

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Investigating Linked Data Usability for Ancient World Research

### Thesis

#### How to cite:

Middle, Sarah (2022). Investigating Linked Data Usability for Ancient World Research. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2022 Sarah Middle



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.00014b1f>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# Investigating Linked Data usability for Ancient World research

Thesis submitted for the degree of Doctor of Philosophy  
in the discipline of Classical Studies

by Sarah Middle

Faculty of Arts and Social Sciences, The Open University

October 2021



## Abstract

Linked Data technologies are used to describe and connect entities, based on features they have in common. Rich semantic descriptions, disambiguation capabilities, and interoperability allow investigation of new research questions and reveal previously undiscovered relationships. However, previous studies have shown that uptake of Linked Data among Humanities researchers has, thus far, been low, partly due to usability issues with the resulting tools and resources. I therefore set out to investigate how their usability might be improved, and how Linked Data technologies might most effectively be integrated with existing research methods. My study focused on the Ancient World, where Linked Data implementation seems to be higher than in other Humanities disciplines, and involved a survey and interviews to elicit user and producer needs from researchers in this subject area.

I start this thesis by introducing and contextualising my research topic in Chapter 1. In Chapter 2, I consult existing literature and datasets to discuss Linked Humanities Data implementation, its advantages, and current barriers. Chapter 3 provides an outline of my survey and interview methodologies, while Chapter 4 presents initial survey analysis and identifies themes for discussion in the subsequent chapters. Chapter 5 focuses on five research methods already embedded in the practices of Ancient World researchers, where Linked Data could effectively be integrated: *Discovering*, *Gathering*, *Data Recognition*, *Annotating*, and *Visualization*. In Chapter 6, I explore the user experience more broadly, including aspects such as interface design, reliability, and data quality. Chapter 7 then discusses areas of the production process that affect Linked Data usability: training, collaboration, user-centred design, documentation, access, and sustainability. My findings form the basis of a series of recommendations in Chapter 8, which focus on teamwork, openness and transparency, extensibility, user consultation, discoverability, sustainability, and communities, culminating in a *Five-Star Model for Linked Humanities Data Usability*.



*To Chris, who supported me from the start  
and Elise, who joined us partway through*



## Acknowledgements

My PhD centres on the importance of community in creating and maintaining digital tools and resources, but community was also vital to the PhD itself...

I'd first like to thank my team of supervisors: Elton Barker, for taking my vaguest notion of a topic and helping me shape it into an actual PhD; Phil Perkins, for providing great ideas, improving my writing, and helping me navigate the murky waters of OU admin, and Alessandro Adamou and Mathieu D'Aquin, for showing me what producing Linked Data really means and continuing to support me despite multiple job and country moves. Each of you brought something unique to this PhD and I'm glad you all joined me on this journey.

Studying this PhD was made possible by receiving AHRC funding through the CHASE DTP. Many thanks are due to CHASE, not only for awarding me this grant, but also providing brilliant training and placement opportunities, as well as a fantastic community. I'd particularly like to thank Rob Witts, Steve Colburn and Clare Hunt for their support.

At the Open University, I'd also like to thank Sara Haslam, Yvonne Bartley, Caitlin Adams and Tracey Debeer for providing the administrative support to help me get the most out of the amazing opportunities I've had during my PhD.

During my PhD, I was lucky enough to secure a place on the Getty-funded Institute in Ancient Itineraries, where I spent four weeks with 20 amazing colleagues (and friends!). I learned so much from all of you – our discussions really helped shape my thinking and opened my mind to different practical and theoretical approaches. Special mentions go to the 'parents' (Stuart Dunn, Graeme Earl, Anna Foka, and Will Wootton), my fellow 'space cadets' (Alex Butterworth, Ryan Horne, David McMeekin, and Chiara Zuanni), and my blog post co-authors, Steph Grimes and Rebecca Levitan.



A massive thank you to everyone in the Linked Ancient World Data and Linked Pasts communities for all the insights I gained from conversations, presentations and publications, as well as for being so friendly, welcoming and ready to listen to a more junior colleague. I'd particularly like to thank my LAWD Catalogue collaborators Gabriel Bodard (who also very kindly pilot tested my survey) and Paula Granados García, as well as Leif Isaksen, the Registry Activity, and the rest of the Pelagios team.

Of course, my thesis would not exist at all without my 212 anonymous participants, particularly the 16 who took part in the interview stage. Thank you all for giving up your time – your insights were absolutely invaluable to me and each one of you made my thesis stronger.

Doing a PhD remotely can be a lonely experience, particularly during a global pandemic, but luckily this was made more enjoyable by my fellow students, who provided daily accountability checks, virtual writing sessions, and space to share each other's highs and lows. A particular thank you to Ashley, Ben, Carmen, Jack, Kat, Louise, Lucinda, Sam, Sophie D and Sophie M – you are all awesome!

It's not *all* about the PhD itself though. So many people outside the academic bubble made it possible for me to study for a PhD without losing myself in the process. Huge thanks to my daughter's nursery; my local parents' group; my theatre group (particularly Hugh, who provided so much encouragement and is now sadly missed); my closest friend Vicki, my mum Fiona (who also assisted with proofreading and transcription) and my dad Jonathan. Last, but by no means least, a heartfelt thank you to Chris and Elise for your constant love and support.

## Table of Contents

<b>Abstract.....</b>	<b>3</b>
<b>Acknowledgements .....</b>	<b>7</b>
<b>Table of Contents.....</b>	<b>9</b>
<b>List of Figures .....</b>	<b>13</b>
<b>List of Tables.....</b>	<b>16</b>
<b>List of Abbreviations .....</b>	<b>17</b>
<b>1 Introduction .....</b>	<b>21</b>
1.1 Hypothesis and Research Questions .....	22
1.2 Humanities Data Approaches: An Overview .....	25
1.3 Linked Data.....	27
1.4 Usability .....	33
1.5 Thesis Structure .....	34
<b>2 Linked Data and Humanities Research .....</b>	<b>36</b>
2.1 Extent of Linked Data in Humanities Research .....	36
2.1.1 Case Study: AHRC-Funded Linked Data Projects .....	37
2.1.2 Case Study: Digital Classicist Wiki.....	41
2.1.3 Summary: Extent of Linked Data in Humanities Research .....	45
2.2 Linked Ancient World Data in Practice .....	45
2.2.1 Place .....	46
2.2.2 Time.....	50
2.2.3 People.....	51
2.2.4 Objects.....	55
2.2.5 Summary: Linked Humanities Data in Practice .....	61
2.3 Advantages of Linked Humanities Data .....	61
2.4 Barriers to Linked Humanities Data.....	66
2.5 Conclusions: Linked Data and Humanities Research .....	74
<b>3 Methodology .....</b>	<b>76</b>
3.1 Describing Digital Humanities Research Methods.....	76

3.2	User Research in the Digital Humanities .....	80
3.3	Survey .....	84
3.4	Demographics.....	90
3.5	Interviews.....	101
3.6	Conclusions: Methodology .....	110
4	<b><i>Tools, Resources and Methods for Ancient World Research .....</i></b>	<b>111</b>
4.1	<b>Tools and Resources.....</b>	<b>111</b>
4.1.1	Using Digital Tools and Resources .....	111
4.1.2	Using Linked Data Tools and Resources.....	123
4.1.3	Producing Digital Tools and Resources .....	134
4.1.4	Summary: Tools and Resources .....	143
4.2	<b>Research Methods .....</b>	<b>145</b>
4.3	<b>Conclusions: Tools, Resources and Methods for Ancient World Research.....</b>	<b>149</b>
5	<b><i>Research Methods for Linked Data Integration.....</i></b>	<b>151</b>
5.1	<b>Discovering .....</b>	<b>151</b>
5.1.1	Search .....	152
5.1.2	Exploration.....	157
5.1.3	Discovering Digital Tools and Resources.....	161
5.1.4	Documentation for Discovery .....	163
5.1.5	Summary: Discovering .....	164
5.2	<b>Gathering .....</b>	<b>165</b>
5.3	<b>Data Recognition .....</b>	<b>168</b>
5.4	<b>Annotating.....</b>	<b>170</b>
5.5	<b>Visualization .....</b>	<b>174</b>
5.6	<b>Case Studies: Combining Research Methods .....</b>	<b>178</b>
5.6.1	Case Study 1: Discovering – Data Recognition – Annotating –Gathering .....	179
5.6.2	Case Study 2: Annotating – Visualization – Discovering .....	180
5.7	<b>Conclusions: Research Methods for Linked Data Integration .....</b>	<b>181</b>
6	<b><i>User Engagement with Linked Data .....</i></b>	<b>183</b>
6.1	<b>Tool and Resource Usability .....</b>	<b>183</b>
6.1.1	User Interfaces.....	184
6.1.2	Reliability .....	188

<b>6.2</b>	<b>Data Usability .....</b>	<b>192</b>
6.2.1	Data Modelling .....	192
6.2.2	Data Quality.....	198
6.2.3	Data Sharing .....	206
<b>6.3</b>	<b>Conclusions: User Engagement with Linked Data .....</b>	<b>209</b>
<b>7</b>	<b><i>Producing Usable Tools and Resources .....</i></b>	<b>211</b>
<b>7.1</b>	<b>Training .....</b>	<b>211</b>
<b>7.2</b>	<b>Collaboration .....</b>	<b>214</b>
<b>7.3</b>	<b>User-Centred Design .....</b>	<b>219</b>
<b>7.4</b>	<b>Documentation .....</b>	<b>222</b>
<b>7.5</b>	<b>Access and Cost.....</b>	<b>226</b>
<b>7.6</b>	<b>Sustainability .....</b>	<b>228</b>
<b>7.7</b>	<b>Conclusions: Producing Usable Tools and Resources.....</b>	<b>234</b>
<b>8</b>	<b><i>Conclusions and Recommendations .....</i></b>	<b>237</b>
<b>8.1</b>	<b>Effective Teams.....</b>	<b>240</b>
<b>8.2</b>	<b>Openness and Transparency.....</b>	<b>244</b>
<b>8.3</b>	<b>Extensibility .....</b>	<b>247</b>
<b>8.4</b>	<b>User Consultation .....</b>	<b>252</b>
<b>8.5</b>	<b>Facilitating Discovery .....</b>	<b>255</b>
<b>8.6</b>	<b>Sustainability, Preservation and Funding .....</b>	<b>257</b>
<b>8.7</b>	<b>Communities of Practice.....</b>	<b>264</b>
<b>8.8</b>	<b>Five-Star Model for Linked Humanities Data Usability .....</b>	<b>267</b>
<b>8.9</b>	<b>Future Work .....</b>	<b>269</b>
8.9.1	Resource Directories .....	269
8.9.2	Digital Annotation Research and Development .....	273
<b>8.10</b>	<b>Conclusions.....</b>	<b>275</b>

<b>References .....</b>	<b>277</b>
<b>Appendix 1. Survey: Using Digital Tools and Resources for Ancient World Research .....</b>	<b>308</b>
<b>Appendix 2. Information Leaflet for Interview Participants .....</b>	<b>320</b>
<b>Appendix 3. Consent Form for Interview Participants .....</b>	<b>322</b>
<b>Appendix 4. Interview Script.....</b>	<b>324</b>
<b>Appendix 5. Documentation Checklist.....</b>	<b>341</b>

## List of Figures

Figure 2.1 Number of ongoing AHRC-funded Linked Data projects over time.....	38
Figure 2.2 Number of projects in each category based on their relationship to Linked Data .....	39
Figure 2.3 Pie chart showing distribution of main subject areas among AHRC-funded Linked Data projects (all subjects with two or more projects are shown; 'Cultural Heritage' comprises projects that relate to libraries, museums or archives without a specific subject topic; 'Other' comprises subjects with only one project) .....	40
Figure 3.1 The Methodological Commons (McCarty & Short, 2002).....	78
Figure 3.2 Pie chart showing participant location by country .....	91
Figure 3.3 Pie chart showing participant roles .....	91
Figure 3.4 Pie chart showing participant age ranges .....	92
Figure 3.5 Pie chart showing participant gender .....	93
Figure 3.6 Pie chart showing participants' answers to the question "Have you ever knowingly used one or more tools/resources that are based on Linked Data?" .....	94
Figure 3.7 Pie chart showing participants' answers to the question "Are you familiar with the term 'Linked Data'?" .....	94
Figure 3.8 Box plot showing distribution of overall digital competence/confidence scores .....	96
Figure 3.9 Box plot showing distribution of digital competence/confidence scores for each statement .....	96
Figure 3.10 Box plots showing distribution of mean digital competence/confidence scores for Linked Data users and non-Linked Data users .....	97
Figure 3.11 Box plots showing distribution of digital competence/confidence scores for each statement for Linked Data users (left) and non-Linked Data users (right) .....	98
Figure 3.12 Pie chart showing whether participants have been involved in the production of a digital tool/resource .....	99
Figure 3.13 Pie chart showing whether producers have been involved in the production of a Linked Data tool/resource .....	100
Figure 3.14 Bar chart showing producers segmented by their use and production of Linked Data.....	100

Figure 3.15 Demographic characteristics of interview participants (a. Linked Data use/production group, b. location, c. role, d. age, e. gender, f. digital competence/confidence).....	104
Figure 4.1 Pie chart showing the number and percentage of digital tools/resources of each type mentioned by participants .....	113
Figure 4.2 Bar chart showing digital Ancient World tools/resources with which participants are most familiar (all tools/resources selected by two or more participants are shown) .....	114
Figure 4.3 Pie charts showing whether users of particular resources (a. Papyri.info, including Papyrological Navigator; b. Perseus) had knowingly used Linked Data .....	117
Figure 4.4 Bar chart showing features of a good digital tool/resource identified by survey participants, segmented by participants' Linked Data usage .....	119
Figure 4.5 Features of digital tools/resources in order of importance as ranked by interview participants: a. non-Linked Data users; b. Linked Data users (8 participants in each group) .....	120
Figure 4.6 Bar chart showing barriers to digital tool/resource use identified by survey participants, segmented by participants' Linked Data usage .....	121
Figure 4.7 Barriers to digital tools/resources in order of the extent to which they affect interview participants: a. non-Linked Data users; b. Linked Data users .....	122
Figure 4.8 Bar chart showing all named Linked Ancient World Data tools/resources mentioned by Linked Data users .....	125
Figure 4.9 Bar chart showing Linked Data tools and resources used most frequently by Linked Data users .....	126
Figure 4.10 Bar chart showing advantages of Linked Data identified by participants from the information about specific tools/resources .....	127
Figure 4.11 Features of Linked Data tools/resources in order of importance as ranked by 7 non-Linked Data user interview participants .....	130
Figure 4.12 Bar chart showing producers' preferred data structures, segmented by their answer to the question of whether they have produced a Linked Data tool/resource .....	135
Figure 4.13 Bar chart showing reasons why producers had not chosen a Linked Data approach .....	138

Figure 4.14 Bar chart showing the number of participants who selected each audience group in relation to the Linked Data tool/resource they most recently produced .....	139
Figure 4.15 Pie chart showing the number of audience groups for which each tool/resource was intended.....	139
Figure 4.16 Bar chart showing the number of participants who selected each of Berners-Lee's five stars in relation to the Linked Data tool/resource they produced most recently .....	141
Figure 4.17 Bar chart showing the number of participants whose tool/resource actually meets the criteria for each of Berners-Lee's five stars.....	142
Figure 4.18 Bar chart showing research activities participants associated with their use of digital tools/resources .....	145
Figure 4.19 Bar chart showing research activities participants associated with their use of Linked Data tools/resources .....	147
Figure 4.20 Bar chart showing research activities participants associated with production of Linked Data tools/resources .....	148
Figure 7.1 Pie charts showing producer demographics (a. location, b. role, c. age, d. gender) .....	215



## List of Tables

Table 2.1 Distribution of 'Linked open data' category pages in Digital Classicist wiki, based on other broad categories to which they belong; some pages belong to more than one category .....	42
Table 2.2 Number of pages in the Digital Classicist wiki containing particular keywords, which do not already appear in the 'Linked open data' category (no such pages were categorised as 'Events') .....	44
Table 3.1 Criteria for each participant group, used for stratified sampling .....	102
Table 4.1 Linked Data "specific features" as included in the follow-up survey, with their relationships to advantages mentioned in the original survey .....	129
Table 4.2 Positive comments from Linked Data users when comparing their chosen Linked Data tool/resource with other digital tools/resources .....	131
Table 4.3 Negative comments from Linked Data users when comparing their chosen Linked Data tool/resource with other digital tools/resources (*participant did not answer Q14 but included these resources in their response to Q20) .....	131
Table 4.4 Improvements to Linked Data tools/resources suggested by Linked Data users (*participant did not answer Q14 but included these resources in their response to Q20; †one of the participants who mentioned this improvement did not name a tool or resource in Q14) .....	133
Table 4.5 Reasons participants provided for preferring particular ways of structuring data (NB: some reasons were provided in relation to more than one type of data structure, so the totals of the numbers in the 'Data Structure(s)' column sometimes exceed the 'Number of Participants') .....	136
Table 4.6 Frequency of selected audience group combinations for which Linked Data tools/resources were intended.....	140

## List of Abbreviations

ADS	Archaeology Data Service
AHDS	Arts and Humanities Data Service
AHRC	Arts and Humanities Research Council
AIP	Association Internationale de Papyrologues
ANS	American Numismatic Society
API	Application Programming Interface
ARIADNE	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe
ASP	American Society of Papyrologists
BRAT	BRAT Rapid Annotation Tool
CDLI	Cuneiform Digital Library Initiative
CHRR	Coin Hoards of the Roman Republic
CIDOC CRM	International Committee for Documentation Conceptual Reference Model
CLTK	Classical Language Toolkit
CMS	Content Management System
CRRO	Coinage of the Roman Republic Online
CSV	Comma Separated Values
CTS	Canonical Text Services
DARE	Digital Atlas of the Roman Empire
DARIAH	Digital Research Infrastructure for Arts and Humanities
DARIAH-DE	Digital Research Infrastructure for Arts and Humanities Germany
DCAT	Data Catalog Vocabulary
DH	Digital Humanities
DHOxSS	Digital Humanities at Oxford Summer School
DHSI	Digital Humanities Summer Institute
DiRT	Digital Research Tools
DOI	Digital Object Identifier
EACM	Epigraphy Aggregation Conceptual Model
EAGLE	Europeana Network for Ancient Greek and Latin Epigraphy
ECHOES	Empowering Communities with a Heritage Open Ecosystem
EDB	Epigraphic Database Bari

EDH	Epigraphische Datenbank Heidelberg
ERC	European Research Council
FAIMS	Federated Archaeological Information Management Systems
FAIR	Findable, Accessible, Interoperable, Reusable
FAQ	Frequently Asked Questions
FOAF	Friend Of A Friend
GeoJSON	Geographic JavaScript Object Notation
GIS	Geographic Information Systems
GODOT	Graph Of Dated Objects and Texts
GtR	Gateway to Research
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ICT	Information and Communications Technology
IDEs	Integrating Digital Epigraphies
IIIF	International Image Interoperability Framework
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
JSTOR	Journal Storage
KOS	Knowledge Organisation System
LAWD	Linked Ancient World Data
LAWDI	Linked Ancient World Data Institute
LOD	Linked Open Data
LOUD	Linked Open Usable Data
LSJ	Liddell-Scott-Jones lexicon
MANTIS	MANTIS: A Numismatic Technologies Integration Service
MANTO	Modelling Ancient Narratives, Territories, Objects
MOOC	Massive Open Online Course
NEH	National Endowment for the Humanities
NER	Named Entity Recognition
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
OCRE	Online Coins of the Roman Empire
OLD	Oxford Latin Dictionary

OLH	Open Library of Humanities
OLiA	Ontologies of Linguistic Annotation
ORBIS	ORBIS: The Stanford Geospatial Network Model of the Roman World
ORCID	Open Researcher and Contributor ID
PART	Participant
PAS	Portable Antiquities Scheme
PDF	Portable Document Format
PeriodO	Periods, Organized
PHI	Packard Humanities Institute
Q	Question
QGIS	Quantum Geographic Information System
RDF	Resource Description Framework
RQ	Research Question
SAWS	Sharing Ancient Wisdoms
SCOTCH	Semantic Collaborative Ontology for Three-dimensional visualisation of Cultural Heritage
SKOS	Simple Knowledge Organization System
SNAP:DRGN	Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names
SPARQL	SPARQL Protocol And RDF Query Language
SRAM	Scholarly Research Activity Model
STELLAR	Semantic Technologies Enhancing Links and Linked data for Archaeological Resources
SUS	System Usability Scale
TaDiRAH	Taxonomy for Digital Research Activities in the Humanities
TAPoR	Text Analysis Portal for Research
TEI	Text Encoding Initiative
TLG	Thesaurus Linguae Graecae
TM	Trismegistos
UKRI	UK Research and Innovation
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

VIAF	Virtual International Authority File
VoID	Vocabulary of Interlinked Datasets
W3C	World Wide Web Consortium
WAI	Web Accessibility Initiative
WHG	World-Historical Gazetteer
WissKI	Wissenschaftliche Kommunikationsinfrastruktur
XML	EXtensible Markup Language

## 1 Introduction

Of the various approaches available for modelling and publishing Humanities data, Linked Data is perhaps the most effective for representing its complexity and nuance, while also facilitating discoverability and reuse. The term ‘Linked Data’ refers to a set of technologies that can be used to describe entities, such as places, people, or objects, and connect them based on features they have in common. Its rich semantic descriptions, disambiguation capabilities, and interoperability can unlock opportunities to address new research questions and reveal previously undiscovered relationships between entities. However, uptake of Linked Data among Humanities researchers has, thus far, been low. While there are various factors that might contribute to this situation (to be explored in Chapter 2), a key reason is likely to be usability issues with existing tools<sup>1</sup> and resources<sup>2</sup>, particularly by researchers with minimal levels of technical skill.

This thesis aims to investigate the issue of Linked Humanities Data usability by focusing on the Ancient World<sup>3</sup>, a subject domain where Linked Data implementation is relatively mature. Study of the Ancient World encompasses multiple disciplines, including Archaeology, Art, History, Literature and Philosophy, and can therefore be considered a microcosm of the Humanities. It therefore provides an excellent case study for Linked Humanities Data usability more generally. As there is currently little information about the use of Linked Ancient World Data tools and resources outside the projects that produced them, my research sought to establish user and producer needs through a survey of the research community and a series of detailed interviews with selected participants. Discussion of their experiences informed a series of recommendations for producers, funders and institutions, to improve future Linked

---

<sup>1</sup> In the context of this thesis, digital ‘tools’ are defined as software that enables the user to carry out a specific function relating to a digital resource (such tools may be online or installed on the user’s computer).

<sup>2</sup> In the context of this thesis, digital resources are defined as any material that can be consumed in an electronic format, including digitised or born-digital texts, images, or artefacts, as well as websites, databases, catalogues, and interactive visualisations.

<sup>3</sup> In the context of this thesis, ‘Ancient World’ is defined as any civilisations existing prior to the end of the Western Roman Empire in 476 CE. Responses to my survey spanned from Mycenaean to Byzantine (approximately 1600 BCE to 1500 CE), with most participants primarily interested in the Classical Greek and/or Roman worlds (approximately 400 BCE to 400 CE).

Ancient World Data usability and facilitate its integration with existing research methods.

In this chapter, I will introduce my research topic by presenting a hypothesis and using it to develop a series of research questions (1.1). To contextualise the subject of Linked Humanities Data, I will give an overview of current modelling and sharing approaches relating to Humanities Data more broadly (1.2), before providing an explanation of Linked Data itself (1.3). I will then turn to another key element of my research topic, that of usability (1.4), before presenting my thesis structure (1.5).

## 1.1 Hypothesis and Research Questions

My research starts with the hypothesis that Linked Data principles and technologies have not yet been sufficiently adopted in Humanities research to achieve their potential benefits, because the conceptual and technological distances from existing research methods are too great. Within this hypothesis are three assumptions:

- i. Linked Data has potential benefits for Humanities research
- ii. Linked Data principles and technologies have not been sufficiently adopted in the Humanities to achieve this potential
- iii. The conceptual and technological distances between Linked Data and existing research methods in the Humanities are too great

Based on these assumptions, I developed a series of research questions (RQs). The first two assumptions require assessment of the current extent and state of Linked Data implementation in Humanities research, the advantages this approach has provided, and possible reasons for lack of uptake. As a result, RQ1 addresses these issues:

*1. How valid are the first two assumptions in the hypothesis?*

*1a. To what extent is Linked Data currently applied in Humanities research?*

*1b. In what ways have Linked Data technologies been implemented thus far, in Humanities research in general and Ancient World research in particular?*

*1c. What are the advantages of applying a Linked Data approach to Humanities research?*

*1d. What challenges and obstacles have previously been identified when considering the application of Linked Data to Humanities research?*

I explore RQ1 using existing literature about Linked Humanities Data projects, while also drawing on lessons learned from the cultural heritage sector (Chapter 2). RQ1a discusses the potential proportion of Humanities research projects that have chosen a Linked Data approach. RQ1b builds on these findings by discussing how Linked Data has been implemented by existing initiatives, with specific reference to the Ancient World. RQ1c relates specifically to the part of the hypothesis that refers to "potential benefits" of a Linked Data approach for Humanities research, by exploring what these potential benefits might be. In response, RQ1d discusses barriers to Linked Data implementation that have previously been identified, which might provide additional reasons for lack of uptake beyond the assertion in the hypothesis that "the conceptual and technological distances from existing research methods in the Humanities are too great".

