access to knowledge, such as the Ptolemaic desire to collect all the books of the inhabited world and commission their translations, not to speak of the contradictory figures preserved by ancient authors who debated the number of texts stored in these libraries ([Parsons 1952](#): 163–203; [Berti and Costa 2010](#): 49–100; [Berti 2016](#)).

We also have information on the role of library directors and their successions. A complex figure is that of Callimachus of Cyrene, who was probably never director of the Alexandrian library, but was a critical editor and commentator on classical authors. Although his famous *Tables* (*Pinakes*) have been improperly referred to as the library catalogue, the fragments of this work attest to a critical activity of collecting biographical and bibliographical data on classical authors. This initiative, accompanied by the contributions of pupils of Callimachus and other Alexandrians, underlies many works by authors who lived after the Cyrenaean poet and had the privilege of accessing the Alexandrian collections ([Pfeiffer 1968](#): 123–51; [Blum 1977](#); [Berti and Costa 2010](#): 101–59). In this regard, authors like Athenaeus, Harpocration and Clement represent interesting examples who still preserve traces of the Callimachean work.

I have chosen these authors for different reasons. The first is that, given that I am using digital data and applying digital methods, the works of these authors are openly accessible in a structured format in the *Perseus Digital Library* (<https://github.com/PerseusDL/canonical-greekLit>) and the *Open Greek and Latin* (OGL) project (<https://github.com/OpenGreekAndLatin>). Moreover, lemmatised and part-of-speech (POS) tagged data of these works are also available in the open repositories of the projects *Opera Graeca Adnotata* (OGA) and *Diorisis* ([Celano 2024](#); [Vatri and McGillivray 2018](#)). The second and most important reason is that the works of these authors are recognised as important collections of myriad references to other authors and works, most of them now lost. The third reason is that, even if these authors produced works belonging to different literary genres, they accessed the same Alexandrian resources, therefore being witnesses of the same cultural environment.

The *Deipnosophists* of Athenaeus and the *Lexicon of the Ten Orators* of Harpocration are different, the first being a scholarly miscellany and the second a lexicon mainly dealing with Athenian legal customs. However, Athenaeus and Harpocration belonged more or less to the same generation and shared the same library resources in Alexandria, which is shown by their knowledge and reference to the same authors. So far, the relationships among the two authors have been analysed in regard to historians, but have to be extended to other genres ([Zecchini 2000](#)).

Considering the Christian perspective, Clement of Alexandria shared with Athenaeus and Harpocration the same cultural context. The importance of a comparison among these authors has been recently affirmed by scholarship, given that juxtaposing the approach of Clement with authors of classical miscellanism could give us a better perspective on differences and similarities in an interdisciplinary dialogue that characterised intellectual circles of Alexandria ([Heath 2020](#)).

In terms of digital accessibility, the language of these authors is significant for performing a wide range of research. My focus is on the analysis of ancient linguistic forms used to refer to the onomastics of authors and the content of their works. The aim is to rediscover the richness of the language of ancient sources, which is usually hidden behind modern conventions and interpretations. This type of analysis benefits from data resulting from two important activities that have characterised the digital

humanities over the last decades. The first is the attempt to digitise ‘all the books of the inhabited world’, to echo the famous passage of the *Epistle to Aristeia* (9–11) in reference to the ambitious cultural programme of the Ptolemies.² The second activity is the linguistic analysis of ancient Greek and Latin sources, which is growing thanks to projects, publications and datasets produced in the field of computational linguistics for historical languages ([Berti 2019](#); [Crane et al. 2009](#)).

Digital linguistic data now form the basis of many critical analyses in philology, literature and historiography. They are also crucial in the face of the spread of AI and LLMs, as both ingest and will ingest more and more data, including those we deal with in the domain of classical antiquity ([Bamman and Burns 2020](#); [Burns 2024](#)). This is why producing more data and caring about their structure and quality are essential for any future preservation of our textual heritage, as will be shown in the following pages.³

To extract the ancient language of bibliographic references and adopt a sustainable digital workflow, my analysis is based on Named Entity Recognition (NER), lemmatisation, disambiguation, data export and stable identifiers. The reasons for this workflow stem from the fact that the core elements of bibliographic references are proper names (that is, named entities), which are still underrepresented digitally, especially for ancient Greek. In terms of stable identifiers, the Canonical Text Services (CTS) of the CITE Architecture provide the syntax of Uniform Resource Names (URNs) to cite instances of author names and work titles, in accordance with one of the main principles of the LOD initiative, which requires stable identifiers for digital resources ([Blackwell and Smith 2019](#)).

NER is a task in Natural Language Processing (NLP) that extracts proper names from modern texts, belonging to a wide variety of categories, such as persons, places, organisations, monetary amounts and many others. NER is also taking its first steps in the domain of historical languages, which are somewhat challenging since they are no longer spoken and the data of their corpora are not as abundant as for contemporary languages ([Berti 2023](#); [Ehrmann et al. 2023](#); [Yousef, Palladino and Jänicke 2023](#)).⁴ My approach is to individuate ancient Greek tokens corresponding to proper names, and to lemmatise and disambiguate them according to NE classes for personal names (PER and PERderiv), placenames (LOC and LOCderiv),

organisations like Panhellenic games and festivals (ORG and ORGderiv) and other entities belonging to subcategories like work titles and chronological expressions (OTH) ([Berti 2024](#)).

Within the group of personal names (PER and PERderiv), it is possible to disambiguate names of authors that function as anchors in the text for identifying other elements of their onomastics and literary activity. If we consider the examples of the introduction, we can annotate the string Μυροῦς Βυζαντίας in the entry of the *Suda* about the grammarian Homer (ο 253) as composed by two named entities: Μυροῦς|PER and Βυζαντίας|LOCderiv. These two entities together correspond to the real entity Moero of Byzantium (Μυρὼ Βυζαντία in the lemmatisation). The expression Μυροῦς Βυζαντίας is annotated as urn:cts:greekLit:tlg9010.tlg001.1st1K-grc1:1.O.253@Μυροῦς[1]-Βυζαντίας[1], which means that this is the string of text from the first occurrence of Μυροῦς to the first occurrence of Βυζαντίας in the *Suda* entry ο 253 of the edition by Ada Adler, which is catalogued under tlg9010.tlg001.1st1K-grc1 in the *Perseus Catalog*. A stable identifier in the form of another CTS URN (urn:cts:greekLit:tlg0220) can be aligned to this annotation of the *Suda* to uniquely identify the string Μυροῦς Βυζαντίας as a reference to the female poet Moero of Byzantium, who is identified with the TLG number 0220 in the *Perseus Catalog* (on TLG and Perseus IDs, see [Pantelia 2022](#) and [Babeu 2019](#)).

Under the CTS URN urn:cts:greekLit:tlg0220, it is possible to group other occurrences of references to Moero in other sources, as for example Μοιρὼ ἡ Βυζαντία and Μοιρὼ in the *Deipnosophists* of Athenaeus (11.80). In the same context, we can extract the reference to one of the works of Moero (Μνημοσύνη) and annotate it as urn:cts:greekLit:tlg0008.tlg001.perseus-grc2:11.80@μνημοσύνη[1], which points to the string of text in chapter 80 of book 11 of the *Deipnosophists* (tlg0008.001) in the edition by Georg Kaibel (perseus-grc2). The inflected form μνημοσύνη can be annotated as corresponding to the personal name Μνημοσύνη (PER) and disambiguated in the context as functioning for the title of the work of Moero that is identifiable with urn:cts:greekLit:tlg0220.ath001, where tlg0220.ath001 is the unique identifier of this work of Moero in the *Deipnosophists* of Athenaeus. In this case, I do not use the work reference number of the TLG *Canon*, because

tlg0220.001 refers to the edition of all the fragments of Moero in the *Collectanea Alexandrina* by John Powell, while tlg0220.ath001 refers explicitly to the citation of the work *Mnemosyne* by Athenaeus ([Berti 2024](#)). The alignment of tlg0220.ath001 with urn:cts:greekLit:tlg0008.tlg001.perseus-grc2:11.80@μνημοσύνη[1] allows the linking of this citation to the edition of the *Deipnosophists* by Kaibel.⁵

As we see through these examples, named entities and unique identifiers in the form of CTS URNs allow us to link texts, editions and catalogues according to the LOD principles. These links are essential for connecting metadata of modern catalogues to the language of ancient sources, including also modern editions that may differ in text reconstruction and interpretation. This possibility highlights the need to digitise more editions of ancient sources and find solutions to overcome the still-existing issues of copyright restrictions that prevent us from utilising the latest published editions ([Cayless 2019](#)).

Furthermore, the aim is not to adopt a traditional approach of sequential close reading and annotate texts word by word, which would require a lot of work and generate dispersed and inconsistent results. In this regard, the use of current technologies such as NER helps to extract entities related to bibliographic citations from our corpora and disambiguate them to model a sustainable workflow, produce consistent annotations and provide classical studies with more data for further collaborative analysis in onomastics, prosopography, historical geography, fragmentary literature and linguistics. Through automation, NER can reduce the manual effort required to annotate bio-bibliographical information in source texts, thus allowing us to uncover the linguistic variety of the original language.

If we consider the example of Moero of Byzantium, named entities related to her bibliographic citations in ancient sources enrich authority lists for onomastics and historical geography. These entities can be used to contribute to prosopographical studies for family relations and their histories, to gazetteers for generating maps and visualising the geographical distribution of proper names and authors in the ancient world, and to philology and linguistics for tracing variants of proper names in ancient sources. As far as literary studies are concerned, these entities can be used as anchors in the text to extract common nouns and adjectives related to literary genres of authors, as for example the form ποιήτρια ἐπῶν καὶ

ἐλεγείων καὶ μελῶν (‘a female epic and elegiac and lyric poet’) referred to Moero of Byzantium (*Suda* μ 1464). Finally, the linguistic forms used to refer to titles or work contents allow us to perceive how rich the literary interests of ancient authors were, which are usually hidden behind modern categories and conventional labels like *fragmenta*.

To cite figures, named entities semi-automatically extracted from the *Deipnosophists* (ed. Kaibel: 273,954 words) include a total of 22,930 occurrences with 8,572 unique occurrences, which are distributed in the following classes: personal names and derivatives (4,455 PER and 219 PERderiv), placenames and derivatives (1,054 LOC and 1,865 LOCderiv), organisations and derivatives (94 ORG and 14 ORGderiv) and other entities (871 OTH). These data have been used to extract named entities from the *Lexicon* of Harpocration (ed. Dindorf: 39,038 words), where we have a total of 6,134 occurrences with 2,188 unique occurrences distributed in the following classes: personal names and derivatives (1,082 PER and 87 PERderiv), placenames and derivatives (457 LOC and 335 LOCderiv), organisations and derivatives (70 ORG and 1 ORGderiv) and others (156 OTH). In these texts, ORG includes festivals, Panhellenic games, magistracies and so on, OTH is used for miscellaneous entities like works, months, constellations, currencies, languages, chronological data and so on, and PER covers gods, persons, authors and personifications.

All these data are used to train an NLP model for NER based on the open-source Python library spaCy (<https://spacy.io/api/entityrecognizer>), which also benefits from the ancient Greek model greCy (<https://spacy.io/universe/project/grecy>). Experimental results are currently available in the *Named Entities Recognizer* of the *Linked Ancient Greek and Latin* (LAGL) project (https://www.lagl.org/entity_recognizer.php) and are used to automatically extract entities from the works of Clement of Alexandria and the lexicon of the *Suda*.

Considering bibliographic references, Athenaeus in the *Deipnosophists* cites 894 authors and 1,978 works, while Harpocration in the *Lexicon* cites 189 authors and 660 works. Linguistic annotations of all these authors and works are available in the catalogues of the *Digital Athenaeus* (<https://www.digitalathenaeus.org/tools/Catalog/index.php>) and the *Digital Harpocration* (<https://www.lagl.org/tools/harpocration/>). Both catalogues are now merged into a single catalogue of the LAGL project, where data from other

works such as the *Suda* (<https://www.lag1.org/tools/suda/>) and those of Clement are progressively added: <https://catalog.lag1.org/>. Open catalogues like *Perseus* and the *Digital Latin Library* (DLL) provide unique identifiers for most of these authors, and missing ones are assigned new numbers with the prefix *lagl*, as for example Mnaseas of Locri or Colophon (lagl0177).

Linked Ancient Greek and Latin LAGL

Catalog of Authors and Works

Ister - urn:cts:greekLit:tlg1450

Ister (Ἰστέρος Ister of Cyrene) - urn:cts:greekLit:tlg1450

Place of birth (Wikidata Statement)

Paphos ([Wikidata](#))

Place of death (Wikidata Statement)

Alexandria ([Wikidata](#)) ([Pleiades](#))

Occupations (Wikidata Statement)

writer ([Wikidata](#))

historian ([Wikidata](#))

librarian ([Wikidata](#))

Ancient Sources

[urn:cts:greekLit:tlg1450](#) in Digital Atheneaus

[urn:cts:greekLit:tlg1450](#) in Digital Harpocraton

[urn:cts:greekLit:tlg1450](#) in Digital Suda

Digital Athenaeus Works

[urn:cts:greekLit:tlg1450.ath001](#) - Ἰστέρος Athides

[urn:cts:greekLit:tlg1450.ath002](#) - Ἰστέρος Argolika

[urn:cts:greekLit:tlg1450.ath003](#) - Ἰστέρος Replies to Timeus

[urn:cts:greekLit:tlg1450.ath004](#) - Ἰστέρος On the city of Ptolemais

Digital Harpocraton Works

[urn:cts:greekLit:tlg1450.harp001](#) - Ἰστέρος The manifestations of Apollo

[urn:cts:greekLit:tlg1450.harp002](#) - Ἰστέρος Atakta

[urn:cts:greekLit:tlg1450.harp003](#) - Ἰστέρος Athides

Digital Suda Works

[urn:cts:greekLit:tlg1450.suda001](#) - Ἰστέρος Athides

Figure 8.1: *Linked Ancient Greek and Latin* (LAGL) project: The Catalog of Authors and Works (CC BY-NC-SA)

In terms of overlapping references, 152 authors are cited by both Athenaeus and Harpocration with significant differences in cited works. For example, as can be seen in Figures 8.1 and 8.2, both authors cite the work on Athenian local history by Istros the Callimachean (tlg1450), but use different forms to refer to it. As for other works, Athenaeus cites the *Argolika*, the *Replies to Timaeus* and the work *On the City of Ptolemais* of Istros, while Harpocration cites his *Manifestations of Apollo* and the *Atakta* (Berti and Jackson in BNJ 334⁶).

As we can see, the extraction of named entities is an effective method to structure texts and individuate linguistic elements for further textual analyses. Lemmata, named entities and bibliographic annotations extracted from the works of Athenaeus and Harpocration are used to automatically extract data from the *Protrepticus* (23,298 words), the *Paedagogus* (56,766 words) and the *Stromata* (162,258 words) of Clement of Alexandria. Considering differences and possible ambiguities, the results are revised and checked to produce consistent annotations.

Works Catalog	
urn:cts:greekLit:tlg1450.harp003 (Ister) Submit	
Ister (ΙΙΙΙ Ister of Cyrene) - urn:cts:greekLit:tlg1450.harp003 (ΙΙΙΙ Atthides)	
A Ανθεστηριών	[Ανθεστηριών: ὄγδοος μὴν οὗτος παρ' Ἀθηναίους, ἱερὸς Διονύσου. Ἰστρος δὲ ἐν τοῖς <u>τῆς συναγωγῆς</u> κεκλησθῆναι φησιν αὐτὸν διὰ τὸ πλεῖστα τῶν ἐκ γῆς ἀνθεῖν τότε.]
E Ἐπενεγκεῖν δόρυ ἐπὶ τῇ ἐκφορᾷ καὶ προαγορεύειν ἐπὶ τῷ μνήματι	[Ἐπενεγκεῖν δόρυ ἐπὶ τῇ ἐκφορᾷ καὶ προαγορεύειν ἐπὶ τῷ μνήματι: <u>Δημοσθένους κατ' Εὐέργου καὶ Μνησιβοῦλου</u> ταῦτά φησιν ἐπὶ τοῦ βιαίως ἀποθανόντος. Ἰστρος δ' ἐν τῇ <u>συναγωγῇ τῶν Ἀτθίδων</u> περὶ Προκρίδους καὶ Κεφάλου εἰπὼν γράφει ταυτί· τινὲς δὲ ἐπὶ τοῦ τάφου δόρυ καταπεπηχότα τὸν Ἐρεχθέα φασὶ πεποιθῆσθαι παρακαθήμενόν τε καὶ τὸ πάθος σημαίνοντα, διὰ τὸ νόμιμον εἶναι τοῖς προσήκουσι τοῦτον τὸν τρόπον μετέρχεσθαι τοὺς φονέας.]
Θ Θεοίνιον	[Θεοίνιον: <u>Δικαῖργος</u> ἐν τῇ <u>διαδικασίᾳ Κροκωνιδῶν πρὸς Κοιρωνίδας</u> . τὰ κατὰ δῆμιους Διονύσια Θεοίνια ἐλέγετο, ἐν οἷς οἱ γεννῆται ἐπέθυσον· τὸν γὰρ Διόνυσσον θεοῖνον ἔλεγον, ὡς δηλοῖ <u>Αἰσχύλος</u> καὶ Ἰστρος ἐν α' <u>Συναγωγῶν</u> .]
K Κοιρωνίδα	[Κοιρωνίδα: ἔστι <u>Δικαῖργος</u> λόγος οὕτως ἐπιγραφόμενος <u>Κροκωνιδῶν</u> [±] <u>διαδικασίᾳ</u> [±] <u>πρὸς</u> [±] <u>Κοιρωνίδας</u> [±], ὃν ἐννοῖ <u>Φιλόνου</u> νομίζουσιν· ἔστι δὲ γένος οἱ Κοιρωνίδα, περὶ ὧν Ἰστρος ἐν τῇ <u>συναγωγῇ τῆς Ἀτθίδος</u> φησιν. ὠνομασμένοι δ' ἂν εἴεν ἀπὸ Κοίρωνος, ὃν νόθον ἀδελφὸν εἶναι φασὶ τοῦ Κρόκωνος, παρ' ὃ καὶ ἐντιμωτέρους εἶναι τοὺς Κροκωνίδας, ὃ δὲ τὸν λόγον γεγραφώς ὅστις ποτ' ἔστιν, τρισὶν ὀνόμασι φησιν αὐτοὺς προσηγορεῖσθαι· καὶ γὰρ Κοιρωνίδας καὶ Φιλίεις καὶ Περιθοίδας.]
Λ Λαμπάς	[Λαμπάς: <u>Λυσίας</u> ἐν τῷ κατ' Εὐέργου, τρεῖς ἄγουσιν Ἀθηναῖοι ἐορτὰς λαμπάδος, Παναθηναίους καὶ Ἡφαιστίους καὶ Προμηθεῖους, ὡς <u>Πολέμων</u> φησιν ἐν τῷ <u>περὶ τῶν ἐν τοῖς προπυλαίοις πάνκων</u> . Ἰστρος δ' ἐν α' τῶν <u>Ἀτθίδων</u> , εἰπὼν ὡς ἐν τῇ τῶν Ἀπατουρίων ἐορτῇ Ἀθηναίων οἱ καλλίστας στολὰς ἐνδεδυκότες, λαβόντες ἡμένας λαμπάδας ἀπὸ τῆς ἐστίας, ὕμνουσι τὸν Ἡφαιστον θύοντες, ὑπόμνημα τοῦ κατανοήσαντα τὴν χρεῖαν τοῦ πυρὸς διδάξαι τοὺς ἄλλους.]
Π Παναθήναια	[Παναθήναια: <u>Δημοσθένους Φιλιππικοῖς</u> , διττὰ Παναθήναια ἤγετο Ἀθήνησι, τὰ μὲν καθ' ἕκαστον ἐνιαυτὸν, τὰ δὲ διὰ πεντετηρίδος, ἅπερ καὶ μεγάλα ἐκάλουν. Ἰσοκράτης <u>Παναθηναϊκῶ</u> φησὶ μικρὸν δὲ πρὸ τῶν μεγάλων Παναθηναίων. ἤγαγε δὲ τὴν ἐορτὴν πρῶτος Ἐριχθόνιος ὁ Ἡφαιστοῦ, καθά φησιν <u>Ἑλλάνικός</u> τε καὶ <u>Ἀνδρῶτις</u> (ὧν, ἑκάτερος ἐν α' <u>Ἀτθίδος</u> [±]), πρὸ τούτου δὲ Ἀθήναια ἐκαλεῖτο, ὡς δεδήλωκεν Ἰστρος ἐν γ' τῶν <u>Ἀττικῶν</u> .]
T Τραπεζοφόρος	[Τραπεζοφόρος: <u>Δικαῖργος</u> ἐν τῷ <u>περὶ τῆς ἱερείας</u> ὅτι ἱερωσύνης ὀνομα ἐστὶν ἡ τραπεζοφόρος, ὅτι δ' αὕτη τε καὶ ἡ κοσμῶ συνδιέπουσι πάντα τῇ τῆς Ἀθηνᾶς ἱερείᾳ αὐτός τε ὁ ρῆτωρ ἐν τῷ αὐτῷ λόγῳ δεδήλωκε καὶ Ἰστρος ἐν ιγ' τῶν <u>Ἀττικῶν συναγωγῶν</u> .]

Figure 8.2: LAGL Catalog: References to the *Atthides* of Istros in the *Lexicon* of Harpocration (CC BY-NC-SA)

If we consider traditional philological tools for Clement, the index of proper names (*Eigennamenregister*) in the edition by Otto Stählin is a fundamental resource, featuring entries in ancient Greek and descriptions in German. Still, it requires a separate close reading of each relevant passage ([Stählin 1905](#)–1936, vol. IV: 70–196). In this respect, data are only partially accessible. Data extraction and annotation allow us to accompany such a traditional resource with a new complete set of all proper names with elements that are missing in the printed edition: inflected forms in ancient Greek; lemmata; named entity labels for disambiguating personal names, placenames and miscellaneous entities; links to external open resources such as dictionaries, lexica, gazetteers, knowledge bases and authority lists; and finally contextualisation of each word in the Greek text, like in the resources currently available in the *Digital Athenaeus* and LAGL projects.

The index of authors and works (*Citatenregister*) of the edition of Clement by Otto Stählin is another invaluable resource of classical

philology, which arranges entries under Ancient and New Testament, Church writers and Heretics (including Apocrypha), and Non-church writers ([Stählin 1905](#)–1936, vol. IV: 1–69). All authors are printed in the Latin form, and titles of extant works are cited in the respective original language, including many fragmentary works. Considering the *addenda* of the edition by Ursula Treu (xxv–xxxiv), the works of Clement contain citations of approximately five hundred Greek and Latin authors. Like for the texts of Athenaeus and Harpocration, named entities related to the onomastics of these authors and the content of their works are currently being disambiguated to populate the text-based catalogue of the LAGL project.

Byzantium

In 1998, thanks to the pioneering work of the late Ross Scaife, the *Stoa Consortium* launched the project *Suda On Line* (SOL) to provide scholars with a freely accessible and keyword-searchable text of the Byzantine lexicon *Suda*: <https://www.cs.uky.edu/~raphael/sol/sol-html/index.html>. Thanks to a successful scholarly crowdsourcing effort, all entries of the *Suda* are now translated into English, and accompanied by at least partial notes and bibliography ([Finkel et al. 2000](#); [Mahoney 2009](#)). This project also inspired technical implementations like *Son of Suda On-Line* (SoSOL), which is a workflow management tool implemented by *papyri.info* and *Perseids* ([Baumann 2013](#); [Almas and Berti 2013](#); [Berti 2021](#): 114).

In 2016 the copyright of the edition of the *Suda* by Ada Adler (614,450 words for a total of 31,342 entries) expired and the entire text, including the critical apparatus, was digitised by the OGL project to be made accessible in an XML format (tlg9010). As the *Indices* of the Adler edition show, the *Suda* is an invaluable source of named entities related to antiquities and to classical authors and works, but we lack a complete analysis of the sources behind it and linguistic annotations of the text ([Adler 1928](#)–1938, vol. 5: 39–133; [Zecchini 1999](#)).

The two *Suda* entries on Callimachus (κ 227 and 228) that differentiate the famous Alexandrian poet and his nephew are an example of the richness of named entities. The entry κ 227 is rich in bio-bibliographical information

on the life and scholarly activity of Callimachus. In this entry, classes of named entities encompass onomastics (with disambiguation of homonyms), prosopography (including family relations), geography (places of origin and intellectual activity), chronology (names of sovereigns and Olympiads) and bibliography (a long list of over twenty works by Callimachus). Another example is the reference to the debate about the first prose work in the entry φ 214 (s.v. Φερεκύδης) that includes a list of the works of Pherecydes of Syros with the three parallel titles Ἑπτάμυχος ἦτοι Θεοκρασία ἢ Θεογονία (*Seven Recesses* or *Divine Mingling* or *Genealogy of the Gods*), which reveal the common ancient practice of using different expressions to refer to the same work ([Castelli 2020](#): 175).

The *Named Entities Recognizer* of the LAGL project currently enables the extraction of named entities from the entire text of the *Suda* and the enrichment of the dataset of named entities and lemmata generated from the works of Athenaeus, Harpocration and Clement. These entities are also further disambiguated to annotate references to the approximately four hundred authors cited in the *Suda*, shedding light on the underlying library resources and allowing for the study of this lexicon in itself, rather than merely as a repository of fragments.

The result will be a complete annotated register of the language used in the *Suda* to refer to the onomastics of authors and to titles/descriptions of their works, as in the expression Νίκανδρος ἐν α' Αἰτωλικῶν for Nicander of Colophon's *Aetolica* in *Suda* τ 677, which can be aligned to the LAGL Catalog entry urn:cts:greekLit:tlg0022.⁷ In this regard, I have already integrated entries of the *Suda On Line* in the search environment of the *Digital Fragmenta Historicorum Graecorum* (DFHG) and the *Digital Athenaeus* projects, where it is possible to automatically detect if words of Greek fragmentary historians and of the *Deipnosophists* correspond to entries of the *Suda* ([Berti 2021](#): 165–69, 311–12).

These annotations open many interesting questions about access to bibliographic resources in Byzantium, and about differences and similarities with Alexandria, considering the geographical and chronological distance between the two scholarly places. These data can be used for comparison with data from Athenaeus, Harpocration and Clement to track citations diachronically and thus the transmission of classical and Christian authors through the centuries.

Wikidata and the Wikimedia universe

As anticipated in the introduction, I am also experimenting with modern metadata about classical antiquity that populate and feed digital libraries and knowledge bases, with a particular focus on Wikidata (<https://www.wikidata.org/>), which offers structured data from the Wikimedia universe (Berti 2025b).

Wikipedia was launched twenty-five years ago and is one of the most widely used resources on the World Wide Web, along with its sister projects under the Wikimedia Foundation umbrella (<https://www.wikimedia.org/>). Many debates have been raised about its use, especially in the fields of scholarship and education (Thieme and Weiß 2020). All these debates underscore one indisputable fact: Wikimedia projects are expanding and offer an incredible amount of data, resources and activities; however, we need to use them critically, contribute to their content actively and teach our students how to manage them.

Although these goals are certainly challenging and still present unresolved critical issues, Wikimedia resources are revealing significant potential for preserving data and metadata on antiquities as well. In this regard, Wikidata, which is the central storage for Wikimedia structured data and a free, collaborative, open knowledge base that can be read and edited by both humans and machines, is proving to be a powerful metadata collector for accessing structured information about classical literature according to LOD principles (Bianchini et al. 2021).

The Catalog of the LAGL project is represented by a Wikidata Property (P12869) to add values and generate statements for author identifiers in the form of Resource Description Framework (RDF) triples: <https://www.wikidata.org/wiki/Property:P12869>. If we consider the examples from the introduction of this chapter, we see that the two Wikidata items for Cleobulina of Rhodes (Q2576040) and Moero of Byzantium (Q541746) include an *LAGL author ID* for the two authors with values expressed as CTS URNs (urn:cts:greekLit:tlg0244 and urn:cts:greekLit:tlg0220) that are directly linked to the LAGL Catalog. Such links enable us to connect linguistic data derived from the annotation of ancient sources to the open collection of Wikidata statements that cover a wide variety of metadata

about the life, intellectual activity and modern editions of ancient authors. In terms of identifiers, Wikidata entries include IDs from national and international library systems and projects such as the TLG *Canon*, the *Perseus Project* and *Trismegistos*, which are crucial for linking the ever-growing metadata on ancient authors. The LAGL project and the Wikidata community are currently discussing the creation of a Wikidata Property for adding work identifiers based on linguistic data (see the WikiProject LAGL: https://www.wikidata.org/wiki/Wikidata:WikiProject_LAGL). In this regard, works are presently assigned experimental CTS URNs in the LAGL Catalog and items in Wikidata, as for example the *Mnemosyne* of Moero of Byzantium, which is retrievable as urn:cts:greekLit:tlg0220.ath001 and as Q125957594.

Metadata queried and extracted from Wikidata can be used for various visualisations and analyses. For example, locations currently connected with LAGL authors in Wikidata permit the creation of maps. [Figure 8.3](#) shows an experimental map where Moero is one of the six authors presently collected in the LAGL Catalog under Byzantium (Q23725 and Pleiades ID 520985). This experimental map is available at <https://catalog.lagl.org/> and shows all Wikidata items with an LAGL author ID (P12869) and their associated Wikidata location properties: place of birth (P19), place of death (P20), country of citizenship (P27), ancestral home (P66), place of burial (P119), residence (P551), location of formation (P740) work location (P937), and significant place (P7153). This map is updated regularly to correct possible mistakes and inconsistencies, and keep track of changes in Wikidata. Moreover, it is possible to visualise all this extracted data in a database with links to the map and vice versa.

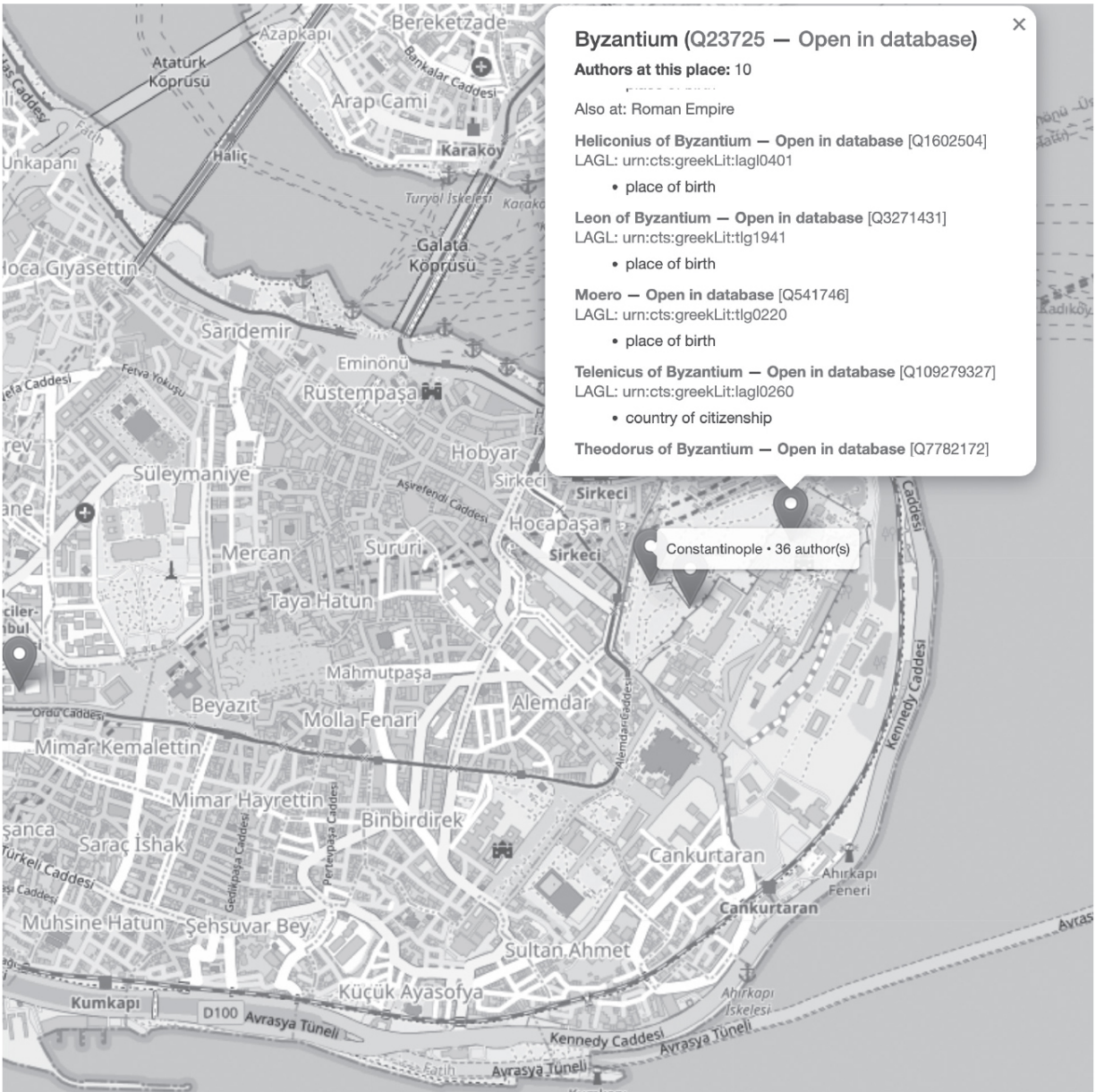


Figure 8.3: LAGL Catalog: Experimental map with Wikidata location properties (CC BY-NC-SA)

In this map other LAGL authors are located in the same area under the Wikidata item for Constantinople (Q16869). This reflects the chronology of the authors and therefore of the history of the city, but in other cases may also underlie inconsistencies in the insertion of data in Wikidata, as with ancient authors who are inconsistently linked to ancient and modern towns or other kinds of toponyms, like the ancient city-state of Samos (Q13580795), the modern city of Samos (Q1018150) or the island of Samos

(Q156882). Moreover, the Wikidata location properties listed in the previous paragraph demonstrate the level of granularity of the knowledge base, but may be challenging for past authors, as ancient sources referring to them generally don't reach this level of granularity. Clear examples in this regard are the place of birth and the place of death, which, in most cases, are difficult to determine from our sources.⁸

Conclusions

The goal of this chapter is to exemplify current historical research that relies on open structured data produced by the principal investigator of the project described in these pages and by third parties. Digital materials include texts digitised from modern editions of ancient sources that preserve bibliographic citations of ancient Greek and Latin authors ranging from classical antiquity to the Byzantine world.

Making use of openly accessible digital data, the chapter shows how it is possible to annotate ancient sources and extract linguistic information thanks to the use of open standards, vocabularies, models and identifiers on the principles of transparency and collaboration that are at the core of the LOD ecosystem. The data derived from linguistic annotations can be utilised by the community in various projects across multiple disciplines. These data can also be enriched with internal and external metadata that will ensure the preservation of both the textual heritage that has been passed on to us and the valuable editorial activities of scholars layered over the past decades and centuries.

Access, sustainability and long-term preservation have been key issues in historical research since ancient times. These issues become more critical and pressing in times of transformation and new media, such as those we are currently experiencing, which are not new in terms of criticality, but in terms of scale and types of technologies that lead us to interact with the machine, which is undoubtedly a new and still largely unknown interlocutor. This reflection is particularly relevant not only to us scholars but primarily to our students, who will increasingly use AI chatbots. Without an understanding of the textual data ingested by LLMs, the risk is that comprehension of AI-generated results will be lost, thereby

compromising the preservation of past textual heritage and its layers of interpretation accumulated over the centuries.

Acknowledgements

This chapter has been written thanks to the generous support of the Deutsche Forschungsgemeinschaft (project no. 434173983). I also thank Camillo Carlo Pellizzari di San Girolamo for his contributions to the discussion on Wikidata properties. I'm also very grateful to Gabriel Bodard, Christopher Ohge, Matteo Romanello and Charlotte Tupman, who have read the chapter and provided me with very helpful comments to improve the text.

Notes

1. Online resources are cited with DOIs whenever possible, otherwise last accessed in June 2026.
2. A comparison between modern massive digitisation projects and ancient collections of texts, such those of the Ptolemies, is interesting to highlight promises, dreams and limits of these enterprises. In spite of the inevitable failures, these initiatives always imply technological revolutions and the creation of massive data not available before.
3. For a useful description of data in the humanities, and a distinction between big data and smart data, cf. [Schöch \(2013\)](#).
4. Further bibliography is available in [Berti \(2026\)](#).
5. This alignment is encoded and stored in the databases behind the *Digital Athenaeus* (<https://www.digitalatheneus.org/>) and the *Linked Ancient Greek and Latin* (LAGL) Catalog (<https://catalog.lagl.org/>).
6. BNJ: *Brill's New Jacoby*. Brill Scholarly Editions. <https://scholarlyeditions.brill.com/bnjo>.
7. See <https://catalog.lagl.org/index.php?what=urn:cts:greekLit:tlg0022>. At the time of writing this paper, annotations of all the entries of the *Suda* are available in the catalogue of the LAGL project: <https://www.lagl.org/tools/suda/>.
8. A similar discussion is valid for Wikidata statements related to the genre and the intellectual activity of ancient authors. A selection of Wikidata properties related to these concepts is experimentally visible in the LAGL Catalog: occupation (P106), field of work (P101), movement (P135), genre (P136). See [Berti \(2025b\)](#).

References

[Adler, A. \(1928–1938\)](#). *Suidae Lexicon*. I–V. Leipzig: Teubner.

- Almas, B. and Berti, M. (2013), ‘Perseids Collaborative Platform for Annotating Text Re-Uses of Fragmentary Authors’, in F. Tomasi and F. Vitali (eds) *DH-Case 2013. Collaborative Annotations in Shared Environments. Metadata, Vocabularies and Techniques in the Digital Humanities*. New York: ACM. <https://doi.org/10.1145/2517978.2517986>.
- Babeu, A. (2019), ‘The Perseus Catalog: of FRBR, Finding Aids, Linked Data, and Open Greek and Latin’, in M. Berti (ed.) *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter, pp. 53–72. <https://doi.org/10.1515/9783110599572-005>.
- Bamman, D. and Burns, P.J. (2020), ‘Latin BERT: A Contextual Language Model for Classical Philology’, in arXiv:2009.10053. <https://doi.org/10.48550/arXiv.2009.10053>.
- Baumann, R. (2013), ‘The “Son of Suda On-Line”’, in S. Dunn and S. Mahony (eds) *The Digital Classicist 2013*. London: University of London, Institute of Classical Studies, pp. 91–106.
- Berti, M. (2016), ‘Greek and Roman Libraries in the Hellenistic Age’, in S. White Crawford and C. Wassen (eds) *The Dead Sea Scrolls at Qumran and the Concept of a Library*. Leiden: Brill, pp. 31–54. https://doi.org/10.1163/9789004305069_005.
- Berti, M. (ed.) (2019) *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110599572>.
- Berti, M. (2021) *Digital Editions of Historical Fragmentary Texts*. Heidelberg: Propylaeum. <https://doi.org/10.11588/propylaeum.898>.
- Berti, M. (2023), ‘Named Entity Recognition for a Text-Based Catalog of Ancient Greek Authors and Works’, in A. Baillot, W. Scholger, T. Tasovac and G. Vogeler (eds) *Digital Humanities 2023. Collaboration as Opportunity*. Graz: University of Graz, Austrian Centre for Digital Humanities, p. 1. <https://doi.org/10.5281/zenodo.8108058>.
- Berti, M. (2024), ‘Digital Canons and Catalogs of Fragmentary Literature’, in F. Neuerburg, T. Tsiampokalos and P. Wozniczka (eds) *Fragmente einer fragmentierten Welt. Zur Problematik des Umgangs mit Fragmenten in der gegenwärtigen klassisch-philologischen Forschung*. Berlin: De Gruyter, pp. 219–238. <https://doi.org/10.1515/9783111508788-009>.

- Berti, M. (2025a), ‘Canoni e cataloghi collaborativi per una filologia sostenibile in ambiente digitale’, in S. Cannavale, V. Casapulla, M. Verstraete and M. Vitali-Rosati (eds) *Orizzonti della filologia digitale. L’Antologia Greca per ripensare formati, paradigmi e collaborazione*. Napoli: Federico II University Press, pp. 131–152.
- Berti, M. (2025b). ‘Linked Ancient Greek and Latin (LAGL) and Wikidata: Structuring and Reusing Data of Classical Literature’ (discussion paper), *Journal of Open Humanities Data*, 11(72), pp. 1-12. <https://doi.org/10.5334/johd.423>.
- Berti, M. (ed.) (2026). *Nomina Omina: Ancient Greek and Latin Proper Names in the Age of Artificial Intelligence*. Digital Classics Online. Themenheft 12(2). Heidelberg: Propylaeum. <https://doi.org/10.11588/dco.2026.12.2>.
- Berti, M. and Costa, V. (2010), *La Biblioteca di Alessandria. Storia di un paradiso perduto*. Tivoli: Edizioni Tored.
- Bianchini, C., Bargioni, S. and Pellizzari di San Girolamo, C.C. (2021), ‘Beyond VIAF: Wikidata as a Complementary Tool for Authority Control in Libraries’, *Information Technology and Libraries*, 40(2), pp. 1–31. <https://doi.org/10.6017/ital.v40i2.12959>.
- Blackwell, C.W. and Smith, N. (2019), ‘The CITE Architecture: A Conceptual and Practical Overview’, in M. Berti (ed.) *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter, pp. 73–94. <https://doi.org/10.1515/9783110599572-006>.
- Blum, R. (1977), *Kallimachos und die Literaturverzeichnung bei den Griechen. Untersuchungen zur Geschichte der Biobibliographie*. Frankfurt am Main: Buchhändler-Vereinigung.
- Burns, P.J. (2024), ‘(Re)Active Latin: Computational Chat as Future Colloquia’, *New England Classical Journal*, 51(1), pp. 1–18. <https://doi.org/10.52284/NECJ.51.1.article.burns>.
- Castelli, E. (2020), *La nascita del titolo nella letteratura greca. Dall’epica arcaica alla prosa di età classica*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110703740>.
- Cayless, H.A. (2019), ‘Sustaining Linked Ancient World Data’, in M. Berti (ed.) *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter, pp. 35–50. <https://doi.org/10.1515/9783110599572-004>.

- Celano, G.G.A. (2024) ‘Opera Graeca Adnotata: Building a 34M+ Token Multilayer Corpus for Ancient Greek’. <https://doi.org/10.48550/arXiv.2404.00739>.
- Crane, G.R., Babeu, A., Bamman, D., Breuel, T., Cerrato, L., Deckers, D., Lüdeling, A., Mimno, D., Singhal, R., Smith, D.A. and Zeldes, A. (2009) ‘Classics in the Million Book Library’, *Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure: Digital Humanities Quarterly*, 3(1). <https://doi.org/10.63744/y4mtrgr32292>.
- Ehrmann, M., Hamdi, A., Linhares Pontes, E., Romanello, M. and Doucet, A. (2023) ‘Named Entity Recognition and Classification in Historical Documents: A Survey’, *ACM Computing Surveys*, 56(2), Article 27. <https://doi.org/10.1145/3604931>.
- Finkel, R., Hutton, W., Rourke, P., Scaife, R. and Vandiver, E. (2000) ‘The Suda On Line (www.stoa.org/sol/)’, *Syllecta Classica*, 11, pp. 178–90. <https://doi.org/10.1353/syl.2000.0005>.
- Heath, J.M.F. (2020) *Clement of Alexandria and the Shaping of Christian Literary Practice. Miscellany and the Transformation of Greco-Roman Writing*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108918640.003>.
- Mahoney, A. (2009) ‘Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia’, *Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure: Digital Humanities Quarterly*, 3(1). <https://doi.org/10.63744/eueck7btfeab>.
- Middle, S. (2024) ‘Linked Ancient World Data: Implementation, Advantages, and Barriers’, *Digital Classics Online*, 10(1), pp. 16–49. <https://doi.org/10.11588/dco.2024.10.104105>.
- Pantelia, M.C. (2022) *Thesaurus Linguae Graecae. A Bibliographic Guide to the Canon of Greek Authors and Works*. Oakland (CA): University of California Press.
- Parsons, E.A. (1952) *The Alexandrian Library, Glory of the Hellenic World. Its Rise, Antiquities, and Destructions*. New York: Elsevier.
- Pfeiffer, R. (1968) *History of Classical Scholarship. From the Beginnings to the End of Classical Scholarship*. Oxford: Clarendon Press.
- Schöch, C. (2013) ‘Big? Smart? Clean? Messy? Data in the Humanities’, *Journal of Digital Humanities*, 2(3), pp. 2–13.

- Stählin, O. (ed.) (1905–1936) *Clemens Alexandrinus*. I–IV. Leipzig: J.C. Hinrichs.
- Thieme, S. and Weiß, J. (2020), ‘Wikipedia & Co.—Was gilt es beim Lesen von Texten aus dem Internet zu beachten?’ in S. Thieme and J. Weiß (eds.) *Lesen im Geschichtsstudium*. Opladen: Verlag Barbara Budrich, pp. 87–98. <https://doi.org/10.36198/9783838553658>.
- Vatri, A. and McGillivray, B. (2018), ‘The Diorisis Ancient Greek Corpus: Linguistics and Literature’, *Research Data Journal for the Humanities and Social Sciences*, 3(1), pp. 55–65. <https://doi.org/10.1163/24523666-01000013>.
- Yousef, T., Palladino, C. and Jänicke, S. (2023), ‘Transformer-Based Named Entity Recognition for Ancient Greek’, in A. Baillot, W. Scholger, T. Tasovac and G. Vogeler (eds) *Digital Humanities 2023. Collaboration as Opportunity*. Graz: University of Graz, Austrian Centre for Digital Humanities. <https://doi.org/10.5281/zenodo.8107629>.
- Zecchini, G. (ed.) (1999) *Il Lessico Suda e la memoria del passato a Bisanzio*. Bari: Edipuglia.
- Zecchini, G. (2000), ‘Harpocration and Athenaeus. Historiographical Relationships’, in D. Braund and J. Wilkins (eds.) *Athenaeus and his World. Reading Greek Culture in the Roman Empire*. Exeter: University of Exeter Press, pp. 153–60.

Chapter 9

Digital epigraphy at Dura-Europos: towards a FAIR-er inscriptional corpus

Anne Hunnell Chen, Carl R. Rice, Ilaria Bucci and J. A. Baird

Introduction

‘Big Dig’ excavations of the late nineteenth and early twentieth centuries, like those at Dura-Europos in mandate-era Syria, bear a complex legacy. On the one hand, the finds uncovered in such expeditions have populated textbooks and museum collections, and the learned study of such objects has shaped insights and debates in a number of humanities disciplines over the last century. On the other hand, imbalances of power and privilege at the time of excavation (and since) have had long-lasting repercussions not only with regard to *who* has access to the physical and intellectual products of such expeditions, but also *whose perspectives* are reflected in the grand narratives about scientific exploration and the interpretation of the ancient past ([Baird 2022b](#); [Baird 2023](#); [Kamash 2023](#)). As recent decades of important research have exposed the complex relationships between Western imperialist interests and early Euro-American archaeology in the Middle East and North Africa (MENA) region, as well as the long-lasting repercussions of those entanglements ([Dyson 1998](#); [Gelin 2002](#); [Scheffler 2003](#); [Díaz-Andreu 2007](#); [Bahrani, Zeynep and Edhem 2011](#); [Starzmann 2012](#); [Kamash 2021](#); [Raja 2023](#)), how do scholars work responsibly with collections (both artefacts and archives) that descend from such classic

sites? How might the digital approaches already gaining ground by serving certain disciplinary interests be harnessed or tweaked to simultaneously help bring about more equitable access to the intellectual products of those excavations and to help rebalance processes of knowledge-making?

These questions stand at the heart of the work of the International [Digital] Dura-Europos Archive (IDEA). In this chapter, we aim to introduce readers to IDEA and its implications for the study of Dura-Europos and analogous sites where descendant collections and specialist insights are fragmented across continents, institutions and disciplinary traditions. We focus on a heavily inscribed area of the site, the monument known as the Palmyrene Gate—marked by texts in multiple languages left by diverse groups over the centuries—and use it to illustrate how Linked Open Data (LOD) methods, including an innovative use of the Wikidata platform, can be harnessed to address both ethical and research-related challenges at sites with long and complex excavation histories. After a brief introduction to the site and its archaeological legacy, we discuss IDEA’s use of Wikidata in more detail and highlight the platform’s utility for epigraphic work. From there, we use the inscriptional data to demonstrate the possibilities of IDEA’s particular blend of LOD methods for making the important data from large-scale archaeological projects like those at Dura more accessible and intelligible to a diverse range of scholars and interested communities, a fundamental step towards more accessible and more collaborative archival interventions.

Dura-Europos: The site and its history

The ancient city of Dura-Europos, situated on an escarpment above the Euphrates River near the village of Salhiyeh, is one of the most important archaeological sites in modern Syria ([Baird 2018a](#)). The city was occupied for nearly six centuries and stood throughout this history on the changing frontier between empires. The Hellenistic settlement was probably founded in the late fourth or early third century B.C.E. as a military garrison to control an important crossing on the Euphrates. The city’s monumental western gate, referred to by scholars as the Palmyrene Gate, was constructed in this earliest phase ([James 2019](#): 232–4; [Baird 2022a](#): 307–9).

In the late second century B.C.E., the Arsacids annexed Dura-Europos and retained it for nearly three hundred years, save for a brief period of Roman occupation during Trajan's Parthian campaigns ([Gregoratti 2016](#); [Baird 2018a](#): 23). The geographer Isidore of Charax named the site as one stop along an itinerary through Mesopotamia at the end of the first century B.C.E., calling the city 'Dura, a city of Nicanor, a foundation of the Macedonians, also called Europos by the Greeks' (Isid. Char. *Mans. Parth.* 1). Isidore's description links the two names by which the city was called in antiquity and which form the hybrid, hyphenated name scholars use for the site ([Baird 2018a](#): 16–19).