Examining the third assumption, however, is more complex, requiring new user research. Through RQ2 and RQ3 I examine existing research methods in the Humanities, with a view to identifying those that might be particularly amenable to integration with Linked Data:

*2. How can "existing research methods in the Humanities" be defined and classified in relation to digital activities?*

*3. Where might Linked Data be integrated with existing Ancient World research methods to demonstrate the benefits of this approach, and how might these findings apply to Humanities research as a whole?*

I address RQ2 using a literature review on frameworks for describing Humanities research methodologies (3.1), with a view to applying the most appropriate framework to projects identified during the user research phase, thereby providing structure for my response to RQ3. RQ3 itself challenges the assumption that "the conceptual...



distances from existing research methods are too great" by identifying specific methods that could be facilitated by a Linked Data approach and suggesting how such technologies might most effectively be integrated in these contexts.

Finally, RQ4 concerns the investigation of Linked Data usability for Ancient World research, to establish how tools and resources that apply this approach might most effectively meet user needs and expand their potential audiences:

*4. How could Linked Data be made more usable by Humanities researchers in general and Ancient World researchers in particular?*

The above question focuses on the assumption that the "technological distances from existing research methods" result in usability issues for the researcher and aims to address how these might be solved. To address RQ3 and RQ4, I first conduct a survey with Ancient World researchers of varying levels of technical ability, to explore their experiences of using digital tools and resources in general and Linked Data in particular (where applicable). I then select a sample of participants to take part in follow-up interviews, to explore their responses in more depth. This methodology is described in more detail in Chapter 3.

My findings from the above research questions (Chapters 4-7) inform recommendations (Chapter 8) for how future Linked Humanities Data resources should be developed, both to optimise usability and to be integrated more clearly with the research process; I will provide an outline of my full chapter structure in 1.5. The contribution of my thesis is to identify where Linked Data might be most effectively applied to Humanities research and to produce recommendations about how the usability of resulting resources might be optimised.

As my research questions demonstrate, Linked Data technologies are a key focus of this thesis. However, before discussing Linked Data itself, I will first contextualise this approach with a brief overview of other Humanities data models, as well as the relatively recent move towards open research in this area.

## 1.2 Humanities Data Approaches: An Overview

Digital research in the Humanities deals with a wealth of different sources, including texts, images, and objects, containing information that can be extracted to produce datasets. Two key considerations in Humanities data production are how that data is modelled and how it might be shared to promote reuse. There are many publications on both topics and a comprehensive review would be beyond the scope of this thesis. Therefore, my intention in this section is to provide a brief overview that situates my discussion of Linked Data (1.3) within the context of the broader Humanities Data landscape.

The process of data modelling involves producing an abstraction from the original sources (Flanders & Jannidis, 2016, p. 230) that will allow machine-readable analysis and comparison, as well as communicating information to other researchers. Considerable thought must be given to how such data should be modelled; for example, what entities might be included, how they might relate to each other, and what level of complexity might be required. The chosen modelling approach can therefore have a significant impact on the extent and nature of the research that might be achieved using the resulting data, as well as its usability. Two modelling approaches commonly applied to Humanities datasets are text encoding and relational databases.

Text encoding usually involves the annotation, or tagging, of a text using a markup language to enhance it with additional information about its content or structure. In the Humanities, the primary standard for text encoding is that set by the Text Encoding Initiative (TEI)<sup>4</sup>, an international consortium that has produced guidelines and schemas (information about what form a document should take and what elements may be included) to encourage consistency across digital projects. These guidelines require the use of an EXtensible Markup Language (XML)<sup>5</sup> format, which is flexible in working with any compatible schema, but requires each document to comprise a single, strict, hierarchy. Text encoding has many advantages (including discoverability, sustainability, and active user communities, particularly where the TEI is concerned), but suffers due

---

<sup>4</sup> <https://tei-c.org/>

<sup>5</sup> <https://www.w3.org/XML/>

to lack of interoperability<sup>6</sup>, as well as limitations imposed by the hierarchical format (Pierazzo, 2016; Renear, 2004).

Another popular approach, that of building a relational database, builds layers of meaning into tabular data by modelling it as a collection of entities and relationships. Separate tables are produced to contain information about each entity type, in the form of attributes (structured as columns), with a unique identifier for each instance of that entity (structured as rows). These identifiers can then be used within other tables to create links between entities, based on their relationships with each other. Relational databases are extremely popular for Humanities research due to their scope for increasing the richness and discoverability of tabular data, while being accessible to researchers with minimal technical skills. However, their potential for data complexity, and the issue that data structures designed for different projects are often incompatible with each other, together reduce the scope for interoperability (Ramsay, 2004; van Hooland & Verborgh, 2014).

As well as affecting both the usefulness and (re)usability of resulting resources, the above interoperability issues are at odds with recent initiatives and encouragement for researchers of all disciplines to produce and share open data. Indeed, interoperability is a cornerstone of the *FAIR* data management principles (Wilkinson et al., 2016), which state that research data should be Findable, Accessible, Interoperable and Reusable. Although these principles were originally developed with the Sciences in mind, their application to Ancient World research (and Humanities disciplines more broadly) are likely to improve usability of the resulting data, as well as that of the tools and resources through which it is accessed.

Findability and Accessibility require persistent identifiers, which provide a unique way of referring to objects (digital or otherwise) that does not change if their location is moved (for digital objects this refers particularly to URL updates due to domain name changes or site restructuring). Common persistent identifiers include Digital Object

---

<sup>6</sup> Interoperability is the ability to use the same data across multiple platforms, thereby maximising the scope for potential reuse by others.

Identifiers (DOI)<sup>7</sup>, used for referring to research publications, and Open Researcher and Contributor IDs (ORCID)<sup>8</sup>, used for referring to the authors of those publications. Persistent identifiers facilitate discoverability of these resources, while also ensuring user trust that they will not lose access over time (Klump & Huber, 2017).

Interoperability and Reusability require open, non-proprietary standards. The Open Data Institute (2018) defines 'open standards' as "reusable agreements that make it easier for people and organisations to publish, access, share and use better quality data". In practice, this means publishing data using formats and vocabularies that facilitate use in other contexts, with no licencing restrictions. Implementing open standards ensures that data can be made openly available, without the need to consider licencing restrictions of proprietary formats or limiting its usefulness to users of specific commercial software. Open standards additionally facilitate transparency, i.e., communicating to users about a dataset's structure, as well as sustainability, ensuring that the data's lifespan will not be limited by the software used to interact with it.

Having provided a brief overview of approaches to modelling and sharing data in the Humanities (and beyond), I have shown that applying the *FAIR* principles, e.g., through the implementation of persistent identifiers and open standards, should promote (re)usability. While many relational databases and encoded texts meet some of these principles, a crucial issue is the lack of interoperability, which might be addressed by applying Linked Data technologies, the subject of my next section and the technological focus of this thesis.

### 1.3 Linked Data

The term 'Linked Data' refers to a set of technologies for describing digital (and physical) resources, which facilitates machine-readable connections between them. Linked Data can be used to connect disparate datasets containing related resources, with the potential to transform the way they are consumed - leading to new insights that would not have been possible previously. Applying a Linked Data approach

---

<sup>7</sup> <https://www.doi.org/>

<sup>8</sup> <https://orcid.org/>

facilitates the implementation of the Semantic Web: a 'Web of Data', where online resources are semantically linked in a machine-readable way, based on the information about them (Berners-Lee, 1998).

For a dataset to be accurately defined as Linked Data, it must comply with Berners-Lee's (2010) Linked Data principles:

- "1. Use URIs as names for things*
- 2. Use HTTP URIs so that people can look up those names.*
- 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)*
- 4. Include links to other URIs. so that they can discover more things."*

Firstly, for a resource to be described using Linked Data, that resource must have a unique and persistent identifier, known as a Uniform Resource Identifier (URI) (point 1) (Berners-Lee, 2010). A URI can refer to any concept or entity in the physical or digital world (i.e., it does not necessarily denote a digital resource), but in a Linked Data representation, a URI must start with 'http://' (point 2). 'HTTP' refers to 'Hypertext Transfer Protocol', which is the standard mechanism by which digital resources are accessed via the World Wide Web (W3C Network Working Group, 2004).

Each piece of information about the resource (e.g., its title, creator, or location) is expressed using three components: a subject (e.g., the resource, expressed using its URI), an object (e.g., the resource's creator, expressed using their URI), and a predicate (the relationship between the subject and object, e.g., 'has creator', expressed using the URI that describes this property). If the subject or object cannot be identified using a URI (e.g. the title of a book, or an object's latitude/longitude), a literal value, such as a number or text string, can be used instead (van Hooland & Verborgh, 2014, p. 48). The three components are expressed as a 'triple' of the form <subject> <predicate> <object>, using the Resource Description Framework (RDF), the standard format for expressing Linked Data (point 3) (Berners-Lee, 1998; Berners-Lee et al., 2001, p. 40; T. Heath & Bizer, 2011, p. 4). Data in RDF format can be queried using the SPARQL Protocol and RDF Query Language (SPARQL) (point 3) (van Hooland & Verborgh, 2014,

p. 48) via a SPARQL endpoint. Both RDF and SPARQL are open standards, as defined in 1.2.

As more resources are described using RDF, connections build up between triples originating from multiple sources, which share the same URI as either the subject or object (point 4) (van Hooland & Verborgh, 2014, p. 49). These URIs can often be found in authority files, usually provided by well-known, trusted institutions, such as national libraries, which provide central and authoritative sources of information about particular topics. For example, a link might be made between a book and its author, whose URI may refer to their entry in the *Virtual International Authority File (VIAF)*<sup>9</sup> (a service that provides URIs for named entities such as people, companies, and places). Once the book and author URIs are linked, every piece of information linked from the author's *VIAF* URI is implicitly linked to the original resource, without explicitly adding extra pieces of information or adjusting the data structure (van Hooland & Verborgh, 2014, p. 47). These connections potentially reduce the duplication of effort involved in inputting data that already exists elsewhere and augment the resource with a much larger amount of information than could be input by a single institution, facilitating discovery by end users.

It is only possible to take full advantage of this interlinking when the datasets concerned are openly available via the World Wide Web. In response, Berners-Lee (2010) developed a five-star model for Linked *Open* Data (LOD):

- “★            *Available on the web (whatever format) but with an open licence, to be Open Data*
- ★★           *Available as machine-readable structured data (e.g. excel instead of image scan of a table)*
- ★★★        *as (2) plus non-proprietary format (e.g. CSV instead of excel)*

---

<sup>9</sup> <https://viaf.org/>

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context"

The five-star model prioritises openness, with full compliance requiring that the dataset meets Berners-Lee's original Linked Data principles.

As Linked Data can be applied in different ways, it is important to be explicit about what is meant by the term 'Linked Data' in a particular context. For the purposes of this study (and following Berners-Lee's principles), a 'Linked Data tool or resource' meets one or more of the following criteria:

- Provides unique identifiers as HTTP URIs
- Provides, consumes, or connects datasets using RDF
- Can be queried via a SPARQL endpoint
- Facilitates the creation or export of data using the above standards

In addition to each individual entity being associated with a URI, it is often helpful to define them in relation to specific classification systems. For example, Euripides might be defined as a person, author, or playwright, while Classical Athens might be defined as a place, settlement, or *polis* (city-state). Such classifications are provided by Knowledge Organisation Systems (KOS), such as ontologies or thesauri, which express a series of conceptual terms and the relationships between them, often organised using some form of hierarchy (International Organization for Standardization, 2011, sec. 2.62; Mayr et al., 2016). Popular thesauri in the cultural heritage domain include the vocabularies developed by the J. Paul Getty Trust, to describe concepts relating to artworks and art historical resources: the *Art and Architecture Thesaurus*<sup>10</sup>, the *Getty Thesaurus of Geographic Names*<sup>11</sup>, and the *Union List of Artist Names*<sup>12</sup>. The Linked Data versions of these vocabularies are based on the *Simple Knowledge Organization*

---

<sup>10</sup> <http://www.getty.edu/research/tools/vocabularies/aat/>

<sup>11</sup> <http://www.getty.edu/research/tools/vocabularies/tgn/>

<sup>12</sup> <http://www.getty.edu/research/tools/vocabularies/ulan/>

*System (SKOS)*<sup>13</sup>, a framework developed for the purpose of representing thesauri (Cobb, 2015, p. 146).

Ontologies are like thesauri, but with a greater level of expressivity. In this context, the word ‘ontology’ refers to a type of controlled vocabulary whose structure facilitates complex relationships between terms, which cannot be represented solely via a hierarchical format (Hughes et al., 2016, p. 163). Ontologies contain classes, used to classify the subjects or objects of triples, as in the Euripides and Athens examples above, as well as properties, used as the predicates that define the relationships between them. An example of a commonly-used ontology is *Friend of a Friend (FOAF)*<sup>14</sup>, used to describe people and the relationships between them, with terms such as ‘Person’, ‘name’, ‘Organisation’, ‘member’ and ‘knows’ (Brickley & Miller, 2014). Using terms from a well-defined ontology such as *FOAF* enhances the machine-readability of resources and enables the computer to search more intelligently based on meanings and relationships rather than arbitrary keywords. Several broader, more extensive, resources provide both authority files and ontologies that describe them, including *DBpedia*<sup>15</sup> and *Wikidata*<sup>16</sup>.

A major ontology in the cultural heritage domain is the *CIDOC Conceptual Reference Model (CIDOC CRM)*<sup>17</sup>, which semantically describes entities, concepts and relationships relating to objects, facilitating interoperability between collections. It is the only cultural heritage ontology to be recognised as an ISO standard<sup>18</sup> (Bruseker et al., 2017, p. 108), indicating its maturity and positive reception. In addition to its extensive core ontology, members of its user community have developed extensions to describe domains such as archaeological excavations<sup>19</sup> and buildings<sup>20</sup>, ancient texts<sup>21</sup>, spacetime<sup>22</sup>, and the provenance of digital objects<sup>23</sup>. Rather than placing the

---

<sup>13</sup> <https://www.w3.org/2004/02/skos/>

<sup>14</sup> <http://xmlns.com/foaf/spec/>

<sup>15</sup> <https://www.dbpedia.org/>

<sup>16</sup> <https://www.wikidata.org/>

<sup>17</sup> <http://www.cidoc-crm.org/>

<sup>18</sup> <https://www.iso.org/obp/ui/#iso:std:iso:21127:ed-2:v1:en>

<sup>19</sup> CRMarchaeo: <http://www.cidoc-crm.org/crmarchaeo/>

<sup>20</sup> CRMba: <http://www.cidoc-crm.org/crmba>

<sup>21</sup> CRMtex: <http://www.cidoc-crm.org/crmtex>

<sup>22</sup> CRMgeo: <http://www.cidoc-crm.org/crmgeo/>

<sup>23</sup> CRMdig: <http://www.cidoc-crm.org/crmDIG>



individual object at the centre of the data model, *CIDOC CRM* models a series of events in which it was involved (e.g., production, acquisition and duplication), thereby providing an informative way of connecting the object to other entities, in relation to a particular point in time (Bruseker et al., 2017, p. 113).

Both Grossner and Hill (2017, p. 9) and Meroño-Peñuela et al. (2014, p. 13) advocate the use of event-based ontologies for representing historical information in an academic research context. However, *CIDOC CRM*'s complexity can impose barriers to its implementation. For example, incorporating multiple terms with similar definitions can cause ambiguity and inconsistency in its application (Gerth, 2016, p. 21; Liu et al., 2017, p. 349) and complicate analysis of the resulting data. Additionally, several projects (including *SNAP:DRGN*, discussed in 2.2.3, below) have chosen not to use *CIDOC CRM* due to the considerable time and resources required to model data in this way (Bodard et al., 2017, p. 35; Kansa et al., 2018, p. 501). The latter is a particular issue for academic projects, due to the short-term nature of research grants.

As demonstrated above, Linked Data principles and technologies are used to connect digital objects by semantically describing their features, and the relationships between them, in a machine-readable way. The benefits of this approach include richer description of digital objects, leading to more effective integration of multiple collections and datasets. Together, these information sources form an infrastructure on which to build and link other digital tools and resources, both within and outside the organisation that produced them. This process facilitates the integration of previously separate datasets, with the potential to transform Humanities research by providing insights that could not have been discovered by looking at each one in isolation (as I will demonstrate further in Chapter 2). However, as stated in the underlying hypothesis for this study (1.1), they have not yet been sufficiently adopted in this context. Usability issues provide a possible explanation for this situation, with my study aiming to establish what these issues might be and how they might be addressed. As such, I will discuss usability in the following section, in terms of its definition and its relationship to Linked Data.

## 1.4 Usability

Although there has been much scholarly debate on the precise definition and parameters of "usability", the definition I will use in this thesis is based on the international standard ISO-9241-11. Here, the International Organization for Standardization (ISO) (2018) defines "usability" (in the context of "human-system interaction") as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Breaking down this definition into its component parts highlights links between the concept of usability and my other research questions.

In this thesis, the "system, product or service" refers to Linked Data and the tools and resources used for its consumption, and "specified users" refers to researchers in the broadest sense of the word (including academics, students, cultural heritage professionals and anyone else who identifies with this term). Their "specified goals" will be determined during my study, in terms of their research methods. The "specified context of use" is Ancient World research, although it is hoped that my findings will apply more broadly to other Humanities disciplines. As part of the same standard, ISO additionally defines "effectiveness" as the "accuracy and completeness with which users achieve specified goals"; "efficiency" as "resources [such as time, effort, cost and materials] used in relation to the results achieved", and "satisfaction" as the "extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user's needs and expectations". All these factors will be considered when investigating Linked Data tool and resource usability in an Ancient World research context.

Linked Data usability is also being studied elsewhere. In parallel with my PhD research, work led by Robert Sanderson and championed by the *Linked Art*<sup>24</sup> initiative has sought to implement and advocate for 'Linked Open *Usable* Data' (LOUD), in response to the complexity and ambiguity that can result from excessive use of ontologies by subject domain experts. The LOUD principles are aimed at data scientists and ontologists, and focus on usability by developers, to achieve a compromise between

---

<sup>24</sup> <https://linked.art/>

rich semantic description and comprehensibility by others, ultimately facilitating data reuse and interoperability. Such a compromise can be assisted by consistent, self-explanatory data models, supported by clear documentation that includes examples and use cases (Delmas-Glass & Sanderson, 2020, p. 22; Linked Art Contributors, n.d.-b). Although there are some clear parallels between development of the LOUD principles and my study, our audiences differ. My focus is on improving Linked Data usability by Ancient World researchers, with my eventual recommendations aimed at all parties involved in Linked Data tool or resource production (project leaders, developers, institutions, and funders), rather than concentrating solely on a specific subset of this group.

I will return to LOUD when discussing my conclusions and recommendations in Chapter 8, while I will discuss usability research in the Digital Humanities more generally in Chapter 3. Having now contextualised my research topics, I will bring this chapter to a close by providing an overview of my thesis structure.

## 1.5 Thesis Structure

Following my hypothesis and research questions set out at the beginning of this chapter, and the background information provided in the subsequent sections, Chapter 2 will explore in greater depth the topic of Linked Humanities Data. Specific reference will be made to tools, resources and initiatives relating to study of the Ancient World, the advantages they provide, and barriers to their use. From here, I will set out the methodology for my user study in Chapter 3, which will involve discussion of both survey and interview research, as well as alternative approaches, with specific reference to their application in the Digital Humanities. With this discussion in mind, I will outline the construction of my own survey and interview scripts and provide an overview of survey participant demographics, leading to interview sample selection.

The following four chapters will discuss my findings. Chapter 4 will contextualise the subsequent chapters by presenting quantitative findings from my survey, predominantly regarding the tools and resources used by participants, along with associated research methods. My discussion will then turn towards the qualitative, with subsequent chapters drawing from my interviews and open-ended survey

questions. Chapter 5 will focus on five key research methods that participants associated with their use of digital resources (*Discovering, Gathering, Data Recognition, Annotation and Visualization*), focusing specifically on Linked Data. The following chapters will explore more general aspects of usability, with Chapter 6 focusing on users' experiences of interacting with Linked Ancient World Data tools and resources, and Chapter 7 identifying key areas in the production process that ultimately affect tool and resource usability.

Finally, my thesis will conclude in Chapter 8 with a summary of the above findings, and a set of recommendations to improve future Linked Ancient World Data usability. These recommendations lead towards my own *Five-Star Model for Linked Humanities Data Usability*, a set of key considerations to be made by producers at the outset of a Linked Ancient World Data project, which would also apply to other Humanities disciplines.

Having introduced my research topic and its technological context, the following chapter will provide a detailed exploration of current initiatives that have applied Linked Data to Humanities research, with the aim of addressing RQ1 through discussion of their extent, implementation, advantages, and barriers.

## 2 Linked Data and Humanities Research

Having introduced the principles and technologies behind Linked Data and significant barriers to its uptake in the previous chapter, in this chapter I will make a more detailed exploration of its application in the Humanities by analysing data and reviewing the literature relating to the four components of RQ1 (outlined in 1.1). I will start by addressing RQ1a, assessing the extent to which Linked Data is currently applied in Humanities research, using publicly-available datasets as case studies (2.1), before providing examples of existing projects predominantly from the domain of Ancient World research (2.2), in response to RQ1b. The following section will explore the advantages of applying Linked Data in a Humanities research context, in relation to RQ1c (2.3). Finally, I will address RQ1d by discussing known barriers to the implementation of Linked Humanities Data, to establish why this might not always be the favoured approach (2.4).

### 2.1 Extent of Linked Data in Humanities Research

In 1.1, I introduced the following hypothesis, assuming that Linked Data is largely underused in Humanities research:

*Linked Data principles and technologies have not yet been sufficiently adopted in Humanities research to achieve their potential benefits because the conceptual and technological distances from existing research methods are too great.*

In this section, I will address this assumption by using existing datasets to better understand the extent of Linked Humanities Data implementation. Firstly, I will analyse data about projects funded by the UK's Arts and Humanities Research Council (AHRC)<sup>25</sup> to establish the proportion that included a Linked Data component, as well as trends over time and across subject areas (2.1.1). Secondly, I will use the *Digital Classicist Wiki's* 'Linked open data' category and Linked Ancient World Data Institute (LAWDI) pages to produce a more geographically diverse list of projects, albeit within a narrower subject area (2.1.2).

---

<sup>25</sup> <https://ahrc.ukri.org/>

### 2.1.1.1 Case Study: AHRC-Funded Linked Data Projects

The AHRC is a major source of funding for Arts and Humanities research projects in the UK, whose data is available via UK Research and Innovation's (UKRI)<sup>26</sup> Gateway to Research (GtR)<sup>27</sup>. To establish the extent of Linked Data projects funded by the AHRC, I exported data on the 5,975 projects with a start date on or after 1 January 2006 and an end date on or before 31 July 2020. Limiting the dataset to projects that ended at least a year before the export took place (August 2021) ensured that records of their outcomes should be complete and therefore more likely to provide accurate information about each project's relationship to Linked Data.

I initially discovered projects involving Linked Data technologies by searching for projects whose title, description, or output information contained any of the following words or phrases:

- linked data
- linked open data
- linking data
- RDF
- semantic web
- semantic links
- semantically linked
- SPARQL
- URI

For each of the 54 projects containing one or more of the above terms, I read its GtR record in full to determine whether it did indeed involve Linked Data. As a result, I identified 33 Linked Data projects<sup>28</sup>, approximately 0.55% of the total. This result implies that Linked Data adoption in the Humanities, at least in the context of AHRC-funded projects, remains minimal. Plotting these projects over time (Figure 2.1), paints

---

<sup>26</sup> <https://www.ukri.org/>

<sup>27</sup> <https://gtr.ukri.org/>

<sup>28</sup> There were several reasons why projects mentioned one or more of my search terms but did not involve Linked Data: some involved linking data via another means (e.g., a relational database); others mentioned Linked Data while listing different topics, approaches and/or standards in the research area, to indicate a knowledge of context, but without incorporating Linked Data into the project, and a small number used URI and/or RDF as acronyms for other concepts and initiatives.

a more nuanced picture, indicating that interest in Linked Data grew during 2012-2015, then seemed to wane. This apparent loss of interest in Linked Data may, however, be due to the omission of ongoing or recently completed projects that started towards the end of the date range. In particular, the AHRC's current *Towards a National Collection*<sup>29</sup> programme is likely to include multiple Linked Data projects.