The city transferred into Roman hands in 165 C.E., ushering in what is perhaps the best understood period of its history ([Baird 2018a](#): 29–37). Dura remained a crucial military outpost for the Romans and the city saw an increasing militarisation, with its northern parts being converted into a military base ([James 2019](#)). During the Arsacid and Roman periods, the city's inhabitants erected a striking array of religious buildings: a series of temples dedicated to Semitic, Mesopotamian, Greek and Roman deities; a Mithraeum; an elaborately decorated synagogue; and a Christian building—one of the earliest such structures ever attested in the archaeological record.

In the mid-250s, the Sasanian king Shapur I besieged the city as tensions along the Roman empire's eastern frontier escalated. The Roman occupants engaged in counter-siege operations, notably the use of backfill to shore up the western wall of the fortifications. This backfill ultimately created specific taphonomic conditions that preserved not only the architectural surfaces of many buildings but also an exceptional amount of organic material (including wooden and leather artefacts, parchments and papyri). As a result, a wide range of inscribed and painted surfaces, together with everyday objects, survived in remarkable condition. Following the Sasanian sack, the settlement was largely abandoned and never fully reoccupied.¹

The academic record associated with the site commences in the late nineteenth century, with references in travel reports. The so-called 'Wolfe Expedition', among whose party was archaeological photographer John Henry Haynes, visited the site in 1885. During his visit, Haynes captured a striking photograph of the Palmyrene Gate, the main gate of the settlement overlooking the steppe to Palmyra, quite possibly the first photograph of Dura ([Ousterhout 2016](#): Figure 71).² European intervention at the site

heightened after 1920 when British troops stationed at Salihiyeh under the command of Captain M.C. Murphy took notice of well-preserved ancient wall paintings in the northwest corner of the fortress ([Breasted 1924](#); [Brody 2011](#)). This chance encounter, a byproduct of power-jockeying between the British and French as they established the limits of their respective mandate territories, would ultimately set in motion the large-scale excavation of the site ([Baird 2014](#); [Baird 2018a](#): 1–16; [Baird 2020](#); [James 2025](#); [Elsner 2025](#)).

The first scientific expeditions at Dura were sponsored by the French Académie des Inscriptions et Belles-lettres from 1922 to 1923 ([Cumont 1926](#)), but were discontinued due to regional instability as Syrians resisted foreign rule ([Thompson 2020](#)). Renewed excavations, jointly sponsored by the French Académie and Yale University, began again in 1928 and continued for ten seasons through 1937.³ In 1986, after a hiatus of nearly fifty years, the French Ministry of Foreign Affairs and the Syrian Directorate of Antiquities and Museums established the Mission Franco-Syrienne de Doura-Europos (MFSDE) to carry out re-examination and conservation of previously excavated remains alongside limited new excavations ([Leriche, Coqueugniot and De Pontbriand 2011](#)).⁴ Archaeological work ceased once again in 2011 with the outbreak of the Syrian Civil War ([Alsaleh 2021](#)). While piling in comparison to the immense human suffering wrought by this conflict, satellite imagery and first-hand accounts confirm that the site has been subject to violent destruction and looting in the course of the war ([Almohamad 2024](#)). These circumstances, coupled with calls to do more with old archaeological data and to re-examine interpretations in light of the historical contexts of their production, have increased the importance of the legacy data and documents related to excavations at Dura prior to 2011. With renewed attention to the legacy datasets and documents comes the opportunity and ethical obligation to explore the use of digital tools to bridge persisting inequalities in access to information descendent of foreign-run excavations.

[The legacies of early archaeology](#)

Sites like Dura with excavation histories that stretch back to archaeology's nascence present a number of practical and ethical challenges. Archaeological excavation is inherently destructive: removal of dirt fill that has preserved buildings and artefacts buried thousands of years ago exposes rare and fragile historical evidence to environmental factors from which they would otherwise be naturally shielded ([Lucas 2001](#); [Baird 2020](#): 8–11).⁵ Well-intended ambitions to responsibly preserve excavated finds by removing them to museum contexts and to share out access to important archaeological resources via the practice of *partage* effectively fragmented related finds from their immediate place of discovery and landed them in museum collections sometimes continents apart ([Marlowe 2024](#): 452).⁶ In the case of Dura, the twentieth-century practice of *partage* has resulted in artefacts from Dura populating many globally significant collections, such as the Yale University Art Gallery in New Haven, the Louvre in Paris and the National Museum in Damascus.⁷

Moreover, the documentary materials associated with the excavations, such as field notebooks and photographs, typically returned with the field directors to their home institutions, where they remain, catalogued in Western languages and according to the conventions and intellectual frames of the collecting institution.⁸ For the contextual data catalogued for images like Wikidata Q124840826 of the Palmyrene Gate, for instance, an initial analogue record was made in the field at the time the image was captured and served as the catalyst from which subsequent new analogue records for the same image descended. Upon the move to digital cataloguing, these analogue records formed the basis for an internal digital database, which was later made public with ArtSTOR. Recently, in 2023, these images and the accompanying metadata were migrated to JSTOR. At each of these steps, new data, metadata and interpretation could be (and sometimes was) introduced, even as old data, metadata and interpretations were likewise perpetuated ([Baird 2023](#): 20–25).

Historically, the metadata developed over objects' lifespans, often in the context of Western institutions, has powered keyword searching and, thus, shaped findability. The title field of the photograph Q124840826, for instance, names the building depicted (the Palmyrene Gate) and a subset of finds (the altar). However, the metadata codified this way also occludes the other subjects this photo contains: the smiling yet unidentified labourer

holding the fragment of a small relief; the inscriptions chiselled into the wall behind him; or the evidence of a variety of local dress present in the region in the early 1930s, to name just a few. Even the names that populate the data that is there are not unproblematic. Typically, metadata entries derived from Western academic contexts inscribe Western academic placenames, while ignoring local or more colloquial ones. As a result, only searches that use those names, such as ‘Palmyrene Gate’, will yield results (and not, say, Bab al-Hawa, one name used among local participants in the excavations). Technological advances and the resources for major cataloguing efforts rarely coincide, and thus there is a high risk for biases or ellipses in the initial cataloguing of physical archival documents to become reinscribed from generation to generation ([Baird 2023](#)). Disciplinary specialists would continue to access these materials to support additional research at a distance,⁹ leading ultimately to the publication of fieldwork reports and catalogues of finds that further reinscribed these biases and ellipses and fragmented contextually related information across publication venues that spoke to the frameworks of contemporary disciplinary specialisations. Power imbalances at the time of excavation combined with these conditions of publication to shape a century-long compounding body of scholarship that still today disadvantages Arabic-speakers and marginalises the contributions and perspectives of local labourers like the man in the photograph mentioned above. More recent interdisciplinary approaches and the burgeoning opportunities of digital connectivity have increased awareness of these and other problems, even as they have unintentionally perpetuated the fragmentation of contextually related information (see, for example, [Kahn and Simon 2022](#)). Foundational work in various cultural heritage-related domains has yielded an ever growing ecosystem of digital resources related to ‘Big Dig’ sites with cross-disciplinary significance. This ecosystem includes collection metadata and images available from institution databases, placename disambiguation in geo-spatial gazetteers, period disambiguation in chronological gazetteers, controlled vocabularies for collection description, expert insights in databases related to specific media like inscriptions, coins and papyri, not to mention the bibliographic records, and now increasingly, open access articles ([Figure 9.1](#)).

Yet, not all of these resources and datasets adhere to the FAIR principles ([GO FAIR 2024](#)).¹⁰ It is an acknowledged problem, moreover, that proliferating FAIR datasets remain ‘islands’ in a sea of data that are not yet as well-networked and interlinked as they could be ([Jackson et al. 2009](#); [Blanke et al. 2012](#); [Pett and Barker 2022](#)). Further, the cultivation and curation of many of these projects and datasets favour speakers of Western European languages and are shaped exclusively according to the disciplinary interests of scholars and curators in the Global North, largely without the input of important stakeholder communities ([Kamash 2021](#); [Kamash 2023](#); [Baird, Kamash and Raja 2023](#)). Indeed, many of these international conversations are currently happening without significant input from Middle Eastern voices. Consequently, there is a distinct danger of both fossilising received one-sided narratives, and developing shared international data standards that inadvertently further marginalise already-marginalised voices, collections and epistemologies. What if we could work together to highlight the pathways between the fragmented but ever-growing resources? And what if we could do it in a way that simultaneously passed those bridges into Arabic, building towards better collaborations with Syrian cultural heritage professionals and other stakeholder communities in Syria and the diaspora fed by years of conflict and displacement?

Linked Open Data in practice at the Palmyrene Gate

IDEA is creating the first multilingual, inter-institutional register of content in the form of a Linked Open Data (LOD) dataset curated through interdisciplinary collaboration and engagement with Syrian communities.¹¹ Central to the project’s approach is the creative use of the Wikimedia ecosystem to provide linguistic and epistemic supports that broaden accessibility, and offer the possibility of extending the digital archive connected with the site to include content and perspectives problematically sidelined in previous generations. IDEA’s use of the Wikidata platform aims to build bridges among ‘islands’ of FAIR data to recontextualise content—like inscriptions—that are often treated apart from the complex assemblages

[illegible]

Today, scholars have begun to recognise the importance of local stakeholder communities' meaningful involvement in archaeological work and curation of sensitive collections ([Atalay 2012](#); [Baird, Kamash and Raja 2023](#); [Chen 2023](#); [Baird and Almohamad 2025](#)). It is not always straightforward, however, to connect with stakeholders in all cases that call

for community engagement, for instance in circumstances where conflict or other crises have driven displaced persons into a diaspora spanning the globe.¹² Moreover, there are additional considerations in reaching out to stakeholders in vulnerable circumstances: is it ethical to call upon the intellectual and/or emotional labour of already strained communities? Emerging LOD methods offer a distinct utility to face these challenges, providing a means by which to transcend language and collaborate across distance, while allowing community members to contribute in ways that respect the time and energy they are able or willing to offer.

Since Tim Berners-Lee first introduced the idea of LOD nearly two decades ago ([Berners-Lee 2006](#)), a growing number of institutions and projects have committed to make the data they manage available for digital discovery and reuse ([Oomen and Baltussen 2012](#); [van Hooland and Verborgh 2014](#); [Mayer 2015](#); [Estermann 2016](#); [Stinson 2017](#); [Davis and Heravi 2021](#); [Nicholson et al. 2023](#)). The overarching vision of LOD aims to create a digital network of relationships among related resources, traversable by humans and computers alike, allowing related datasets and digital resources to speak the same computer-based language behind the scenes, no matter the language of content curation.¹³

But to avoid the pitfalls of reifying existing accessibility and perspectival biases, it is important that those datasets and digital resources are structured for future use by a diverse range of stakeholders. While the existing pillars driving cultural heritage's digital information landscape linguistically and epistemically favour Europeans and Americans ([Napoli 2024](#); [Nurmikko-Fuller 2023](#): 70–71), the good news is that many of these resources are, more often than not, aware of their incompleteness and the need for more diverse perspectives weighing into the shaping of the content they manage. In fact, many are malleable and even invite help from professionals with relevant knowledge bases (Pleiades geo-spatial gazetteer: <https://pleiades.stoa.org>; PeriodO chronological gazetteer: <https://perio.do/en/>; Getty Vocabularies: <https://www.getty.edu/research/tools/vocabularies/>). There is a looming problem, however, in that the tools and methods of LOD, including 'Authority' resources' malleability, are not equally well-known across global professional communities. A key part of our work, therefore, *must be* to seek ways to bring a more diverse range of 'Authorities' and experts—like the many Syrian heritage professionals who

have been on the front lines protecting sites and collections—into international conversations about shared digital futures that in large part continue to exclude them ([Almohamad 2021](#)).¹⁴

This chapter focuses on the Palmyrene Gate—also known as the Palmyra Gate, Porta Praetoria or Main Gate—likely built in the mid-second century B.C.E. and serving as Dura’s main western entrance until the city’s fall in the mid-third century C.E. Its passageway and adjacent rooms preserve an extraordinary concentration of artefacts, including altars, figurative sculpture and some 200 inscriptions—painted, carved or scratched—in languages ranging from Greek and Latin to Aramaic dialects and Safaitic. While many were left by the Roman military, others record the names and commemorations of people who passed through or lingered at the gate, offering a vivid glimpse into the city’s cultural and linguistic diversity ([Baird 2018b](#)). But the Palmyrene Gate also bears more recent stories, with graffiti left by people seeking to inscribe themselves into the place to remember and be remembered ([Baird and Almohamad 2025](#)). The diversity of people, languages and artefacts preserved at the Palmyrene Gate are just one example of the complex web of relationships that define Dura-Europos’s material and social history. This multiplicity of actors, objects and textual evidence resonates with LOD principles, where archaeological knowledge is conceptualised as a network of relationships among people, places, things, events and bibliography.

In practical terms, therefore, the IDEA team is working from high-quality print and digital resources to systematically establish Wikidata records—or ‘items’, to use WD’s nomenclature—for buildings, objects, inscriptions, people and key archival documentation related to Dura. These records become a place to collect and align existing unique identifiers for related digital content maintained in repositories across the web (often monolingual) to enable cross-walking between separately maintained ‘Authority’ resources with closed contributing communities. The language agnosticism of Wikidata’s design effectively passes the discipline-specific knowledge and intellectual networks required to track down relevant existing identifiers across language barriers with minimal bespoke translation work thanks to the ability to piggyback off thousands of existing community-contributed translations for concepts and metadata fields shared by numerous cultural heritage projects and institutions. Upon the creation of

an item, Wikidata assigns an autogenerated Q-number (QID) that serves as a language-agnostic signifier for the concept represented by the item. Variant spellings, language differences and divergent interpretational traditions can be paired with this number to assert equivalencies that improve discoverability. For example, the site of Dura-Europos itself has been assigned the number Q464266, and (so far) community members have linked more than fifty language labels (ranging from English to Azerbaijani to Ukrainian to Arabic).

Making use of the same principle, based on conversations with members of the Salihyeh community, Adnan Almohamad, an IDEA team member, has also asserted a local tradition for the site's name (الصور 'the wall'). Significantly, this colloquial name is an indication of a different—but equally valid—epistemic tradition that co-exists alongside the scholarly tradition, one rooted in the immediate environment rather than the frameworks of ancient empires. Importantly, mapping different language and epistemic traditions to a language-agnostic signifier effectively allows the same concept to be searchable according to the labels preferred by different communities. Parallel, or even conflicting, traditions can co-exist in such an environment. For example, multiple contributors or communities may hold diverging understandings of the site's identity or history; these can be stored as distinct statements (see below) with qualifiers and references, making it possible to surface the differences rather than smoothing them over. Thoughtful deployment of such a feature can make communities that subscribe to conflicting interpretations aware of one another where distance and linguistic barriers may have previously stood in the way.

Wikidata also allows editors to map the network of relationships among elemental components—items for people, buildings, archival documents, objects, inscriptions—by assigning syntactic relationships between individual items. This is accomplished by articulating a 'statement', effectively a metadata field, on individual items. Statements are typically comprised of a 'property' and a 'value', which can be either another item (with its own Q-ID) or a data value (depending upon the type of property in question). For example, the statement in [Table 9.1](#) communicates in both English and Arabic that a specific ancient graffito was discovered at the Palmyrene Gate.

In this statement, the item ‘Inscription PR2.D27 from Palmyrene Gate, Dura-Europos’ (‘من البوابة التدمرية، دورا أوروبوس PR2.D27 النقش’) is assigned a ‘location of discovery’ (مكان الاكتشاف) with a value ‘Palmyrene Gate’ (البوابة التدمرية). Put another way, one Wikidata item (‘Inscription PR2.D27’ Q111931996) is linked to another Wikidata item (‘Palmyrene Gate’, Q98930725) through a Wikidata property (‘location of discovery’, P189). Moreover, because the data is linked through the QIDs, the relationships between the item, property and value transcend linguistic barriers. In other words, a user searching in, say, Mandarin could query data uploaded in English by an English-speaking Wikidata editor. For the first time, the metadata for each building, artefact and archival document the team has added became searchable and readable in Arabic—thus, now accessible to new communities with interests in the site. And while semantic meanings of the signifiers have been historically predominantly shaped by Western sensibilities, broader and more inclusive participation, especially from marginalised communities, can help introduce alternative perspectives and gradually balance this bias.

Wikidata statements can express other information, both at a broader and more specific level. For example, the statement in [Table 9.2](#) expresses the broad classificatory schema to which an item, Inscription PR2.D27, belongs (again, both in English and Arabic):

Whereas the statement in [Table 9.2](#) showed a broader generic category, the statement in [Table 9.3](#) connects the inscription to a specific archival photograph of it taken in the 1920s.

Table 9.1: Example Statement 1

Item	Property	Value
English: Inscription PR2.D27 from Palmyrene Gate, Dura-Europos	location of discovery	Palmyrene Gate
Arabic: التدمرية، دورا أوروبوس PR2.D 27 النقش من البوابة	مكان الاكتشاف	البوابة التدمرية
WD Q- Q111931996 ID	P189	Q98930725

Table 9.2: Example Statement 2

Item	Property	Value
English: Inscription PR2.D27 from Palmyrene Gate, Dura-Europos	instance of	graffiti
Arabic: التدمرية، دورا أوروبوس 27 من PR2.D النقش البوابة	نموذج من	مخرشة / غرافيتو
WD Q- Q111931974 ID	P31	Q3567529

The effect of the individual statements on the record for this inscription is to digitally network it with other contextually related details, including an existing record for the architectural feature it is scratched into, and archival photographs that depict it in situ, among other details. Tools built into the Wikidata environment, such as the Image Positions annotation tool that allows editors to label depicted features by establishing coordinates internal to a pictorial space, can be put to use to further clarify the relationships between items ([Figure 9.2](#)).

Moreover, through the series of links established among related records, it is possible to source the spatial coordinates stored on the record for the place of discovery (Palmyrene Gate) such that the findspot of the graffito can be visualised in map view, as well as to return a list of records for physical objects found at the same location but today preserved in collections thousands of miles apart. Importantly, end-users can make such inquiries of the dataset in their native language thanks once again to the multilingual environment.¹⁵

Table 9.3: Example Statement 3

Item	Property	Value
English: Inscription PR2.D27 from Palmyrene Gate, Dura-Europos	depicted by	Dura-Europos archival photograph, YUAG Negative number dura-b109~01

Item	Property	Value
Arabic: التدمرية، دورا PR2.D النقش أوروبوس 27 من البوابة	صور من قبل	صورة أرشيفية لدورا أوروبوس، معرض فنون جامعة بيل رقم dura-b109~01 نيجاتيف الصورة
WD Q- Q111931974 ID	P1299	Q124841766

Through these means, content held in separate institutions, published in different languages, or fragmented across disciplinary silos comes together via the web of relationships in the Wikidata environment. To facilitate multilingual multivocality in shaping the dataset, and as a basis for future computer-aided data alignment checks as part of periodic data maintenance in Wikidata, IDEA contributors collectively shape ShEx Schemas for people, places, things, events and bibliography related to Dura ([Thornton, Seals-Nutt and Chen 2024b](#)). The benefits of computer-aided consolidation of multidisciplinary and multilinguistic information are myriad for researchers, instructors and students. Perhaps most importantly, however, this particular approach to drawing fragmented content together prepares the way for a more diverse range of editors to weigh into the shaping of the archival record connected with the site in ways that have never been possible before because of collection dispersal and linguistic barriers.¹⁶

A further benefit to making use of the Wikidata environment is that it offers an accessible means for extension of the archival record to include new contributions that speak to a wider range of perspectives. In combination, the Wikidata and Wikimedia Commons platforms can be harnessed to network new archival content in a variety of media, including audio, video, 3D models and more, into the LOD record, offering a means by which to begin filling in longstanding archival elisions that effectively write out the lives of contemporary Syrians. For example, while Dura was conceived of as an abandoned site by its excavators—a view that clearly privileged its Hellenistic to Late Roman phases over those that came after—it actually had a long history of human presence after the Sasanian siege, including the presence of modern and contemporary Syrians: over the twentieth century, that included people employed as labourers for various

expeditions, many visitors and tourists, as well as people who lived locally to the site who used it to graze animals, for recreation and other purposes ([Baird and Almohamad 2025](#)). Some of these people have left a material trace at the site, including within the Palmyrene Gate. There, alongside the messages in Greek left by Roman military personnel two millennia ago, are modern Arabic graffiti ([Baird and Almohamad 2025](#): Figure 13.4). Marks like that made by a contemporary visitor named Ahmad Hamid inscribe their own legacy alongside those of the ancient occupants of the site, attesting to their presence and also to a shared human impulse to make a mark that links them to this specific place. The open editorial principle of Wikidata means that the record of the Palmyrene Gate can be expanded to include contemporary material, not usually catalogued by archaeologists, and that the ancient and modern graffiti can be treated on the same terms, speaking back to disciplinary blind-spots that give preference to particular periods or categories of material (Q133809452).¹⁷ Such inclusions enable a step towards a more level playing field, not only for users in multiple languages, but also for subverting the chronological privilege (for example, of classical languages), treating different traditions in the same way.

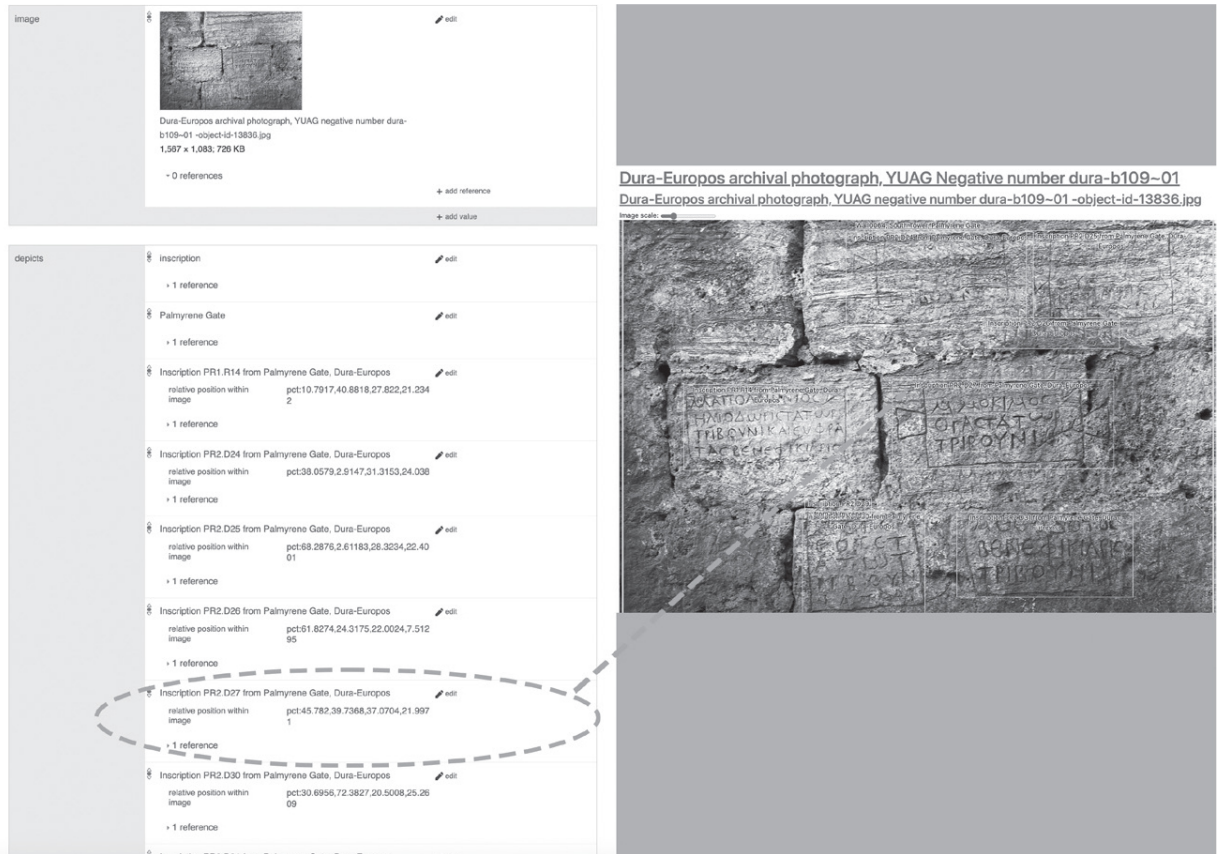


Figure 9.2: Archival photograph (YUAG dura-b109~01; Q124841766) annotated by members of the IDEA team to clarify location of inscriptions on the walls of the city gate. Yale University Art Gallery (CC BY-SA 4.0)

Ultimately, the Wikimedia component of IDEA’s collectively shaped dataset establishes a reusable and extensible multilingual work-product that can be built on for future research. It also dynamically powers an English/Arabic web application for user-friendly browsing of information related to the buildings, artefacts and archival documents related to the site (Thornton, Seals-Nutt and Chen 2024a; <https://dura-europos.stories.k2.services/>).

Conclusion

IDEA aims to transform colonially entangled artefact and archive collections into intercultural and intergenerational learning spaces, where students, professionals and stakeholder communities worldwide can grapple inclusively with archaeology’s inequitable legacies. These digital methods

also de-silo insights across disciplines such as archaeology, art history, classics and religious studies. As more Global North collections become machine-readable, there is a risk of fossilising one-sided narratives and marginalising already underrepresented voices. Thoughtful design can counter this. Given the realities of the recent war and existing linguistic and epistemic biases in Dura-related datasets, IDEA prioritises tools that make it possible to work around issues of community displacement, inability to travel and language barriers. Using the Wikimedia environment, the project bridges ‘islands’ of FAIR data, recontextualising dispersed archaeological content and fostering collaboration with marginalised scholars and descendant communities. As such, it establishes an innovative methodological model applicable to the many cultural heritage contexts similarly entangled with legacies of colonialism.

Acknowledgements

IDEA builds on FAIR data provided by institutional and discipline-specific digital resources, and we are especially grateful for the partnership and support of the Yale University Art Gallery staff. All Wikidata generated by this project is available under a CC0 Public Domain Dedication. Wikimedia Commons contributions mirror institutionally generated FAIR resources and are available under CC BY licences.

The work of IDEA is made possible through the generous support of the National Endowment for the Humanities (2022–25) and the American Council of Learned Societies (2025–27). We are grateful to the community of Wikidata editors and the members of the IDEA team for their dedication and support. We also extend our heartfelt thanks to colleagues who contributed to ‘Lives in Ruins’, an initiative made possible with funding from the OSUN/Talioires Network of Engaged Universities.

Notes

1. Ammianus Marcellinus referred to Dura as a ‘deserted town’ in his account of Julian’s eastward marches in the 360s: Amm. Marc. 23.5.8, 24.1.5. On the gradual abandonment of the site and evidence for ephemeral occupations after 256, see [Baird 2012](#).

2. The photograph is now at Cornell University, J.R. Sitlington Sterrett Collection of Archaeological Photographs, collection ID: Sterrett_018, collection number: 4776, available at: <https://digital.library>

[.cornell.edu/catalog/ss:12513113](https://cornell.edu/catalog/ss:12513113); Wikidata <https://www.wikidata.org/wiki/Q130215869>.

3. For a complete list of the main studies arising from these expeditions, including both preliminary and final reports, see [Baird \(2018a\)](#): xiv–xvi).

4. Results of the MFSDE were presented in the volumes *Doura-Europos études* and *Europos-Doura Varia 1*. Note that the Franco-Syrian mission later adopted the revised Mission Franco-Syrienne de Europos-Doura (MFSED) to align with the team’s re-prioritisation of the placename associated with the Hellenistic city.

5. Less invasive modern methods include use of ground penetrating radar and LiDAR visualisations to target the deployment of more destructive exploratory methods.

6. Under French Mandate, the system of *partage* governed how antiquities from excavations were divided. Foreign teams were allowed to dig on condition that the finds were shared between the excavators and the Syrian state. Typically, objects deemed duplicates or less significant went to the excavators’ home institutions, while the most important or unique pieces were retained for Syrian museums, above all the National Museum in Damascus. Although in theory *partage* offered a less exploitative model for sharing an excavation’s spoils that was an improvement on earlier practices, as Marlowe has neatly summarised the matter: ‘*partage* rarely lived up to the model of equality implied by its name (from the French “partager”, “to share”). Excavators frequently kept more than their share by under-reporting the quantity or quality of their finds and some sponsoring institutions retained finds that were intended to be temporary loans.’ On the latter issue in relation to the papyri and parchment documents recovered at Dura, see [Baird \(2018a\)](#): 60). On *partage*’s entanglement with extractive imperialist ambitions to build national and institutional prestige with impressive museum collections, see also [Dyson \(1998\)](#); [Bahrani, Zeynep and Edhem \(2011\)](#); [Stevenson \(2019\)](#); and [Griswold \(2020\)](#).

7. Artefacts from Dura are attested in at least nine additional collections. Many such objects were brokered as trades between peer institutions, or collected by foreign travellers to the site and later bequeathed to academic collections. On analogous fragmenting of contextually related artefacts from early excavations in Egypt into collections around the world, see [Stevenson \(2019\)](#).

8. Only portions of Yale University Art Gallery’s archive related to the Dura excavations, primarily photographs, have so far been digitised and made available as digital surrogates. Large proportions of the archive’s holdings, including handwritten field notebooks, correspondence and analogue artefact records with handwritten annotations supplemental to published accounts, have yet to be systematically digitised.

9. Following the volumes summarising the results of the excavations by Franz Cumont and the Preliminary Reports of the Yale-French expedition, fifteen Final Reports were published in English between 1943 and 2004, focusing on specific classes of objects or contexts. Works published in recent decades drawing on archival documentation have provided valuable syntheses of key topics, including a site overview ([Baird 2018a](#)), domestic contexts ([Baird 2014](#)), the military sector ([James 2004](#); [James 2019](#)), connections to Palmyra ([Dirven 1999](#)), the Mithraeum ([McCarty and Dirven 2020](#)) and a critical historiographical overview of Cumont’s work related to Dura ([Kaizer 2020](#)).

10. FAIR principles are currently considered a standard for works in epigraphy (see, for instance, the project FAIR Epigraphy, [Heřmánková, Horster and Prag 2022](#)) and graffiti, both pictorial and textual ([Trentin 2021](#)).

11. IDEA team members’ contributions are made directly to open knowledge bases (in particular, Wikidata and Wikimedia Commons) hosted via the Wikimedia Foundation. Currently, the register of content IDEA manages includes approximately 25,000 records for artefacts, buildings, archival documents, people, events and bibliography related to Dura, as well as a daily growing human- and

computer-readable network of assertions about relationships among those components (for example, object X → found at → place Y; photograph A → depicts → place Y). Public-facing records and the work-in-progress of mapping the network of relationships can be browsed via Wikidata, Pleiades and the Getty Art and Architecture Vocabularies.

12. The United Nations estimated 350,209 civilian casualties between 2011 and June 2022 due to the conflict in Syria. As of December 2022, 6.8 million people were reported to be internally displaced, and as of June 2023, 5.3 million Syrian refugees and asylum seekers had registered globally. For statistics and a timeline of key events in the Syrian conflict, see [Loft and Kirk-Wade \(2024\)](#).

13. For a gentle introduction to the concept of LOD, see [The Fitzwilliam Museum \(2022a\)](#); [The Fitzwilliam Museum \(2022b\)](#).

14. Work on the part of Syrian-led cultural heritage organisations like Heritage for Peace (HfP), The Day After Heritage Protection Initiative (TDA-HPI) and the ATHAR Project has helped to document and examine the motivations and mechanics behind recent destructions in Syria. For a more comprehensive discussion of initiatives concerned with cultural heritage destruction in Syria and Iraq, see [Munawar \(2019\)](#).

15. SPARQL query and results in English: <https://w.wiki/BnSG>; in Arabic: <https://w.wiki/FVbv>.

16. In spring 2024, as part of the ‘Lives in Ruins’ outreach initiative funded by the OSUN/Talloeires Network of Engaged Universities and held by Baird at Birkbeck College, the IDEA dataset served as a hands-on training space for familiarising a group of Syrian cultural heritage professionals with the possibilities offered by LOD methods. Born of collaborative discussions in these workshops, participants in the pilot workshop designed targeted projects for data enrichment (that is, LOD annotation, etc.) carried out in the 2024–2025 academic year.

17. Wikidata: <https://www.wikidata.org/wiki/Q133809452>. The differential treatment of inscriptions including graffiti of different periods on particular monuments is a broader issue, with decisions about what is worth recording and what is considered vandalism persisting ([Merrill 2011](#)).

References

[Almohamad, A. \(2021\)](#) ‘Community Responses to Protecting Archaeological Sites During Conflict: Evidence from Northwest Syria’, *CARA*. Available at: <https://thearchaeologist.news/thearch/en/community-responses-to-protecting-archaeological-sites-during-conflict/>.

[Almohamad, A. \(2024\)](#) ‘Syria and Dura-Europos Today’, *Pasts Imperfect*, 5(June). Available at: <https://pasts-imperfect.ghost.io/pasts-imperfect-special-issue/>.

[Alsaleh, A. \(2021\)](#) *Historical Dictionary of the Syrian Uprising and Civil War*. Lanham: Rowman & Littlefield Publishers.

Ammianus Marcellinus (2019) *History*. Translated by J.C. Rolfe. *ToposText*. Available at: <https://topostext.org/work/493>.

[Atalay, S. \(2012\)](#) *Community-Based Archaeology: Research with, by, and for Indigenous and Local Communities*. Berkeley: University of

California Press.

Bahrani, Z., Zeynep, Ç. and Edhem, E. (2011), *Scramble for the Past: A Story of Archaeology in the Ottoman Empire, 1753–1914*. Istanbul: SALT.

Baird, J.A. (2012), ‘Dura Deserta: The Death and Afterlife of Dura-Europos’, in N. Christie and A. Augenti (eds) *Urbes Extinctae: Archaeologies of Abandoned Classical Towns*. Farnham: Ashgate, pp. 307–29.

Baird, J.A. (2014), *The Inner Lives of Ancient Houses: An Archaeology of Dura-Europos*. New York: Oxford University Press.

Baird, J.A. (2018a), *Dura-Europos*. London: Bloomsbury Academic.

Baird, J.A. (2018b), ‘In Small Scratches Forgotten: New Perspectives on Graffiti from Ancient Dura-Europos’, *Tijdschrift Voor Geschiedenis*, 131(1), pp. 17–33. <https://eprints.bbk.ac.uk/id/eprint/19481/>.

Baird, J.A. (2020), ‘The Ruination of Dura-Europos’, *Theoretical Roman Archaeology Journal*, 3(1), pp. 1–20. <https://doi.org/10.16995/traj.421>.

Baird, J.A. (2022a), ‘Dura-Europos’, in T. Kaizer (ed.) *A Companion to the Hellenistic and Roman Near East*. Hoboken: Wiley-Blackwell, pp. 305–314.

Baird, J.A. (2022b), ‘The Site of the Archive: Responsibility and Rhetoric in Archival Archaeology of the Middle East’, in O. Boubou, A.C. Miranda and R. Raja (eds) *Archival Historiographies. The Impact of Twentieth-Century Legacy Data on Archaeological Investigations*. Turnhout: Brepols, pp. 161–73.

Baird, J.A. (2023), ‘Unclassified: Structured Silences in the Archaeological Archive’, in R. Raja, (ed.) *Shaping Archaeological Archives: Dialogues between Fieldwork, Museum Collections, and Private Archives*. Turnhout: Brepols Publishers, pp. 19–32.

Baird, J.A. and Almohamad, A. (2025), ‘The Ruins That Remain: Remembering Dura-Europos in Salhiyeh’, in A. H. Chen and L.R. Brody (eds) *Dura-Europos Past, Present, Future*. Turnhout: Brepols Publishers, pp. 200–211.

Baird, J.A., Kamash, Z. and Raja, R. (2023), ‘Knowing Palmyra: Mandatory Production of Archaeological Knowledge’, *Journal of Social Archaeology*, 23(1), pp. 76–98.

- Berners-Lee, T. (2006), 'Linked Data—Design Issues'. Available at: <https://www.w3.org/DesignIssues/LinkedData.html> [last modified 18 June 2009, at 18:24:33]
- Blanke, T., Bodard, G., Bryant, M., Dunn, S., Hedges, M., Jackson, M. and Scott, D. (2012), 'Linked Data for Humanities Research—The SPQR Experiment', in *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, pp. 1–6. <https://doi.org/10.1109/DEST.2012.6227932>.
- Breasted, J.H. (1924), *Oriental Forerunners of Byzantine Painting: First-Century Wall Paintings from the Fortress of Dura on the Middle Euphrates*. Chicago: University of Chicago Press.
- Brody, L. (2011), 'Yale University and Dura-Europos: From Excavation to Installation', in L.R. Brody and G.L. Hoffman (eds) *Dura-Europos: Crossroads of Antiquity*. Chestnut Hill: McMullen Museum of Art, pp. 17–32.
- Chen, A.H. (2023), 'Collaborative Curation of Digital Archaeological Archives: Promise, Prospects, and Challenges', in R. Raja (ed.) *Shaping Archaeological Archives: Dialogues between Fieldwork, Museum Collections, and Private Archives*. Turnhout: Brepols Publishers, pp. 33–46.
- Cumont, F. (1926), *Fouilles de Doura-Europos (1922–1923)*. Paris: P. Geuthner.
- Davis, E. and Heravi, B. (2021), 'Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation', *Journal on Computing and Cultural Heritage*, 14(2), pp. 1–18. <https://doi.org/10.1145/3429458>.
- Díaz-Andreu, M. (2007), *A World History of Nineteenth-Century Archaeology. Nationalism, Colonialism, and the Past*. New York: Oxford University Press.
- Dirven, L. (1999), *The Palmyrenes of Dura-Europos: A Study of Religious Interaction in Roman Syria*. Leiden: Brill.
- Dyson, S.L. (1998), *Ancient Marbles to American Shores: Classical Archaeology in the United States*. Philadelphia: University of Pennsylvania Press.
- Elsner, J. (2025), 'Rostovtzeff's Dura', in A. H. Chen and L.R. Brody (eds) *Dura-Europos Past, Present, Future*. Turnhout: Brepols Publishers, pp.

22–40.

Estermann, B. (2016), ‘OpenGLAM Benchmark Survey—Measuring the Advancement of Open Data / Open Content in the Heritage Sector’, in *Proceedings of the International Symposium on the Measurement of Digital Cultural Products*, 9–11 May 2016, Montréal, Canada. https://www.bfh.ch/dam/jcr%3A10e986d4-0f54-4565-bbd4-6e21e3120ac6/Estermann_B.Openglam_Benchmark_Survey.pdf.

The Fitzwilliam Museum (2022a), *Linked Open Data, What on Earth Is That?* Available at: <https://www.youtube.com/watch?v=mMR6JQ1M6qE>.

The Fitzwilliam Museum (2022b), *Linked Open Data: How Does It Work?* Available at: <https://www.youtube.com/watch?v=0m79yDb4AzE&t=0s>.

Gelin, M. (2002), *L’archéologie en Syrie et au Liban à l’époque du mandat, 1919–1946: histoire et organisation*. Paris: P. Geuthner.

GO FAIR (2024), ‘FAIR Principles’. Available at: <https://www.go-fair.org/fair-principles/>.

Gregoratti, L. (2016), ‘Dura-Europos: A Greek Town of the Parthian Empire’, in T. Kaizer (ed.) *Religion, Society, and Culture at Dura Europos*. Cambridge: Cambridge University Press, pp. 16–29.

Griswold, S. (2020), ‘Locating Archaeological Expertise: Debating Antiquities Norms in the A Mandates, 1918–1926’, in P. Bourmaud, N. Neveu and C. Verdeil (eds) *Experts et Expertise Dans Les Mandats de La Société Des Nations: Figures, Champs, Outils*. Paris: Presses de l’Inalco, pp. 141–58. <https://books.openedition.org/pressesinalco/38328?lang=en>.

Heřmanková, P., Horster, M. and Prag, J. (2022), ‘Digital Epigraphy in 2022: A Report from the Scoping Survey of the FAIR Epigraphy Project’, *Zenodo*, 3 June. <https://doi.org/10.5281/ZENODO.6610696>.

van Hooland, S. and Verborgh, R. (2014), *Linked Data for Libraries, Archives, and Museums. How to Clean, Link and Publish Your Metadata*. Chicago: Neal-Schuman.

Isidorus of Charax (2014) *Parthian Stations*. Translated by W.H. Schoff. *ToposText*. Available at: <https://topostext.org/work/217>.

Jackson, M., Antonioletti, M., Hume, A., Blanke, T., Bodard, G. and Hedges, M. (2009), ‘Building Bridges between Islands of Data—An Investigation into Distributed Data Management in the Humanities’,

- 2009 Fifth IEEE International Conference on E-Science, pp. 33–9. <https://doi.org/10.1109/e-Science.2009.13>.
- [James, S. \(2004\)](#), *The Arms and Armour and Other Military Equipment. The Excavations at Dura-Europos Conducted by Yale University and the French Academy of Inscriptions and Letters, 1928 to 1937. Final Report VII*. Oxford: Oxbow Books.
- [James, S. \(2019\)](#), *The Roman Military Base at Dura-Europos, Syria: An Archaeological Visualisation*. New York: Oxford University Press.
- [James, S. \(2025\)](#), ‘The Bloody Genesis of Dura-Europos Studies. Archaeology, Colonialism, and Violence at Salhiyeh, 1916–1920 and Beyond’, in A. H. Chen and L.R. Brody (eds) *Dura-Europos Past, Present, Future*. Turnhout: Brepols Publishers, pp. 58–79.
- [Kahn, R. and Simon, R. \(2022\)](#), ‘Feast and Famine: The Problem of Sources for Linked Data Creation’, in T. Andrews, F. Diehr, T. Efer, A. Kuczera and J. van Zundert (eds) *GRAPH 2020. Graph Technologies in the Humanities, Vienna, Austria, 21–22 February 2020*, pp. 86–100. <https://ceur-ws.org/Vol-3110/#paper5>.
- [Kaizer, T. \(2020\)](#), ‘Introduction’, in D. Praet, T. Kaizer and A. Lannoy (eds) *Dura-Europos, Franz Cumont. Bibliotheca Cumontiana—Scripta Minora*, 7. Turnhout: Brepols—Academia Belgica, pp. xi–xcviii.
- [Kamash, Z. \(2021\)](#), ‘Rebalancing Roman Archaeology: From Disciplinary Inertia to Decolonial and Inclusive Action’, *Theoretical Roman Archaeology Journal*, 4(1): 4, pp. 1–41. <https://doi.org/10.16995/traj.4330>.
- [Kamash, Z. \(2023\)](#), ‘Who Can Access the Past? Archives, Technological Solutionism and Digital Colonialism in (Post-)Conflict Syria’, in R. Raja (ed.) *Shaping Archaeological Archives: Dialogues between Fieldwork, Museum Collections, and Private Archives*. Turnhout: Brepols Publishers, pp. 9–18.
- [Leriche, P., Coqueugniot, G. and De Pontbriand, S. \(2011\)](#), ‘New Research by the French-Syrian Archaeological Expedition to Europos-Dura and New Data on the Polytheistic Sanctuaries in Europos-Dura’, in J.Y. Chi and S. Heath (eds) *Edge of Empires: Pagans, Jews, and Christians at Roman Dura-Europos*. New York: Institute for the Study of the Ancient World, New York University, pp. 15–38.

- Loft, P. and Kirk-Wade, E. (2024), 'The Syrian Civil War: Timeline, UK Aid and Statistics', House of Commons Library. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-9381/>.
- Lucas, G. (2001), 'Destruction and the Rhetoric of Excavation', *Norwegian Archaeological Review*, 34(1), pp. 35–46.
- Marlowe, E. (2024), 'The Long, Winding, and Bumpy Road. Seeing Museum Antiquities as Colonial Legacies', in K. Blouin and B. Akrigg (eds) *The Routledge Handbook of Classics, Colonialism, and Postcolonial Theory*. Abingdon: Routledge, pp. 444–65.
- Mayer, A. (2015), 'Linked Open Data for Artistic and Cultural Resources', *Art Documentation: Journal of the Art Libraries Society of North America*, 34(1), pp. 2–14.
- McCarty, M. and Dirven, L. (2020), 'Rethinking the Dura-Europos Mithraeum: Diversification and Stabilization in a Mithraic Community', in M. McCarty and M. Egri (eds) *The Archaeology of Mithraism: New Finds and Approaches to Mithras-Worship*. Leuven: Peeters, pp. 165–82.
- Merrill, S.O.C. (2011), 'Graffiti at Heritage Places: Vandalism as Cultural Significance or Conservation Sacrilege?' *Time and Mind*, 4(1), pp. 59–75.
- Munawar, N.A. (2019), 'Competing Heritage: Curating the Post-Conflict Heritage of Roman Syria', *Bulletin of the Institute of Classical Studies*, 62(1), pp. 142–65.
- Napoli, P.M. (2024), 'Epistemic Rights, Information Inequalities, and Public Policy', in M.A. Horowitz, H. Nieminen, K. Lehtisaari, and A. D'Arma (eds.) *Epistemic Rights in the Era of Digital Disruption*. Cham: Springer International Publishing, pp. 47–62. https://doi.org/10.1007/978-3-031-45976-4_4.
- Nicholson, C., Kansa, S., Gupta, N. and Fernandez, R. (2023), 'Will It Ever Be FAIR? Making Archaeological Data Findable, Accessible, Interoperable, and Reusable', *Advances in Archaeological Practice*, 11(1), pp. 63–75. <https://doi.org/10.1017/aap.2022.40>.
- Nurmikko-Fuller, T. (2023), *Linked Data for Digital Humanities*. New York: Routledge.
- Oomen, J. and Baltussen, L.B. (2012), 'Sharing Cultural Heritage the Linked Open Data Way: Why You Should Sign up', in *Museums and the Web, 11–14 April 2012, San Diego, CA, USA*. <https://www>

- [.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data.html](https://museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data.html).
- Ousterhout, R.G. (2016), *John Henry Haynes: A Photographer and Archaeologist in the Ottoman Empire 1881–1900*. Hawick: Cornucopia Books.
- Pett, D. and Barker, E. (2022), ‘Linking Islands of Data’, The Fitzwilliam Museum. Available at: <https://fitzmuseum.cam.ac.uk/research/projects/linking-islands-of-data>.
- Raja, R. (2023), ‘Shaping Archaeological Archives: Fieldwork, Collections, and Private Archives—Issues of Curation and Accessibility’, in R. Raja (ed.) *Shaping Archaeological Archives: Dialogues between Fieldwork, Museum Collections, and Private Archives*. Turnhout: Brepols Publishers, pp. 1–8.
- Scheffler, T. (2003), ‘“Fertile Crescent”, “Orient”, “Middle East”: The Changing Mental Maps of Southwest Asia’, *European Review of History: Revue Européenne d’histoire*, 10(2), pp. 253–72.
- Starzmann, M.T. (2012), ‘Archaeological Fieldwork in the Middle East: Academic Agendas, Labour Politics and Neo-Colonialism’, in S.J. van der Linde, M.H. van den Dries, N. Schlanger and C.G. Slappendel (eds) *European Archaeology Abroad: Global Settings, Comparative Perspectives*. Leiden: Sidestone Press, pp. 401–414.
- Stevenson, A. (2019), *Scattered Finds: Archaeology, Egyptology and Museums*. London: UCL Press. <http://gen.lib.rus.ec/book/index.php?md5=570b8836c40d9a23f5c2141191085599>.
- Stinson, A. (2017), ‘Wikidata in Collections: Building a Universal Language for Connecting GLAM Catalogs’, *Down the Rabbit Hole*, 13 December. <https://medium.com/freely-sharing-the-sum-of-all-knowledge/wikidata-in-collections-building-a-universal-language-for-connecting-glam-catalogs-59b14aa3214c>.
- Thompson, E.F. (2020), *How the West Stole Democracy from the Arabs: The Syrian Congress of 1920 and the Destruction of Its Historic Liberal-Islamic Alliance*. New York: Atlantic Monthly Press.
- Thornton, K., Seals-Nutt, K. and Chen, A. (2024a), ‘Dura-Europos Stories: Developing Interactive Storytelling Applications Using Knowledge Graphs for Cultural Heritage Exploration’, *Semantic Web*, 15(6), pp. 2237–2251. <https://doi.org/10.3233/SW-243552>.

- Thornton, K., Seals-Nutt, K. and Chen, A. (2024b), 'Encoding Archaeological Data Models as Wikidata Schemas: Utilizing Shape Expressions to Structure Collaborative Linked Open Data for Digital Storytelling within the International Dura-Europos Archive,' *The International Journal of Technology, Knowledge, and Society*, 21(1), pp. 69–83. <https://doi.org/10.18848/1832-3669/CGP/v21i01/69-83>.
- Trentin, M.G. (2021), 'Form, Content, and Space: Methodological Challenges in the Study of Medieval and Early Modern European Graffiti', *Papers from the Institute of Archaeology*, 31(1). <https://doi.org/10.14324/111.444.2041-9015.1283>.

Chapter 10

Reception of Sapphic language in Hadrianic verse epigraphy: building on open philological data

Marja Vierros and Gabriel Bodard

Introduction

The ‘corpus’ examined in this chapter is a tiny quantity of text: four poems by Julia Balbilla, constituting a total of fifty-three lines of verse, inscribed in Ancient Greek in the Roman period on a Dynastic Egyptian stone monument. We apply a computational linguistic approach to this small corpus, and this allows us to study both the language use and the reception and imitation of Sapphic poetry in particular, and other Classical Greek literature in general, by utilising open data produced earlier with similar linguistic annotation. We will conclude with some observations on Balbilla’s imitation of Sapphic Greek being more convincing morphologically than syntactically, as well as on the validity of quantitative approaches on the small dataset, and the value of building on earlier, open licensed work.