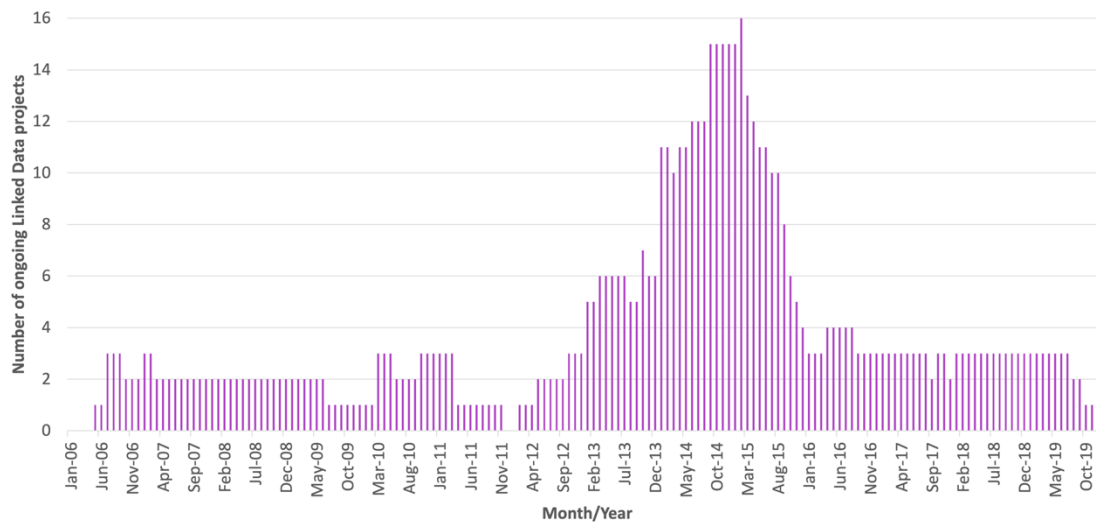


Figure 2.1 Number of ongoing AHRC-funded Linked Data projects over time

To further explore each project's relationship to Linked Data, I used the information provided in their GtR records to divide them into the following broad categories:

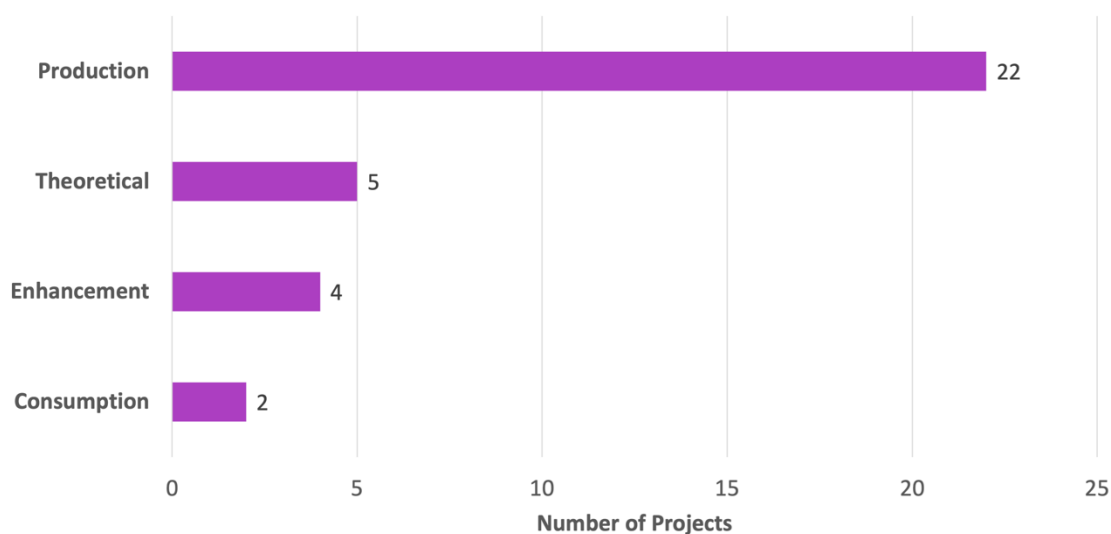
**Production:** the project involved Linked Data production, either creating completely new data or converting existing data; there may have been some consumption or enhancement of this data following its production.

**Enhancement:** the project focused on enhancement of Linked Data produced prior to the start of the project; no production of new Linked Data was involved.

**Consumption:** the project focused on consumption of Linked Data produced prior to the start of the project; no production of new Linked Data was involved.

<sup>29</sup> <https://www.nationalcollection.org.uk/>

**Theoretical:** the project focused on how Linked Data might be implemented in a particular Arts/Humanities context, but there was no practical component.



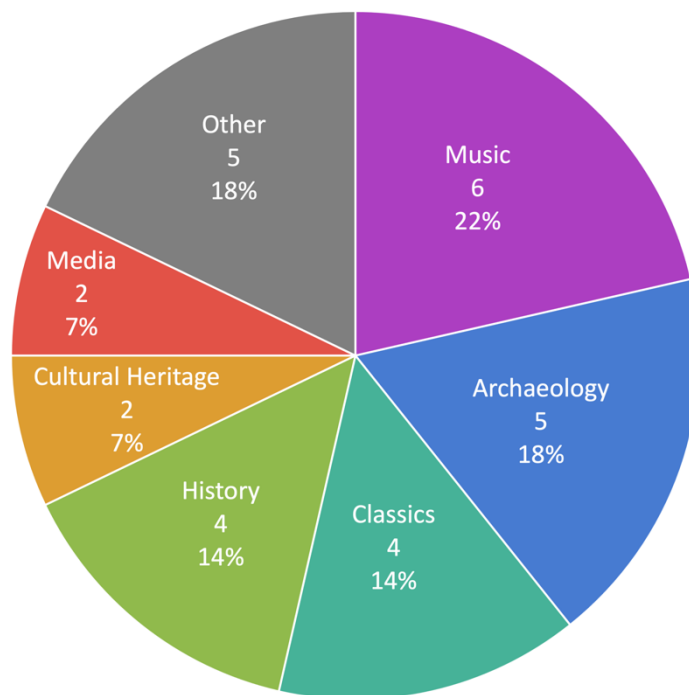
*Figure 2.2 Number of projects in each category based on their relationship to Linked Data*

Results are shown in Figure 2.2 and indicate that while some projects may involve consumption or enhancement of Linked Data resources in addition to production, few projects were primarily enhancement or consumption based. Even then, most of these projects involved working with data that had previously been produced by the same team of people. This could imply a real or perceived funder bias towards projects that appear to demonstrate greater innovation, over those that seek to improve, reuse, or sustain existing resources – either researchers assume that producing new data would increase the probability of securing a grant, or that the AHRC favours data production projects. Alternatively, we might anticipate that the more recent projects not included in the dataset could incorporate a greater degree of enhancement and/or consumption once these resources have been disseminated sufficiently widely to become more embedded into the existing work of researchers.

Focusing solely on the production, enhancement, and consumption projects, I then ascertained their subject areas. As the diagram in Figure 2.3 illustrates, a significant proportion are based on subjects relating to the Ancient World, with Archaeology and Classics projects comprising nearly a third of the total. This relatively high proportion confirms my assertion in the previous chapter that Linked Data uptake has tended to



be greater for Ancient World research than other Humanities disciplines. Ancient World data might be particularly amenable to this approach as it often involves linkable entities such as places, people, and events, and is not subject to the same copyright or ethical restrictions as more recently produced data.



*Figure 2.3 Pie chart showing distribution of main subject areas among AHRC-funded Linked Data projects (all subjects with two or more projects are shown; 'Cultural Heritage' comprises projects that relate to libraries, museums or archives without a specific subject topic; 'Other' comprises subjects with only one project)*

While the above findings begin to convey the extent of Linked Data implementation among Humanities research projects, there are some inherent limitations to the AHRC dataset. Firstly, the AHRC as an organisation does not exist specifically to fund digital or infrastructure-based research; therefore, a higher proportion of relevant projects might be funded by other sources. Secondly, there is considerable variation in the quality of data about these projects, as it is input by project researchers, rather than the AHRC themselves. Although those closest to the project are theoretically likely to provide the most accurate information, they are also likely to have multiple competing priorities that reduce the amount of time available for this task. As such, some relevant projects may not contain sufficient information to be identified through my queries.

Finally, considering the Arts and Humanities in general provides an extremely broad scope in terms of subject, but studying these projects through the lens of one funder is relatively narrow. As a result, conclusions may be drawn about Linked Humanities Data projects that would not apply outside of the UK, or even outside of the AHRC. To alleviate these issues, it will be beneficial to consider the AHRC projects alongside those from a different kind of dataset, discussed in the following section.

### 2.1.2 Case Study: Digital Classicist Wiki

In this second case study, I will investigate the extent of Linked Data implementation for the study of the Ancient World by exploring three curated lists of relevant projects. These lists appear on the *Digital Classicist Wiki*<sup>30</sup>, a community-maintained resource for digital research on the Ancient World that provides lists of projects, tools, and techniques, organised into categories. Since its inception in 2004 (Mahony, 2017, para. 3), the Digital Classicist has provided a central web resource to promote collaboration among its global community (Mahony, 2017, para. 4; Mahony & Bodard, 2010, p. 2). The wiki is just one component of a broader network that incorporates seminars and mailing lists, as well as participation at relevant events relating to Digital Humanities and/or the Ancient World. Users can contribute wiki entries about their areas of interest and expertise, which are made available for editing and updating by other members of the community, providing a form of "peer review" (Mahony, 2011, para. 20, 2017, para. 10). The *Digital Classicist Wiki* can therefore be considered a reliable, current source of information about digital tools and resources for the study of the Ancient World, with no funder restrictions. Its geographical scope is greater than that of the AHRC, as it is theoretically an international resource; however, as the content is predominantly Anglophone in nature, it cannot be considered completely representative of global digital Classics.

The first curated list comprises wiki pages with the category 'Linked open data'<sup>31</sup>. In August 2017, there were 45 pages in this category, which had increased to 79 by August 2021. Although this increase could indicate a rise in the number of Linked Data initiatives, it could also have occurred in response to encouragement from the

---

<sup>30</sup> [https://wiki.digitalclassicist.org/Main\\_Page](https://wiki.digitalclassicist.org/Main_Page)

<sup>31</sup> [https://wiki.digitalclassicist.org/Category:Linked\\_open\\_data](https://wiki.digitalclassicist.org/Category:Linked_open_data)

organisers to add more content to the wiki. Pages listed in this category have been designated by at least one Digital Classicist community member as involving ‘Linked open data’, defined as "projects, tools and other resources that use, work with, or offer information about linked open data, including URIs, RDF, vocabularies and ontologies" (Digital Classicist, 2019b). The principal categories on the wiki, used as part of its primary navigation menu, are:

- **FAQ:** "Discussions or advice around digital practices or methods relating to the study of the ancient world. Some may take the form of questions, other as nascent "how to" or overview pages" (Digital Classicist, 2019c)
- **Projects:** "Projects applying computing technologies to Classical/Ancient Historical research" (Digital Classicist, 2016b)
- **Tools:** "Tools... of special interest for members of the Classics community" (Digital Classicist, 2014)
- **Events:** "Conferences, seminars, training courses, and other events that are described in the Wiki" (Digital Classicist, 2019a)

Table 2.1 indicates that Linked Ancient World Data initiatives are more likely to take the form of projects that produce, enhance, or consume data, rather than developing tools. Overall, ‘Linked open data’ pages make up a small but significant proportion of the Digital Classicist wiki, suggesting that a Linked Data approach is becoming increasingly established in Ancient World research.

Category	Number of ‘Linked open data’ pages with this category	Category as a % of ‘Linked open data’ pages	Number of pages in wiki overall with this category	‘Linked open data’ pages as a % of this category
Project	57	72%	534	11%
Tool	21	27%	224	9.4%
FAQ	3	3.8%	53	5.7%
Event	1	1.3%	15	6.7%
[None of the above]	5	6.3%	N/A	N/A

*Table 2.1 Distribution of ‘Linked open data’ category pages in Digital Classicist wiki, based on other broad categories to which they belong; some pages belong to more than one category*

It should be noted, however, that these categories were not applied systematically. Instead, they have been added by multiple different members of the Digital Classicist community over time, each of whom has their own unique perspective. To assess how many Linked Data projects might be missing from the 'Linked open data' category list as a result, I performed keyword searches on the full text of all wiki pages; results are shown in Table 2.2. These searches identified a further 24 pages that could be categorised as 'Linked open data'. There may yet be more pages that should belong in the 'Linked open data' category, but where the authors have not included relevant keywords in their descriptions.

In addition to its list of pages, the 'Linked open data' category contains a subcategory, the Linked Ancient World Data Institute (LAWDI)<sup>32</sup>. LAWDI events were held in 2012 and 2013 to encourage the sharing of Ancient World research data using Linked Data principles and technologies, as well as to discuss how this might be achieved and where the challenges lay (Elliott et al., 2012). In doing so, LAWDI additionally facilitated the development of communities of practice for working with Linked Ancient World Data. The LAWDI page on the *Digital Classicist Wiki* links to lists of web resources associated with 2012<sup>33</sup> and 2013<sup>34</sup> attendees, which number 33 and 34 resources respectively: due to overlap between the two events, the combined list numbers 47 resources. Attendees were selected based on application to the Institute organisers, which demonstrates some subjectivity about which resources were and were not included. Additionally, as both events took place in the USA (Digital Classicist, 2016a), there might be a disproportionate number of American projects. That said, these lists form a useful point of comparison with the Digital Classicist wiki's 'Linked open data' category.

---

<sup>32</sup> [https://wiki.digitalclassicist.org/Linked\\_Ancient\\_World\\_Data\\_Institute](https://wiki.digitalclassicist.org/Linked_Ancient_World_Data_Institute)

<sup>33</sup> [https://wiki.digitalclassicist.org/LAWDI\\_2012\\_Websites](https://wiki.digitalclassicist.org/LAWDI_2012_Websites)

<sup>34</sup> [https://wiki.digitalclassicist.org/LAWDI\\_2013\\_Websites](https://wiki.digitalclassicist.org/LAWDI_2013_Websites)

Keyword(s)	Number of pages	Number of 'Projects'	Number of 'Tools'	Number of 'FAQ'
"linked data"	9	5	0	0
"linked open data"	10	7	1	1
"semantic web"	1	0	1	0
"semantic technologies"	0	0	0	0
"semantically linked"	0	0	0	0
"RDF"	6	3	2	0
"SPARQL"	1	0	0	0
"URI"	5	1	0	0

Table 2.2 Number of pages in the Digital Classicist wiki containing particular keywords, which do not already appear in the 'Linked open data' category (no such pages were categorised as 'Events')

It should firstly be noted that while all LAWDI participants were interested in the potential of Linked Ancient World Data, their projects did not necessarily involve Linked Data technologies; therefore, the combined list does not represent 47 Linked Data projects. Assessing which resources did in fact involve Linked Data proved to be a challenging task, as not all documentation is transparent about technologies or data structures. However, using the definition provided in 1.3, I was able to confidently identify 21 resources. Of these, eight also had *Digital Classicist Wiki* pages in the 'Linked open data' category and eight appeared under different categories (the five that did not appear in the wiki at all tended to be non-Ancient World specific resources, such as Dublin Core<sup>35</sup> and the British Museum catalogue<sup>36</sup>). The eight that appeared in the wiki but not the 'Linked open data' category predominantly involve Linked Data, but it is not the primary focus, such as *Papyri.info*<sup>37</sup> (below, 2.2.4) and *Perseus*<sup>38</sup> (below, 4.1.1).

While none of these three lists from the Digital Classicist wiki can be considered exhaustive, I can be reasonably confident that together they provide a strong indication of the extent of Linked Data implementation in digital projects for Ancient

<sup>35</sup> <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

<sup>36</sup> [http://www.britishmuseum.org/research/search\\_the\\_collection\\_database.aspx](http://www.britishmuseum.org/research/search_the_collection_database.aspx)

<sup>37</sup> <https://wiki.digitalclassicist.org/Papyri.info>

<sup>38</sup> [https://wiki.digitalclassicist.org/Perseus\\_Digital\\_Library](https://wiki.digitalclassicist.org/Perseus_Digital_Library)

World research. Looking at 'Projects' and 'Tools' in the wiki, I can infer that approximately 10% of digital initiatives for Ancient World research involve Linked Data, using the broad definition provided in 1.3. This number is likely to be higher, based on LAWDI resources described in the Digital Classicist wiki but which do not appear in the 'Linked open data' category. The recent rise in 'Linked open data' pages additionally suggests that this proportion is continuing to increase.

### 2.1.3 Summary: Extent of Linked Data in Humanities Research

In this section, I have suggested the possible extent of Linked Data implementation in Humanities research, with particular focus on the Ancient World. My findings, from analysing AHRC and *Digital Classicist Wiki* data, indicate that a Linked Data approach is still rare among Humanities research projects, even those with a digital remit.

However, a small but significant proportion of digital initiatives for Ancient World research involve Linked Data, with both datasets indicating that this is increasing.

These findings should be treated with caution, however, as there is no single source that lists every Humanities project, tool, or resource; even those datasets describing specific subsets, based on discipline or funder, are inconsistent in their terminology, definitions, and data quality. Therefore, it is not possible to determine an exact figure for the overall extent of Linked Data implementation, beyond a general sense that this remains an unusual approach in the Humanities. This issue might be mitigated by production of Linked Data-specific directory resources, a suggestion to which I will return when discussing my findings relating to digital tool and resource discovery (5.1.3) and in relation to possible future work in this area (8.9.1).

To continue my exploration of the application of Linked Data to Humanities research, the following section will look in more detail at some specific initiatives.

## 2.2 Linked Ancient World Data in Practice

When Humanities researchers are exploring a topic using a digital tool or resource, their search terms usually revolve around particular concepts, or "contextual entities", categorised by Lee (2011, p. 106) as Object, Agent, Occurrence, Purpose, Time, Place, Form of expression, Concept or Abstraction, and Relationship. Accurate identification and description of these concepts is intrinsic to effective digital representation that facilitates their discovery. Linked Humanities Data tools, resources, and initiatives

often revolve around one of these contextual entities, while incorporating their relationships to others.

In this section, I will explore initiatives that have sought to identify, define, and describe a particular concept using a Linked Data approach. I will start by discussing relatively mature initiatives that focus on place (2.2.1), before moving to the concepts of time (2.2.2) and people (2.2.3), which require greater complexity when represented as Linked Data. Finally, I will explore several initiatives that describe objects (2.2.4) and their relationships to the other contextual entities. Initiatives included in this section were selected due to their relatively high uptake by the research community, as evidenced by their scale, as well as the frequency with which they were mentioned during my user research (on which I will expand in 4.1.1). The following discussion will provide concrete examples to demonstrate where Linked Data has been applied to Humanities research, as well as introducing their advantages and barriers. I will primarily focus on the Ancient World, although some initiatives will have broader applicability, particularly those from the cultural heritage domain.

#### 2.2.1 Place

Place is a key component of Humanities research that involves the study of historical events or people, or the movement of objects and materials. Real and mythological places appear in art or are mentioned in literature, demonstrating that place additionally permeates many research topics with a less obvious geographic component. Place entities can often be identified with relative ease from a source's content or metadata; there also exist numerous resources, both print and digital, that contain information about these places.

Such information resources often take the form of gazetteers. At its simplest, a gazetteer provides a list of place names, but may additionally include geographic coordinates or relationships to other places. These relationships can be described using RDF, with each place entity represented by a URI<sup>39</sup>. Incorporating a system of unique identifiers aligns alternative spellings and translations, while ensuring disambiguation from other places with similar names. Using Linked Data to connect a

---

<sup>39</sup> The terms RDF and URI were introduced in 1.3.

digital object with one or more gazetteer URIs enriches that object with information about these place(s), which is further enhanced if the gazetteer(s) contain external links, e.g. to other gazetteers (Berman, Mostern, et al., 2016; Grossner et al., 2016; Horne, 2020b). Resulting connections can be visualised as maps or networks and incorporated into discovery tools, to enable powerful cross-collection searching and identifying key entities and relationships (Isaksen et al., 2018). Such tools and resources have the potential to break down the barriers between datasets, improve the efficiency of the research process, and provide a more holistic view of a subject domain in terms of its geography.

*GeoNames*<sup>40</sup> provides URIs for places throughout the world, integrating geographic data from multiple official sources, such as the National Geospatial-Intelligence Agency and Ordnance Survey (Geonames, n.d.). Its near-comprehensive coverage and open licence have made it attractive to many Digital Humanities projects that incorporate a spatial element; however, it was designed to represent the world in its current state. Although alternative names for a place are permitted (and can include historic names), their inclusion cannot be associated with dates or boundary changes. Berman et al. (2016, p. 124) and Simon et al. (2016a, p. 107) suggest that one way to address this issue is to produce suitable historical gazetteers that link places to their modern equivalents in *GeoNames*. This recommendation emphasises the importance of producing specialist resources that are semantically rich enough for Humanities research, while recognising the value of connection to major information sources.

One such specialist resource is *Pleiades*<sup>41</sup>, a gazetteer providing persistent URIs for ancient places, which stemmed originally from the digitisation of the *Barrington Atlas of the Greek and Roman World* (Elliott & Gillies, 2009, para. 41), but has evolved into a community-driven resource with an increasingly broad geographical scope. *Pleiades* URIs are additionally linked to *GeoNames* URIs, where an appropriate place record is available (Simon et al., 2016a, p. 102); however, *Pleiades* is more fluid in its definition of 'place'. Its creators were influenced by Tuan's (1975, p. 152) conception of 'place' as "a center of meaning constructed by experience", with most places lying on a

---

<sup>40</sup> <https://www.GeoNames.org/>

<sup>41</sup> <https://pleiades.stoa.org/>



spectrum between "points in a spatial system" and "strong visceral feelings". Unlike *GeoNames*, a *Pleiades* place need not always correspond to a physical location, which provides the flexibility to include mythological places, or those whose location is unknown (Gillies, 2015). *Pleiades* is therefore better suited to represent the multiplicity of places encountered in Humanities research. Furthermore, to provide historical nuance and disambiguation, *Pleiades* permits multiple names to be associated with each place, with each name having a start and end date (Schneider et al., 2018, p. 15). Researchers are therefore able to identify historical places with greater precision, rather than approximating to their counterparts in the contemporary world.

*Pelagios*<sup>42</sup> is one of the best known Linked Ancient World Data initiatives. Its original goal was to bring together online resources that mention ancient places, to facilitate data sharing, discovery and visualisation for academic researchers and the wider public (Simon et al., 2012, p. 1). Like *Pleiades*, it has subsequently increased its temporal and geographical scope (Simon et al., 2014, p. 105), as well as producing tools, such as the *Recogito*<sup>43</sup> annotation platform, and the *Peripleo*<sup>44</sup> visualisation and discovery tool. In *Pelagios*, place names in a digitised text or image are identified and annotated with relevant gazetteer URIs, using the W3C Web Annotation Data Model<sup>45</sup>. Rather than expecting contributing organisations to adjust their data structures, *Pelagios* provides interconnectivity by hosting only these annotations as "stand-off markup", linking to the record for each object in its original dataset. In this way, the gazetteers provide a "central backbone" to connect multiple datasets, based on their relationships to place (Simon et al., 2017, p. 114). To ensure interoperability between different gazetteers, *Pelagios* developed a Gazetteer Interconnection Format<sup>46</sup>, which has since been superseded by the Linked Places format<sup>47</sup>. Its implementation enables gazetteers with very different approaches to defining and representing the concept of 'place' to be reconciled and used alongside each other (Simon et al., 2016a, p. 106). In reviewing the literature, I found *Pelagios* to be the most regularly cited initiative of those

---

<sup>42</sup> <https://pelagios.org/>

<sup>43</sup> <http://recogito.pelagios.org/>

<sup>44</sup> <http://pelagios.org/peripleo/map>

<sup>45</sup> <https://www.w3.org/TR/annotation-model/>

<sup>46</sup> <https://github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format>

<sup>47</sup> <https://github.com/LinkedPasts/linked-places>

included in the current section, supporting Berman et al.'s (2016, pp. 123–124) assertion that it now provides a "de facto standard" for "exchange of historical place name data". Its relative success in this regard could be due to several factors, such as openness, low barrier to entry, and active community – themes to which I will return during discussion of my findings.

Another initiative that uses the Linked Places format is the *World-Historical Gazetteer (WHG)*<sup>48</sup>, which identifies and describes places from 1500 CE to the present day, with a global scope (World-Historical Gazetteer, 2017). Like *Pleiades*, the *WHG* used places from a print publication<sup>49</sup> as a foundation, which has been supplemented by a specialist gazetteer of colonial Latin America, *HGIS de las Indias*. Additionally, users are encouraged to link to authorities such as *GeoNames* and the *Getty Thesaurus of Geographic Names* where possible (Grossner, 2019; World-Historical Gazetteer, 2018). From an early stage in the *WHG*'s conception, it was emphasised that contributors should retain ownership of their data, and that instructions should be provided to ensure that users cite contributed datasets correctly (Manning & Mostern, 2015, p. 7).

This section has discussed different approaches to representing place using Linked Data technologies. *GeoNames* and *Pleiades* exemplify the value of both a broad, global gazetteer and a narrower, historical approach, while simultaneously demonstrating that Linked Data about ancient places should not rely on physical coordinates and must effectively represent change over time. *Pelagios* and *WHG* provide examples of how Linked Data from gazetteers can be brought together into an infrastructure and made available to the research community via user-friendly tools and resources. Overall, these resources show that the representation of ancient places using Linked Data is relatively mature; the following sections will discuss representation of time and person entities, currently at an earlier stage in their development, potentially due to the lack of existing resources that might readily be repurposed for a Linked Data context.

---

<sup>48</sup> <https://whgazetteer.org/about/>

<sup>49</sup> Dorling Kindersley's *Atlas of World History*, edited by Jeremy Black

### 2.2.2 Time

While many disciplines incorporate the modelling of time as an absolute concept, where exact days, hours, minutes, seconds and beyond can be identified with absolute precision, this is not the case when representing information about the premodern world. Dates must often be reconciled to different calendrical systems (if they are provided at all). More often, dates are unclear or unavailable, in which case a relative chronology must be applied, based on aspects such as an object's style, typology or context. Such definitions are frequently contested and are subject to change if new evidence comes to light. As these characteristics vary by geographical region, each chronological period defined in this way is bound up with the concept of place, further demonstrating the link between place and time. For example, the Greek Bronze Age spanned approximately 3200 to 1050 BCE, while in Britain the Bronze Age is dated to between 2500 and 700 BCE.

While several initiatives seek to represent time using a Linked Data approach, *Periods, Organized (PeriodO)*<sup>50</sup>, a gazetteer of chronological periods, appears to be the most advanced in terms of development, and appears most frequently in the literature. Rather than seek to provide an authoritative identifier for each chronological period, *PeriodO* provides a URI for each *assertion* of a chronological period (Buchanan et al., 2016, p. 3). Each assertion includes the period name, date range and the geographical area to which it applies, linked to a URI from a spatial gazetteer such as *Pleiades* (2.2.1) (Shaw et al., 2018). In this way, *PeriodO* "attempt[s] to mirror scholarly practice" (Rabinowitz, 2014) by representing the 'fuzziness' and disagreements in this area of scholarship. *PeriodO* has therefore been developed with existing research processes and the research community in mind: acknowledging that much Humanities data cannot be considered authoritative, and that representing it in a way that implies otherwise would deter potential users (Rabinowitz et al., 2016, p. 51). Assertions are linked based on relationships between them, e.g. whether a particular term provides a broader or narrower definition than another for a similar chronological period (Rabinowitz et al., 2018, p. 207). Users can therefore maintain consistency by linking to periods that fall within the same sequence, as well as comparing similar definitions.