The Colossi of Memnon

The two colossal, monolithic statues of Pharaoh Amenhotep III, covered in hieroglyphic inscriptions and reliefs, have stood at the gates to the mortuary temple in Egyptian Thebes (modern Luxor) since about 1350 B.C.E. ([Rosenmeyer 2018](#): 1–9). They are now almost the only part of that gargantuan complex still standing, although parts of the temple, including many other statues, have been excavated in modern times. The statues have been visible and famous for well over 3000 years, including the Roman and early modern periods, when they were much visited and written about. Much of the rest of the mortuary temple was destroyed by a major earthquake some time between 1200 and 900 B.C.E.; apparently a later—although otherwise unknown—quake of c. 27 B.C.E. toppled the upper half of the northernmost statue ([Karakhanyan et al. 2010](#)). According to Roman sources, the first of which being Strabo in the early first century C.E., the broken statue would let out a notable sound around dawn.¹ This may have been due to air or water vapour trapped in the damaged stone warming at the start of the day, but since the Romans now identified these statues as the fallen hero of the Trojan War, Memnon, son of the dawn goddess Eos, the sound was also associated with the mourning of Dawn.

The statues, the ‘Colossi of Memnon’, were much visited by Roman tourists and other travellers in Egypt, and in the first couple of centuries C.E. there was a tradition of adding short secondary inscriptions or graffiti, especially to the legs of the broken Memnon, dozens of which survive. When the Emperor Hadrian and Empress Sabina visited the area in 130 C.E. accompanied by the highly educated noblewoman Julia Balbilla, the latter wrote four short poems in an archaising Aeolic Greek dialect commemorating their interactions with Memnon, and had them inscribed on the statue ([Cirio 2011](#): 45–28, [Rosenmeyer 2018](#): 141–5). The first text describes Hadrian’s first visit to Memnon, at which the statue ‘cried out’ not once but three times out of respect for the emperor. The second poem recounts some of the story of Memnon, in mythology and the history of the statues, and also gives Balbilla’s own royal genealogy. The third uses the occasion of the statue not giving voice as an excuse to praise the Empress Sabina’s beauty in effusive, almost erotic terms. The fourth also tells of a visit to the statue by Balbilla and Sabina, and gives the exact date of the inscription (25th of Hathyr = 21 November 130).²

The Colossi of Memnon have always been known to European visitors, and to antiquarian and classical scholarship. The first coherent epigraphic treatment of the texts was that of Letronne, although he points out that Friedrich Jacobs half a century earlier included the first poem in his *Greek Anthology* ([Letronne 1848](#): 351–68).³ Autopsy editions followed with Peek and the Bernand brothers, whose text was basically followed by Cirio and

Rosenmeyer ([Peek 1934](#): 95; SEG 1937: 124; [Bernand and Bernand 1960](#): 80–97; [Cirio 2011](#): 48–52; [Rosenmeyer 2018](#): 145–8).⁴

Commentary on the inscriptions has not dwelt on the literary content of the poems, whose narrative is pretty straightforward, but rather on the choice of archaising dialect, vocabulary, morphology and metre—variously Homeric, Sapphic and generically ‘archaic’ in influence.⁵ Many modern scholars have engaged in speculation over whether we should read any homoerotic intent in the choice to evoke Sapphic verse in these texts, especially given their mention of Sabina’s beauty. Bowie went so far as to suggest the texts implied Balbilla may have been ‘the empress’s companion, perhaps her answer to Hadrian’s Antinous’ ([Bowie 1990](#): 62); Rosenmeyer dismissed the idea that the Sapphic echoes suggest Balbilla had, or even meant playfully to imply, an erotic relationship to the Empress Sabina ([Rosenmeyer 2018](#): 159–62); McArdle’s more nuanced argument shows that the specific echoes of Sappho’s language in the poem are better understood through a ‘queer reading’, while avoiding ‘clumsy claims of an actual lesbian relationship between poetess and empress’ ([McArdle 2020](#)).⁶

Digitising Balbilla and open data

Digital versions of the text of Balbilla’s poems have been available for some time through the TLG (based on the non-epigraphic text of Kaibel) and the PHI (based on Bernand’s editions).⁷ A collaboratively designed project to encode and annotate the four inscriptions was begun in 2019, intended to demonstrate a range of digital and epigraphic approaches to classical texts, and discussed with cohorts of masters students over the next few years.⁸ The texts were marked up in TEI EpiDoc XML, including indication of personal names and places, all words lemmatised (i.e. dictionary headwords added by hand, but also by cross-reference to the morphological parsing in the treebank, discussed below), both epigraphic and verse lineation and encoding of metre; Greek texts and French and English translations were aligned using the Ugarit text and translation alignment interface; morphology and syntax were collectively annotated (‘treebanked’) using the Arethusa annotation framework from Perseids.⁹ Descriptive and historical data for the epigraphic editions were enhanced with metadata and significant entity URIs from Linked Open Data resources such as the Pleiades Gazetteer, Wikidata, SNAP and other prosopographical databases, and the EAGLE Vocabularies.¹⁰

For the purposes of this paper, the poems of Balbilla were also treebanked anew in the PapyGreek portal, which is especially designed for syntactic annotation of papyri and inscriptions (providing the possibility to annotate the original and standardised layers separately), and includes a search interface.¹¹ The poems of Sappho have been treebanked within the Pedalion project.¹² Both authors’ treebanked literary production was then ingested into the *Kiln for Treebanking* (KTB) query tool within the corpus that also includes other Greek poetry, in order to be able to query them in tandem and also in comparison to other poetic texts.¹³

Treebanking is a powerful method for morphosyntactic annotation and lemmatisation for ancient texts ([Bamman et al. 2009](#); [Bamman and Crane 2011](#)). There are several projects that have applied it to different types of texts and there are also several steps whereby the interpretation of the annotator can result in a different outcome even with similar structures and forms (see, for example, [Mambrini 2016](#)). The guidelines used in Ancient Greek and Latin Dependency Treebanks exist in a couple of versions that differ in minor points, and the different projects using them may have made their own internal decisions to resolve some questions, where the guidelines offer no clear instruction, or diverge from guidelines for strategic, linguistic or other reasons.¹⁴

Despite detailed guidelines, different human annotators can end up with different results at several levels in the annotation process. We may exemplify this to some extent with one sentence from Balbilla’s first poem (= Letronne 343 (G.22)/Bernand 28/Cirio 1 (p. 76)), ll. 4–5, sentence 3:

Μέμνονα πυνθανόμαν Αἰγύπτιον, ἀλίω αὔγαι αἰθόμενον, φώνην Θηβαῖ<κ>ω ’πυ λίθω.

I’ve heard tell that Memnon the Egyptian, warmed by the sun’s rays, utters a loud sound from the Theban stone. (transl. Rosenmeyer)

I had learned that the Egyptian Memnon, when ignited by the rays of the sun, released a voice from the Theban stone. (transl. Bodard)

The sentence has several archaic dialectal features Balbilla has used to give the poem an Aeolic flavour: the masculine singular genitive ending in -ω instead of Attic -ου, unaugmented imperfect form πυνθανόμαν with the ending in -αν rather than -ην, present infinitive form ending in -ην rather than -εῖν, the preposition ἀπό ending in grapheme <υ>, the long vowel <α> instead of <η> in ἡλιος (cf. [Rosenmeyer 2018](#): 155).¹⁵ Most of these forms had to be manually inserted within the Arethusa treebanking editor, since the morphological form recognition tool is

not familiar with all dialectal forms (also the accentuation differences caused the tool not to recognise otherwise familiar forms).

The two authors of this contribution have some differences in their syntactic trees (see [Figure 10.1](#)). The first difference comes from the understanding of φώνην as a noun by Bodard et al. and as an infinitive by Vierros; this means that in the tree, Vierros placed the infinitive as the object of πυνθανόμαν, while Bodard, who took it as a noun, created an artificial token as the object. For the infinitive, the accusative subject is then Memnon. One of the most common problems with which an annotator is faced meeting a participle is whether to interpret it as attributive or adverbial. Here we see the two options taken: Vierros considered the participle αἰθόμενον to be attributive and placed it under Memnon with the syntactic label ATR, whereas Bodard took it as adverbial and placed it under his artificial token (OBJ) with the label ADV. Both interpretations are viable: as ATR the participle is thought to be a feature more closely related to the noun it defines as a property of the noun, whereas the ADV option considers the action indicated by the participle to signify the circumstances of the verb it is attached to; it is an optional modification of the verbal action (called also a circumstantial participle). Vierros's interpretation considers that the kindling or lighting up of Memnon[’s statue] is a feature of this statue alone when hit by the rays of the sun, while the adverbial option takes the warming up as a more common phenomenon that affects the action of uttering a sound (assumed meaning of the artificial token, marked as [1] in the tree, [Figure 10.1](#)).

Table 10.1: Balbilla 1, sentence 3 with lemmata and morphological annotations

Μέμνονα	πυνθανόμαν	Αἰγύπτιον	ἄλιω	αὐγαι
Μέμνων- Memnon:n.masc.sg.acc	πυνθάνομαι- learn:v.mp.ind.1sg.impf	Αἰγύπτιος- Egyptian:adj.masc.sg.acc	ἥλιος- sun:n.masc.sg.gen	αὐγή- ray:n.fem.sg.dat.
αἰθόμενον	φώνην	Θηβαῖ<κ>ω	’πυ	λίθω
αἶθω-kindle:v.mp. pr.ptcpl.masc.sg.acc	φωνέω-utter a sound:v.pr.inf.	Θηβαϊκός-Theban: adj.masc.sg.gen	ἀπό-from:adp	λίθος- stone:n.masc.sg.gen

It can be noted, then, that both syntactic and morphological interpretations may differ, but for the most part the annotations are usually similar, when following the guidelines. In the case of inter-annotator disagreement, the differences should be discussed and a possible solution reached; however, if differing annotations are all well argued and not against the guidelines, one can be considered primary and the other stored separately for possible later use. In this study we used Vierros's interpretations, which follow the guidelines of the PapyGreek project (see [Table 10.1](#)).

The decisions made by different projects on how to treat, for instance, some particles (always as adverbs or sometimes also as conjunctions)¹⁶ can have an effect on our statistics when we start counting attestations of, for example, conjunctions in a bigger corpus. When one is aware of some of the likely differences, the queries can be formulated in such a way that the dangers of mismatch are avoided.

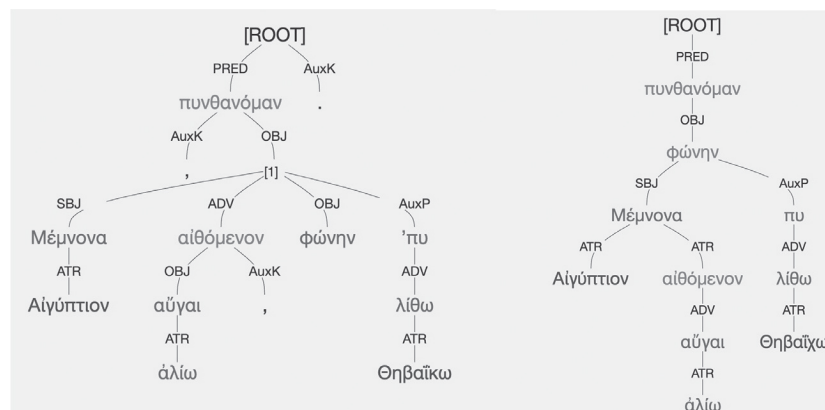


Figure 10.1: The trees of Balbilla 1, sentence 3 by Bodard et al. (left: <https://www.perseids.org/tools/arethusa/app/#/perseids?chunk=3&doc=61737>) and Vierros (right: <https://papygreek.com/text/466666/arethusa>) ([Description](#))

The potential of treebanked texts for text analysis is huge, and can already be seen from the one short sentence treated above. We can query features including predicates (and forms of verbs are used as predicates), infinitival and participial complements (here, for example, the accusative-and-infinitive construction), or the ways in which attributes are used in the noun phrases or prepositional phrases (also considering the word order) or which pre- or postpositions or pronouns are used and how.

Research questions

The computational linguistic methods discussed in this chapter are generally applied to large or huge bodies of text, on the order of at least thousands, if not many millions, of words. The goal is not only to find patterns in the algorithmic treatment of data, but to do so at a scale that would be unfeasible for traditional human linguistic analysis, and such that results may be considered statistically significant across language as a whole. Clearly the statistical results of a computational analysis of fifty-three lines of verse, a total of 335 words (tokens) of Greek, will not have the same scope for extrapolation.

We believe there is value to this project, however. To start with, an advantage of working with such a small corpus is that we can control the quality of the digital data (as discussed above with respect to comparing and tidying pre-existing treebanks), and look in detail at individual results rather than rely on sampling; for example, the three infinitives in Balbilla's poems, as mentioned in our discussion below, constitute 6 per cent of her verb forms and this percentage *can* be compared to other poetry—and it is necessary to pinpoint the small percentage—but it is also very easy to analyse each individual infinitive and their function in Balbilla's language. Even with these limited results, we occasionally see interesting patterns in the linguistic analysis of the two authors, which could valuably be tested and further explored in a larger project with more ancient authors and later imitators. We should also not be surprised to note that the very acts of preparing, correcting, annotating and analysing these bodies of text are acts of close reading, throwing up new questions and ideas for digital comparison, as well as familiarising the researchers with the language and register of the texts in question.

This chapter therefore will address four main research questions, each drawing to different degrees on the core issue of building on reusable open data centred by this volume.

1. Firstly, the overt topic of the treebank research itself: comparing linguistic patterns in the Balbilla poems that might be relevant to the question of imitation or evocation of Sappho's poetry or ancient Greek literature more broadly. This approach goes beyond the discussion of lexicon and dialect characterised for instance by García Teijeiro and Rosenmeyer,¹⁷ to address morphology and syntax. We consider to what degree (conscious or otherwise) imitation of syntax and style played a role in Balbilla's echoes, influence and intertext across such long timespans, and explore the difference between simple frequency of use of grammatical cases, tenses and voices, and structural syntax, characterised by the relationships between such parts of speech.
2. What features of reception and conspicuous intertextuality are evidenced in the morpho-syntactic features of target and source texts? Are any archaic or classical influences visible against the morpho-syntax of the Imperial Greek context of the authorship of the Balbilla poems? We offer some preliminary thoughts on whether treebanks in general are a promising medium for the analysis of influence and imitation in antiquity.
3. We will also draw some conclusions about the utility of computational linguistic methods when applied to smaller texts and corpora. Are the predicted benefits such as control of quality and manual checking of results valid in the face of the loss of scale and statistical analysis? Are there any advantages to this method over manually counting grammatical forms as a nineteenth-century linguist might have done?
4. Finally, and especially pertinent to the theme of open data, we have been forced to think about how to communicate the results of linguistic research carried out with software such as KTB Lite in a replicable manner and in accordance with the principles of open science. Fundamentally this means that not only the written outcomes of and reasoning behind our research be published in a Findable, Accessible, Interoperable and Reproducible form, but also the data, code and indeed paradata—the parameters and settings that led to our specific results (Chue Hong et al. 2022). KTB Lite is open source and includes all the parsed texts needed by this project; a reader installing the tagged version of the tool may follow the links to search results in this chapter and find our exact results. We have also extracted the key results into a data table that is deposited alongside this publication, for ease of access by linguists who may not have the capacity or the desire to install and execute open source software on their local system.

Observations

In this section we will look at what morpho-syntactic differences we can discern from treebanked data of Balbilla's and Sappho's texts. Due to the small token number of Balbilla's corpus the results may occasionally merely be due to chance, but it still presents a closed corpus of 'language of Julia Balbilla in the Colossi of Memnon' and either reveals or not those kinds of linguistic traits that have led researchers to consider Balbilla is imitating Sappho, and from this angle we are trying to lure more out of the texts with new methods.

Verbs

Starting with verbal morphology,¹⁸ we can see that the favourite tense of the poets is different: Balbilla uses the aorist the most (in 28 out of 52 of her verbs, that is ca. 54%), whereas in Sappho the present tense prevails (ca. 45%, that is 191 of her total of 420 verbs are in present tense and ca. 32%, that is 134 instances, are in the aorist). Sappho's range of tenses is also wider than Balbilla's, who only has aorist, present and imperfect forms; Sappho's texts also include verbs in perfect and future tenses, and even one in pluperfect. The verb's mood also varies more in Sappho's poems, although both authors have over 50 per cent of indicative forms. Sappho uses slightly more other moods, like the imperative, subjunctive and optative, than Balbilla. In fact, the only instance of the optative in Balbilla is restored: μό[λοι] in poem 3.¹⁹ The only use of subjunctive by Balbilla is in the same poem, but that word is actually readable on the stone; the form appears in a negative final clause: μή καί τοι βασίλευς κοτέσῃ, 'so the king won't be angry with you' (tr. Rosenmeyer).

A potentially exciting difference can be found within the non-finite forms, but their number is small, even minuscule when discussing the infinitives; Sappho uses more infinitives than Balbilla, and Balbilla has a larger percentage of participles than Sappho (Balbilla's 14 participles form 27% of her verbs vs 19% (78) in Sappho; Balbilla has only three infinitives, which is 6% of her 52 verbs vs 51 infinitives in Sappho, which makes 12% of her verbs). The diminishing use of infinitives in Greek was a well-known phenomenon ([Bentein 2018](#); [Horrocks 2010](#): 94 and *passim*; [Joseph 1983](#)). Without delving into the complexities of this issue more deeply, Balbilla's language still seems to follow the general line of development in postclassical Greek (note that the tense and aspect of the infinitive also play a role; cf. for example [Kavčič 2016](#)).

It seems that verbal morphology does not confirm the idea of Balbilla imitating Sappho very profoundly. For example, the use of present tense in Sappho can be interpreted as an indication of experiencing feelings caused by events or people around the writer, whereas Balbilla is using her words much more for the purpose of narrating events in the past tense around the Memnon statue, the history of the people involved and even her own genealogy. Sappho, on the other hand, uses the (historical) present tense also for narrating events.

Personal pronouns

Sappho's texts include clearly more of the first person pronoun ἐγώ 'I' than Balbilla's (5 instances of 27 pronouns in Balbilla (18.5%), 67 out of 244 in Sappho (27.5%)). Also, the use of the second person singular σύ 'you' is more popular in Sappho (ca. 15% in Balbilla (4/27) and ca. 20% in Sappho (49/244)). The overt mentioning of a personal pronoun is not of course necessary in Greek, and a different view point is gained by looking at the finite verbs' person marking; the third person usage is much more common in the verb forms than the first or second both in Balbilla's and Sappho's poems; slightly more so in Balbilla's (ca. 63% of finite verbs are in third person, singular + plural, in Balbilla and 52% in Sappho). Therefore, the emphasis of first person by way of pronouns is even more striking in Sappho, and a sign that Balbilla's possible imitation of her did not really reach this aspect.

A clear difference is seen in which third person pronoun was selected: Sappho's pronouns include, for example, οὗτος and ἐκεῖνος, but Balbilla uses most often the pronoun αὐτός (4 occurrences in total from 27 pronouns, lemmatised three times as αὐτός and for some reason once as καὶαὐτός). This pronoun does appear in Sappho, too, but much less so (only 6 instances of the lemma αὐτός from 244 pronouns, and it is used mostly as 'self' when referring to a person, and a couple of times it is a neuter pronoun for non-animate referents). In Balbilla's poems the pronoun αὐτός is always used for persons; but in the nominative only once, where the meaning 'self' can be understood (poem 1, sent. 7: κοίρανος Ἀδριανὸς[ς] ... δ' ἀσπάζεσθαι κ' αὐτὸς Μέμνονα, 'and Emperor Hadrian himself gave many greetings in turn to Memnon', transl. Bodard). Thus, we have no clear sign of postclassical use of αὐτός as a 'common' third person pronoun in the nominative in Balbilla. However, the use of αὐτός comes close to a demonstrative in poem 2, sent. 2: κ[αὶ] αὐτ[αν] τὰν σέμναν ἄλοχον, 'the noble spouse' (see [Horrocks 2010](#): 128; 148–9 on this type of postclassical usage of αὐτός).

Grammatical case

A notable feature in Balbilla's poems is the high number of genitives used within everything where case-marking is applied. Her percentage of genitives is ca. 23 per cent (47/205), while in Sappho the number is 12 per cent (162/1,319). When compared to other literary texts, the prose corpus has a similar amount (ca. 24%) of genitive usage as Balbilla, while Sappho is closer to the poetry corpus (ca. 16%). Sappho uses the nominative and accusative clearly more than the genitive and dative (nominative ca. 35%, accusative 37%, dative 12%); Balbilla's nominative and accusative usage are only somewhat more frequent than the genitive, and dative has the lowest figure of these four cases (nominative ca. 31%, accusative 24%, dative 19%). Vocative appears in both, but unsurprisingly only a few times. The genitives in Balbilla are in higher proportion attributive (34 ATR: 7 ADV: 2 OBJ: 3 SBJ) than in Sappho (85 ATR: 33 ADV: 21 OBJ: 3 SBJ); the ratio of genitive objects is remarkable in Sappho and could be worth further study. The style of Balbilla then comes closer to prose texts in this respect; her attributive genitives are at least partly caused by her descriptions of herself and the Emperor Hadrian and Empress Sabina.

Sentence complexity (visuality of trees and subordination)

When browsing through the sentences in their tree visualisation form, one gets the impression that Sappho's sentences are often longer and more complex, both in subordinated structures as well as in the amount of coordinated predicates, than what Balbilla's small corpus offers. Naturally, many of Sappho's sentences are fragmentary resulting either in flat trees, where syntactic connections between words are missing, or even in word groups where we do not really know if they belonged to the same sentence or not.

Visual effect can be confirmed by some numbers. The maximum token count in Balbilla's sentences is thirty-four, while Sappho's longest sentence contains eighty-one tokens (sentence 44, the famous φαίνεται μοι κῆνος ἴσος θεοῖσιν ἔμμεν' ὄννηρ ... ['He appears godlike to me, that man']). Sappho also has more subordinate structures in some of her sentences, the outlier including seventeen (the same longest sentence just mentioned), but otherwise ranging from zero to seven, while the maximum number of subordinated structures per one sentence in Balbilla is four (in one sentence: poem 1, sent. 7). Balbilla reaches two levels of subordination depth, while Sappho has one more level.

Conclusions

Balbilla's imitation of Sappho

As can be seen from Rosenmeyer's detailed analysis, and from comparing the full linguistic analysis in treebank form of morphology and lemmas of the poems in Julia Balbilla's name at the leg of the colossus, Balbilla has successfully imitated the Aeolic and archaic morphology and used many such lexemes. Our morpho-syntactic annotations have helped to compare the frequency of syntactic structures as well as grammatical forms to similar annotations of the Sapphic corpus. The comparison showed that Balbilla's syntax is clearly less affected by Sapphic language than the forms and choice of words. Balbilla's language follows more its own time, and prose literature, in the choice of verbal tense and non-finite verbs (participles pro infinitives) as well as in favouring the genitive case and certain pronouns. Balbilla's sentences are also shorter and less complex. This relatively short and simple analysis points to the direction that Balbilla's interest was not to write poetry matching Sappho in a profound way, but rather to show that she was aware of the dialectal and epic words and their special forms. Her poems mirror Roman-period Greek usage in their way of formulating sentences. The picture could be more nuanced with deeper analysis. The fact that Balbilla did not use the same metre as the Aeolic monodic lyrics, but the elegiac, suggests that she addressed the epigraphic audience more familiar with that metre in epigrams. She is somewhere in between the worlds of epic and Aeolic lyric and her own day's honorary texts as well as recording historical events.

Methodological assessment

We are conscious, of course, that the very small scale of Balbilla's (surviving) poetic corpus means one could have performed a thorough linguistic analysis of Balbilla, at least, in an entirely traditional, manual manner based on the printed edition. Digital annotation and the application of computational linguistic tools and statistical analysis, our internal critic argues, may be overkill for the study of fifty-three lines of verse. Aside from the fact that the very

suitability of quantitative methods to such a small body of text is one of our research questions, our work is not an entirely arbitrary application of technology to this text.

For one thing, morpho-syntactically annotating the entire corpus of poems and fragments of Sappho in order to compare her use of language to Balbilla's would have been a significant individual effort. The fact that previous projects, combining automated and manual processes, have produced openly licensed texts and treebank data, has allowed us to quantitatively compare the texts of Sappho, and to some degree also the wider corpus of Classical Greek literature, to the patterns we find in our smaller corpus. Working with treebanked texts enabled the identification and counting of certain linguistic features, especially morpho-syntactic relations, in a more consistent and objective way. Most importantly, this meant we were treating the larger, comparative corpus and our smaller, subject text in the same way, with the same annotations and technologies, and therefore comparing them in consistent and reproducible ways, that we can share in turn as open data (on which more below).