---

<sup>50</sup> <https://perio.do/>

Similar initiatives include the *Graph of Dated Objects and Texts (GODOT)*<sup>51</sup>, which aligns different calendrical systems and provides URIs for dates mentioned in digitised ancient texts; where possible, dates are also standardised using the Julian calendar system (Grieshaber, 2016; Kuczera, 2017, pp. 182–185). *ChronOntology*<sup>52</sup> is a similar initiative to *PeriodO* that additionally applies type categorisation (Deutsches Archäologisches Institut, n.d.). *Topotime*'s<sup>53</sup> data model assigns place names and geographical coordinates to events, while representing the uncertainty of historical date ranges; it additionally provides a tool for visualising timelines (Grossner et al., 2016, p. 94; Grossner & Meeks, 2014).

Using *PeriodO* as a case study, I have discussed the complexities of capturing the 'fuzziness' of time when representing aspects of the Ancient World and demonstrated how this has been addressed using a Linked Data solution. *PeriodO*'s assertion-based model mirrors scholarly dialogue surrounding chronology and incorporates the additional concept of place to provide key contextual information, thereby mitigating any potential doubts about its suitability for use in academic research. Both time and place are crucial to the subject of the next section: using Linked Data to represent people.

### 2.2.3 People

People appear in a wide variety of contexts in ancient sources. These include creators of objects and authors of texts, as well as those depicted or mentioned within them. Some might have existed in real life, while others are either fictional, or their existence cannot be confirmed with certainty. Information about people, or person-like entities, is available via authority files and prosopographies. 'Prosopography' refers to the study of people within a defined population (aiming to identify connections between them based on biographical information), as well as to the tools and resources produced to facilitate this process (Bodard et al., 2017, p. 28). A prosopography includes (at a minimum) people's names, but can also include other information, such as places and dates for their birth and/or death, alternative names, and related people. Various

---

<sup>51</sup> <https://godot.date/>

<sup>52</sup> <https://chronontology.dainst.org/>

<sup>53</sup> <http://dh.stanford.edu/topotime/>

online prosopographies exist, often from digitising printed books, with many individuals appearing in multiple databases due to population overlaps. As with geographical and temporal gazetteers, this information can be represented particularly effectively as Linked Data, incorporating connections between different records for the same person.

There are inherent complexities in modelling data about people, for which Linked Data can be a particularly effective solution. For example, multiple people often have the same name, an issue which can be addressed by assigning a distinct URI to each individual (Bodard et al., 2017, p. 29). However, it is not always possible to determine with certainty whether multiple mentions of the same name refer to the same person (Broux, 2017c, pp. 348–349). The more cautious approach of identifying each instance with a separate URI can therefore result in multiple URIs for the same person (Varga et al., 2018, p. 39), which can be aligned if further evidence is discovered. Similarly, an individual can have many names, or epithets, all of which can be linked to the same URI, potentially with further information about the context in which each name is used. Person URIs can additionally be linked to relevant places, chronological periods, and events, as well as other people.

The *Virtual International Authority File (VIAF)*<sup>54</sup> aggregates information about people from authority files produced by libraries and cultural heritage institutions throughout the world; some organisations, events and places are also included, albeit to a lesser extent (Angjeli et al., 2014, pp. 2–3; Hickey & Toves, 2014). In recent years there has been increasing engagement with the research community, with a view to enriching the information available. In particular, an initiative called *Scholars' Contributions to VIAF* (OCLC, 2019) resulted in the incorporation of Ancient Greek data from the *Perseus Digital Library*<sup>55</sup> (Smith-Yoshimura, 2013) and Syriac data from *Syriaca.org: the Syriac Reference Portal*<sup>56</sup> (Smith-Yoshimura, 2014). As well as broadening VIAF's linguistic diversity, these additions have increased its relevance to Ancient World researchers. However, like *GeoNames* (2.2.1), VIAF does not provide the degree of

---

<sup>54</sup> <http://viaf.org/>; introduced in 1.3

<sup>55</sup> <http://www.perseus.tufts.edu/hopper/>

<sup>56</sup> <http://syriaca.org/>

nuance required for many Ancient World projects; for example, it contains relatively little information about mythological entities (Gerth et al., 2016, p. 15). Again, more domain-specific resources are needed, ideally linked to *VIAF* to ensure connection to the wider web of data.

Perhaps the most extensive and well-known resource for ancient people is *Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names (SNAP:DRGN)*<sup>57</sup>. Like *VIAF*, *SNAP:DRGN* provides authority files for people, but with a focus on the Ancient World. Existing catalogues of personal names (such as that provided by *Trismegistos* – 2.2.4, below) are aligned and combined into one searchable dataset, and connected based on their relationships with each other, using the SNAP data model (Bodard et al., 2017, p. 31; Lawrence & Bodard, 2015). Similarly to *Pelagios* (2.2.1), *SNAP:DRGN* has opted for relatively lightweight methods for data linking, to best reflect the ambiguities and incompleteness inherent in Ancient World person data, particularly where mythological entities are concerned (Bodard et al., 2017, pp. 31–36).

However, the funded period for *SNAP:DRGN* came to an end before a query interface could be developed that would have ensured accessibility to users unfamiliar with SPARQL. Prag and Chartrand (2018, p. 248) describe *SNAP:DRGN*, and digital prosopographical work more generally, as a "work-in-progress", indicating that despite its adoption by various initiatives, it is not generally considered to be as mature as *Pelagios* or *PeriodO*, for example. While the data has the potential to be extremely useful, the proportion of interested parties who might be able to access it is relatively small, limiting the scope for its consumption. This situation additionally demonstrates the importance of funding projects to enhance and consume existing Linked Data, rather than prioritising the production of new datasets, mentioned in relation to AHRC projects in 2.1.1.

A more recent initiative, *Modelling Ancient Narratives, Territories, Objects (MANTO)*<sup>58</sup>, took a different approach. In creating their dataset of Greek mythological people (and

---

<sup>57</sup> <http://snapdrgn.net>

<sup>58</sup> <https://www.manto-myth.org/manto>

person-like entities), they started by consulting original texts<sup>59</sup>, extracting passages that mentioned these entities, and using this information to model data about their interactions with places, objects, and other people (Hawes & Smith, 2020). Data is managed using the *Nodegoat*<sup>60</sup> platform, which provides flexible data modelling and visualisation options for Humanities projects. As such, existing data in *MANTO* can already be displayed as maps and networks, without requiring specific technical skills. The project team additionally consulted their potential user community during development, by asking them to test a prototype version of the tool and complete an online questionnaire (Hawes, 2020). Although the data is currently incomplete, the project remains in progress, with a large community of data contributors. In the spirit of building on previous work, *Pleiades* URIs are used to refer to places, while *MANTO* developed new URIs for person-like entities (Hawes & Smith, 2021; *MANTO*, 2020).

*SNAP:DRGN* and *MANTO* have demonstrated how Linked Data technologies can be implemented to model and disseminate data about people in general and within an Ancient World context specifically. Key considerations include disambiguation, the accurate representation of relationships, and the inclusion of person-like mythological entities. I also emphasised the importance of working with the research community for effective development of domain-specific resources. A particular concern in this section was the impact of short-term funding models on sustainability and usability, which might be mitigated by taking an incremental, topic-specific approach, while taking advantage of existing technologies where possible. Next, I will explore the representation of objects using a Linked Data approach, incorporating data models and discovery platforms.

---

<sup>59</sup> Including Homer's *Iliad*, Hesiod's *Theogony*, Apollodoros' *Library of Greek Mythology*, and Pausanias' *Description of Greece*

<sup>60</sup> <https://nodegoat.net/>

#### 2.2.4 Objects

Humanities research in general, and Ancient World research in particular, involves the study of objects, often held in the collections of cultural heritage institutions. These objects include texts, artworks, and material culture, with many occupying more than one of these categories. Modelling object data is complex: each one is a product of a series of interventions (usually by people) over time, often in multiple places; objects can be classified into different types, with increasing levels of granularity, and their contents are often rich with textual and/or visual information. Linked Data technologies are ideally poised to capture the layered and multi-faceted nature of such objects in a way that would be difficult (if not impossible) to achieve accurately using tabular formats. In turn, object-based data is particularly amenable to a Linked Data approach, largely because it is often already available as a structured format, in collections or archaeological databases.

As discussed in 1.3, many initiatives, particularly in the cultural heritage domain, have applied a Linked Data approach by implementing the *CIDOC CRM* ontology. These include the British Museum, who mapped their catalogue metadata to *CIDOC CRM* to make it available via a SPARQL endpoint<sup>61</sup> (Gerth et al., 2016, p. 12). Representing the collection data as *CIDOC CRM* additionally enabled the development of *ResearchSpace*<sup>62</sup>, a tool that allows researchers to collate, visualise, and explore resources (themes to which I will return in Chapter 5), facilitating the conceptualisation of a research area. *ResearchSpace* is aimed at domain experts and does not require its users to possess specific technical skills (Oldman & Tanase, 2018). Another cultural heritage application of *CIDOC CRM* is *Linked Art*<sup>63</sup>, a community-led data model for describing art objects. Acknowledging the complexity issues outlined in 1.3, *Linked Art* uses a simplified version of *CIDOC CRM*, to minimise confusion and ambiguity (Linked Art Contributors, n.d.-a). Such an approach might therefore be a potential solution to *CIDOC CRM* implementation in a Humanities research context.

---

<sup>61</sup> At the time of my study, this service was accessible at <https://collection.britishmuseum.org/> but it has since become unavailable.

<sup>62</sup> <https://researchspace.org/>

<sup>63</sup> <https://linked.art/>



*CIDOC CRM* has also been applied to archaeological infrastructures. For example, its implementation in *Arachne*<sup>64</sup>, the object database of the German Archaeological Institute, has greatly facilitated its interoperability and potential for information exchange (Scriba & Stockinger, 2016). Indeed, the *CIDOC CRM* framework was used to link *Arachne* data with that of *Perseus*' art and archaeology collection as part of the *Hellespont Project* (G. R. Crane, 2014). Additionally, the *Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE)* project<sup>65</sup> integrated datasets from European national repositories by mapping the entities within them to *CIDOC CRM*. Users can access this data via the *ARIADNE Portal*, which includes visualisation and query functionality (Aloia et al., 2017; Meghini et al., 2017), demonstrating the potential for ontologies such as *CIDOC CRM* to form the basis of Linked Ancient World Data access via a usable interface.

Elsewhere in the domain of Ancient World research, there are relatively mature Linked Data solutions for the representation and discovery of specific objects such as coins, inscriptions, and papyri. These objects form particularly interesting use cases due to their combination of textual, visual, and material culture characteristics. A complex aspect of many papyri and some inscriptions is their fragmentary nature. Parts of the same text can appear on multiple physical objects, many are incomplete, and scholars must often use their best judgement regarding missing words. Integration of multiple collections is therefore essential for a more comprehensive understanding of their contents and meaning (Celano, 2018, p. 139). Contrastingly, coins are mass-produced objects with a specific set of attributes that applies to all instances, irrespective of chronological period or geographical area, which can often be classified using established typologies (S. Heath, 2018, p. 36). For example, every coin has obverse (heads) and reverse (tails) sides, and every coin was minted in a particular place. This predictable structure, incorporating discrete entities and concepts that can be identified with persistent URIs, as well as the numerous potential connections to external resources, makes Linked Data particularly appropriate for modelling numismatic data.

---

<sup>64</sup> <https://arachne.dainst.org/>

<sup>65</sup> <https://ariadne-infrastructure.eu/>

*Nomisma*<sup>66</sup> is a Linked Data ontology and resource provided by the American Numismatic Society (ANS)<sup>67</sup> to facilitate integration and discovery of data about coins, predominantly from the Ancient World. *Nomisma* links coin hoards, types and specific instances (S. Heath, 2018, p. 41), using terms from established ontologies and authority files, such as *CIDOC CRM* and *Pleiades* (2.2.1), to facilitate integration with external datasets (Gruber, 2016, p. 100; Gruber & Meadows, 2021). Like many of the resources discussed in this section, it was developed as a result of international collaboration between experts in the field, ensuring accurate representation of information and acceptance by the research community (Gruber, 2018, p. 55). In OCLC's recent Linked Data implementation survey (Smith-Yoshimura, 2018), *Nomisma* was among the eight most frequently used linked datasets in the cultural heritage domain, all of which received over 100,000 requests per day.

To include their data in *Nomisma*, contributors need not structure it as Linked Data themselves. Instead, they provide a Google spreadsheet, with headings that can be directly mapped to *Nomisma* terms, which is then validated and converted to RDF (Gruber, 2016, pp. 104–105). The resulting data can then be queried via a SPARQL endpoint, with the option to export in CSV format, which facilitates analysis using external software and tools (S. Heath, 2018, p. 50). The ANS also provides discovery mechanisms via resources such as *Online Coins of the Roman Empire (OCRE)*<sup>68</sup>, *Coin Hoards of the Roman Republic (CHRR)*<sup>69</sup>, *Coinage of the Roman Republic Online (CRRO)*<sup>70</sup> and *PELLA: Coinage of the Kings of Macedonia*<sup>71</sup>, as well as its overarching catalogue, *MANTIS*<sup>72</sup>. Providing multiple methods for searching *Nomisma* data ensures its consumption is neither restricted to more technically advanced users, nor limited by the confines of a visual interface.

---

<sup>66</sup> <http://nomisma.org/>

<sup>67</sup> <http://numismatics.org/>

<sup>68</sup> <http://numismatics.org/ocre/>

<sup>69</sup> <http://numismatics.org/chrr/>

<sup>70</sup> <http://numismatics.org/crro/>

<sup>71</sup> <http://numismatics.org/pella/>

<sup>72</sup> <http://numismatics.org/search/>

*Trismegistos*<sup>73</sup> provides access to metadata about documentary texts from the Ancient World, while linking externally to the texts themselves. These texts were predominantly written on papyri, but also include ostraca (pottery sherds), wooden tablets, and inscriptions; their information and contents were previously contained in other databases and print publications (Depauw, 2018, pp. 193–195; Depauw & Gheldof, 2014). Its remit was originally restricted to Egypt between 800 BCE and 800 CE but has since expanded geographically and is beginning to expand chronologically (Broux, 2017b). In addition to descriptive metadata about the texts, entities mentioned within them, such as places, people and chronological periods have been extracted using a combination of named entity recognition (NER) and manual editing (Broux, 2017a, p. 13; Broux & Depauw, 2015; Depauw, 2018, pp. 196–197). Although *Trismegistos* data is modelled using relational databases (Trismegistos, n.d.; Verreth, 2017, p. 202), it provides URIs in the form of TM-numbers (Depauw, 2018, p. 199), which are used in Linked Data resources such as *EAGLE* (discussed below), and *Pelagios'* (2.2.1) *Periplo* resource (Simon et al., 2016b).

In addition to providing metadata and URIs, *Trismegistos* has consumed this data to perform statistical analyses, uncovering trends over time (Depauw & Stolk, 2015), and to produce visualisations, particularly in the area of social network analysis (Broux, 2017b, 2017c; Depauw, 2018, pp. 197–198). However, while *Trismegistos* data remains openly available, since January 2020 this additional functionality, as well as performing more complex queries via the user interface, requires users (or their institutions) to pay a subscription fee (Trismegistos, 2019). Although this move was deemed necessary to ensure long-term sustainability of the resource, it is unfortunate that features with the potential to open up *Trismegistos* to a wider audience are now only available to a restricted subset of users.

*Papyri.info*<sup>74</sup> is an aggregated collection of digitised documentary papyri that integrates several online resources<sup>75</sup>, with disambiguation provided by *Trismegistos* URIs (Reggiani, 2017, p. 227). Its tools comprise the *Papyrological Navigator*, for

---

<sup>73</sup> <https://www.trismegistos.org/>

<sup>74</sup> <http://papyri.info/>

<sup>75</sup> The Duke Databank of Documentary Papyri, the Advanced Papyrological Information System, and the Heidelberg Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens

searching via a user interface with advanced query options, and the *Papyrological Editor*, a collaborative editing tool. Texts are encoded using EpiDoc<sup>76</sup>, and connected using RDF (Papyri.info, n.d.). The resulting resource not only comprises a searchable collection of texts, it also provides open, dynamic and collaborative digital critical editions, to which any user can contribute (Reggiani, 2017, pp. 241, 269). A recent review of Papyri.info (Vannini, 2018) praised its comprehensiveness, as well as its commitment to openness and sustainability, but there were criticisms relating to usability and documentation. Elsewhere, Bambaci et al. (2019, p. 28) have praised the relative ease with which users can encode text in the *Papyrological Editor*, using familiar papyrological conventions without compromising data quality and richness.

It is clear from a recent fundraising campaign to ensure *Papyri.info*'s sustainability (Torallas Tovar & Schubert, 2019), led by the Association Internationale de Papyrologues (AIP) and the American Society of Papyrologists (ASP), that *Papyri.info* now occupies a position of fundamental importance to the wider papyrological community, rather than appealing only to those who consider themselves digital humanists. These scholarly organisations have recognised that, despite considerable contributions by volunteers, a dedicated role is required for effectively managing the resource, which should not be funded by introducing subscription fees, ensuring that *Papyri.info* will remain openly available.

The *Europeana Network for Ancient Greek and Latin Epigraphy* (EAGLE)<sup>77</sup> is an initiative by Europeana<sup>78</sup>, a digital library based on a Linked Data model that aggregates multiple European digital collections. EAGLE provides access to images and metadata for ancient inscriptions by aggregating multiple existing databases. Contributing partners include the *Epigraphic Database Bari* (EDB), *Epigraphische Datenbank Heidelberg* (EDH) and *Trismegistos*. EAGLE facilitates discovery of these databases' contents via a single search interface that incorporates filtering, faceting, and image recognition capabilities (Prandoni et al., 2017). Its development took a collaborative approach, with feedback on models and processes actively sought from contributors to optimise

---

<sup>76</sup> EpiDoc (<https://sourceforge.net/p/epidoc/wiki/Home/>) is an extension of TEI (1.2), initially developed for encoding epigraphic documents, but which has since been applied to papyri and manuscripts.

<sup>77</sup> <https://www.eagle-network.eu/>

<sup>78</sup> <http://www.europeana.eu/portal/en>

the end result (Rocco, 2017, p. 128; Santucci et al., 2016, p. 53). Inscription texts are structured using EpiDoc, with artefacts modelled using the *EAGLE Common Metadata Model and Schema*, predominantly comprising selected terms from *CIDOC CRM*. These approaches are united using *EAGLE's Epigraphy Aggregation Conceptual Model (EACM)* (Mannocci et al., 2014, p. 291), with duplicate entries disambiguated using *Trismegistos* identifiers.

Central to *EAGLE's* objectives was broadening access to inscriptions to wider audiences beyond academia, which has included developing mobile and storytelling applications. The former provides visual search capabilities to enable in situ identification of inscriptions, while the latter enables users to view and create narratives using *EAGLE* resources, with the option to incorporate maps from *Pelagios* (2.2.1) (Liuzzo et al., 2017; Orlandi et al., 2014). In a recent review, Hedrick (2018) praised the storytelling app for its "elegant way" of integrating *EAGLE* materials, but found that *EAGLE's* main search interface can be difficult to use effectively for users unfamiliar with its constituent databases. These comments suggest that *EAGLE's* search interface (aimed predominantly at researchers) might benefit from some of the usability measures incorporated into its more creative resources (aimed at a broader public).

The above discussion has demonstrated how Linked Data has been used to represent different aspects of complex objects, using a variety of approaches. The examples discussed in this section demonstrate the potential for using Linked Data as the basis for creative and engaging outputs for a wide variety of users, while remaining academically rigorous. Related functionality might include discovery, analysis, and visualisation. The interlinked themes of funding and sustainability persist, as well as their actual and potential impacts on tool and resource usability.

### 2.2.5 Summary: Linked Humanities Data in Practice

In the section above, I have explored a series of key initiatives, tools and resources involved in the production and consumption of Linked Data for Ancient World research. Each took a collaborative approach to development, with the aim of ensuring usability and usefulness of the resulting tools and resources. Although they primarily focus on separate entity types, their producers have recognised the impact of these entities on each other; as such, there are many connections between them. Some resources are broad and multidisciplinary, and therefore ideal for making connections to the wider Linked Data ecosystem, with domain-specific initiatives often required to provide more accurate representations. In some cases, these representations have the potential to become quite complex, an issue that must often be balanced with the practicalities of a fixed-term research project. In many cases, it can be preferable to start with a simpler, incremental, approach to linking data, with scope to increase the complexity at a later stage. This approach additionally allows more time to develop a usable interface; however, even where such interfaces have been developed and are well-used, it is extremely difficult to guarantee their long-term sustainability while simultaneously ensuring that they are free to access.

Using the Ancient World as a case study, this section has demonstrated the breadth of Linked Humanities Data resources available. In doing so, I have additionally introduced some of the advantages of this approach, as well as barriers to these resources' implementation and use – themes to be explored in the following sections.

### 2.3 Advantages of Linked Humanities Data

In discussing various examples in 2.2 to demonstrate how Linked Data technologies have been implemented in a Humanities research context, I have also introduced several actual and potential advantages that this approach can bring. This section will explore these advantages in more detail, drawing on other resources where appropriate. I will start by considering the advantages of using URIs to identify concepts and align datasets, before describing how modelling data in this way can result in new knowledge via inferencing. To continue, I will explore how Linked Data technologies can represent complexity in digital objects, as well as how the resulting tools and resources facilitate the research process. Finally, I will demonstrate that the

openness of Linked Data standards and formats enables collaboration and sustainability.

Section 2.2 contained numerous examples where key concepts and entities are identified using URIs, providing a unique and consistent mechanism for referring to them, as well as disambiguating those with similar names and uniting equivalent terms in multiple languages. Assigning URIs to digital objects is recommended by the first of Berners-Lee's Linked Data principles (1.3); even if this is the only Linked Data principle followed by a project, it can nonetheless assist interoperability, as well as the understanding of terminology by both humans and machines.

Where appropriate vocabularies and authority files already exist, their URIs can be used by other initiatives, to ensure clarity and consistency of meaning and to connect their resources to machine-readable information from external sources. For example, rather than creating a new resource-specific list of places, place information in an Ancient World resource can be represented using the relevant *Pleiades* (2.2.1) URI(s) (Horne, 2020b). Additionally, Prag and Chartrand (2018, p. 248) attest that the approach of linking to established URIs can bring much-needed consistency to resources that rely heavily on project-specific implementations of standards that are more open to interpretation, such as EpiDoc. As well as providing efficiency by avoiding duplication of effort and potential redundancy (Buzi et al., 2018, p. 40), linking to established URIs (where possible) enables integration. Integrating related resources in this way leads to richer descriptions of concepts and entities throughout interlinked datasets, enhancing discoverability.

Once a dataset is modelled using established ontologies, a reasoning system can be used to automatically produce new knowledge, inferred from the relationships between entities, which would be difficult or impossible for a single user to discover manually. For example, in the *Sharing Ancient Wisdoms (SAWS)* project, automated reasoning was used to detect texts that were linked to the same translation but were not linked to each other (Tupman & Jordanous, 2014). Once the reasoner has elicited this information, it is the researcher's task to verify its likelihood based on any known gaps or issues inherent in the data (Meroño-Peñuela et al., 2014, pp. 20–21). Datasets

aligned using Linked Data technologies therefore have the potential to become more than the sum of their parts. However, few publications relating to Linked Humanities Data refer to the application of these capabilities, perhaps because thus far, emphasis has been placed on building infrastructures; substantial work on inferencing might then form the next stage of the process.

While any form of data modelling about real-life concepts and entities is an approximation, Linked Data can provide a more accurate means of representing these complexities, facilitating the process of finding commonalities and eliciting meaning. Modelling complexity is particularly important where the information about an entity comprises a collection of relationships that do not fall into a neat, hierarchical structure. For example, several of the object-based resources discussed in 2.2.4 use models that represent both objects themselves and any visual or textual elements they contain. In many cases, objects can often be most accurately represented using an event-based model like *CIDOC CRM* (1.3), rather than a table of characteristics or an increasingly complicated set of relational databases. Linked Data can therefore be a particularly effective approach for modelling complexity inherent in concepts and entities relevant to Humanities research.

Humanities research topics often involve the analysis and interpretation of evidence from multiple sources of different types; using a Linked Data approach brings these research objects together, allowing them to be explored in the same virtual space. Once datasets are connected using Linked Data techniques, multiple collections and repositories that previously existed in separate silos can be searched via a single federated query, with the potential to visualise their combined data (Bagnall & Heath, 2018, p. 184; Geser, 2016, p. 26; Meroño-Peñuela et al., 2014, p. 19). The ability to conduct one search instead of many reduces the time required for the discovery/data collection phase of a project, allowing more time to be spent on analysis and interpretation (Gruber, 2018, p. 59); in particular, SPARQL allows complex queries that interrogate multiple characteristics of a digital object simultaneously (Frunzeanu et al., 2016, p. 120). Therefore, Linked Data tools and resources have the potential to facilitate investigation into new research questions that would either be difficult or impossible to address using other technologies.



Linked Data tools and resources also have great potential for serendipity, exposing collections and datasets previously unknown to the researcher (Gruber, 2018, p. 62), such as those from relevant, but distantly related, subject domains (Geser, 2016, p. 50; Pagé-Perron, 2017, p. 11), or those in other languages (Koch & Koch, 2017, p. 171). As these datasets are often produced by a variety of organisations, for a wide range of purposes, they complement each other's information and provide the user with broader context for their object of study, potentially resulting in more accurate interpretations (Kansa et al., 2018, p. 503). Searching multiple datasets simultaneously can additionally reveal relationships between objects held in separate collections (Geser, 2016, p. 12; Rabinowitz et al., 2018, pp. 212–213), as well as facilitating virtual reunification of objects from the same place. Linked Data therefore facilitates interaction with multiple datasets simultaneously, for a variety of research purposes.