Future work

If one were to push this project further, with a view to assessing more comprehensively the quality of Balbilla's pastiche of Sapphic poetry, some larger-scale analysis—requiring more time-consuming text-preparation—would be called for. The most obvious starting point would be to conduct a thorough comparison of the syntax of Archaic and Classical Greek literature with texts composed in the early Imperial period, more or less contemporary with the second-century C.E. Balbilla. Ideally this huge task would include a general study of word use and other morpho-syntactic features across the two eras, quantifying some core aspects of differences between Classical and Second Sophistic Greek ([Bozia 2018](#); [Tribulato 2022](#)). In particular, however, we should focus on the stylistic and grammatical habits of poetry, especially lyric poetry and epigraphic verse across the two eras, enabling us to more securely place Balbilla's small poetic corpus on a spectrum between the two. One obvious feature to begin with would be to look at sentence length and sentence complexity in different time periods and text types (also taking the writing surface into account), since there might exist presuppositions along the lines of more simple archaic language versus more complex Hellenistic or Roman poetry, which might not hold true.

A more comprehensive study would also compare to Balbilla, Sappho and Greek poetry as a whole the work of Damo, a four-line verse also inscribed on the statue of Memnon in Aeolic dialect perhaps slightly earlier than, or contemporary with, Balbilla ([Rosenmeyer 2018](#): 108 and 164–6).²⁰ Although also very small in scale, inclusion of this text will help to contextualise assumptions made about the uniqueness of the language and style of Balbilla.

The current project was not able, due to the status of the data and tool at the time of writing, and the time available on our part for learning to use it effectively, to analyse and compare elements of the Ancient Greek word order of the two authors. This would be a promising method for looking at both style and especially language variation over time ([Manolessou 2000](#); [Viti 2008](#); [Celano 2014](#); [Guardiano and Stavrou 2019](#)), with some enhancement and processing of the digital texts in KTB Lite. Of course, the bulk of very fragmentary texts in the Sapphic corpus would be unsuitable for word-order analysis, but even focusing on the longer, complete or near-complete poems would offer a useful contrast and comparison to Balbilla. Word order tends to be more flexible in verse (by necessity), so it would be useful to analyse whether Balbilla follows Sappho in this trend, or is more consistent with epigraphic and Second Sophistic prose. While probably a small part of a more thorough future study of Balbilla's imitation of Sappho, taking account of word order would certainly be low-hanging fruit.²¹

The most important control data for an assessment of Balbilla's Sapphic imitation would be to contrast it with linguistic features of the larger Greek epigraphic verse corpus, including that of the Roman period. This endeavour would require more intense data preparation, as a coherent database of Greek metrical inscriptions does not exist, much less with manual morphological and syntactic analysis. There has, however, been a flowering of interest in Greek and Latin verse epigraphy in the last couple of decades,²² so a project to create and annotate such a dataset would be pushing at an open door, and quite feasible as a short-to-medium term objective. Identifying the strongest points of linguistic divergence between epigraphic verse of the Roman East, and the literary poetic genre of which Sappho is a part, might suggest some features to look for in the texts of Balbilla and other explicit poetic imitators.

Acknowledgements

The project described in this chapter has benefited at all stages from open texts, data and tools, licensed for reuse and well-documented in all cases. It is therefore fundamental that we in turn make our texts, data and documentation equally available and open. The GitHub repositories for the Balbilla project include all of the

EpiDoc files and treebanked XML discussed above.²³ To this we have added CSV exports of the tabular data from our morphosyntactic queries, and our fork of KTB Lite containing the treebanked Balbilla texts and Pedalion's Sappho, along with Yordanova's 'Greek literature' sub-corpus. These are all of course works in progress, in particular those annotated and aligned texts that are used for collaborative practice with students, and so we have archived a snapshot of the tabular analyses, aligned to this publication, in the Zenodo open research repository under a CC BY 4.0 licence.²⁴

The authors would like to thank the developers of the Kiln Treebank (KTB) tool, Polina Yordanova and the late Jamie Norrish, for infrastructural work, without which this project would have been impossible. We also acknowledge the generosity of Alek Keersmaeckers and Yoran Joosten of the Pedalion Project, on whose linguistically annotated text of Sappho we relied. We are very grateful to our editor, Valeria Vitale, to the keen reviewers, Francesco Mambrini, Mar A. Rodda and Polina Yordanova (again!), and to the faculty and students of SunoikisisDC and the Digital Classics MA module in London in 2021, 2023 and 2025. The PapyGreek project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 758481), with invaluable development work by Erik Henriksson, and support from the Perseids team with the incorporation of the open source Arethusa annotation tool.

Notes

1. Strabo 17.1.46 describes the sound of the singing statue as ὡς πληγῆς οὐ μεγάλης ('like a small strike'); cf. Pausanias 1.42.3: κιθάρας ἢ λύρας ῥαγεῖσιν χαριδῆς ('a kithara or lyre string breaking').
2. Balbilla 4, 7–8: εἰκόστω πέμπτω | δ' ἄματι μήνος Ἄθω. Date conversion via <https://godot.date/>.
3. Balbilla's inscriptions are his 343 (G.22), 344 (G.23), 345 (G.25) and 347 (G.24).
4. SEG: *Supplementum Epigraphicum Graecum*, vol 8 (1937), nos. 715–718 (pp. 124–26). See also <https://scholarlyeditions.brill.com/sego/>.
5. For example, Carandini (1969: 83), 'l'erudita poetessa usa forme poetiche archaiche e forme lessicali coliche'; García Teijeiro (1984: 99–102), who observes that while the poems evoke explicitly Aeolic language, the strict metrical scheme of Sapphic verse is not observed; detailed analysis in Rosenmeyer (2018: 155–9).
6. Brennan (1998: 221–2) was quietly sceptical; Auanger (2002: 216–22) discusses the nuanced Roman receptions of Sappho as homosexual or homoromantic, suggesting that such imitations and references certainly could have been intended to imply love between women, although this is not always explicit in literary sources; Yatromanolakis (2007: 254–8), who takes an anthropological approach to the ancient reception of Sappho, shows that ἐράομαι and ἐρατός may well, but do not necessarily, have for an ancient reader the connotation of erotic attraction.
7. *Thesaurus Linguae Graecae* author 1213; according to TLG Canon [direct link not possible], work 001 based on Kaibel (1878: 414–17, fr. 988–92). Packard Humanities Institute, *Searchable Greek Inscriptions*, text <https://epigraphy.packhum.org/text/227949> (and three following numbers) based on Bernard and Bernard (1960 nos. 28–31).
8. Preliminary versions of parts of this project may still be found on GitHub at <https://github.com/gabrielbodard/Balbilla>.
9. EpiDoc: Epigraphic Documents in TEI XML, <https://epidoc.stoa.org/>; Tariq Yousef's Ugarit, <https://ugarit.ialigner.com/>; Perseids and Arethusa, <https://www.perseids.org/tools/arethusa/app/#/>, with thanks to the Greek and Latin Treebanking community and especially to Polina Yordanova for support with interpreting and following the rules of the Ancient Greek Dependency Treebank (Celano 2014).
10. For example, Pleiades, Colossi of Memnon: <https://pleiades.stoa.org/places/786066>; Wikidata, Colossi of Memnon: <https://www.wikidata.org/wiki/Q750209>; Wikidata, Julia Balbilla: <https://www.wikidata.org/wiki/Q441786>; PIR² I 650 (Iulia Balbilla): <https://pir.bbaw.de/id/7153>; EAGLE Vocabulary, statue: <https://www.eagle-network.eu/voc/objtyp/iod/241>.
11. The PapyGreek portal, <https://papygreek.com/>, uses Arethusa for treebanking (n. 9). For the corpus and its annotation style, see Vierros and Henriksson (2021). For the search interface, see Henriksson and Vierros (2025).
12. For the Pedalion project, see Keersmaeckers et al. (2019). The main annotator of Sappho's poems was Yoran Joosten (browsable treebanks available at: <https://perseids-publications.github.io/pedalion-trees/sappho/1>). The source text is that of Lobel-Page from https://www.greek-language.gr/digitalResources/ancient_greek/anthology/poetry/browse.html?text_id=277 (Keersmaeckers pers. comm.).
13. KTB Lite was developed by Polina Yordanova and Jamie Norrish for Yordanova's forthcoming PhD thesis. It has already been used, for example, in syntactic queries presented by Vierros and Yordanova (2022) and Yordanova-Nanev (2026). The official release of the full KTB is forthcoming; KTB Lite is available at <https://gitlab.com/polinayordanova/ktb-lite>. A snapshot of the version used in this chapter is available for replication and testing of our results (n. 18 below).
14. Vanessa Gorman has documented her annotation decisions in 'Tips for treebanking', <https://github.com/vgorman1/Greek-Dependency-Trees/blob/master/Treebanking%20Tips%20for%20Dependency%20Syntax.pdf> (2021); PapyGreek has its usage described in these Guidelines (Google Doc of the Sematia Guidelines 2026 snapshot – <https://preview.tinyurl.com/PapyGreekGL>); Pedalion project only mentions that they follow the guidelines of Perseus Dependency Treebanks, without clarifying the version; see <https://github.com/perseids-publications/pedalion-trees> and <https://www.relicta.org/pedalion/glaux/>. One major difference is that Gorman's trees, <https://vgorman1.github.io/>, are annotated using the earlier version 1.1 (Bamman and Crane 2008) and PapyGreek and (most likely) Pedalion using the version 2.0. (Celano 2014) of the AGDT guidelines. Some of the most notable differences concern the use of the label ATV (verbal attribute), which is replaced by the more common ADV (adverbial) in the 2.0 version (see 3.6. in Celano 2014), and the part-of-speech tag for the discourse connector δέ when it is used as a coordinator (labelled as conjunction then, but as adverb when not).
15. Some dialectal features, like the lack of initial aspiration and the placement of the accent, are not visible in the inscriptions themselves, naturally.
16. For example, the second position δέ is taken as a conjunction and placed above the predicate as a coordinator (relation label COORD) in Guidelines 2.0 and PapyGreek, while in the Pedalion trees it depends on the predicate with the relation label AuxY and is considered an adverb. This difference is programmatically corrected in KTB, so the trees there are comparable in this respect. Some other conjunctions, like ὡς, can result in different

interpretations as well (either as a subordinating conjunction, AuxC, or as an adverbial, ADV). A somewhat more significant problem is the part-of-speech adjective that can include quantifiers, numerals and even pronouns; see the discussion in [Vierros and Yordanova, forthcoming](#).

17. n. 5 above.

18. If a reader chooses to download and install the Balbilla fork of KTB Lite, following the instructions at <https://www.github.com/Balbilla/KTB Lite>, they will be able to view our search results in the browser using the following links and modifying the search parameters; one can start, for example, with Balbilla's verbs: http://127.0.0.1:9999/search/word.html?text_corpus=literature_verse&text_author=Balbilla&word_pos=Verb&start=0 and compare them to Sappho's verbs: http://127.0.0.1:9999/search/word.html?text_corpus=literature_verse&text_author=Sappho&word_pos=Verb&start=0.

19. Optative has generally been said to have very little use in postclassical Greek ([Horrocks 2010](#): 102), but see [La Roi \(2025\)](#) on a more nuanced picture and thorough bibliography.

20. Rosenmeyer proposes Claudia Damo, another member of Hadrian's circle, accompanied them to Egypt in 130; cf. [Cirio \(2011\)](#): 119–26, who called Damo 'del periodo dei Flavi'; Bernard and Bernard (1960: 179); [Peek \(1934\)](#): 105, no. 9).

21. KTB Lite is especially well equipped for the study of word order, including the word order within the noun phrase, position of the modifiers and the basic sentence word order (subject-object-verb), where the focus can be on any of the sentence constituents.

22. Greek Verse Inscriptions of Cyrenaica: <https://igcyr2.unibo.it/>; Musisque Deoque: A digital archive of Latin poetry from its origins to the Italian Renaissance: <https://www.mqdg.it/>; MAPPOLA: Mapping out the Poetic Landscape(s) of the Roman Empire: <https://db.mappola.eu/>; CARMEN: Communal Art—Reconceptualising Metrical Epigraphy Network: <https://carmen-itn.eu/>. Linguistic annotation of epigraphic texts generally has been excellently prototyped in WoPoss: A corpus to analyse the evolution of modality in the diachrony of the Latin language: <https://woposs.unine.ch/>.

23. 'Balbilla' GitHub organisation: <https://github.com/Balbilla>.

24. DOI: <https://doi.org/10.5281/zenodo.18892628>.

References

- [Auanger, L. \(2002\)](#), 'Glimpses through a Window: An Approach to Roman Female Homoeroticism through Art Historical and Literary Evidence', in A. Richlin and L. Auanger (eds) *Among Women: From the Homosocial to the Homoerotic in the Ancient World*. Austin: University of Texas Press, pp. 211–55.
- [Bamman, D. and Crane, G. \(2008\)](#), 'Guidelines for the Syntactic Annotation of the Ancient Greek Dependency Treebank (1.1)'. Available at: <https://static.perseids.org/guidelines-syntactic-annotation-greek-1-1.pdf>.
- [Bamman, D. and Crane, G. \(2011\)](#), 'The Ancient Greek and Latin Dependency Treebanks', in C. Sporleder, A. Van den Bosch and K. Zervanou (eds) *Language Technology for Cultural Heritage. Theory and Applications of Natural Language Processing*. Berlin, Heidelberg: Springer, pp. 79–98. https://doi.org/10.1007/978-3-642-20227-8_5.
- [Bamman, D., Mambrini, F. and Crane, G. \(2009\)](#), 'An Ownership Model of Annotation: The Ancient Greek Dependency Treebank'. Available at: <http://www.perseus.tufts.edu/~ababeu/tlt8.pdf>.
- [Bentein, K. \(2018\)](#), 'The Decline of Infinitival Complementation in Ancient Greek. A Case of Diachronic Ambiguity Resolution', *Glotta*, 94, pp. 82–108.
- [Bernand, A. and Bernard, É. \(1960\)](#), *Les inscriptions grecques et latines du Colosse de Memnon*. Paris: Institut Français d'Archéologie Orientale.
- [Bowie, E.L. \(1990\)](#), 'Greek Poetry in the Antonine Age', in D.A. Russell (ed.) *Antonine Literature*. Oxford: Oxford University Press, pp. 53–90.
- [Bozia, E. \(2018\)](#), *Measuring Greekness: A novel computational methodology to analyze syntactical constructions and quantify the stylistic phenomenon of Attic oratory*. PhD thesis, University of Leipzig.
- [Brennan, T.C. \(1998\)](#), 'The Poets Julia Balbilla and Damo at the Colossus of Memnon', *Classical World*, 91–4, pp. 215–34.
- [Carandini, A. \(1969\)](#), *Vibia Sabina: Funzione politica, iconografia e il problema del classicismo Adrianeo*. Florence: Leo Olschki.
- [Celano, G. \(2014\)](#), 'Guidelines for the annotation of the Ancient Greek Dependency Treebank 2.0'. Available at: https://github.com/PerseusDL/treebank_data/edit/master/AGDT2/guidelines.
- [Chue Hong, N.P., Katz, D.S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F.E., Harrow, J., Castro, L.J., Gruenpeter, M., Martinez, P.A., Honeyman, T., Struck, A., Lee, A., Loewe, A., van Werkhoven, B., Jones, C., Garijo, D., Plomp, E., Genova, F., RDA FAIR4RS WG. \(2022\)](#), 'FAIR Principles for Research Software (FAIR4RS Principles) (1.0)'. Zenodo. <https://doi.org/10.15497/RDA00068>.
- [Cirio, A.M. \(2011\)](#), *Gli epigrammi di Giulia Balbilla: (ricordi di una dama di corte) e altri testi al femminile sul colosso di Memnone*. Satura: Testi e Studi di Letteratura antica 9. Pensa Multimedia.
- [García Teijeiro, M. \(1984\)](#), 'Notas sobre el vocabulario de los epigramas de Julia Balbilla' in L. Gil and R.M. Aguilar (eds) *Apophoreta Philologica Manuel Fernández-Galiano a sodalibus oblata I*. Madrid: Estudios Clásicos XXIV-1, pp. 99–102. Available at: <https://www.estudiosclasicos.org/wp-content/uploads/EClas-087.pdf>.

- Guardiano, C. and Stavrou, M. (2019), ‘Comparing Patterns of Adjectival Modification in Greek: A Diachronic Approach’, *Quaderni Di Linguistica E Studi Orientali*, 5, pp. 135–73. <https://doi.org/10.13128/QUISO-2421-7220-25964>.
- Henriksson, E. and Vierros, M. (2025), ‘PapyGreek Search: Exploring the Language of Greek Papyri’, in N. Reggiani (ed.) *Digital Papyrology III: The Digital Critical Edition of Greek Papyri: Issues, Projects, and Perspectives*. Berlin; Boston: De Gruyter, pp. 163–84. <https://doi.org/10.1515/9783111070162-011>.
- Horrocks, G.C. (2010), *Greek: A History of the Language and Its Speakers*. 2nd ed. Oxford: Wiley-Blackwell.
- Joseph, B.D. (1983), *The Synchrony and Diachrony of the Balkan Infinitive*. Cambridge: Cambridge University Press.
- Kaibel, G. (1878), *Epigrammata Graeca ex lapidibus conlecta*. Berlin: Reimer.
- Karakhanyan, A., Avagyan, A. and Sourouzian, H. (2010), ‘Archaeoseismological Studies at the Temple of Amenhotep III, Luxor, Egypt’, *Ancient Earthquakes*. The Geological Society of America Special Paper 471. [https://doi.org/10.1130/2010.2471\(17\)](https://doi.org/10.1130/2010.2471(17)).
- Kavčič, J. (2016), ‘The Decline of the Aorist Infinitive in Ancient Greek Declarative Infinitive Clauses’, *Journal of Greek Linguistics*, 16(2), pp. 266–311. <https://doi.org/10.1163/15699846-01602004>.
- Keersmaekers, A., Mercelis, W., Saelens, C. and Van Hal, T. (2019), ‘Creating, Enriching and Valorizing Treebanks of Ancient Greek’, in *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, Paris, France, pp. 109–117. Association for Computational Linguistics. Available at: <https://aclanthology.org/W19-7812/>.
- La Roi, E. (2025), ‘Rewriting the History of the Optative in the Post-Classical Greek Papyri and Literary Texts (III BCE–VI CE): Persistence, Formulae and Innovation’, in K. Bentein and M. Vierros (eds) *Rays of Language: Linguistic Perspectives on Non-Literary Papyri and Related Sources*. Helsinki: Helsinki University Press.
- Letronne, A.-J. (1848), *Recueil des inscriptions grecques et latines de l’Égypte*. Paris: L’Imprimerie royale.
- Mambrini, F. (2016), ‘The Ancient Greek Dependency Treebank: Linguistic Annotation in a Teaching Environment’, in G. Bodard and M. Romanello (eds) *Digital Classics Outside the Echo-Chamber*. London: Ubiquity Press. Available at: <https://doi.org/10.5334/bat.f>.
- Manolessou, I. (2000), *Greek Noun Phrase Structure: A Study in Syntactic Evolution*. Cambridge: Cambridge University Press. <https://doi.org/10.17863/CAM.31279>.
- McArdle, K. (2020), ‘Rethinking Julia Balbilla: Queer Poetics on the Memnon Colossus’, in *Society for Classical Studies Conference 2020*, paper 19.2. Abstract available: <https://classicalstudies.org/annual-meeting/151/abstract/rethinking-julia-balbilla-queer-poetics-memnon-colossus>.
- Peek, W. (1934), ‘Zu den Gedichten auf dem Memnonskoloss von Theben’, *Mitteilungen des deutschen Instituts für Ägyptische Altertumskunde in Kairo*, 5–1, pp. 95–127.
- Rosenmeyer, P.A. (2018), *The Language of Ruins: Greek and Latin Inscriptions on the Memnon Colossus*. Oxford: Oxford University Press.
- Tribulato, O. (ed.) with Merisio, E.N. (2022–) *Digital Encyclopedia of Atticism*. Venice: Edizioni Ca’ Foscari. Available at: <https://atticism.eu/>.
- Vierros, M. and Henriksson, E. (2021), ‘PapyGreek Treebanks: A Dataset of Linguistically Annotated Greek Documentary Papyri’, *Journal of Open Humanities Data*, 7, p. 26. <https://doi.org/10.5334/johd.55>.
- Vierros, M. and Yordanova, P. (2022), ‘Querying Syntactic Constructions in Ancient Greek Parsed Corpora: A Case Study on the Genitive Absolute in Literature and Documentary Papyri’, in S. Chronopoulos, F.K. Maier and A. Novokhatko (eds) *Digital Text Analysis of Greek and Latin Sources: Methods, Tools, Perspectives*. Classics@ 20. Cambridge, MA: Harvard Center for Hellenic Studies. Available at: <https://classics-at.chs.harvard.edu/querying-syntactic-constructions-in-ancient-greek-parsed-corpora-a-case-study-on-the-genitive-absolute-in-literature-and-documentary-papyri/>.
- Vierros, M. and Yordanova, P. (forthcoming), ‘Scaled Linguistic Analysis of Papyri’, in O. Bobou, R. Raja and E. Barker (eds) *Oxford Handbook for Digital Classical Studies*. Oxford: Oxford University Press.
- Viti, C. (2008), ‘Genitive Word Order in Ancient Greek: A Functional Analysis of Word Order Freedom in the Noun Phrase’, *Glotta*, 84(1–4), pp. 203–238. <http://www.jstor.org/stable/40730997>.
- Yatromanolakis, D. (2007), *Sappho in the Making: The Early Reception*. Hellenic Studies 28. Cambridge, MA: Harvard University Press.
- Yordanova-Nancy, P. (2026), ‘“Tolerable Fluency and Grace ... and Occasionally an Interesting Word Order”: Quantifying Language Proficiency for the Study of Word Order Variation in Documentary Papyri’, in G. di

Bartolo, D. Kölligan, P. Filos and G.K. Giannakis (eds) *Trends in Classics—Supplementary Volumes*. Berlin: De Gruyter, pp. 257–82.

Chapter 11

The next decade of Byzantine sigillography: the impact of open data as collaboration

Martina Filosa, Jonathan Shea and Alessio Sopracasa

Introduction

Byzantine sigillography is a relatively recent discipline, born in the nineteenth century, and has never been able to count on a very large number of practitioners ([Sode 2001](#); [Prigent 2017](#)): unlike sister disciplines such as epigraphy, papyrology or numismatics, there has been little reflection on its foundations as a discipline and the development of shared practices and standards.¹ Similarly, it is impossible to speak of schools of sigillography, as it was often the working habits of individuals that prevailed, with a very low rate of transmission of knowledge to other generations. This has contributed to isolating the study of Byzantine seals from the study of seals of other periods and cultural areas. Worse still, this situation even tends to isolate the various editing ventures carried out within Byzantine sigillography itself, since the same features are not necessarily edited in the same way. The challenge is therefore to make Byzantine sigillography more open; in other words to develop standardised practices, enhancing more transparent, verifiable and replicable methods and results, to enable better interaction between scientific endeavours in Byzantine sigillography, and between the latter and scholarly activity within Byzantine studies in the broad sense and beyond.

The purpose of this chapter is to highlight the work in progress and the results achieved by examining the French-German project ‘Unlocking the Hidden Value of Seals: New Methodologies for Historical Research in Byzantine Studies—DigiByzSeal’ (2022–2026) (henceforth DigiByzSeal), funded by the Agence Nationale de la Recherche and the Deutsche Forschungsgemeinschaft, and the activities of Dumbarton Oaks research centre (Washington, DC).

We firmly believe that the best way to address these shortcomings and meet this challenge is through embracing digital solutions. This includes providing researchers and the public with freely accessible tools, practices and recommendations that are easy to adopt and share. One such solution is SigiDoc, an XML-based and TEI and EpiDoc-compliant data model.² The development of its 2.0 version is currently underway as part of the DigiByzSeal project.³

The SigiDoc initiative originated with Charlotte Roueché, who was among the first to consider applying digital methods to Byzantine seals. This idea was particularly inspired by her experience with EpiDoc, and it was Roueché who gave the initiative its name, SigiDoc. This concept dates back to the early 2000s, and between 2006 and 2011, numerous sigillographers were invited to discuss it.⁴ The goal was to gather as many experts as possible to address the lack of shared rules in the field. This collective effort aimed to establish a common framework that the scientific community could adopt and identify with. Interestingly, the initial difficulty in reaching a consensus highlighted both the urgency and the importance of this endeavour. In 2015, Roueché encouraged Alessio Sopracasa to revive the project, and by 2017, the Department of Byzantine and Modern Greek Studies at the University of Cologne had joined the effort. While sigillography is only now beginning its digital transformation, disciplines such as epigraphy, numismatics and papyrology have long embraced their digital potential. SigiDoc and the DigiByzSeal project, which has been built around it, are rising to this challenge and are now developing solutions.