Use of Linked Data resources for discovery and analysis need not be restricted to researchers with an understanding of SPARQL. *Pelagios* (2.2.1), *EAGLE* and *Nomisma* (both 2.2.4) have demonstrated that such resources can be usable even for those with limited technical expertise or subject knowledge. Queries can be facilitated via detailed filters and faceting options (Simou et al., 2017, p. 221), with data made meaningful to a variety of audiences by building narratives around curated subsets of digital objects (Liuzzo et al., 2017, p. 521). There also exist user-friendly resources for creating and editing Linked Data, like *Pelagios'* annotation platform *Recogito*, rendering Linked Data production (at least on a small scale) possible for researchers without specific training. As with any type of technology or data structure, users with the requisite skills and experience are likely to derive the most benefit from Linked Data technologies; however, if resources are carefully designed with potential audiences in mind, all interested Humanities researchers should be able to engage with them.

Even if Linked Data producers are not able to build a resource themselves, the fact that their data uses open standards and vocabularies means it can more easily be accessed and used by others; for example, data structured using a Linked Data approach is more likely to be discoverable by "major search engines" (Koch & Koch, 2017, p. 171).

Similarly, providing an open dataset using Linked Data formats can encourage new collaboration opportunities that might result in the funding and expertise required to build a user-friendly interface (Buzi et al., 2018, p. 50). Therefore, as well as connecting datasets, Linked Data has the potential to bring Humanities researchers together and facilitates the process of building on previous work in overlapping subject areas.

The openness of Linked Data technologies additionally has positive implications for sustainability, both in terms of their immediate reuse potential and the relative ease of preserving data that uses open standards. In the longer term, as no specific software is required to engage with Linked Data formats, the data itself should continue to be usable (and reusable) for a variety of purposes (Biston-Moulin & Thiers, 2018, p. 160; Jordanous et al., 2012), even if the interface for its consumption reaches the end of its life. Although issues remain with the sustainability of the tools and interfaces for consuming Linked Data (as demonstrated by *Trismegistos* and *Papyri.info*, 2.2.4), choosing a Linked Data format is a positive step towards ensuring longevity and reuse of the underlying data.

The above discussion addresses RQ1c by demonstrating the advantages that Linked Data can provide to Humanities research. These include consistent and unambiguous identification, alignment of different data models and vocabularies, the production of new knowledge via inferencing, and the effective modelling of complex concepts and entities. Once the data has been produced, there are lightweight mechanisms to connect multiple datasets, on which usable resources can be built. Incorporation of open standards facilitates collaboration and ensures sustainability of the data (if not the tools or interfaces). All the above advantages result in Linked Data facilitating research processes such as discovery, annotation, and visualisation.

However, there has been less uptake of Linked Data for Humanities research than one might expect after exploring these advantages. This discrepancy is due to challenges and obstacles to its implementation, many of which have implications for its subsequent use. Such barriers will be the subject of the following section.

## 2.4 Barriers to Linked Humanities Data

As mentioned above, the current section will examine the barriers to Linked Data production and implementation. I will start by exploring technical barriers, including difficulties in representing concepts and entities accurately, challenges in aligning models that use different approaches, limitations of both simple and complex data models, problems when encountering uncertainty, and issues inherent in relying on external data sources. My discussion will then consider wider issues, including potential lack of awareness and training for data producers, the lack of usable tools and resources for producing and consuming Linked Data, and situations where enhanced discoverability of data would not be desirable.

Although linking to established URIs is generally advisable to ensure interoperability with existing datasets, in many cases they do not sufficiently represent what needs to be described. For example, when focusing on place (2.2.1), *GeoNames* is often inadequate for describing historical places, while locations in *Pleiades* do not have sufficient precision to allow detailed spatial analysis (Horne, 2020a, p. 218). As a result, data producers are often left with the choice between linking to established URIs that may not accurately represent the entity or concept they wish to describe or creating new URIs that link to their more established (but not exactly equivalent) counterparts. Where no suitable existing URIs are available, creating new ones would usually be more desirable in a scholarly context; however, it requires considerably more time than using existing resources and is often not a realistic prospect for fixed-term projects, due to their need for long-term sustainability.

Additionally, people differ in their interpretation of the meaning of terms used to represent entities and concepts, resulting in inconsistent application. For example, Prag and Chartrand (2018, p. 248) highlight the potential ambiguities caused by the *EAGLE* (2.2.4) vocabularies incorporating terms from existing sources. Even where clear definitions are available, the suitability of a particular term in a particular context remains subject to the implementer's interpretation of its meaning, which may differ from that of the user. This discrepancy could result in misunderstandings, although these can be mitigated with clear documentation to describe the producer's understanding of their terminology.

Similarly, there are often subtle disciplinary or theoretical differences between definitions of terms that might initially appear to be equivalent, which can cause difficulties in aligning ontologies, affecting subsequent information retrieval (Gerth, 2016, pp. 31–32). Such differences potentially lead to incompatibility with some researchers' views, or require them to adjust existing ways of working (Geser, 2016, pp. 12, 73; Meroño-Peñuela et al., 2014, p. 20). To compound this issue, Limp (2011, p. 278) states that in the archaeology domain, individual ontology development is seen as an integral part of scholarship, even when suitable terms and vocabularies already exist. Without some level of cooperation between data producers, this situation potentially results in project-specific data silos with models that are more difficult to align (Geser, 2016, p. 17). Projects such as *Federated Archaeological Information Management Systems (FAIMS)* have sought to mitigate issues of conflicting terminology by mapping local terms for particular concepts and entities to a core ontology (Ross et al., 2015, pp. 126–127); however, if there is not exact alignment between term definitions, mapping is either not possible or requires extensive documentation to advise users of potential inconsistencies.

As mentioned above, Linked Data provides many ways to model data about complex objects; however, difficulties can arise in aligning datasets modelled using different approaches. For example, papyri often exist in the form of multiple fragments that originally belonged to the same text; simultaneously, the same writing surface may have been reused for two or more texts. This situation has resulted in different approaches to identifying and numbering texts, fragments and writing surfaces (Ast & Essler, 2018, p. 69; Polis & Razanajao, 2016, p. 25). In the case of *Papyri.info* (2.2.4), some source collections choose to assign each text by a different scribe to a different URI, while others choose to identify all texts written on the same surface with the same URI (Baumann, 2013, p. 97; Cayless, 2011, p. 32; Reggiani, 2017, pp. 74–75). Therefore, using Linked Data technologies to aggregate multiple collections does not always provide a straightforward means of aligning data models, potentially resulting in inconsistency and confusion for the user.

As demonstrated in 2.2, different approaches to modelling Linked Humanities Data have varying levels of complexity, with accompanying advantages and disadvantages. Simple data models might provide efficiency and a low barrier to implementation, but often do not capture the degree of nuance required for research purposes (Liu et al., 2017, p. 350). Conversely, complex models like *CIDOC CRM* (1.3) provide a rich level of detail; however, their implementation often requires considerable time, and the multiple subtly different terms for similar concepts and entities cause inconsistency, reducing the scope for interoperability (Cayless, 2019, p. 46). Additionally, once such a complex ontology has been applied, it becomes more difficult to make inferences from the data because the number of conditions required for an inference to be confirmed increases with the number of terms used (Isaksen, 2011, p. 155). Due to these issues, producing and consuming data involving complex ontologies like *CIDOC CRM* might deter researchers from working with Linked Data at all in future. A compromise could involve implementing a simpler approach at the outset, while ensuring there is scope for adding further complexity. Such an approach might be achieved by encouraging producers to map some or all of their data to a set template (Binding et al., 2019, p. 371) or by applying a simplified version of the ontology (Linked Art Contributors, n.d.-a).

Much of the data that supports Humanities research is based on interpretation rather than fact. Some data will never reach the point where it can be considered ‘factual’; for example, representations of mythological entities about whom there is conflicting information from different sources (Bodard et al., 2017, p. 36). To take advantage of the benefits Linked Data provides, producers must consider how uncertainty can best be modelled within the data structure. For cases where several alternative values are possible, Thaller (2020) advocates a mathematical approach, where a probability is assigned to each value; Niccolucci and Hermon (2017) demonstrate how a similar approach might be modelled using *CIDOC CRM*. However, in many cases it is not feasible to perform a reliable calculation of probability (e.g., if the relative frequencies of different possible values cannot readily be estimated), and assigning probability based on the data producer’s level of certainty would be extremely subjective. As an alternative, *PeriodO* (2.2.2), *Pleiades* and *WHG* (both 2.2.1) mirror more traditional scholarly practice by modelling information as cited assertions that link to the original

source, which Golden and Shaw (2016) refer to as nanopublications. Despite these efforts, there is still no universally agreed method for expressing and communicating uncertainty in Linked Humanities Data and this is likely to remain the case. Humanities researchers should always treat both digital and physical sources with an element of caution; their critical evaluation can be assisted by appropriate information about these sources' limitations.

In addition to the uncertainty issues described above, another barrier to incorporating data from external sources is the lack of control over its availability (Geser, 2016, p. 15). Some resources rely on querying an external SPARQL endpoint; however, their availability can be unpredictable. For example, Gerth et al. (2016, p. 17) had originally intended to incorporate data from the British Museum in their experiments with integrating sculpture datasets, but were prevented from doing so due to unavailability of the SPARQL endpoint. Additionally, in their 27-month study of 427 SPARQL endpoints encompassing different subject areas, Buil-Aranda et al. (2013, pp. 289–290) found that only 32.2% were available 99-100% of the time, while 29.3% were available less than 5% of the time. They note that many endpoints in the latter category were produced through experimentation with the technology, rather than long-term resource provision, and are now permanently unavailable. As well as causing functional issues, resource unavailability can affect citations, a particular concern when referring to online sources in scholarly publications (Hannemann & Kett, 2010, p. 2), and may perpetuate the idea that online material is not stable enough to be cited.

Similarly, it is often unclear whether data held in external resources is accurate and up to date (Calvanese et al., 2016, p. 213); therefore, any inaccuracies could potentially be reproduced across multiple resources (Angjeli et al., 2014, p. 2). A 2013 survey by AthenaPlus found that participants were concerned about the implications for data quality when relying on external content (Geser, 2016, p. 11). Such reliability issues might explain why many 'linked datasets' contain only internal links, rather than connecting to external resources (Isaksen, 2011, p. 64); however, this phenomenon may also be due to an actual or perceived lack of relevant external datasets and ontologies, or not being aware of the benefits of linking to more general resources. Quality issues can be mitigated with appropriate documentation, including provenance

information, to assure potential users that the data is trustworthy (Geser, 2016, pp. 69–71).

The above barriers are technical in nature and predominantly apply to tool and resource producers who are already familiar with Linked Data. However, many more researchers planning the production of digital tools and resources are new to this approach. Others are aware of its existence but remain to be convinced about the benefits and whether they outweigh the challenges. Before these researchers even consider the technical obstacles to Linked Data implementation, there are wider issues that must be addressed.

Many Humanities researchers are not aware of Linked Data technologies, the potential benefits of this approach in the context of their research topics, or how it might be implemented. This situation is likely due to Linked Data rarely being covered in standard institutional training offered to staff and students, which usually focuses on spreadsheets and relational databases. As a result, researchers often tend to think in a tabular format, without considering that their data might be better represented by a networked, graph structure (Barbera, 2013, p. 96; Ross et al., 2015, p. 118). Even for those researchers who are aware of the potential benefits of Linked Data technologies, substantial training is often required for their effective implementation, which may not seem a realistic approach to take within the tight time constraints of funded projects (Isaksen, 2011, pp. 153–154; van Hooland & Verborgh, 2014, p. 51). For example, Granados-García (2020, pp. 261–264) found that using a Linked Open Data approach required far more time and training than initially anticipated and recommended that future data producers ensure they have sufficient technical knowledge and support (or the time to acquire them) before embarking on similar projects. In fact, Smith-Yoshimura (2018) found when surveying Linked Data producers from the cultural heritage domain that the main barrier to producing Linked Data was a "steep learning curve for staff", indicating the importance of addressing this fundamental obstacle.

While I have noted implications for project timescales throughout the above discussion, time constraints are a particular concern for researchers who are unfamiliar with Linked Data and its production. This situation is compounded by

difficulty in accessing, or unawareness of, adequate training, and often results in the prioritisation of immediate research objectives rather than longer-term usage of the dataset, tool, or resource beyond the lifetime of the project. To avoid these issues, and ensure that high quality, reusable data is produced, Gerth (2016, p. 14) recommends that Linked Data should be produced via a collaborative process involving both technology experts and domain specialists. However, Geser (2016, pp. 12, 56) acknowledges that, while ideal, expert support is not always available, and that the development of usable tools to support Linked Data production could be a more effective solution to minimise the need for training.

Although the need for training would be partially addressed by user-friendly tools that allow non-technical researchers to produce or enhance Linked Humanities Data, such tools are few in number (Barbera, 2013, p. 98; Thiery & Engel, 2016, p. 259). Additionally, as these tools are often produced with the goal of academic experimentation rather than long-term usability, they have rarely been tested sufficiently to ensure consistency and reliability, and are often not adequately maintained (Geser, 2016, pp. 55–56). *Pelagios'* (2.2.1) *Recogito* platform (Simon et al., 2017, 2019) and the Semantic Technologies Enhancing Links and Linked data for Archaeological Resources (STELLAR) applications<sup>79</sup> (Geser, 2016, pp. 12, 58) are notable exceptions. Until the learning curve for developing Linked Data resources is reduced, the temptation will likely be to work with familiar data structures; a potentially 'safer' option that ensures development of a usable resource within the funded period, while maximising the amount of time available for research.

In addition to requiring effective tools to produce Linked Data, the resulting data should be made accessible to end users via a usable interface. However, due to its relatively complex structure (and often, producers' lack of familiarity thereof), building a Linked Data driven resource and ensuring its usability by non-technical researchers can be more difficult and time-consuming than a similar resource based on tabular data. If insufficient development time is available within the funded period of a project to produce a usable interface, its absence significantly limits the number of people

---

<sup>79</sup> <https://hypermedia.research.southwales.ac.uk/kos/stellar/stellar-applications/>



who could potentially engage with and benefit from the dataset. As a result, some resources, such as *SNAP:DRGN* (2.2.3), exist in the form of rich but largely inaccessible datasets that must be queried via a SPARQL endpoint.

Even users who are familiar with SPARQL might not always be able to make effective use of a SPARQL endpoint, however, as they first need to be familiar with the way in which new and existing ontologies and data models have been implemented in this particular context (Calvanese et al., 2016, p. 214). The W3C recommends that providers describe linked datasets using the Vocabulary of Interlinked Datasets (VoID), which includes specifying ontologies used (Alexander et al., 2011). However, such a description is not always included; out of 427 SPARQL endpoints, Buil-Aranda et al. (2013, p. 280) found that approximately two thirds were not accompanied by a VoID description. Additionally, a single query has the potential to generate huge amounts of data, which is both unhelpful to the consumer and computationally intensive for the host. Therefore, even for technically experienced users, direct access to a SPARQL endpoint is not always advisable; providing access via an interface or API is usually preferable (Gunter et al., 2019, pp. 4–5; Schweizer & Geer, 2019, p. 1).

In cases where a usable interface can be produced in the time available, producers should acknowledge that methods for presenting and visualising Linked Data may be unfamiliar to Humanities researchers. For example, many potential users may not have prior experience or training in network analysis (Barker, 2020), which may result in an inability to use the resource to its full potential, or to inaccurate interpretations of their query results. Accurate interpretation by the user is particularly important for resources with inferencing capabilities; as Hickey and Toves (2014) warn, results will reflect any ambiguities or inconsistencies in the original data, and as such should always be checked by domain experts.

Creating a usable resource for exploring Linked Data can, however, have some negative implications; making this data more discoverable can amplify information now known to be incorrect, terminology now considered to be offensive (Gibson & Kahn, 2016; Modest & Lelijveld, 2018), or potentially distressing images (Holterhoff, 2017). This issue is common to all data types, but its effects can be more pronounced

where discovery is aided via semantic enrichment. Avoiding or mitigating this issue often requires extensive work to update the dataset before linking can be considered, which poses a particular barrier for smaller institutions with limited funding. Otherwise, publishing the original dataset as Linked Data could have harmful consequences for users, as well as potentially perpetuating outdated language and interpretations in subsequent academic research.

Similarly, the exact locations of heritage sites are often considered sensitive information due to the potential for looting or vandalism (Tolle & Wigg-Wolf, 2016, p. 277). To mitigate this, while allowing some discoverability, Kansa et al. (2018, p. 494) reduced the level of precision to which locations are described in the *Digital Index of North American Archaeology*. As well as altering newly published datasets in this way, similar actions may need to be taken for data that already exists online, but where integration with other datasets or richer description may significantly increase its visibility.

In this section, I have discussed known barriers to Linked Data implementation, addressing RQ1d. Many are due to technological challenges, such as the accurate representation of concepts and entities, the difficulty in aligning data from sources structured in different ways, the disadvantages of selecting either a simple or complex data model, the particular importance in Humanities research of effectively communicating uncertainty, and the reliability issues inherent in full or partial dependence on external sources. Even researchers with significant familiarity in Linked Data production will need to assess the potential impacts of these barriers on project timescales before opting to implement this approach.

The final three barriers encompass much broader challenges, particularly for researchers who are unfamiliar with Linked Data or inexperienced in its production. The first concerns a general lack of awareness among Humanities researchers of the benefits of Linked Data, with little training available to rectify this, and few use cases demonstrating its effective implementation. The second is the relatively small number of usable tools or resources for producing or consuming Linked Data, potentially due to their being produced for experimentation with the technology, rather than long-

term availability. The third concerns instances where enhanced data discoverability is undesirable, e.g., for collections containing sensitive material. These higher-level barriers must be addressed before potential producers can even consider the above technological challenges.

Although the challenges and obstacles to production and consumption of Linked Data are substantial, they are not insurmountable. While expert consultation is usually an ideal solution, many barriers can be mitigated with access to information, training, and usable tools and resources; however, time and money continue to be the ultimate obstacles to Linked Data implementation. In the above discussion, I have largely focused on barriers to Linked Data production; however, the way in which each of these barriers is addressed (or not) has significant implications for the user. Insights into users' experiences of Linked Data tools and resources in a Humanities context could therefore be extremely valuable for determining how best to address such barriers; however, there is little existing research in this area.

## 2.5 Conclusions: Linked Data and Humanities Research

In this chapter, I have addressed RQ1 by exploring the current situation with regard to Linked Humanities Data, including the extent to which this has occurred (2.1); some examples of relevant tools, resources and data structures (2.2); the advantages of this approach (2.3), and potential barriers to its implementation (2.4). Analysing data on existing projects demonstrated that Linked Data implementation is still rare in Humanities research, but that a relatively high proportion of Linked Humanities Data projects relate to the Ancient World. These are often focused on representation and discovery of different entity types, such as place, time, people, and objects, and take different data modelling approaches with varying levels of complexity and granularity. Exploring these examples highlighted how implementing Linked Data can improve the quality, discoverability, and interoperability of Humanities datasets.

During my discussion of the advantages of, and barriers to, production and consumption of Linked Data in Humanities research, I have identified several key themes, which both affect and determine the usability of the resulting tools and resources:

- Awareness
- Interface design
- Discovery
- Openness
- Reliability
- Data quality
- Analysis
- Documentation
- Collaboration
- Sustainability

To better understand, and intervene in, the community of practice made up of Ancient World researchers engaging with Linked Data, further research is required to explore the user's perspective in relation to the above themes. In the following chapters, I aim to address this gap in the existing literature by discussing my study of Ancient World researchers' experiences with digital tools and resources in general, and Linked Data in particular. This user research should either confirm the above advantages and barriers, or offer differing views, as well as providing an indication of areas to prioritise in future developments.

While the main body of this thesis will comprise my findings and analysis, in the next chapter I will outline my survey and interview methodologies.

### 3 Methodology

After addressing my first set of research questions in the previous chapter, my remaining questions relate to methods used in Humanities research, their potential for Linked Data integration, and the usability of the resulting tools and resources. While I will address the first of these in reviewing the literature on Digital Humanities research methods (3.1), the remaining two issues require a new study that involves consulting the researchers themselves. I will outline various user research methods that have been utilised in the Digital Humanities (3.2) before exploring my chosen methods in sufficient detail to ensure transparency, as well as comparability with similar projects in future. The first component of my user research was a survey of Ancient World researchers (3.3), whose demographic characteristics are presented in 3.4. Using these characteristics, I selected a sample of participants to interview about their experiences, thereby obtaining more detailed information; the interview phase is outlined in 3.5. This chapter will then conclude (3.6) by looking forward to the discussion of my findings.

#### 3.1 Describing Digital Humanities Research Methods

Alongside an increase in the application of digital technologies to Humanities research since the latter part of the 20<sup>th</sup> century, various initiatives have focused on the classification of research methods – both those relating to the Digital Humanities specifically, as well as Humanities research more broadly. In discussing these frameworks below, I will begin with Unsworth's (2000) Scholarly Primitives and McCarty and Short's (2002) Methodological Commons. These ideas provided a foundation for several later frameworks, including the Arts and Humanities ICT Methods Taxonomy, which had a specific Digital Humanities focus. More recently, *TaDiRAH* has used these past frameworks to produce a detailed vocabulary in response to user needs.

The Scholarly Primitives were introduced by Unsworth (2000), who suggested that many research activities, regardless of discipline, could be classified under one or more of the seven basic functions of 'Discovering', 'Annotating', 'Comparing', 'Referring', 'Sampling', 'Illustrating', and 'Representing'. With this framework, Unsworth aimed to provide a focus for future tool development in the domain of Digital Humanities, by

ensuring that these tools addressed the needs of the relevant primitive rather than those of a specific project (to promote sustainability and minimise duplication). Examining Humanities research methods from their most basic functions before building up the level of detail in their description can therefore have an impact on how Humanities research is conducted in future. Unsworth's Scholarly Primitives have been cited by several more recent initiatives, such as *TaDiRAH* (below), as their fundamental influence (Borek et al., 2016, para. 5), as well as being revised and expanded by Palmer et al. (2009).

Another format for describing Humanities research was developed by McCarty and Short (2002) and is shown in Figure 3.1. This diagram visualises the interface between 'traditional' Humanities research methodologies and the application of digital tools and methods. The central Methodological Commons represents digital techniques that can be applied across traditional disciplinary boundaries and incorporates the "major data-types encountered in the humanities". McCarty and Short are keen to point out that this model is not intended to be exhaustive, and that it is expected to evolve over time.

To explore specific use cases and related issues, a more granular approach to research methods classification will be required. A detailed taxonomy<sup>80</sup>, initially focused on digitisation methods, was developed by the UK's Arts and Humanities Data Service (AHDS) in 2003. This taxonomy was later expanded by the AHRC ICT Methods Network during 2005-2008 and used to classify approximately 400 AHRC-funded projects with a digital output (AHRC ICT Methods Network, 2008). Once this project ended, the taxonomy was hosted by the Centre for e-Research at King's College London as part of the arts-humanities.net website, and continued to be enhanced and updated until funding ceased in 2011 (Hughes et al., 2016, pp. 155–156).

---

<sup>80</sup> A controlled vocabulary structured in a hierarchical format

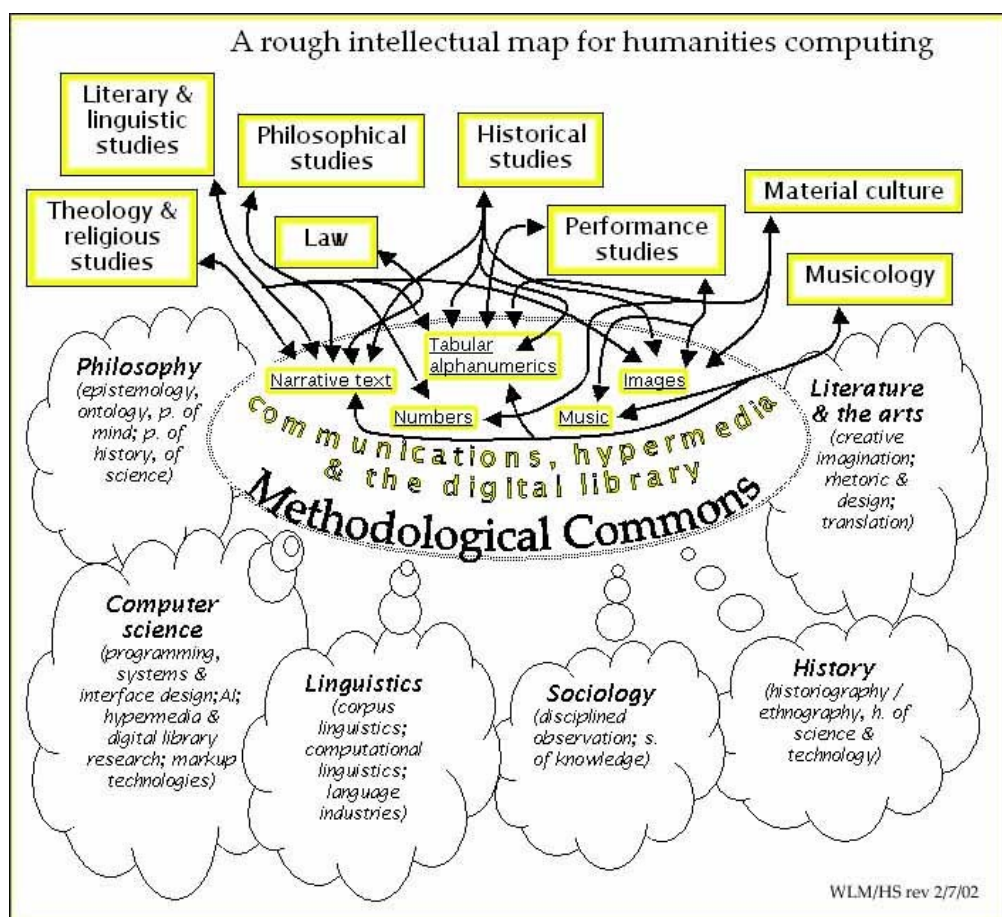


Figure 3.1 The Methodological Commons (McCarty & Short, 2002)

The Scholarly Primitives and the arts-humanities.net taxonomy have together formed the basis for a recent ontology, the *Taxonomy for Digital Research Activities in the Humanities* (TaDiRAH)<sup>81</sup>. TaDiRAH was developed by members of the *Digital Research Tools* (DiRT)<sup>82</sup> Steering Committee and the Digital Research Infrastructure for Arts and Humanities in Germany (DARIAH-DE) (Borek et al., 2016, para. 1; Dombrowski & Perkins, 2014). The short-term goal was to improve discoverability within DiRT and DARIAH's *Doing Digital Humanities* bibliography<sup>83</sup> by bringing together concepts from existing vocabularies relating to digital tools and methods (Borek et al., 2016, para. 8; Dombrowski & Perkins, 2014; Perkins et al., 2014, pp. 181–182). Prior to the inception of TaDiRAH, it was difficult to find individual records within these resources because users first had to know the specific category to which they had been assigned. As many tools and resources spanned multiple research activities, it was decided that a new

<sup>81</sup> <http://tadirah.dariah.eu/vocab/index.php>

<sup>82</sup> Now integrated with the *Text Analysis Portal for Research* (TAPoR), <http://tapor.ca/home>

<sup>83</sup> [https://www.zotero.org/groups/doing\\_digital\\_humanities\\_-\\_a\\_dariah\\_bibliography/items/order/creator/sort/asc](https://www.zotero.org/groups/doing_digital_humanities_-_a_dariah_bibliography/items/order/creator/sort/asc)

taxonomy would provide a better context in which to classify them (Borek et al., 2016, paras 2–3).