The Byzantine seals project at Dumbarton Oaks began with the appointment of Nicolas Oikonomides as Advisor in Sigillography in 1973. He was joined by John Nesbitt in 1973 and Eric McGeer in 1990. Two significant products of the project up to the retirement of John Nesbitt in 2010 are the first six volumes of the print catalogue *Catalogue of Byzantine*

Seals at Dumbarton Oaks and in the Fogg Museum of Art and the creation of the Athena inscription font. In 2010, when Jonathan Shea joined the project, it was decided to focus on an online catalogue. After much deliberation this was built directly into the Dumbarton Oaks website using a Plone content management system (<https://plone.org/>). Over the following years the focus of cataloguing moved away from the comprehensive entries and commentaries of the print volumes towards entries that included the basic physical data for the seals, quality photography, descriptions, transcriptions in Greek or Latin, using the font Athena Ruby, and then an English translation. As this last point makes clear, the driving goal was to publish online as many seals as quickly as possible, and crucially in a way that made the material as accessible as possible. Sigillographic publications that lack a modern language translation of the inscriptions can never be useful to any but trained Byzantinists. Proceeding in this way has meant that the entire Dumbarton Oaks collection has been freely available for many years longer than if we had waited on publication until the perfect and complete entry of each seal had been written. Furthermore, one of the many benefits of the digital medium over print is that now that this first phase of cataloguing is complete, those missing detailed commentaries can be added, and possibly updated, as the scholarship is conducted.

Edition of the sigillographic legend

The lack of standardised practices is evident in various aspects of editing Byzantine seals. For instance, the information provided about the seal's materiality can vary significantly. Something as basic as the weight is often missing. Descriptions of the seal's preservation state are frequently limited because sigillographers tend to rely on photographs and feel they can omit detailed verbal descriptions, even though these descriptions have scholarly value, make the resource more accessible and allow for disability accommodation (a verbal description can be used with text-to-speech programs) and might prove significant for the history of the discipline ([Alexander 2020](#); [Wallace 2021](#); [Filosa et al. 2023](#)). Similarly, the depiction of iconography can range from a simple identification of the subject to an extensive description of details like clothing and attributes.

Diplomatic transcription

With regard to diplomatic transcription, Byzantine sigillography (and numismatics) do have a transcription standard. Unlike Greco-Roman epigraphy on stone, where the diplomatic transcription usually features capital letters in *scriptio continua*, the transcription standard for Byzantine sigillography and numismatics consists of a font that schematically represents the features of the characters found on the written support. The challenge here is ensuring this standard is used consistently across its broad range of applications; initiatives such as SigiDoc are, we shall argue, best equipped to encourage the widespread, consistent and informed use of this font.

When preparing material for his 1986 study *A Collection of Dated Byzantine Lead Seals*, Nicolas Oikonomides realised that simply using a mixture of upper- and lower-case Greek and Latin characters was not an ideal solution to the problem of representing seal inscriptions ([Oikonomides 1986](#): 15). He developed a True Type font called Seal Short, at the time the only font for Byzantine inscriptions of any kind, to represent the nuances of the characters found on Byzantine seals, including letter variants, ligatures, abbreviations and some decorations. After a rebuild in the early 1990s, this font became known as Athena and was joined by five separate fonts intended for numismatic publication, also developed at Dumbarton Oaks. In 2010 Joel Kalvesmaki (Dumbarton Oaks) and John Hudson (Tiro Typeworks) began working on an updated and expanded font not only combining all of the above but also adhering to Unicode standards, which is today's Athena Ruby (named for Glen Ruby d. 2004, Dumbarton Oaks Publications Manager). Athena Ruby was intended as a Greek and Latin font with characters found on coins, seals, weights, mosaics and frescoes. It was also created with web-based projects using XML and JSON in mind ([Kalvesmaki 2015](#); [Trécourt-Codine and Sarah 2012](#)).

As a True Type-flavoured Open Type and Unicode-compliant font, Athena Ruby was designed with features that are only beginning to be supported by accessible software and web browsers; for example, the Open Type feature that allows characters in a ligature to be typed and machine read as individual letters but displayed as a single character, or recognising that the font's twenty-nine alphas are variants of the same character

represented by a single Unicode codepoint. Outside of InDesign, few programs can currently handle the way Athena Ruby reads and understands characters and their variants. The shortcomings of common browsers and word processors are overcome through Private Use Area (PUA) duplicates for variant letterforms, whereby each individual variant is assigned its own unique codepoint. While less than ideal because of the way it mixes fully Unicode-compliant characters with PUA letterforms, this allowed particularly print publications of seals to proceed. Many projects using Athena Ruby since 2010 are thus in a mixture of Unicode-compliant and PUA characters.

Over the next decade we will enter a period of transition that will see the PUA characters become obsolete. In anticipation, Dumbarton Oaks is developing a new version of Athena Ruby that cuts out the PUA, leaving the Unicode-compliant characters that will be optimised for web publishing and advanced word processing and design software. The difficulties of typing in such a large font will be overcome by adapting the existing online keyboard to the new, clean version. At the same time, Dumbarton Oaks will maintain the existing PUA version, with its own keyboard, for anyone using software more suited to it. To facilitate moving back and forth between the two there will be a conversion tool where clean Unicode inscriptions can be turned into PUA using text and vice versa.

The font will grow as colleagues suggest new variants and letterforms. While consistent use poses challenges—such as which ligatures are seen differently—these are minor compared to Athena Ruby’s advantages over regular fonts, including the original Athena. Cataloguers and SigiDoc will play a key role in identifying additions and guiding consistent use. Athena Ruby should remain central to digital work, offering a wide, machine-readable set of free characters tailored to our field.

Interpretive edition

The area where editing practices diverge the most is in the transcription of the legend’s text. Here, the absence of common rules is glaringly apparent, revealing the high degree of personal interpretation. Paradoxically, even within the same publication, the methods used to handle similar features

can vary from one seal to another, highlighting the inconsistency in editorial practices.

The most significant consequence of this inconsistent editing work is the lack of interoperability: first, ‘internal’ interoperability within a single corpus and more broadly within sigillographic production as a whole. This means that the legends cannot be explored in a reliable, systematic and exhaustive manner, sometimes even within a single publication or across different published corpora. Additionally, these data lack ‘external’ interoperability because their inconsistent production methods prevent them from being related to work in other disciplines and fields, ultimately hindering their transformation into Linked Open Data (LOD).

The inconsistencies in the transcription of legends arise not only because the criteria are not universally shared but also because the commonly used criteria are very limited and do not cover all possible cases. The so-called Leiden conventions are only marginally used in the editing of Byzantine seals.

The need to improve and standardise the transcription of the legend is undoubtedly one of the main reasons for the existence of SigiDoc and it has been a primary focus of its development, particularly within the DigiByzSeal project. Currently, the Guidelines are heavily centred on this aspect:⁵ they encourage editors to make more consistent editorial choices, adopting criteria they are not traditionally accustomed to, but without imposing any rigid editorial style and providing flexible recommendations.

Although the legend is undoubtedly the core of the sigillographic data contained in an edition, it is far from being the only kind; data on the materiality of the seal, its history, the bibliography relating to it and the commentary dedicated to it are all valuable pieces of information. The key is to ensure that the encoding is completed and the data is made available, while the visualisation of this encoding can be tailored by the editors according to their preferences. DigiByzSeal wants to make available not only a publication interface that allows researchers (or users) to view the editions of the seals, but also access to the encoded source files that are the real ‘long term outcome of the project’ ([Turska et al. 2016](#)–2017: §2).

[Indexing and searching named entities](#)

Indexing and searching are the most effective ways to make sigillographic data accessible and, more importantly, reusable. The development of a seal-centric search interface in the framework of the DigiByzSeal project marks a significant advancement in the use and reuse of this data. This interface emphasises tokenisation (explicit markup of words) as a means of fostering user-driven exploration. This design philosophy ensures that scholars can engage with the material on their own terms, without the influence of developers' or Byzantine researchers' biases. By offering data in a tokenised format, the interface empowers users to identify and analyse patterns independently, promoting a more flexible and open-ended approach to the investigation. This methodology reflects the broader objectives of digital humanities, where the value is placed on interpretative pluralism and the tools employed are intended to facilitate, rather than restrict, scholarly inquiry.

Towards an atomised search form

In order to continue opening up Byzantine sigillography, we need to make better use of all the main data (aular dignities, functions in civil, military or ecclesiastical administration, geography, iconography and so on) derived from seals, which means making it reusable, while integrating it into the mass of data available to researchers and other interested parties. To manage the data, SigiDoc uses EpiDoc Front-End Services (EFES),⁶ the customisation and publication platform developed since 2017 for inscriptions edited with EpiDoc, of which SigiDoc is a subset. To generate indexes and search forms, EFES relies on authority lists—XML files that organise data by category and serve as thesauri for each corpus published using SigiDoc. While the use of EFES is strongly recommended within the DigiByzSeal project, it is not mandatory. What matters most is that these authority lists exist, they follow a straightforward structure and are easy to maintain.

These lists are the true treasure trove shared by all editorial initiatives using SigiDoc. They enable data to circulate across projects and serve as a bridge between sigillography and broader fields—enriching both Byzantine studies and research beyond the Byzantine world, while also allowing sigillography to benefit from related scholarly work in other domains.

They also embody the choice made by the DigiByzSeal project to introduce an atomic encoding of the seal legends. To give an example, if we take the function of judge of the Hippodrome and of Paphlagonia inscribed on a tenth to eleventh century seal held in the Dumbarton Oaks collection (BZS.1951.31.5.420),⁷ the encoding recommended by SigiDoc is as follows:

```
<rs type="officest" subtype="civil" ref="#of0112">κριτῆ</rs> <rs  
type="institutions" ref="#org0017">ἐπὶ τοῦ Ἱπποδρόμου</rs> καὶ  
<rs type="places" ref="#geo0043">Παφλαγονίας</rs>
```

Where `ref="#of0112"`, `ref="#org0017"` and `ref="#geo0043"` refer to entries in the authority lists of, respectively, civil offices, institutions and places, as made clear by the `@type` attributes.

Rather than compiling exhaustive lists that attempt to account for every possible variation of judicial functions as exercised across different institutions and locations, a more efficient strategy involves disaggregating each component of the expression and assigning it a distinct encoding. This not only facilitates the management and maintenance of authority lists, but also significantly enhances interoperability, enabling more seamless integration with both internal datasets and external resources. For instance, placenames—primarily reflecting the administrative and ecclesiastical geography of the Byzantine empire—are linked to established geographical gazetteers such as Pleiades, GeoNames, the Digital Tabula Imperii Byzantini or more regionally focused repertories.

In the case of the same seal, the Dumbarton Oaks catalogue adopts a similar—though somewhat less granular—approach. Metadata is entered into distinct fields via dropdown menus, which are then aggregated on the public-facing webpage. For this particular example, the categorisation would be: Office = judge of the Hippodrome, Place/Location = Paphlagonia. Currently, Dumbarton Oaks does not further subdivide offices into specific subtypes. Although the system was originally designed to support integration with geographic gazetteers, the inclusion of persistent identifiers (URIs) is planned for a later phase, alongside the publication of the accompanying commentaries.

Cases requiring special attention

The processing of certain types of data, such as personal names and identifiers attributed to seals and their matrices, required specific solutions. While the granular approach used for other data is still applied here, it doesn't necessarily result in the creation of authority lists. This is mainly because maintaining such lists would be impractical, especially since this data is expected to be used across multiple projects and is more likely to evolve and expand over time. Instead, a different but equally straightforward method is used to make the data available for reuse, both within and outside the project.

Personal names

Traditionally, the most sought-after information on seals has been the names of their owners (or issuers) and today sigillography plays a crucial role in the prosopography of the Byzantine empire. Prosopographical repertories now span almost the entire Byzantine period, many of which are available online, including the *Prosopographie chrétienne du Bas-Empire* (PcBE, fourth to seventh century), the *Prosopography of the Later Roman Empire* (PLRE, 260–641), the *Prosopography of the Byzantine Empire* (PBE, 641–867), the *Prosopographie der mittelbyzantinischen Zeit Online* (PmbZ, 641–1025), the *Prosopography of the Byzantine World* (PBW, 1025–1180) and the *Prosopographisches Lexikon der Palaiologenzeit* (PLP, 1261–1453).

Some of these repertories, such as the PBE, PmbZ and PBW, routinely incorporate sigillographic data. These works span the period when seals were becoming widely used in Byzantium, making it natural to draw on this rich source. Among them, the PBW stands out for its particular focus on seals and it would be impossible to speak of this without recalling, with deep gratitude, Charlotte Roueché's instrumental role in shaping this monumental project.

The DigiByzSeal project, built around SigiDoc, is not designed to create a new prosopography of the Byzantine period, as it is a sigillographic, not prosopographic, endeavour.

There are two main stages in the study of the issuer of a seal: firstly, the person is identified; secondly, the individual must be associated with or differentiated from his or her homonyms. The information obtained during

the first stage is brought together in the XML edition file within a `<listPerson>` element (or `<listOrg>` if the seal belongs to an institution). However, DigiByzSeal does not aim to compile a unified authority list of all individuals mentioned on the seals, as would be done in a prosopographical repertory, because, being an ongoing project, the definition of this list and of the identifiers associated with the names could not be maintained efficiently.

For this reason, a SigiDoc edition will not assign unique identifiers to individuals on its own. Instead, the responsibility for disambiguating or identifying individuals with the same name will fall to the editors or the curators of the prosopographies that incorporate and use this data and by which these identifiers will eventually be provided. Instead, the data will be searchable through the various corpora published using SigiDoc and can be extracted as a raw list containing at least the names and unique identifiers of the seals on which these names appear. This allows the information to be openly accessible and usable for integration into existing prosopographies.

Dumbarton Oaks has adopted a data structuring approach closely aligned with that of the DigiByzSeal team, guided by similar considerations. Personal names are recorded in two distinct fields—*Name* and *Family Name(s)*—while institutional affiliations, such as monasteries or comparable entities, are captured separately in an *Institution* field.

Unique identifiers for seals and matrices

One of the primary objectives of Byzantine sigillography is to identify and group together all seals issued from a single matrix (known as a *boulloterion*). This involves both identifying different series of seals produced by the same matrix and virtually reconstructing the matrices themselves, of which only half a dozen have survived.⁸ Dumbarton Oaks and SigiDoc have explored complementary ways to achieve this goal.

Dumbarton Oaks' approach to grouping seals issued from the same *boulloterion* has evolved over time. In its print publications, seals believed to originate from a single matrix were treated as parts of a unified entry. For instance, in *DOSeals* 4, entry 11.19, shown in [Figure 11.1](#) (left), two individual seals, one broken in half—BZS.1958.106.5462 and BZS.1947.2.195—are combined into a single entry for John, imperial

spatharios and komes tes kortes of Pamphlagonia, reflecting a reconstruction of the original boulloterion.

11.19 John imperial spatharios and komes tes kortes of Pamphlagonia (X c.)

(a) DO 58.106.5462—D. 21 mm. W. 6.40 g.

(b) DO 47.2.195.—D. 21 mm. W. 3.44 g. Half missing.

Both specimens come from the same boulloterion, and their readings are combined below.

Unpublished.

Obv: Bust of the Virgin holding Christ before her. Circular inscription within a border of dots:

+ΚΕΡΟΗΘΕΙΤΟΥΔΟΝΑ.

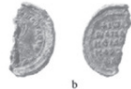
Rev: Inscription of five lines. Border of dots.

+ΙΩΑΝΝ|Ρ·CΠΑΘ·S|K·OΠ·T·T·T·T|K·OΠ·T·T·T·T|Φ·Λ·A·Γ·

Κ(ύρι)ε βοήθει τῷ σῷ δοῦλ(ῳ) Ἰωάν(νῃ) β(ασιλικῷ) σπαθ(αρίῳ) (καὶ) κόμητ(ι) τ(ῆς)
κόρτ(ης) Παμφλαγ(ονίας).



a



b

Obverse

Bust of the Virgin holding Christ before her. Circular inscription within a border of dots:

...ΤΩCΩΔΟΝΑ,

[Κ(ύρι)ε βοήθει τῷ σῷ δοῦλ(ῳ)]

Reverse

Inscription of five lines. Border of dots.

+ΙΩΑ

Ρ·CΠΑ

ΚΟΜΗ..

ΚΟΡΤ...

ΦΛ..

Ἰωάν(νῃ) β(ασιλικῷ) σπαθ(αρίῳ) (καὶ) κόμητ(ι)
τ(ῆς) κόρτ(ης) Παμφλαγ(ονίας)

Figure 11.1: DOSeals 4, entry 11.19 (left) and Dumbarton Oaks Online Catalogue, DO BZS.1947.2.195 (right)

© 2001 Dumbarton Oaks Research Library and Collection, Trustees for Harvard University. Originally published in *Catalogue of Byzantine Seals at Dumbarton Oaks and in the Fogg Museum of Art*, vol. 4, The East, edited by Eric McGeer, John Nesbitt, and Nicolas Oikonomides. Available at: <https://www.doaks.org/resources/seals/byzantine-seals/BZS.1947.2.195>

In contrast, the online catalogue—functioning as a digital inventory of museum holdings—presents each seal as a distinct object. Each seal is documented individually, including details such as missing or damaged letters as shown in [Figure 11.1](#) (right). A composite inscription is then reconstructed based on the combined evidence from all three items. These seals are linked using a ‘Parallel seals’ tag, connecting their separate entries within the database. This same backend architecture also supports other forms of relationships. For example, seals that share a similar design and inscription but are not exact matches struck by the same boulloterion can be linked under the ‘Similar seals’ tag. Likewise, seals that differ visually but are known to have been created by the same individual may also be grouped in this way, allowing for nuanced connections between objects across the collection.

The DigiByzSeal project has taken a step further towards transforming Byzantine seals into fully digital objects, making them easily findable and linkable by assigning each seal a unique, automatically generated identifier, which also serves as the bibliographic reference for the seal as published in

SigiDoc. To facilitate the creation of series of parallels, SigiDoc also assigns a unique, reusable identifier to each (virtual) matrix.

These unique identifiers enable seamless connections between different corpora or within the same corpus, making it easier to integrate and utilise the information beyond the realm of sigillography and Byzantine studies. This approach also supports the development of other electronic resources by ensuring that the data is accessible and reusable across various platforms. Only one other digital initiative assigns IDs to matrices: PBW. It is therefore essential to link the IDs given by SigiDoc with the latter, when the same matrix has been clearly identified.

A sustainable publishing infrastructure

Ensuring the discipline's continuity requires more efficient use of seal edition data. Traditional print methods are increasingly unsuitable, as the volume of published seals grows. Though print quality has improved—with better formats, layouts and photos—costs have soared, making this model economically unsustainable and dependent on external funding. To ease this burden, researchers need accessible, reusable resources. A shared digital infrastructure, centralised yet distributed across SigiDoc-based projects, promotes data interoperability and collaboration, aiming to build an open-ended, evolving virtual corpus of Byzantine sigillography.

And this is the major challenge for Byzantine sigillography: while it is recognised as a distinct and important field, with unique characteristics compared to sigillographies of other periods or regions, there is still much to be done. If we assume that a discipline, in order to be truly defined as such, must fulfil two conditions—having a textbook and being able to draw on the widest possible corpus of material—then we can clearly see that the sigillography of the empire falls short compared to better-resourced fields, such as epigraphy, numismatics and palaeography ([Rhoby 2015](#): 17–18). Of course, reference corpora have existed for some time, such as the George Zacos collection ([Zacos and Vegler 1972](#); [Zacos 1984](#)) and the Dumbarton Oaks collection ([Oikonomides et al. 1991–2026](#)), not to mention Vitalien Laurent's attempt, unfortunately left unfinished, to establish a sigillographic corpus of the Byzantine empire ([Laurent 1965–1981](#)). However, these

publications do not fully meet the needs of a modern, comprehensive textbook and some are becoming outdated.

The DigiByzSeal project is taking crucial steps to address these gaps. By prioritising the sharing of tools and results, the project aims not only to adhere to contemporary standards of open research but also to bridge the ‘communication gap’ within the Byzantine sigillography community. This approach is designed to lay the essential groundwork for advancing the field and ensuring its future growth.

In establishing this digital ecosystem, DigiByzSeal has adopted the GitHub version control and collaboration platform to disseminate the resources developed within its framework. At the heart of this system is the SigiDoc organisation, which serves as a hub for the repositories that house the foundational components essential for producing, managing and sharing digital editions of Byzantine seals.⁹ This central structure acts as the core repository for all projects utilising the SigiDoc framework.

The foundational components, as visualised in [Table 11.1](#) (column 1), include the SigiDoc 2.0 schema, which ensures the validation of XML files containing sigillographic editions; the XML template defining the structure of a SigiDoc edition; the stylesheets that manage the transformation and display of seals’ legends; the standard version of EFES, adapted for SigiDoc; and the authority lists. These resources form the backbone of the system, enabling seamless interaction and collaboration between individual edition projects. Together with the Guidelines,¹⁰ they represent an ‘encoding policy’ for Byzantine sigillography and ‘are the most important acts that make all further research and long-term preservation of editors’ wealth of knowledge (not to mention publication) possible’ ([Turska et al. 2016](#)–2017: § 2).

Table 11.1: Overview of the GitHub repository structure behind the SigiDoc portal (from [Sopracasa et al. 2024](#): § 44)

1. Central SigiDoc repository		2. Collection repositories	
1a. Unique repository for unchanging code	1b. Forkable code	2a. Forked code	2b. Raw XML data

1. Central SigiDoc repository		2. Collection repositories	
schema	central EFES- SigiDoc instance	individual EFES-SigiDoc instances	XML files containing the edition of the seals
full template	authority files	authority files	
EpiDoc stylesheets			

The various SigiDoc-based editing projects, also hosted on GitHub, are directly linked to this foundational material. DigiByzSeal currently oversees two projects, based in Paris¹¹ and Cologne.¹² As depicted in [Table 11.1](#) (column 2), each project contains three primary repositories: a fork¹³ of the SigiDoc authority lists, a fork of the EFES platform and XML files for the editions of seals from distinct collections. The XML files representing the seal editions are the responsibility of the individual editing projects. Authority lists can be modified locally within each project and may later be submitted via pull requests to the central SigiDoc repository for integration. Similarly, while the EFES fork can be customised to meet project-specific needs, fundamental changes can be proposed through pull requests for inclusion in the standard EFES version within the SigiDoc organisation. Conversely, updates made to the central repository are propagated to the forks to ensure synchronisation.

This collaborative model is bidirectional: updates made to the central SigiDoc repository are immediately, but optionally, accessible to all participating projects, which, in turn, may propose changes, which may possibly be integrated into the central repository. This two-way exchange primarily affects authority lists and EFES, facilitating day-to-day work. However, the stylesheets, templates and schema are also designed to evolve in response to contributions from SigiDoc-based projects, as new requirements emerge. These updates can be managed through pull requests to the central SigiDoc organisation, ensuring that the system remains flexible and responsive to the scholarly community's needs.

The infrastructure proposed for digital Byzantine sigillography serves a dual purpose. On the one hand, individual projects will be able to publish their sigillographic corpora online using EFES, or alternative publication platforms. These platforms will generate websites—hosted on servers chosen by the respective project leaders—where the edited seals, along with their indexes and search functionalities, will be accessible to users. On the other hand, thanks to the git-based architecture that links these projects, a collective virtual repository will emerge, enabling a unified search engine to query all corpora edited through the SigiDoc data model. Notably, this applies even to projects that do not use EFES or any customisation platform but still contribute XML edition files and participate in the authority lists.

We are aware that the use of GitHub (or another git) and EFES can raise problems concerning their own sustainability. By the end of the project (March 2026) we plan to deposit all the XML edition files in institutional repositories both in France¹⁴ and in Germany;¹⁵ in this way, even if the platform becomes unsustainable or a better one comes along, the XML data will continue to be available and is in the best form to be reused and redeployed; the same is planned for the transformation tools in use, according to the LOCKSS principle.¹⁶

In addition, since EFES is built upon various independent software components—some of which are no longer actively maintained (for example, Kiln, which relies on Cocoon)—it is only a matter of time before projects employing EFES encounter issues of maintainability, thereby jeopardising both accessibility and usability. Rather than waiting for these technologies to become obsolete and risk collapse, it is necessary to anticipate the inevitable ageing of the software by proactively preparing a solution that will ensure the continuation of research and outputs derived from the project. The aim, therefore, is to take an initial step towards a more sustainable and durable version of EFES, one that will employ lightweight and long-lasting technologies such as XSLT, XML and JavaScript. By refactoring EFES's key components, the intention is to provide a prototype solution—available by the end of the project—that can be deployed immediately while also serving as the foundation for the development of a more resilient successor to the EFES platform.

Nevertheless, although we have devoted significant effort to the online presentation of our data, the encoding itself was not designed to conform to

the EFES display—in fact, the latter was developed to accommodate the former. While presentation may be the most ephemeral aspect of the work done so far, it should not be underestimated. Most users of the seal editions and the SigiDoc portal will not want to engage with the underlying XML or raw data. Moreover, the essential task we now face—convincing specialists in Byzantine sigillography to adopt SigiDoc—requires us to demonstrate what their encoding efforts could yield in practice. Data encoding and its web presentation are two distinct yet equally important components of the editorial process and it is our responsibility to emphasise and communicate the value of both.