Terms are divided into ‘Research Activities’, ‘Research Objects’ and ‘Research Techniques’, with ‘Research Activities’ subdivided into eight top-level goals (‘Capture’, ‘Creation’, ‘Enrichment’, ‘Analysis’, ‘Interpretation’, ‘Storage’, ‘Dissemination’ and ‘Meta-Activities’ (TaDiRAH, 2014e)), influenced by the Scholarly Primitives (Borek et al., 2016, para. 5). Incorporating terms from multiple sources and situating them within a broader, overarching framework ensures that all tools and resources can be described at a detailed level, while providing a structure that relates to more general research methods. One reason for using existing vocabularies as the foundation for *TaDiRAH* was its focus on potential users, who are more likely to adopt a scheme with which they have some familiarity (Borek et al., 2016, para. 14). To enhance this structure further, *TaDiRAH* has recently been made available as Linked Data (Tóth-Czifra, 2021). User testing was pivotal to the development process (Perkins et al., 2014, p. 182), which should ensure wide adoption and sufficient appetite for it to be maintained in the future. Using a framework like *TaDiRAH* could facilitate the evaluation of future projects, to ensure that the user’s needs are being met for the planned research goals of the resource.

In exploring initiatives for describing Digital Humanities research methods, I found that the Scholarly Primitives provide a generic, overarching approach to describing research methods, which demonstrates continued applicability, where a more specific vocabulary may have become deprecated. The Methodological Commons similarly illustrates the blurring of disciplinary boundaries, while additionally making explicit the data types and underlying theories involved. Of particular importance where Linked Data is concerned, these entities and concepts are not only listed, but are shown in context with the connections between them. The more granular vocabulary of the AHRC ICT Methods Network/arts-humanities.net has been practically applied to the classification of methods from past research projects, although it has not been updated since 2011 (which calls its accuracy and currency into question).



I therefore concluded that a similar but more recent framework, *TaDiRAH*, would be the most effective means of classifying Humanities research methods in my own study, before assessing their potential for Linked Data integration. While the broad concepts of the Scholarly Primitives might be sufficient for identifying overarching activities in Humanities Linked Data projects, *TaDiRAH* provides a more specific indication of areas on which to focus – nuances that might be missed if using a more general vocabulary. Furthermore, exploring the development of *TaDiRAH* has provided insights into the connection between usability and sustainability identified in the previous chapter, as well as the benefits of user-centred design, which relates to user research, the topic of the following section.

### 3.2 User Research in the Digital Humanities

While user research in the Digital Humanities is often concerned with the evaluation of a specific tool or resource as part of its development process, several recent projects have conducted user research into how such tools and resources are being used as part of the wider digital ecosystem. This research often involved a combination of quantitative and qualitative approaches, to achieve both breadth and depth. Learning from these initiatives has been instrumental in informing my methodology, enabling me to employ appropriate user research methods for eliciting information about Linked Data resource usability and to explore where such resources might best be integrated with existing research methods.

The quantitative component of user research projects in Digital Humanities often comprises a survey (although other methods, such as server log analysis (Warwick et al., 2012), have been applied in some studies). Broadly, a survey involves a series of structured questions to obtain information and opinions about real world issue(s) from a wide variety of people. Results can be readily analysed and compared both within the survey itself and with similar research. If the target population is well-defined and the response rate is sufficient, results can be generalised across the population. Additionally, surveys are cheaper to administer to large numbers of participants than other user research methods.

Isaksen's (2011) survey on the implementation of semantic technologies for publishing cultural heritage data is particularly pertinent to this thesis. 67 participants were selected due to their involvement in 57 relevant projects (Isaksen, 2011, p. 49). In analysing their responses, Isaksen found that only one project met all four of Berners-Lee's original Linked Data principles (1.3). Furthermore, many datasets were neither openly available nor linked to external resources, despite many participants citing data sharing and integration capabilities as major advantages of semantic technologies. Disadvantages mentioned by Isaksen's participants included the difficulty of understanding such technologies, insufficient training and documentation, and poor user experience (Isaksen, 2011, pp. 64–66).

A broader example of a Digital Humanities survey, the DARIAH Survey on Digital Practices in the Arts and Humanities, aimed to explore how European Humanities researchers were using digital technologies, to inform future infrastructure developments. Methods they hoped to capture included "how [humanists] organize, manage, enrich, annotate, use and disseminate research resources", in addition to search and discovery (Papaki et al., 2015) – a list reminiscent of the Scholarly Primitives discussed in 3.1. The survey received 2,177 responses; when asked to rate their different needs in order of importance, participants almost unanimously placed "improved findability and access to existing digital research resources or data" as the top priority, with an average score of 9.5 out of 10 (Dallas et al., 2017).

While surveys can be an efficient means of obtaining information from a large population in a short period of time, they often do not provide the degree of nuance or flexibility inherent in qualitative approaches. In Digital Humanities, this often takes the form of user observations when evaluating an individual tool or resource (e.g., Wusteman's (2017) study on the *Letters of 1916*<sup>84</sup> resource). However, for studies with a broader remit that encompasses multiple tools and resources, interviews tend to be preferred. An interview is a specific type of conversation where one person (the researcher) aims to elicit information from the other person (the participant) (della Porta, 2014, p. 228; Oates, 2005, p. 186; Vogt et al., 2017, p. 32). Interviews can be an

---

<sup>84</sup> <http://letters1916.maynoothuniversity.ie/>

extremely valuable way of eliciting participants' feelings or experiences about a particular phenomenon or event (Vogt et al., 2017, p. 39), or (in human-computer interaction research) their experiences of interacting with a particular type of digital technology (Portigal, 2013, pp. 3–6).

In a recent study focusing on digital methods used by Humanities researchers, Ridge (2016) conducted 29 semi-structured interviews of between one and two hours to gain insights into the digital research practices of academic, local and family historians, selected through purposive sampling. Discussion was structured based on an adapted version of the Scholarly Primitives, informed by the work of Palmer et al. (2009) and the development of *TaDiRAH*. Ridge found that digital tools and resources, particularly those involving crowdsourcing, had made historical research processes more accessible to a wider audience; however, these processes could be improved via a usable, automated means of producing structured, semantic, data from strings of text.

Elsewhere, development of the *Scholarly Research Activity Model (SRAM)* was informed by interview research. *SRAM* is a framework based on a holistic understanding of Humanities research practices and user experiences, with the purpose of informing future digital resource development (Benardou et al., 2013, p. 105). The project incorporated 15 semi-structured interviews with Humanities researchers of between 45 minutes and two hours duration, which discussed how participants approached research topics and searched for relevant sources (Benardou et al., 2013, pp. 113–115). Findings were used to develop a series of requirements for future research infrastructures. These include managing multiple types of digital object (including primary and secondary sources), providing clarity regarding access conditions, "semantic interoperability" (incorporating Linked Data), collaboration, and including a "registry of resources, services and tools" (Benardou et al., 2013, pp. 120–123). While the third requirement relates most obviously to the topic of this thesis, the others relate to themes that emerged from my findings (Chapters 5-7) and informed my recommendations (Chapter 8).

The above studies have demonstrated how interviews can effectively be used to elicit information about participants' digital research activities and user experience.

However, the following examples will demonstrate how survey and interview research can be combined to ensure that the relevant initiatives benefit from the advantages of both quantitative and qualitative approaches. Kemman et al. (2014) conducted research into searching behaviours of scholars using electronic databases and archives, which similarly involved both quantitative and qualitative methods. Initially, they conducted an online survey, achieving 288 respondents, before interviewing three participants in more detail. They found that the most frequently used search resources were *Google* and *JSTOR*, but that the range of resources used increased with participants' information retrieval self-efficacy, concluding that these findings indicate a preference for efficiency and ease of use over trust in a resource's context and provenance (Kemman et al., 2014). This study demonstrates an effective combination of methods to produce results, but the small number of interviews might be one reason why the authors highlight the need for more detailed research when discussing their future plans.

Another key initiative was conducted by Europeana Cloud in 2015, to explore how the content and structure of *Europeana* might be improved. As part of this work, 65 Humanities scholars completed an online survey to share their experiences of using digital resources (Angelis et al., 2015, p. 17). The survey found that search queries tend to focus on specific features like those identified in 2.2, i.e., people, places, and time periods (Angelis et al., 2015, p. 25). However, usage of *Europeana* itself was surprisingly low - only 10% of respondents used it at least once a month, and 41% had never used it at all (Angelis et al., 2015, p. 21). This corresponds with findings by Kemman et al. (above), where *Europeana* was found to be one of the least used resources by Humanities researchers. Recommendations included improving the coherence of distinct collections and themes by combining individual records under overarching descriptions (Angelis et al., 2015, pp. 222–223).

The *Europeana* survey was followed by a series of case studies, each of which was focused on a digital research tool in the subject domain of History, Sociology, or Education, and explored how this tool was used by individuals or groups of researchers. Desk research and observations were supplemented by six semi-structured one-hour interviews, which informed the *Europeana* team about which

features of each tool tended to be used, researchers' experiences of using these features, and tasks for which other tools or resources were preferred. Responses were used to identify where integration of the tool with *Europeana* might be most effective (Angelis et al., 2015; Benardou & Dunning, 2018).

Reviewing previous user research in Digital Humanities has provided further evidence for the crucial role of discoverability, while additionally shedding light on some of the usability issues in Linked Data tools and resources. These include deficiencies in the areas of documentation and training, as well as the need for curated content to facilitate access by users unfamiliar with a resource's structure. The effectiveness of surveys and interviews in previous Digital Humanities projects, alongside the implementation of frameworks for describing research activities (in some cases), indicated that these would be appropriate methods to apply in my study. Using both quantitative and qualitative approaches in this way enabled me to build up a broad picture of Linked Data use and production in Ancient World research, supplemented by more detailed information on participants' experiences. My application of these methods is discussed in detail in the following sections.

### 3.3 Survey

The first, quantitative, stage of my study comprised a survey of Ancient World researchers about their experiences with digital tools and resources in general and their use and/or production of Linked Data in particular. Construction of this survey was informed by the Digital Humanities examples above, as well as literature from the Social Sciences domain. In this section, I will discuss design and implementation of my survey, starting with an overview of the target audience, before considering question formats and content, as well as factors relating to survey publication and participant recruitment.

Based on my initial hypothesis that Linked Data use among Humanities researchers is relatively low, I anticipated that conducting a survey focused solely on Linked Data would yield few responses. Additionally, when considering usability of future Linked Data resources and their integration with existing research methods, it is also important to look at the wider context of digital resource use, particularly among

researchers unfamiliar with Linked Data, or who have encountered barriers when attempting to use such resources. I therefore aimed the survey at all Ancient World researchers to encompass the full spectrum of digital experience and competence. I defined 'Ancient World researchers' as anyone who performs any kind of research activity in relation to this subject area, which could potentially include e.g., cultural heritage practitioners and field archaeologists, as well as academic staff and students. Response rates to online surveys tend to be higher among participants who have a personal interest in the survey topic (Keusch, 2015, p. 200). Targeting a survey at participants with a particular research interest (i.e., the Ancient World) should therefore have boosted the response rate; however, it was also important to describe the survey in a way that piqued the interest of those who may be less digitally inclined.

Bearing this broad target audience in mind, the survey had three main sections aimed at different segments of the population:

- 1) All participants (general questions on digital tool or resource use and associated methods)
- 2) Participants who have knowingly used one or more Linked Data tools or resources
- 3) Participants who have produced digital tools or resources, with additional questions for those who have produced Linked Data

Categorising potential participants in this way produced results that were applicable to a broad range of Ancient World researchers, including more specialist groups with differing goals and levels of technical expertise.

The majority of questions in the survey (Appendix 1) relate directly to one or more of my wider research questions (1.1), as recommended by Vogt et al. (2017, p. 18). Following advice from Adams and Cox (2008, p. 19), I grouped survey questions into themed sections using clear headings, and based on the audience categories listed above. As per Dillman, Smyth and Christian's (2014, pp. 320–321) recommendation, all questions were optional, to ensure that participants were not blocked from completing the survey if they were unable to or chose not to answer particular questions.

My survey incorporated both open and closed question types. Closed questions (where the participant must select their answer from a series of options) can be advantageous for ease and speed of survey completion and analysis; however, they can also be restrictive. Participants may not feel that any of the options accurately represents them, and if they do not have the opportunity or the time to explain their answer, the resulting data is unreliable. Open questions (where the participant writes their answer in a text box) resolve this issue by allowing a potentially infinite variety of responses. However, open questions are more difficult and time consuming to answer, leading to rushed and/or incomplete responses, skipping questions, or withdrawal from the survey. Additionally, considerably more time and resources are required to analyse data from open questions (Kelley et al., 2003, p. 263; Lauer et al., 2013, p. 344; Oates, 2005, p. 223). I therefore aimed to have an appropriate balance between both question formats; however, there was a noticeable drop in participation from the first open question onwards.

Early in the survey, participants were asked to identify positive features of existing tools and resources (Q10), and barriers to their use (Q11). Participants were then asked if they were familiar with the term 'Linked Data' (Q12) and if they had knowingly used one or more Linked Data resources (Q13). If a participant replied 'No' or 'Unsure' to the latter question, the survey used conditional logic to skip the next section; this approach is recommended by Lauer, McLeod and Blythe (2013, p. 342) to make the user's experience as efficient as possible and meets Dillman, Smyth and Christian's (2014, p. 126) recommendation, "Make sure the question applies to the respondent".

Those who had knowingly used at least one Linked Data tool or resource were then presented with the Linked Data users' section. Participants were asked to list all such tools or resources they had used (Q14) before selecting the one with which they were most familiar (Q15-16). The remaining questions in this section focused on this one tool or resource, with the assumption that answering on the most familiar tool or resource would yield the most comprehensive and useful answers. Participants were asked to state any advantages that Linked Data brought to the tool or resource (Q17a), as well as commenting on its usability (Q20-22).

All participants were then asked if they had been involved in the production of one or more digital tools or resources (Q23); those who responded in the negative then skipped to the end of the survey. To gain an insight into current alternatives to Linked Data, participants were asked for their preferred approach to structuring data (Q23a). The following question asked if they had produced one or more tools or resources based on Linked Data (Q23b); those who responded 'No' were asked why (Q24), before skipping to the end of the survey. Participants who had produced one or more Linked Data tools or resources were then asked a series of questions about the one they had produced most recently (assuming development of this tool or resource would incorporate lessons learned from previous projects). Participants were asked about its target audience (Q28), as well as the data's compliance with Berners-Lee's (2010) five-star model (Q29). Finally, participants were asked whether they would use a Linked Data approach again (Q30) and if they had any further comments (Q31).

All participants were additionally asked for a small amount of demographic information (Q1-5), as recommended by Green (2014, p. 38) to facilitate comparison of results across different groups, such as career stages. Q7 additionally asked participants to gauge their digital competence by placing their agreement with each of a series of statements on a five-point Likert scale. This is similar to the "information retrieval self-efficacy" questions asked in Kemman et al.'s (2014) survey, which assisted analysis by facilitating the division of participants into groups based on their perceived level of expertise. However, some studies suggest participants may provide the response they perceive as being the most desirable, rather than the most accurate (Adams & Cox, 2008, p. 20; Vogt et al., 2017, p. 17). It was therefore important to consider that the digital focus of the survey might have encouraged participants to overestimate their digital competence; it might be more accurate to describe their responses as assessing their level of digital confidence.

A key aim of the survey was to establish which research methods have the strongest association with Linked Data, thereby demonstrating where it might best be integrated. Following my discussion of initiatives for Humanities research methods classification (3.1), I opted to incorporate *TaDIRAH* into the survey, with the aim of



using consistent, structured terminology that would facilitate my analysis, resulting in reliable conclusions. I asked participants to select one or more *TaDiRAH* research activities that they associated with their use of digital tools and resources in general (Q6); relevant participants were later asked to do the same regarding their use (Q18) and production (Q27) of Linked Data.

Another area of the survey where I attempted to provide additional structure to participant responses was Linked Data tool and resource usability. While user experiences are highly subjective, making them a prime candidate for open questions, it can be difficult to draw comparisons from the resulting data. I therefore decided to supplement open questions with a standard means of measuring usability, the System Usability Scale (SUS) (Brooke, 1996). SUS consists of a series of ten statements to which users express their level of (dis)agreement on a five-point Likert scale (Q19), from which a score is calculated. These statements were selected by testing users with a wider pool of 50 statements and determining which provoked the most extreme responses (Brooke, 2013, p. 34). SUS is system-agnostic, i.e. the statements are general enough to refer to any kind of technology (Brooke, 2013, p. 36), facilitating comparison between different types of tool or resource and between the results of different surveys.

In hindsight, while SUS would have been useful for a survey that asked participants to assess one or more named tools or resources, my survey asked participants to choose the Linked Data tool or resource with which they were most familiar (Q15). As a result, I received between one and three responses for each tool or resource. Although Sauro (2011, pp. 131–132) suggests that the lowest possible sample size could be two, I did not consider the results I received from this question to be sufficiently reliable to warrant further discussion.

Following pilot testing and minor amendments to the survey, I used a sampling frame (Adams & Cox, 2008, p. 25) provided by mailing lists and social media to target the study population. In addition to posting the survey on *Twitter*, I contacted the

following mailing lists: Antiquist<sup>85</sup>, Britarch<sup>86</sup>, Classics-L<sup>87</sup>, ClassicsGrads<sup>88</sup>, Classicists<sup>89</sup>, Digital Classicist<sup>90</sup>, and the UK Museums Computer Group<sup>91</sup>. As a specific population of Ancient World researchers was being targeted, this approach could be described as purposive sampling (Kelley et al., 2003, p. 264). Contacting these digital channels seemed to be the most appropriate means of asking potential participants to share information about their engagement with digital resources. However, using this method meant that I was not able to calculate the response rate; making the link accessible to anyone online meant that it was impossible to determine the number of people who saw the link and did not respond.

Measures I implemented to maximise the number of responses included publicising the survey on a Wednesday morning (Faught et al., 2004), ensuring a clear subject line containing a topic of interest to the target audience (Keusch, 2015, p. 193), conveying authority by providing details of my institution, funder and supervisor in the email message (Keusch, 2015, p. 192; Vogt et al., 2017, p. 19), and reassuring participants about response confidentiality, as well as including information about how to withdraw if desired (Adams & Cox, 2008, p. 26; Keusch, 2015, pp. 200–201; Vogt et al., 2017, p. 19). As recommended by Keusch (2015, pp. 196–197), I additionally provided estimates of time taken to complete the survey based on the results of pilot testing: 15 minutes for participants unfamiliar with Linked Data, and 30 minutes for those with experience of Linked Data use or production. On clicking a link to the survey, participants were presented with an information page that included details about the survey content and its relationship to my PhD research. As recommended by Adams and Cox (2008, p. 20) the page provided definitions of key terms as they are used in the survey, to aid participants' comprehension and ensure that responses were provided within a consistent frame of reference.

---

<sup>85</sup> <https://groups.google.com/forum/#!forum/antiquist>

<sup>86</sup> <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=BRITARCH>

<sup>87</sup> <https://lsv.uky.edu/scripts/wa.exe?A0=CLASSICS-L&A0=CLASSICS-L>

<sup>88</sup> <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=CLASSICSGRADS>

<sup>89</sup> <https://listserv.liv.ac.uk/cgi-bin/wa?A0=CLASSICISTS>

<sup>90</sup> <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=DIGITALCLASSICIST>

<sup>91</sup> <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=mcg>

The survey was open between 21 March and 2 May 2018. When it closed, it had received 259 responses; however, further inspection showed that 47 responses were empty. The analysis will therefore consider only those 212 participants who answered at least one question. In comparison with the Digital Humanities surveys discussed above (3.2), this is a similar magnitude to Kemman et al.'s (2014) survey, albeit less than their 288 responses, possibly due to their wider subject area remit. Elsewhere, my response figure is considerably more than Isaksen's (2011) or Europeana's (Angelis et al., 2015), probably due to their use of more selective recruitment methods. Conversely, DARIAH's survey (Dallas et al., 2017) achieved a considerably higher response rate, with over 2000 participants, which is likely due to a combination of their broader remit and their organisational reputation. Using these figures as a guide, I can be confident that my survey achieved sufficient responses to provide useful and reliable results in this context.

### 3.4 Demographics

Having briefly mentioned the presence of demographic questions above, this section will now turn to discussing responses to these questions in detail, to provide an overview of the survey population characteristics. In addition to their location, role, age, and gender, I will present responses relating to participants' relationships to Linked Data and digital tool and resource production. One of the initial demographic questions asked participants for the country in which they are located (Figure 3.2). The most common locations were the UK, the US, and Germany, which is more likely to be indicative of the channels used to distribute the survey, than an accurate representation of the global population of Ancient World researchers. This phenomenon can be illustrated with location information from two of the other surveys discussed in 3.2: the DARIAH survey received the most responses from DARIAH member countries (with the highest proportions from France, Germany and Serbia) (Dallas et al., 2017), while the overwhelming majority of Kemman et al.'s (2014) participants were located in the Netherlands and Belgium, due to targeting specific individuals in these countries.

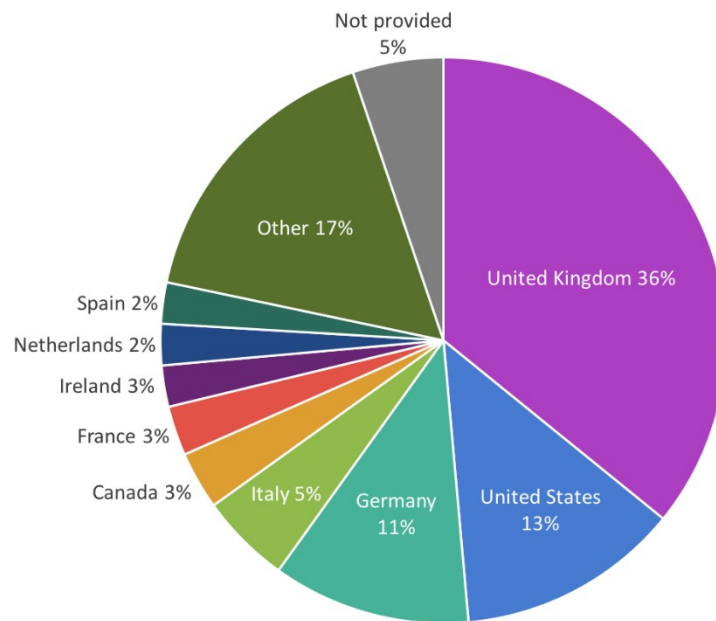


Figure 3.2 Pie chart showing participant location by country

Participants were also asked for their current role (Figure 3.3). Most were associated with academia, representing a variety of career stages, with a small number of participants from each of the other categories, representing a similar distribution to the surveys by DARIAH (Dallas et al., 2017), *Europeana* (Angelis et al., 2015) and Kemman et al. (2014). It is therefore clear that my findings are based predominantly on the use of digital tools or resources for Ancient World research within academia.

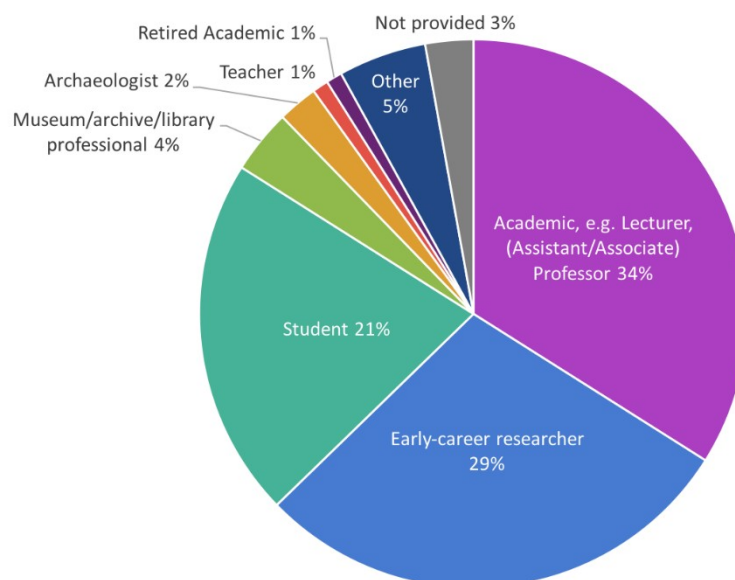
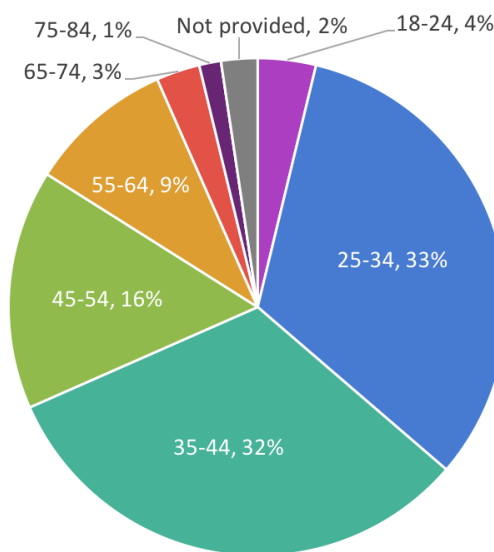


Figure 3.3 Pie chart showing participant roles

Participant age range is shown in Figure 3.4; all but one age range (85+) was represented. Most participants were aged between 25 and 44, with fewer participants at older and younger ranges. As these figures are broadly similar to those in DARIAH's (Dallas et al., 2017) and Kemman et al.'s (2014) surveys, these results are likely to represent the age ranges of those researchers who are more digitally engaged, rather than solely indicating the demographics of the mailing lists and social media channels where the survey was shared.



*Figure 3.4 Pie chart showing participant age ranges*

In terms of gender (Figure 3.5), there is a more even distribution, with 50% participants identifying as female, 45% male and 5% who did not answer the question or selected 'Prefer not to say' or 'Prefer to self-describe'. DARIAH's survey population demographics were also slightly more skewed towards female participants (57%) (Dallas et al., 2017), with both sets of results supporting the finding from previous research that women are more likely to respond to surveys than men (Keusch, 2015, p. 189). Conversely the majority (57%) of Kemman et al.'s (2014) participants were male, perhaps because they were individually targeted, rather than self-selecting from a wider potential population.

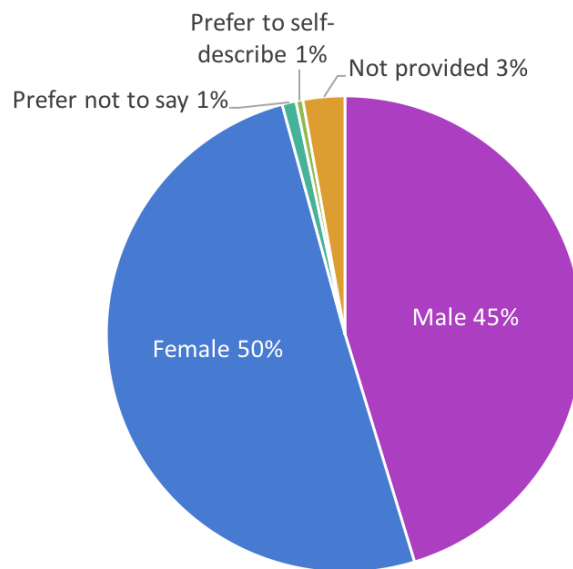


Figure 3.5 Pie chart showing participant gender

A key question later in the survey asked participants if they had knowingly used one or more tools or resources based on Linked Data; results are shown in Figure 3.6. Of the 133 participants who answered this question, the majority responded ‘No’ or ‘Unsure’, with 49 participants (37%) responding ‘Yes’ (I will refer to these participants as ‘Linked Data users’). The fact that Linked Data users are in the minority is hardly surprising, based on the assumption in the first part of my original hypothesis (1.1), *“Linked Data principles and technologies have not yet been sufficiently adopted in Humanities research to achieve their potential benefits”*. I would argue that the proportion of Linked Data users among the global population of Ancient World researchers would in reality be even lower. Survey participants are likely to represent a sample of this population with relatively high levels of digital engagement in general and are therefore potentially more likely to be existing Linked Data users.

However, despite the predictably low usage of Linked Data, the term is starting to creep into the consciousness of the Ancient World researcher; as shown in Figure 3.7, 47% of the 134 participants who answered the preceding question said that they were familiar with the term ‘Linked Data’. In terms of participant numbers, this equates to 15 participants who were aware of the existence of Linked Data but had perhaps not yet found an application of this approach that was relevant to their work. For increased Linked Data adoption to occur, resources must be useful and usable for potential as well as actual users.

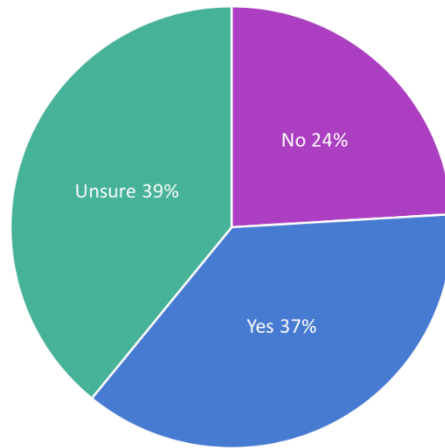


Figure 3.6 Pie chart showing participants' answers to the question "Have you ever knowingly used one or more tools/resources that are based on Linked Data?"

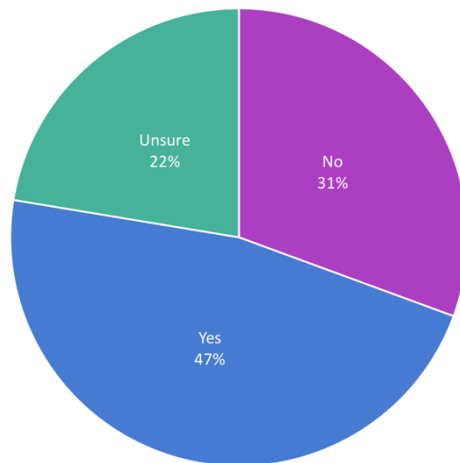


Figure 3.7 Pie chart showing participants' answers to the question "Are you familiar with the term 'Linked Data'?"

While identifying potential users poses more challenges than identifying actual users, the willingness of non-Linked Data users to try a new type of digital tool or resource might be inferred from their existing levels of technical engagement and skill. These characteristics can be determined from an early question in the survey asking participants to assess their digital competence and confidence. Participants selected their level of agreement (Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, or Strongly Disagree) with a series of statements:

1. I regularly use digital tools/resources
2. I am confident in using digital tools/resources
3. I actively seek out new digital tools/resources
4. My research would not be possible without digital tools/resources
5. I am aware of the underlying data structures behind those digital tools/resources I regularly use
6. I have sufficient skills/experience to teach others about digital tools/resources
7. I have the ability to create my own digital tools/resources

To effectively analyse participants' digital competence/confidence levels, I converted each response to a score between 1 (Strongly Disagree) and 5 (Strongly Agree).

Calculating the mean of each participant's scores resulted in an overall digital competence/confidence score for that participant. The distribution of overall scores is shown in Figure 3.8<sup>92</sup>, and is positively skewed, i.e., a greater number of participants scored towards the upper end of the scale than the lower end. This indicates that, generally, participants were confident about their levels of digital competence, with the vast majority having an average score above 3. Rather than assuming Ancient World researchers in general tend to have a high level of confidence in their digital abilities, these figures might be more reflective of the type of people who chose to take part in the survey. That said, a small proportion of participants did have a lower overall score; therefore, almost the full spectrum of scores between 1 and 5 has been represented.

A more detailed picture emerges when considering the scores for each individual statement (Figure 3.9). This diagram shows that most participants scored either 4 or 5 for statements 1 and 2, between 3 and 5 for statements 3 and 4, and between 2 and 4 for statements 5-7. Therefore, while most participants were familiar with digital tools

---

<sup>92</sup> Box plots represent the distribution of a set of numbers. The outer points represent the minimum and maximum values in the range, while the central line represents the median. The upper boundary of the box represents the upper quartile (or 75<sup>th</sup> percentile), the median of the top 50% of values in the set, while the lower boundary of the box represents the lower quartile (or 25<sup>th</sup> percentile), the median of the bottom 50% of values in the set. If the box appears towards the top of the diagram, as in Figure 3.8, this indicates that more values in the set are towards the higher end of the range than towards the lower end.



and resources and depended on them for their research to some extent, fewer participants had a deeper understanding of the more technical aspects.

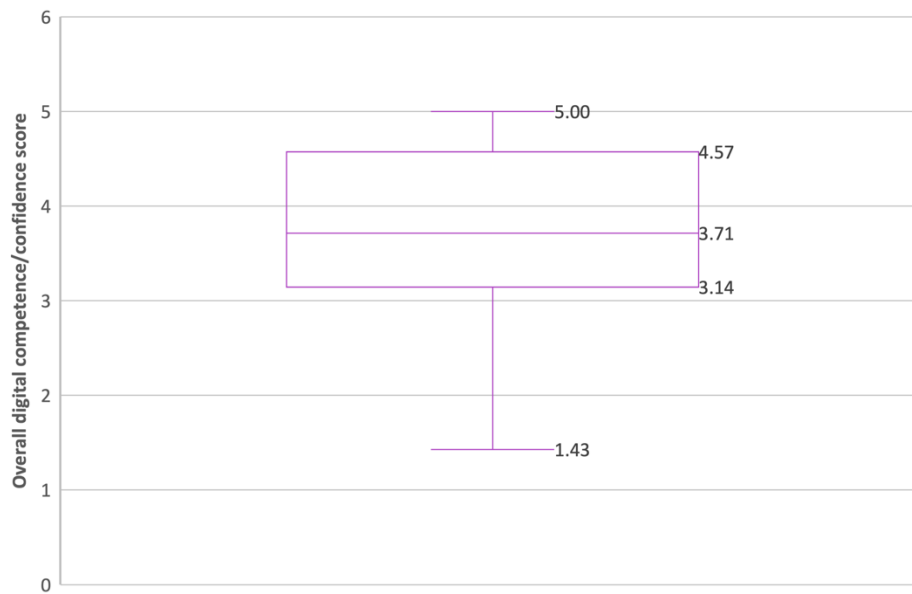


Figure 3.8 Box plot showing distribution of overall digital competence/confidence scores



Figure 3.9 Box plot showing distribution of digital competence/confidence scores for each statement

In addition to analysing these overall results, I also compared responses from Linked Data and non-Linked Data users. Using the same method as before, but segmenting the responses by Linked Data usage, I calculated the scores for each question and each participant's overall mean. Figure 3.10 shows the distribution of mean digital confidence/competence scores in Linked Data users and non-Linked Data users.

These diagrams indicate that Linked Data users tended in general to have a higher level of technical skill than non-Linked Data users, and that there is less variation in digital competence/confidence among Linked Data users. They also demonstrate that non-Linked Data users represent the full spectrum of digital competence/confidence, with distribution still skewed (albeit to a lesser extent) towards higher scores; over half the participants have an average score greater than 3.

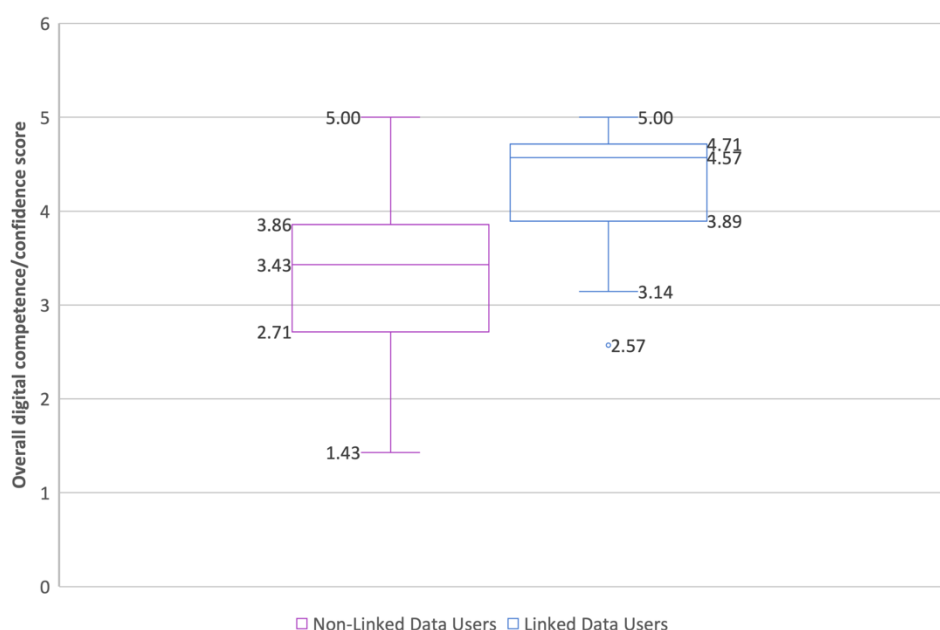


Figure 3.10 Box plots showing distribution of mean digital competence/confidence scores for Linked Data users and non-Linked Data users

Looking at the individual questions (Figure 3.11), it is clear that Linked Data users' scores are almost entirely skewed towards the higher end; however, there is more variation in responses to statement 7. Non-Linked Data users' scores are more evenly distributed, although there is a clear transition from predominantly higher scores for statement 1 to predominantly lower scores for statement 7. From these distributions, it can be inferred that while these non-Linked Data users have a range of technical

skills, the majority are likely to be interested in and experienced enough with digital tools and resources to be willing to try out new ones (of relevance to their research). They could therefore be considered as a sample of potential Linked Data users, who could provide insights into how such resources might be made more usable.

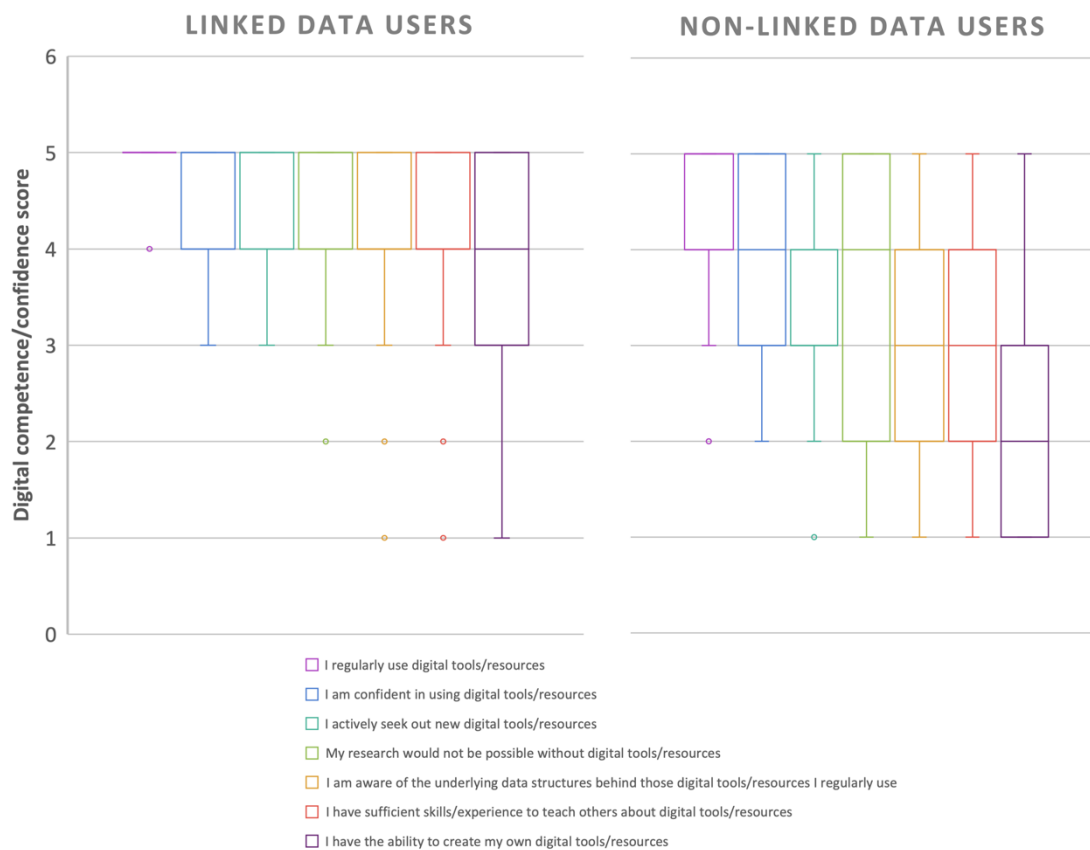
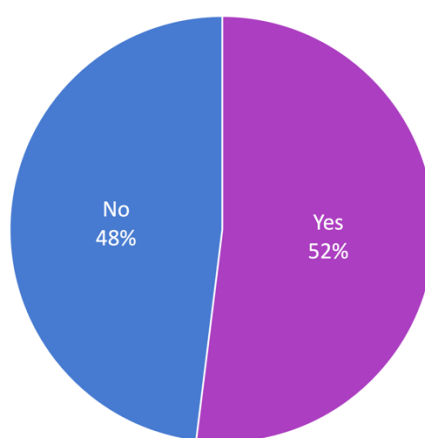


Figure 3.11 Box plots showing distribution of digital competence/confidence scores for each statement for Linked Data users (left) and non-Linked Data users (right)

In the final section of the survey, participants were asked if they had been involved in the production of a digital tool or resource; responses are shown in Figure 3.12. Of the 127 participants who answered this question, over half (52%) of participants answered 'Yes'. This relatively high proportion could be due to the breadth of the question; as well as those participants involved in major Digital Humanities projects, I found during the interview phase that this also included individuals creating their own research databases or websites; therefore, these findings represent a range of experience levels and scales of resource. For the remainder of this thesis, I will refer to those participants who answered positively to this question as producers.

Finally, we turn to Linked Data production; Figure 3.13 shows that 35% of producers have been involved in producing a Linked Data tool or resource, which equates to 22 participants. A logical assumption would be that these producers represent slightly less than half of the 49 Linked Data users identified above, which implies that Linked Data resources are indeed being used by people outside the immediate projects that produced them. However, if we look in more detail at producers, segmented by whether they used and produced Linked Data (Figure 3.14), there are some unexpected results. Although most Linked Data producers were also Linked Data users, a small number were not, or were unsure if they had used a Linked Data tool or resource. This apparent inexperience of using the type of resource they are involved in producing could result in usability issues, unless they are working with colleagues who have used such resources. Speaking with these participants during the interview phase could have been extremely informative to determine whether non-use of Linked Data affected their production processes; however, the one participant from this group who volunteered to be interviewed did not respond to their invitation.



*Figure 3.12 Pie chart showing whether participants have been involved in the production of a digital tool/resource*

Responses to demographic questions indicate that survey findings will most closely represent the experiences of people aged between 25 and 44, who work within academia in the UK or US. With the exception of location, these characteristics were broadly consistent with those of other comparable survey populations, where this information was available. Linked Data users were in a minority and tended to have a higher level of digital competence/confidence when compared with non-Linked Data users. However, many non-Linked Data users could be considered as potential Linked

Data users, as their scores generally demonstrate a relatively high level of digital engagement, indicating a willingness to try new tools or resources. A sizeable proportion of participants had produced digital tools or resources, some of whom had produced Linked Data; however, not all Linked Data producers were also Linked Data users.

Having conducted the survey and assessed participant demographics, I was then able to proceed with the interview phase of my user study.

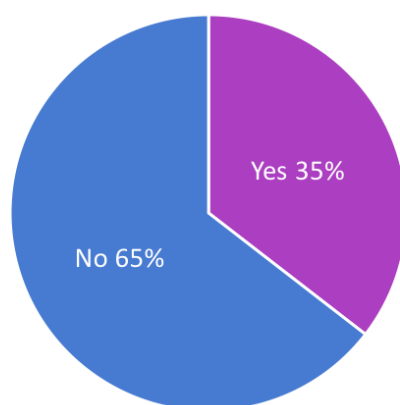


Figure 3.13 Pie chart showing whether producers have been involved in the production of a Linked Data tool/resource

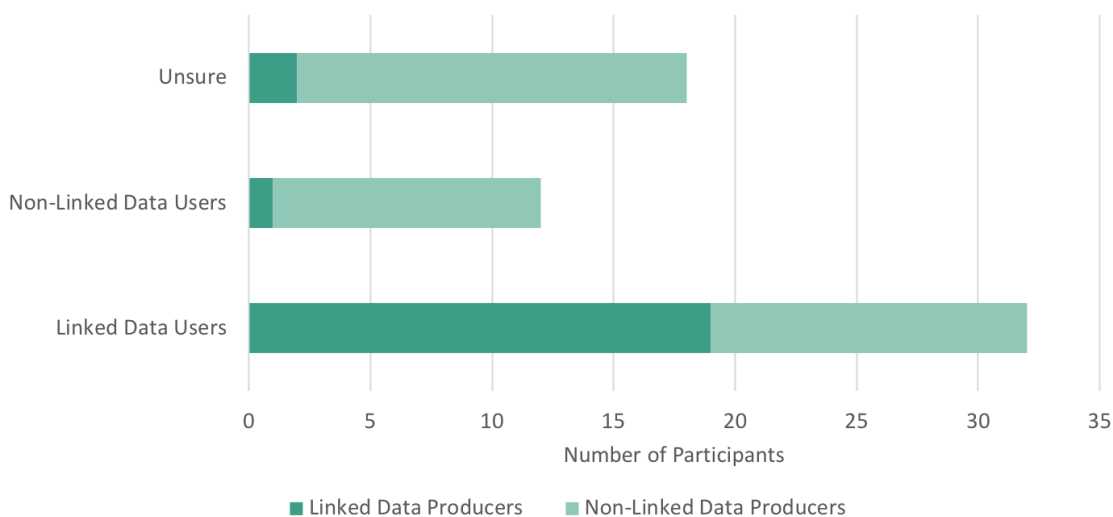


Figure 3.14 Bar chart showing producers segmented by their use and production of Linked Data

### 3.5 Interviews

The second, qualitative component of my study involved interviews with selected survey participants to explore their experiences in more depth. As discussed above (3.2), interviews complement survey research well, by facilitating more detailed discussion of participants' responses without the constraints imposed by survey questions. In planning these interviews, it was important to achieve the optimal sample group for the most interesting and representative results and to design questions that would best complement the survey. Various factors also had to be considered regarding the interview process itself, as well as the transcription that followed each interview.

At the end of the survey (Appendix 1, Q33-34), participants could optionally indicate their willingness to take part in a follow-up interview. As I received 52 positive responses, an appropriate strategy was required to select a sample of participants that could be interviewed in the time available. There are many approaches to sampling, most of which can be categorised as either 'random' or 'purposive' sampling. Random, or probabilistic, sampling is where participants are selected completely at random and should provide the most representative (and therefore generalisable) coverage of the research population (Lucas, 2014, p. 393). However, the researcher cannot be confident that a random sample will yield the most valuable responses (Miller & Crabtree, 2004, p. 191). For my interviews, I aimed to select a range of participants with different levels of experience and skill in the use and production of digital resources in general and Linked Data in particular. As these participants were represented in very different proportions among those who had expressed their willingness to be interviewed (e.g., 56% had neither used nor produced a Linked Data resource), a random sample may well have missed some interesting and informative cases.

Purposive, or non-probabilistic, sampling is where the researcher selects participants based on particular characteristics that should ensure the most "information-rich" responses (Patton, 2002, p. 230; Portugal, 2013, p. 36), while being mindful that this sample will not be representative of the population as a whole (Lucas, 2014, p. 394; Vogt et al., 2017, p. 33). One type of purposive sampling is stratified sampling, where

the target population is divided into groups (strata) based on specific characteristics; a small sample is then selected from each group, either using random sampling (to reduce any further selection bias), as recommended by Lucas (2014, pp. 396–397) or another purposive method (Patton, 2002, p. 240). One such approach is maximum variation, or heterogeneity, sampling, which ensures that participants in the sample are as different from each other as possible. This approach can be particularly useful when seeking cases of particular interest, or identifying common themes across a diverse population (Patton, 2002, pp. 234–235).

Group Number	Group Characteristics	Number of Participants
1	Not used a Linked Data resource; not produced any digital resource	15
2	Not used a Linked Data resource; produced a non-Linked Data digital resource	14
3	Not used a Linked Data resource; produced a Linked Data resource	1
4	Used a Linked Data resource; not produced any digital resource	4
5	Used a Linked Data resource; produced a non-Linked Data digital resource	8
6	Used a Linked Data resource; produced a Linked Data resource	10

*Table 3.1 Criteria for each participant group, used for stratified sampling*

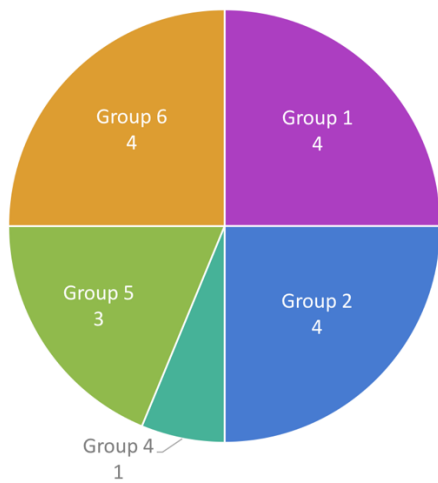
In my study, I chose to initially use stratified sampling by dividing participants into six groups based on their experiences in the use and production of Linked Data resources, as shown in Table 3.1. I aimed to select an equal number of participants for each group (apart from group 3, which contained only one participant). Although it was not possible to determine the characteristics of a completely representative sample of all Ancient World researchers, it was important to ensure a diverse range of participants to avoid findings being heavily influenced by demographic factors. I therefore used maximum variation sampling to select participants from each group with contrasting ages, genders, roles, locations, and digital confidence/competence scores. I additionally ensured that the overall sample of interview participants was broadly representative of the survey population, resulting in an initial sample size of 15.