A core component of this infrastructure is a central portal, developed as part of the DigiByzSeal project and hosted at the University of Cologne.¹⁷ This ‘SigiDoc portal’—as it has been called¹⁸—will act as a meta-search engine, capable of querying all participating SigiDoc-based corpora simultaneously.¹⁹ The precision and scope of search results will depend on the granularity of the encoding done by the various projects. The long-term goal of DigiByzSeal, and its most significant legacy, is to lay the groundwork for the *Corpus Sigillorum Ævi Byzantini*, an urgently needed resource in Byzantine sigillography.²⁰

This infrastructure offers a robust, collaborative solution for managing open data in Byzantine sigillography. It encourages collective engagement while allowing individual projects to maintain autonomy in publishing their datasets. A centralised, top-down effort—what might be considered a ‘Napoleonic’ approach—to gather and curate all sigillographic material under one team and location would be unrealistic and bound to fail. Such an endeavour would face too many practical hurdles: Who would lead such an effort? Where would it be based? How long would it take, and with what funding? Instead, building the *Corpus Sigillorum Ævi Byzantini* requires a joint, multilateral effort, pooling the expertise and resources of many. The proposed infrastructure embodies this collaborative spirit, facilitating shared efforts while still allowing individual projects the freedom to contribute at their own pace.

A key challenge ahead is encouraging adoption of the SigiDoc recommendations and their supporting infrastructure, essential for enabling interoperability across sigillographic projects. But interoperability must serve real scholarly purposes, not remain a theoretical goal. The SigiDoc

portal aims to make this potential tangible through shared standards and integration. Digital tools in the humanities often face resistance, especially from scholars used to traditional methods. This scepticism, tied to fears of losing control or disrupting established practices, must be addressed. A generational shift may help ease a transition towards a collaborative digital endeavour, but it would be a real loss if this were to happen without the participation of the senior scholars; the widest possible collaboration will greatly benefit the development and the success of the *Corpus Sigillorum Aevi Byzantini*.

Teaching and outreach

The greatest challenge facing Byzantine sigillography is ensuring continuity, which is essential for laying strong foundations. Although a new generation is currently joining the field, there are very few training centres and specialists. Due to the near absence of structured programmes within universities or research centres, sigillographers are often trained in an individual manner, from ‘teacher to pupil’, on the periphery of recognised diploma-granting courses. Teaching and training a new generation of students in Byzantine sigillography has been an ongoing struggle for decades. Within the already limited pool of students in Byzantine studies—one of the *Orchideenfächer* par excellence—it is challenging to find those interested in the study of seals. Even among established researchers, one of the chief barriers to using seals is that they are simply unaware that the material can have applications and implications in their area of study. This shortage of awareness and expertise in Byzantine sigillography means that this invaluable resource remains under-utilised by scholars in Byzantine history, philology and art history ([Jeffreys 2009](#)). SigiDoc, online cataloguing and teaching all share the same goal of broadening access to sigillographic material and helping scholars from all areas of Byzantine studies realise its value and integrate seals into their work. This final point—namely, integration—is particularly significant, as few students today set out to become sigillographers. Realistically, this is unlikely to change in the coming decade. However, it remains highly valuable to promote the use of seals as a source among all scholars of Byzantium. What we can actively do

is to take more seals out of the scholarly niche and place them on public display and into the hands of students.

The authors, with institutional support, are addressing current challenges by offering training in Byzantine sigillography through seminars and schools. Dumbarton Oaks has hosted biennial summer schools since the 1980s, and since 2015, the University of Cologne has run a seminar led by Prof. Claudia Sode on the Feind Collection. These programmes teach scholars how to read, identify and date seals. Students engage directly with seals—an uncommon experience—creating a physical link to the past. Each seal offers insights into identity, faith, profession and social ties. This hands-on approach appeals to diverse audiences, from students to museum visitors. Both institutions have successfully trained new sigillographers and sparked wider public interest through community outreach.

However, it is only thanks to the contribution of digital solutions that the efforts made by these institutions and their researchers have been brought to the attention of an audience beyond the circle of specialists, thereby enabling them to meet the challenge of openness. The basis of all Dumbarton Oaks's outreach is the scholarship that has gone into its catalogues, particularly the online version, which is itself a form of outreach. More people view that version, between 2,000 and 5,000 unique page views a month in recent years, than will ever look at the print publications, valuable as those are. The digital format, with its possibility for virtually unlimited space, modern language translations of inscriptions, large images and constantly updated entries and commentaries, and the possibilities that it has opened up in exhibitions and outreach, has allowed Dumbarton Oaks to reach an audience far larger and far more diverse than ever before.²¹ On its side, if the University of Cologne has emerged as a key centre for the study and teaching of Byzantine sigillography, this is particularly true within the context of digital humanities. Through projects like DigiByzSeal and even more so 'Creating a Sustainable Digital Infrastructure for Research-Based Teaching in Byzantine Studies' (henceforth DiBS), supported by the VolkswagenStiftung 2022–2028, the university is advancing both research and pedagogy in this field.²² A cornerstone of this effort is the SigiDoc training weeks, which are designed to equip participants—ranging from students and early career researchers to museum curators and established scholars—with practical skills in XML

encoding and data modelling ([Drach et al. 2024](#): 5–12). These training sessions emphasise hands-on learning, enabling participants to work directly with Byzantine seals and implement digital tools like SigiDoc for creating and managing scholarly editions.²³

If the narrative becomes less about a small piece of lead and instead focuses on the person who owned it, seals become the ideal medium to connect modern audiences with the empire and its people. Although their size, language and state of preservation can present challenges, as tactile objects with engaging images and the words of their owners engraved on their surface, seals are perfect candidates for exhibition and educational outreach initiatives, such as ‘Lasting Impressions: People, Power, Piety’, recently organised by Dumbarton Oaks.²⁴ Similarly, a key outcome of the efforts of the DiBS project is the creation of a digital exhibition, originating from the international seminar ‘Approaching Byzantium Through Seals: A Digital Exhibition’, led by Dr Christos Malatras at the University of Cologne in the Fall term of 2023/24.²⁵ This exhibition serves as both a teaching tool and a public-facing platform, highlighting how seals can reveal insights into Byzantine society and culture ([Drach et al. 2024](#): 14–18).

A key strength of these efforts is the dialogue between scholarly research and public engagement. This ensures accurate, respectful dissemination of sound knowledge while also helping scholars improve communication and adapt their work to varied audiences.

Conclusions

The developments described in this chapter have all happened in the last ten years. It has become clear that Byzantine sigillography as a whole must embrace several critical shifts: it needs to expand its working methods and tools, and foster the sharing of results and knowledge; it must strengthen internal connections and collaborations among its specialists, building a cohesive community; it should engage more deeply with other fields of Byzantine studies and contribute to positioning Byzantium within medieval studies, moving it beyond the often inappropriate confines of ‘classics’; finally, it should reach out to a broader audience, including enthusiasts and

the curious. While medieval Western seals frequently capture public interest, there is no reason—aside from a lack of will—that the same cannot be true for the bullæ of the Eastern Roman empire. Digital (to use a deliberately vague expression) is the way forward; the first goal is achievable in the near future, provided we maintain the momentum to adopt these emerging solutions. The remaining three objectives may require more patience, with their success likely hinging on the achievement of the first. The next ten years promise to be just as exciting as we take the lessons we have learnt and move forward.

Acknowledgements

The authors would like to extend their warmest thanks to Elli Mylonas, Tara Andrews and Carl Rice for helping to improve this contribution with their valuable comments and suggestions.

The research presented in this paper was supported by the Agence Nationale de la Recherche (France) and the Deutsche Forschungsgemeinschaft (project ‘DigiByzSeal’: <https://anr.fr/Project-ANR-21-FRAL-0008>; <https://gepris.dfg.de/gepris/projekt/469385434>), by the VolkswagenStiftung 2022–2028 (project ‘DiBS’: <https://uni.koeln/NXPQU>) and by the Dumbarton Oaks research centre.

Notes

1. There is no handbook of Byzantine sigillography: some overviews, without the ambition of fulfilling this function, include those of [Tsougarakis \(1999\)](#) and [Cheynet \(2008\)](#).
2. For the TEI (Text Encoding Initiative), see <https://tei-c.org/>; for EpiDoc, the recommendations for the scholarly edition of ancient documents, see <https://epidoc.stoa.org/>.
3. On SigiDoc 2.0, see most recently [Sopracasa et al. \(2024\)](#). Source code available from: <https://github.com/SigiDoc>. All links and digital resources cited in this article, unless otherwise stated, were last accessed on 11 May 2025.
4. A summary of the discussions during meetings held in the late 2000s and early 2010s can be found at <http://sigidoc.wikidot.com>.
5. Available from: <https://github.com/SigiDoc/Guidelines/wiki>.
6. Available from: <https://github.com/EpiDoc/EFES/wiki>.
7. Available from: <https://www.doaks.org/resources/seals/byzantine-seals/BZS.1951.31.5.420>.
8. One of them is held by the Harvard Art Museum: Object no. 1951.31.6, Harvard Art Museum’s collections online, <https://hvrd.art/o/291105>.

9. Available from: <https://github.com/SigiDoc>.
10. Available from: <https://github.com/SigiDoc/Guidelines/wiki>.
11. Available from: <https://github.com/sceaux-byzantins-paris>.
12. Available from: <https://github.com/byzantinistik-koeln>.
13. A fork is a copy of a repository that allows us to freely experiment with changes without affecting the original project but that can still receive updates from the original repository.
14. The Huma-Num infrastructure: <https://www.huma-num.fr/>.
15. The Data Center for the Humanities at the University of Cologne: <https://dch.phil-fak.uni-koeln.de/>.
16. Lots of Copies Keep Stuff Safe: <https://www.lockss.org/>.
17. Available from: <https://sigidoc.cceh.uni-koeln.de>.
18. We are aware that this name is misleading, SigiDoc being a data-model, not a search engine: we decided to adopt it because the name ‘SigiDoc’ is already well established among sigillographers and digital humanists.
19. An overview of its structure can be seen at <https://github.com/SigiDoc>.
20. We borrow, modifying it, the expression ‘Inscriptiones Græcæ Ævi Byzantini’ found in [Rhoby \(2015: 29\)](#). In the field of numismatics, similar endeavours include projects such as *Online Coinage of the Roman Empire* (<https://numismatics.org/ocre/>) and *Seleucid Coins Online* (<https://numismatics.org/sco/>), which offer extensive digital collections of coins for scholars and enthusiasts alike. Although they are based on different technologies and starting points, they share the same goal as our project: making these resources—be they seals or coins—widely accessible and enhancing their study.
21. For an overview of the various activities, see <https://www.doaks.org/resources/educational-resources/lasting-impressions-module/>.
22. More information available from <https://uni.koeln/NXPQU>.
23. The SigiDoc training weeks follow the well-established example of the EpiDoc training weeks, which have been successfully conducted over the past decade. A comprehensive analysis of the theoretical and practical aspects of the pedagogy behind the EpiDoc training weeks can be found in [Bodard and Stoyanova \(2016\)](#) and [Bodard and Vagionakis \(2022\)](#).
24. Dumbarton Oaks, 23 November 2021 to 4 December 2022. Available from: <https://www.doaks.org/resources/online-exhibits/lasting-impressions>.
25. Available from: <https://byzantine-seals-exhibition.dibs.uni-koeln.de>.

References

- [Alexander, G. \(2020\)](#) ‘Mining Digital Humanities Archives to Produce Accessible Material Editions of Scholarly Texts’, *IDEAH*, 1(1). <https://doi.org/10.21428/f1f23564.4e680428>.
- [Bodard, G. and Stoyanova, S. \(2016\)](#) ‘Epigraphers and Encoders: Strategies for Teaching and Learning Digital Epigraphy’, in G. Bodard and M.

- Romanello (eds) *Digital Classics Outside the Echo-Chamber*. London: Ubiquity Press, pp. 51–68. <https://doi.org/10.5334/bat.d>.
- [Bodard, G. and Vagionakis, I. \(2022\)](#), ‘EpiDoc and Epigraphic Training in the Era of Remote and Hybrid Teaching’, *Digital Classics Online*, 8, pp. 106–121. <https://doi.org/10.11588/dco.2022.8.90358>.
- [Cheynet, J.-C. \(2008\)](#), ‘Introduction à la sigillographie byzantine’, in J.-C. Cheynet (ed.) *La société byzantine. L’apport des sceaux*, 1. Paris: Association des amis du Centre d’histoire et civilisation de Byzance, pp. 1–82.
- [Drach, S., Filosa, M., Malatras, C. and Sode, C. \(2024\)](#), ‘Teaching Digital Byzantine Sigillography: First Experiences and Future Strategies’, in M. Filosa, C. Neuefeind and C. Sode (eds) *Digital Approaches to Medieval Sigillography. Digital Medievalist*, 17(1), pp. 1–25. <https://journals.openedition.org/digitalmedievalist/391>.
- [Filosa, M., Bodard, G. and Gad, U. \(2023\)](#), ‘Description, Translation and Process: Making the Implicit Explicit in Digital Editions of Ancient Text-Bearing Objects’, in G. Bodard and C. Palladino (eds) *Can’t Touch This: Digital Approaches to Materiality in Cultural Heritage*. London: Ubiquity Press, pp. 51–75. <https://doi.org/10.5334/bcv.d>.
- [Jeffreys, M. \(2009\)](#), ‘The Seals Module of the Prosopography of the Byzantine World’, *Byzantinoslavica*, 67, pp. 17–23.
- [Kalvesmaki, J. \(2015\)](#), ‘Introducing Athena Ruby, Dumbarton Oaks’ New Font for Byzantine Inscriptions’, in A. Rhoby (ed.) *Inscriptions in Byzantium and Beyond: Methods – Projects – Case Studies*. Vienna: Österreichische Akademie der Wissenschaften, pp. 121–6.
- [Laurent, V. \(1965–1981\)](#), *Le Corpus des sceaux de l’empire byzantin*. Paris: Centre National de la Recherche Scientifique.
- [Oikonomides, N. \(1986\)](#), *A Collection of Dated Byzantine Lead Seals*. Washington, DC: Dumbarton Oaks Research Library and Collection.
- [Oikonomides, N., Nesbitt, J., McGeer, E., Cotsonis, J. and Seibt, W. \(1991–2026\)](#), *Catalogue of Byzantine Seals at Dumbarton Oaks and in the Fogg Museum of Art*, 1–8. Washington, DC: Dumbarton Oaks Research Library and Collection.
- [Prigent, V. \(2017\)](#), ‘De l’apport des “petits monuments”: faire de l’or avec du plomb’, in M.-H. Blanchet and I.-A. Tudorie (eds) *L’apport des*

Assomptionnistes français aux études byzantines: une approche critique.
Paris: De Boccard, pp. 386–452.

Rhoby, A. (2015), ‘A Short History of Byzantine Epigraphy’, in A. Rhoby (ed.) *Inscriptions in Byzantium and Beyond: Methods – Projects – Case Studies*. Vienna: Österreichische Akademie der Wissenschaften, pp. 17–30. <https://doi.org/10.2307/j.ctv8d5td5.6>.

Sode, C. (2001), ‘Andreas David Mordtman mladšij (1837–1912) i načalo vizantijskoj sigilografii’, *Vizantijskij Vremennik*, 85, pp. 178–83.

Sopracasa, A., Bigalke, J., Buchs, N. and Meziridou, S. (2024), ‘Creating a Sigillographic Search Engine for Byzantium: Preliminary Results’, in M. Filosa, C. Neufeind and C. Sode (eds) *Digital Approaches to Medieval Sigillography. Digital Medievalist*, 17(1), pp. 1–32. <https://journals.openedition.org/digitalmedievalist/399>.

Trécourt-Codine, F. and Sarah, G. (2012), ‘Du plomb au pixel. Transcrire les légendes des monnaies du haut Moyen Âge’, *Revue Numismatique*, 169, pp. 261–277. <https://doi.org/10.3406/numi.2012.3183>.

Tsougarakis, D. (1999), *Εισαγωγή στη βυζαντινή σφραγιδογραφία*. Athens: Kanakis.

Turska, M., Cummings, J. and Rahtz, S. (2016–2017), ‘Challenging the Myth of Presentation in Digital Editions’, *Journal of the Text Encoding Initiative*, 9. <https://doi.org/10.4000/jtei.1453>.

Wallace, A. (2021), ‘Accessibility and Open GLAM’, in J. McCutcheon and A. Ramalho (eds) *International Perspectives on Disability Exceptions in Copyright and the Visual Arts: Feeling Art*. Abingdon and New York: Routledge, pp. 24–43. <https://doi.org/10.2139/ssrn.3615749>.

Zacos, G. (1984), *Byzantine Lead Seals, 2, Compiled and Edited by J. Nesbitt*. Bern: Benteli.

Zacos, G. and Veglery, A. (1972), *Byzantine Lead Seals, 1*. Basel: J.J. Augustin.

Index

accessibility, [7](#), [15](#), [62](#), [85–6](#), [148](#), [169](#), [186](#), [190–91](#), [229](#)

artificial intelligence (AI), [138](#), [169](#)

ancient and medieval languages

Arabic, [133](#), [197](#)

Aramaic, [192](#)

Armenian, [22](#)

Greek, [32–3](#), [66](#), [154–6](#), [169–71](#), [197](#), [210–11](#)

Latin, [65](#), [166](#)

annotation, [112](#), [135–6](#), [148–150](#), [165](#), [171–9](#), [195](#), [209–11](#), [217–8](#)

British International Research Institutes (BIRI)

British Institute at Ankara (BIAA), [84–5](#), [91](#), [94–5](#)

British Institute of Libyan and Northern African Studies (BILNAS), [84](#), [86](#), [93–5](#), [98n6](#), [130–31](#), [138–9](#), [141n23](#)

British School at Athens (BSA), [84](#), [92](#), [98](#)

British School at Rome (BSR), [84–6](#), [89–91](#), [97](#), [133](#), [138](#)

Canonical Text Services (CTS), [150–51](#), [160n9](#), [169–71](#)

CARE principles, [5](#), [7–8](#), [131](#)

community responsibility, [192–4](#), [197–8](#)

Creative Commons licences, [5](#), [97](#), [140](#), [199](#), [220](#)

cultural heritage collections

BILNAS Archive, [86](#), [92–5](#), [130](#)

Dumbarton Oaks, [233–6](#)

IRT Photographic Collection, [86–9](#)

Manar al Athar Digital Archive, [91](#), [133](#)

Yale University Art Gallery, [188](#), [196–7](#)

data formats

CSV, [6–7](#), [220](#)

database, [20](#), [32](#), [44](#), [47](#), [93](#), [178](#), [189–90](#), [236](#)

GIS, [107](#), [117–19](#), [135](#)

RDF, [7](#), [46](#), [177](#)

txt, [3](#)

XML, [6](#), [50](#), [55](#), [126](#), [158](#), [175](#), [209](#), [230](#), [235](#), [238–40](#)

EpiDoc, [2](#), [14](#), [126](#), [209](#), [228](#). *See also* [Kiln-based tools → EFES](#)

See also SigiDoc, [227–47](#) *passim*

epigraphic digital corpora

Electronic Archive of Greek and Latin Epigraphy (EAGLE), [209](#)

Epigraphic Database Heidelberg (EDH), [45–6](#)

Inscriptions of Aphrodisias (InsAph), [102](#), [141n16](#)

Inscriptions of Greek Cyrenaica (IGCyr), [222n22](#)
 Inscriptions of Roman Cyrenaica (IRCyr), [125–7](#), [137](#)
 Inscriptions of Roman Tripolitania (IRT), [86–91](#), [125–7](#)
 Inscriptions of Sicily (I.Sicily), [129](#)
 Packard Humanities Institute Greek Epigraphy (PHI), [209](#)
 factoid prosopography, [20–22](#), [44](#), [128](#)
 FAIR principles, [7](#), [75](#), [85](#), [190–91](#), [200n10](#)
 findability, [7](#), [54](#), [75](#), [85](#), [189](#), [236](#)
 gazetteers
 GeoNames, [116](#), [126–9](#), [132–3](#), [233](#)
 Heritage Gazetteer of Cyprus (HGC), [14](#), [106–7](#), [127](#)
 Heritage Gazetteer of Libya (HGL), [14](#), [85](#), [127–9](#)
 Pleiades, [14](#), [54](#), [88](#), [106](#), [126–8](#), [132–3](#), [177](#), [192](#), [233](#)
 World Historical Gazetteer (WHG), [100n39](#), [132](#), [137](#)
 interoperability, [4](#), [7](#), [92](#), [179](#), [231](#), [241](#)
 Kiln-based tools
 EpiDoc Front-End Services (EFES), [233](#), [238–40](#)
 Kiln for Treebanking (KTB), [209–10](#), [221n13](#)
 Leiden conventions, [2–4](#), [231](#)
 linguistics, [149–51](#), [169](#), [174](#), [210–11](#), [214–17](#)
 linked open data (LOD), [6–7](#), [21–3](#), [33–5](#), [45–6](#), [85](#), [121](#), [167](#), [190–6](#), [209](#), [231](#). *See also* [ontologies](#)
 metadata, [6–7](#), [26](#), [31](#), [85](#), [90–1](#), [97](#), [131–3](#), [166–7](#), [176–7](#), [189–190](#), [193–5](#), [209](#), [233](#)
 named entities and NER, [158](#), [171–6](#), [232](#)
 natural language processing (NLP), [170–2](#), [149](#)
 ontologies, [6](#), [21–4](#), [45](#), [118](#)
 CIDOC-CRM, [23–6](#), [33](#), [45](#)
 PeriodO, [141n15](#), [192](#)
 SNAP:DRGN, [45](#), [92](#), [209](#)
 See also [Wikidata](#)
 optical character recognition (OCR), [6](#), [47–8](#), [148–150](#), [165](#)
 paradata, [7](#), [214](#)
 Pelagios Network, [46](#), [54](#), [122n3](#), [132–3](#)
 Peripleo, [122n3](#), [132](#),
 prosopography, [3](#), [13](#), [19–20](#), [42–6](#), [61–4](#), [91–2](#), [126](#), [171](#), [234–5](#)
 Prosopographia Imperii Romani, [42](#), [45](#), [221n10](#)
 Prosopography of the Byzantine World, [19–40](#) *passim*, [43–4](#), [75](#), [234](#)
 Prosopography of the Later Roman Empire, [36n2](#), [41–60](#) *passim*, [75](#), [234](#)
 See also [factoid prosopography](#)
 See also [ontologies](#) → [SNAP:DRGN](#)
 Roueché, Charlotte M., [25](#), [56](#), [76](#), [84](#), [97](#), [121](#), [127](#), [140](#), [228](#), [234](#)
 sustainability, [15](#), [53–4](#), [169–71](#), [237–40](#)
 Text Encoding Initiative (TEI), [6](#), [47–9](#), [55](#), [111](#), [158–9](#), [244n2](#)
 See also [EpiDoc](#)

teaching and pedagogy, [133–6](#), [165](#), [241–3](#), [245](#)

training and workshops, [52](#), [92–3](#), [95](#), [100](#), [137–9](#), [142](#), [200](#), [242](#), [245](#)

translation and transliteration, [120](#), [166](#), [175](#), [193](#), [209](#), [229](#), [242](#)

vocabularies. *See also* [ontologies](#)

EAGLE vocabularies, [209](#)

FAIR Epigraphy, [200n10](#)

FOAF, [53](#)

Getty, [133](#), [192](#)

VIAF, [93](#)

volunteer, community and student work, [48](#), [134–6](#), [141n29](#), [209](#), [220](#), [241–2](#)

Wikidata, [12](#), [91](#), [94–5](#), [137](#), [176–9](#), [193–6](#), [209](#)

Overview: A hierarchical tree diagram representing two versions of the syntactic structure of a sentence. The trees begin at the top with the root node, branching downwards into multiple syntactic categories and words.

Detailed structure of the left tree (Bodard et al.):

- Top level:
 - The root node connects to PRED (main predicate) with the word form πυνθανόμαν.
 - Below that, the object node OBJ contains an artificial token marked as [1].
- Branches from OBJ:
 - SBJ (subject) Μέμνονα, which governs
 - ATR (attribute) Αιγύπτιον.
 - ADV (adverbial) αιθόμενον, which governs two branches; see below.
 - OBJ (object) φώνην.
 - AuxP (adposition) ’πυ, which governs
 - ADV (adverbial) λίθω, which governs
 - ATR (attribute) Θηβαϊκω.
- Two further expansions from ADV αιθόμενον:
 - OBJ (object) αὔγαι, which governs:
 - ATR (attribute) ἄλιω.
 - AuxK (auxiliary punctuation) comma (,).

Detailed structure of the tree on the right side (Vierros)

- Top level:
 - The root node connects to PRED (main predicate) with the word form πυνθανόμαν.
 - Below that, the object node OBJ contains φώνην.
- The OBJ node divides into two branches:
 - the first branch is headed by SBJ (subject) Μέμνονα, which governs
 - ATR (attribute) Αιγύπτιον.
 - ATR (attribute) αιθόμενον, which governs:
 - ADV (adverbial) αὔγαι, which governs:
 - ATR (attribute) ἄλιω.
 - the second branch is headed by AuxP (adposition) πυ, which governs
 - ADV (adverbial) λίθω, which governs

- ATR (attribute) Θηβαῖχῶ.

[Return to text](#)