I approached the initial 15 participants via email with an information leaflet about the project (Appendix 2) and a copy of the consent form (Appendix 3). I adapted these documents for Linked Data producers to provide the option to waive their anonymity if I included any quotes from them that refer to named tools or resources they have produced. After two months, eight of these participants had either been interviewed or agreed to take part, with the remaining seven either not responding (two participants) or declining the invitation (five participants). As a result, I selected seven other participants who shared similar demographic criteria and approached them in the same way, as well as sending follow-up invitations to those who had not responded. I interviewed 16 participants in total (7.5% survey participants); their demographic characteristics are shown in Figure 3.15.

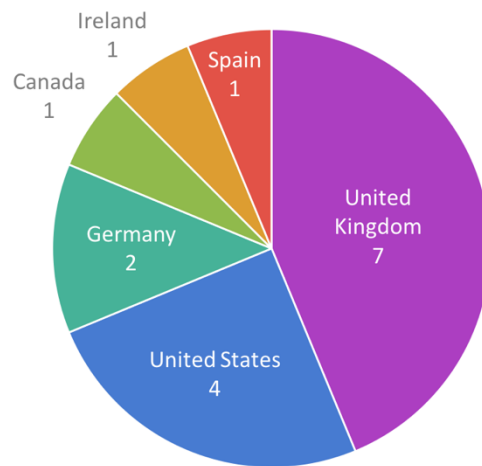
Figure 3.15a illustrates that, despite my efforts to ensure equal group sizes, there was some imbalance. In particular, the one participant from group 3 did not respond to the interview invitation, and only one participant from group 4 responded positively. The resulting proportions are, however, potentially indicative of the most likely combinations of Linked Data usage and digital tool or resource production among Ancient World researchers, i.e., people who have used Linked Data are more likely to have also produced a digital tool or resource (Linked Data or otherwise) and people who have produced Linked Data are more likely to have used it. The locations shown in Figure 3.15b and the age ranges in Figure 3.15d are broadly representative of the overall survey population, discussed in 3.4; however, the case is slightly different for role (Figure 3.15c) and gender (Figure 3.15e).

In Figure 3.15c, a disproportionately high number of interview participants were in academic roles, while a disproportionately low number were early-career researchers. This is partly due to academics being spread more evenly across the six Linked Data groups and partly because academics were both more likely to volunteer to be interviewed during the survey stage *and* to respond positively to an interview invitation. The most frequent reason given for declining an interview invitation was heavy workload and lack of time, which seems to have had more effect on the decision to take part for early-career researchers, who are likely to be in more precarious roles.

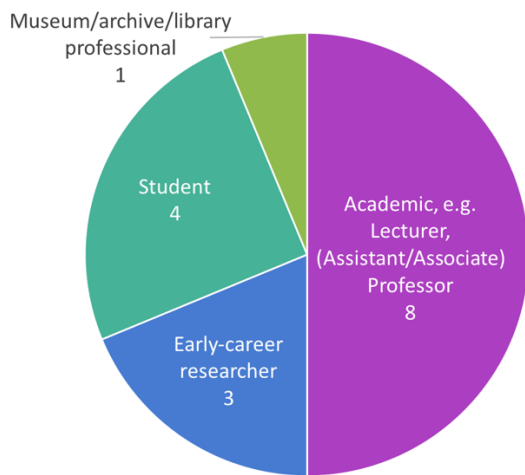




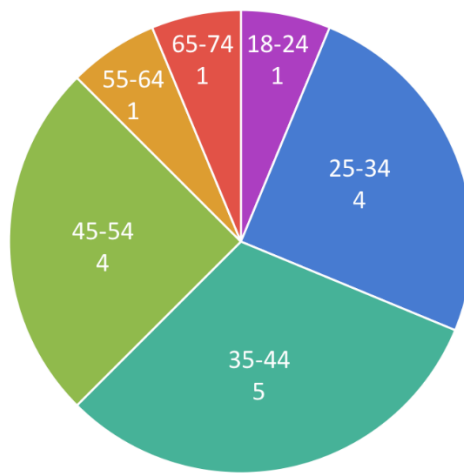
a.



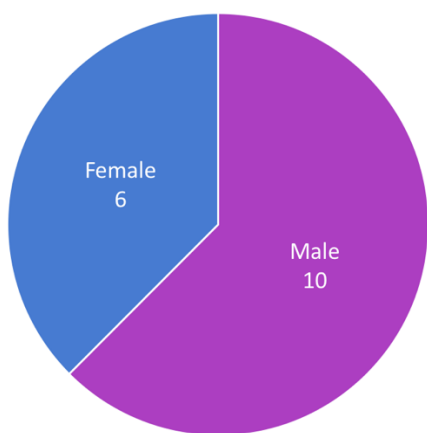
b.



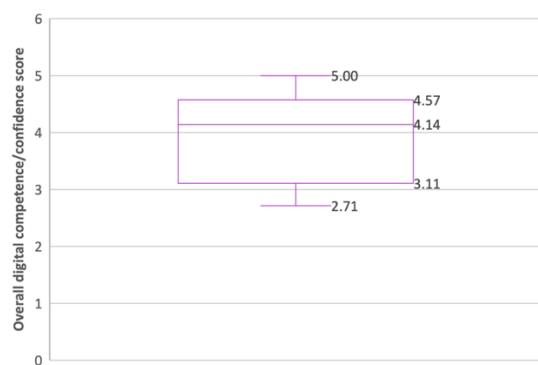
c.



d.



e.



f.

Figure 3.15 Demographic characteristics of interview participants (a. Linked Data use/production group, b. location, c. role, d. age, e. gender, f. digital competence/confidence)

Similarly, Figure 3.15e shows that a disproportionately high number of men took part in the interview stage, and a disproportionately low number of women. This was the case for several reasons. Firstly, although a slightly higher proportion of women answered the survey, more men than women volunteered to be interviewed. Secondly, among those women who volunteered to be interviewed, the distribution of Linked Data groups was much more uneven than for the men – for example, nearly half of the women who volunteered to be interviewed were in group 1 and only one woman (who accepted the interview invitation) was in group 6. Thirdly, women were far more likely than men to decline an interview invitation; in all cases this was due to heavy workload and lack of time, as with the early-career researchers.

The distribution of digital competence/confidence scores (Figure 3.15f) is broadly similar to that for the whole survey population shown in Figure 3.8, although the lowest score of 2.71 is higher than the lowest score in the population, 1.43. Among those who volunteered to be interviewed, two participants scored lower than 2.71, the lowest being 2.43, but neither accepted the invitation; therefore 2.71 was the lowest possible score for an interview participant. This slightly higher distribution of scores indicates that participants with a higher level of digital competence/confidence were more likely to volunteer to take part in the interview stage. My experience of conducting the interviews indicates that this was the characteristic most likely to affect participants' responses. However, several participants mentioned before or during the interview that they had completed the survey at a time when they were interested in starting to use digital tools and resources, and that their experience and skill levels had increased during the intervening months. In some cases, there may therefore be a slight mismatch between participants' digital competence/confidence at the point of taking the survey, compared to how they would have scored at the time of their interview. Therefore, while I can be confident that the interviews cover a broad range of digital skills and experience, the findings may be less applicable to those with scores towards the lower end of the spectrum.

Turning to interview construction, research interviews can be completely structured or completely unstructured, but most lie on a spectrum between the two. Highly structured interviews are effectively researcher-administered surveys, where each

participant is asked the same questions in the same order, without digressions. This provides consistency and control; however, researchers do not have the opportunity to explore participants' responses in more depth, seek additional information, or explore what is most important to the participant (Esterberg, 2002, pp. 86–87; Oates, 2005, pp. 187–188; Vogt et al., 2017, p. 40). At the opposite extreme, unstructured interviews have few or no questions at all, and are more akin to free-flowing conversations, albeit with some prompts to discuss particular topics. This approach may feel more natural and put participants more at ease, and can work particularly well for ethnographic research; however, it renders comparison of responses more difficult (Esterberg, 2002, p. 89; Oates, 2005, p. 288; Vogt et al., 2017, p. 40). Like many researchers, I decided to take a semi-structured approach – a 'middle ground' where the questions are scripted, but there is flexibility to diverge from the script to ask follow-up questions if required (Adams & Cox, 2008, p. 22; Oates, 2005, p. 188).

In a similar way to the survey questions, each interview question related to one or more of my research questions as recommended by Vogt et al. (2017, p. 37) and Portugal (2013, p. 31); the full interview script can be found in Appendix 4 and is divided into sections according to the participant's experience with Linked Data use or production, using the six groups described above. The script was tailored to each participant, as recommended by della Porta (2014, p. 238), based on which sections were relevant to them. For those who had been involved in the production of one or more digital (including Linked Data) resources, production-related questions were included first as a priority; use-related questions could then be cut if there was insufficient time remaining.

Early questions were not particularly challenging and were designed to put the participant at ease, as recommended by Esterberg (2002, p. 96); language was kept simple and unambiguous throughout. Some questions were expansions of the survey questions; for example, the Linked Data and non-Linked Data Tool/Resource Production sections went into much more detail about the tool or resource produced than would have been possible in the survey, and Q41-43 asked participants about their experiences of conducting methods selected during the survey. Other questions arose from the survey findings; for example, I asked participants to rank frequently

mentioned features or barriers in a follow-up survey, then discussed their ranking in Q44-51. Each question was initially relatively open, to ensure that participants focused on what was important to them, without any assumptions about what would be most relevant. I used sub-questions and prompts, both scripted and spontaneous, to encourage participants to expand on shorter answers. Some of the wording was adapted for each participant, e.g., when discussing specific methods or resources they had selected or mentioned in the survey.

Once the interview questions had been approved by the Open University's Human Research Ethics Committee<sup>93</sup>, I performed a pilot interview, as recommended by della Porta (2014, p. 237), Esterberg (2002, p. 100) and Oates (2005, p. 189). I chose a participant from group 6, to test the full range of interview questions (Linked Data production, digital tool/resource use, Linked Data use) and confirmed that interviews for this group could be completed within an hour, but that slightly more time would be preferable to allow the participant to provide more detailed responses. The participant was positive about their interview experience and suggested that I asked about participants' production of other tools or resources, rather than focusing solely on the one they produced most recently, which resulted in Q28, "How did your experiences of producing [RESOURCE] compare with your experiences of producing other digital tools or resources, either using Linked Data or other technologies?". No other changes to the script were requested, although during the interviews themselves, I adapted question ordering (and sometimes wording) according to the direction of the conversation. For example, if a participant mentioned a specific tool, resource, or method as part of their response to an early question, I asked for more detail at that point in the interview, rather than waiting until the time I would usually have asked that question.

I conducted eight interviews in person and eight online, due to budget and time constraints prohibiting me from visiting each participant's geographic location. There is some debate in the literature regarding the value of online interviews. Vogt et al. (2017, p. 43) are in favour and suggest that "privileging one form of communication

---

<sup>93</sup> Reference HREC/2018/2807/Middle

seems archaic". However, della Porta (2014, p. 248) cautions that "technology-mediated contacts can make the building of trust with the interlocutors more difficult and distraction easier". While building rapport with participants happened more naturally in person, meaning that face-to-face interviews tended to be more relaxed, I found that the online interviews provided valuable insights that could not have been obtained by other means. Additionally, online interviews tended to produce clearer recordings as both the participant and I were positioned closer to our microphones. The main disadvantage of online interviews was occasional connectivity issues, which affected the flow of the conversation and sometimes resulted in asking the participant to repeat what they had been saying. However, as the affected participants were generally familiar with attending online meetings, such issues were not surprising, and no one appeared visibly or audibly frustrated as a result.

Before each interview took place, the participant signed a form (Appendix 3) to confirm that they had given their informed consent to take part in this phase of the research, and for their interview to be audio recorded. Linked Data producers (group 6) gave their consent to be mentioned by name in relation to specific tools or resources they had produced; the remaining participants were assured of their anonymity, i.e., they would not be identifiable from any transcripts or information included in subsequent reports, such as this thesis. Participants were also informed that anonymised interview transcripts would not automatically be made available to other researchers, but that they could be requested via a metadata record in an online repository. Both della Porta (2014, p. 242) and Adams and Cox (2008, p. 23) emphasise the importance of making such ethical issues clear from the outset of the interview.

To ensure no information was missed, all interviews were audio-recorded using both Piezo<sup>94</sup> on a laptop and Samsung Voice Recorder<sup>95</sup> on a mobile device, providing redundancy in case of unforeseen technical issues with either piece of software. I additionally took notes during each interview to keep track of any changes to question ordering, and as an aide memoire to ask about any remarks that would be interesting to explore in more detail. This combined audio-recording and note-taking approach is

---

<sup>94</sup> <https://rogueamoeba.com/piezo/>

<sup>95</sup> <https://play.google.com/store/apps/details?id=com.sec.android.app.voicenote>

recommended by both Esterberg (2002, p. 106) and Oates (2005, pp. 190–191). As suggested by Portugal (2013, p. 115), after each interview I made further notes to reflect on the process, and to provide any non-verbal information that would not be obvious from the transcripts (e.g. mood, atmosphere, or tone of voice).

Several participants remarked following the interview that taking part had helped them to reflect on their own practice. One said that the discussion raised questions about their research methodology, which they would follow up; another, who was largely concerned with the production of digital tools and resources, said it made them realise how rarely they use tools and resources developed by others, and that they would seek to change their ways of working as a result. This finding was an unforeseen benefit of my study and demonstrates the power of interviews as a reflective exercise to facilitate the participant's own development.

I transcribed the audio recordings as soon as was feasible after each interview took place, using denaturalised transcription (UK Data Service, n.d.), where words are transcribed using standard written English, rather than attempting to record each individual sound. As such, my transcripts resemble playscripts, with all punctuation used in the same way as the reader would expect from a written publication (although I have not attempted to correct grammatical errors). O'Connell and Kowal (1995) recommend this approach because the varying uses of standard punctuation symbols in other transcription systems can be confusing and potentially misleading to the reader. My approach is similar to Williamson's (2014) 'broad transcription' example, although I have additionally conveyed emphasis by using italics, and have enclosed comments such as "[unclear]" or "[connection break]", in square brackets. In subsequent chapters, where I have quoted from interviews, I have also used square brackets to insert additional words in cases where they might provide clarity for the reader. In other cases, I have removed words from the original transcript (e.g., due to repetition), which is represented using an ellipsis '...'. As recommended by Oates (2005, p. 194), I gave each participant the opportunity to check the transcript of their interview and amend any statements where the meaning might not be clear in written form (e.g. where their tone of voice, expression or gesture was not adequately

captured in the transcript). Most participants approved their transcripts with no changes, or minimal alterations.

In this section, I have detailed the design and application of interviews with selected survey participants, to form the second stage of my study. I selected a sample of 16 participants with differing levels of digital/Linked Data experience, while being broadly representative of the survey demographics, albeit with some imbalances in gender and role. Interview questions emerged from survey data analysis and were tailored to participants based on their sample group and survey responses. The interviews themselves took place in person and online and were audio-recorded, followed by opportunities for reflection. I found the interview phase invaluable for gaining deeper insights into participants' experiences than could be gleaned from the survey alone. As such, the two approaches complimented each other well.

### 3.6 Conclusions: Methodology

I started this chapter by assessing various frameworks for classifying Humanities research methods, particularly regarding digital activities, and found that *TaDiRAH* provided the most appropriate level of granularity and consistency for incorporation in my study. I then explained my decision to use a combined survey and interview approach, based on their effectiveness in other Digital Humanities projects, and discussed how I applied these methods to my study. In particular, this involved the inclusion of *TaDiRAH* terms to classify participants' research methods, to which I will return in the following two chapters.

Combining a survey and interviews provided a broad overview of digital engagement by over 200 Ancient World researchers, as well as the opportunity for more detailed discussions with a sample of 16. Together, these approaches provided significant insights that enabled me to address my research questions, which I will present in the following chapters. Although my discussion will predominantly focus on qualitative findings gained from interviews and open-ended survey questions (Chapters 5-7), the following chapter will provide key contextual information based on quantitative analysis from my survey.

## 4 Tools, Resources and Methods for Ancient World Research

Following the previous chapter's explanation of my survey and interview methodology for studying the usability of Linked Ancient World Data, I will now move on to discuss my findings from this research. While subsequent chapters involve detailed exploration of my interview findings, the current chapter will provide a contextual foundation for these discussions by analysing quantitative data from survey responses. In doing so, I will begin to address RQ3, on the integration of Linked Data with existing research methods, and RQ4, on improving Linked Data usability, by identifying key themes for further exploration. Firstly (4.1), I will consider factors affecting the usability of digital tools and resources for studying the Ancient World. Due to the relatively large number of survey questions relating to this topic, this section forms the majority of this chapter. Secondly (4.2), I will identify digital methods conducted by Ancient World researchers (based on a relatively small number of survey questions), highlighting those where Linked Data might most effectively be integrated.

### 4.1 Tools and Resources

As explained in 3.3, the survey was intended for all Ancient World researchers regardless of their familiarity with Linked Data. The first set of findings to be discussed in this chapter relate to the use and production of digital tools and resources for Ancient World research, to provide some initial insights into how Linked Data usability might be improved, in response to RQ4. I will start by exploring participants' use of digital tools and resources in general (4.1.1), then look in more detail at their use of Linked Data resources in particular (4.1.2). Finally, I will discuss participants' production of digital tools and resources (4.1.3), incorporating responses to Linked Data-specific questions.

#### 4.1.1 Using Digital Tools and Resources

Initial questions focused broadly on participants' experiences of using digital tools and resources in their research, with the aim that these findings might be applicable to the specific case of Linked Data. In this section, I will first identify the most popular tools and resources named by survey participants, before discussing positive features of digital tools and resources, as well as potential barriers to their use.



An early question (Appendix 1, Q9) asked participants to name the digital tools or resources with which they are most familiar. It did not specify that these should relate to the Ancient World in particular, and responses showed that participants used a wide variety of digital tools and resources, with 148 named in total. To illustrate the breadth of examples provided, I categorised responses into the following types:

**Ancient World:** tool/resource relates entirely or predominantly to the Ancient World

**Medieval:** tool/resource relates entirely or predominantly to the Medieval period

**Early Modern:** tool/resource relates entirely or predominantly to the Early Modern period

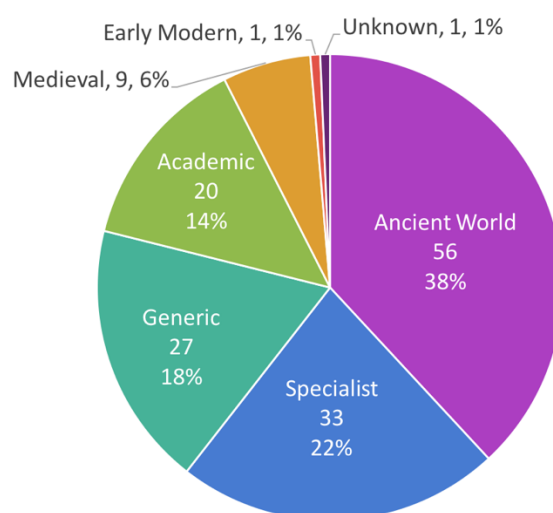
**Specialist:** software or service developed to perform a specific technical task that is not restricted to academia; includes visualisation software, code libraries, text analysis tools, geographic information systems (GIS)

**Academic:** tool/resource developed for use by academic researchers of any discipline; includes virtual research environments, library catalogues, journal databases, academic social networks, reference management

**Generic:** tool/resource used for non-specialist purposes by a wide variety of user groups; includes office software, major social media channels, file sharing/cloud storage platforms, general search engines

The results of this categorisation are illustrated in Figure 4.1, which shows the number and proportion of tools and resources from each type. Although the highest proportion were specific to the Ancient World, some participants additionally use tools and resources that relate to later historical periods (possibly for studying Classical Receptions), indicating that boundaries between the Ancient World and other Humanities subjects can be quite fluid. Many participants also mentioned Academic

and Generic tools and resources (31% and 37% of those who answered this question, respectively). However, the majority of participants did not include tools or resources from these categories, despite the likelihood that they do indeed use (for example) search engines, office software, reference management systems and library catalogues. One explanation could be that these tools and resources are so ubiquitous that their use is felt to be implicit; alternatively, as the survey focused on Ancient World research, participants may have considered only tools and resources relating to this subject domain.



*Figure 4.1 Pie chart showing the number and percentage of digital tools/resources of each type mentioned by participants*

Specialist tools and resources provide an insight into specific research activities conducted by participants. In particular, many of the examples relate to aspects of archaeological research, including GIS, 3D modelling and image editing software. These responses indicate that archaeologists are likely to use specialist tools without a disciplinary remit, or those created for other disciplines, rather than restricting themselves to archaeology-specific tools or resources. Interview responses confirmed that archaeological researchers tended to prefer selecting software based on desired activities and outcomes rather than disciplinary scope, while researchers who focused more on textual sources predominantly used tools and resources with the Ancient World (or a particular aspect thereof) at the forefront.

As usage information for Ancient World tools and resources is most pertinent to this thesis, and likely to be the most comprehensive, I will now look at these findings in more detail. 56 Ancient World tools and resources were mentioned in total, with 21 mentioned by two or more participants, shown in Figure 4.2. Those mentioned by most participants include *Thesaurus Linguae Graecae* (TLG)<sup>96</sup>, *Perseus*<sup>97</sup>, *L'Année Philologique*<sup>98</sup>, *Papyri.info* (2.2.4), *Packard Humanities Institute (PHI) Greek Inscriptions*<sup>99</sup>, and *Loeb Classical Library*<sup>100</sup>, indicating that a significant proportion of participants were interested in text-focused research. Of these six resources, three are freely available, while the other three require a paid subscription for full access.

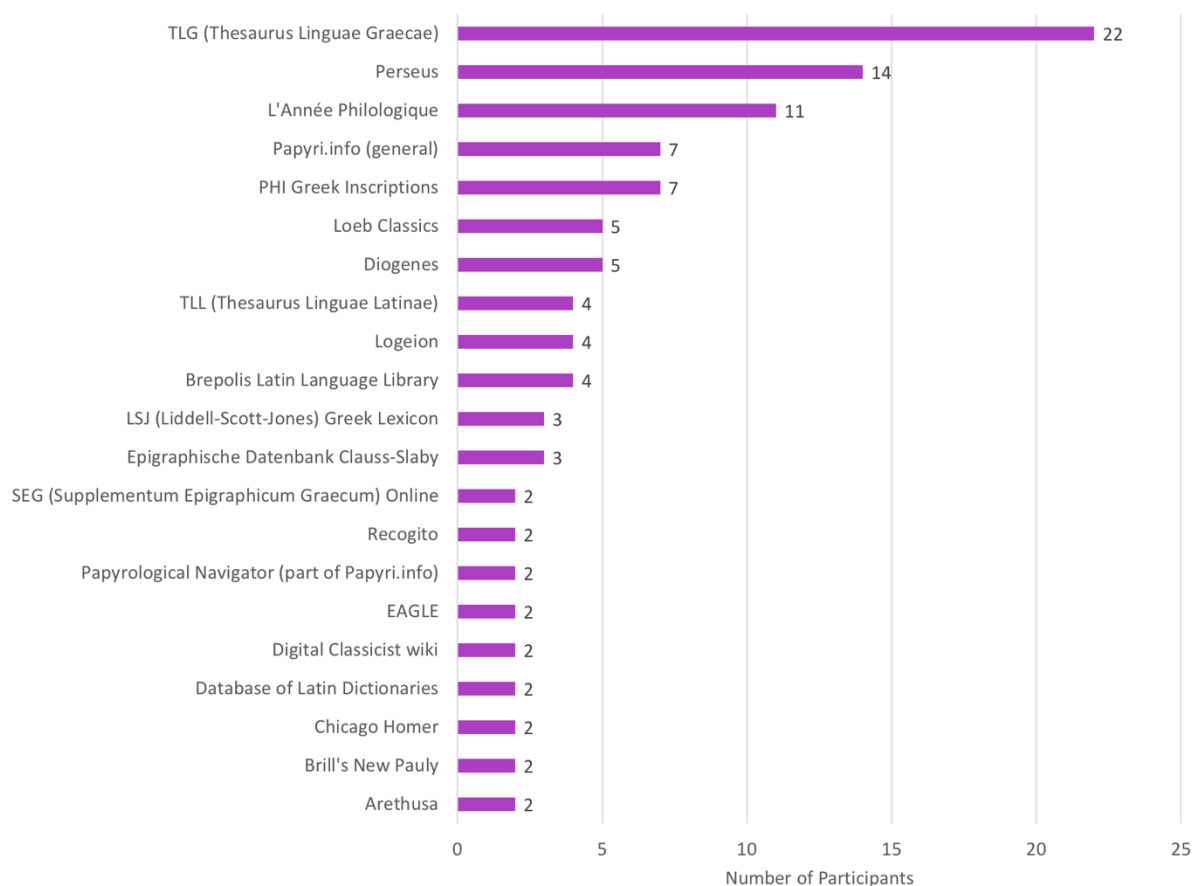


Figure 4.2 Bar chart showing digital Ancient World tools/resources with which participants are most familiar (all tools/resources selected by two or more participants are shown)

<sup>96</sup> <http://stephanus.tlg.uci.edu/>

<sup>97</sup> <http://www.perseus.tufts.edu/hopper/>

<sup>98</sup> <http://www.annee-philologique.com/>

<sup>99</sup> <https://inscriptions.packhum.org/>

<sup>100</sup> <https://www.loebclassics.com/>

Of the free resources, I have already discussed *Papyri.info* (2.2.4), which leaves *Perseus* and *PHI Greek Inscriptions*. The latter is described by its creators as "a comprehensive database of all ancient Greek inscriptions" (Packard Humanities Institute, n.d.).

Inscriptions can be browsed by region or retrieved via keyword searches, which have a geographical filter option. Documentation within the resource is minimal and does not appear to include information about the data model or standards used. While it is free to search and access the *PHI* texts via the resource itself, users must first agree to use the data for personal/research purposes only; reuse and sharing elsewhere are not permitted (Packard Humanities Institute, 2017).

*Perseus* was originally conceived as a digital library of ancient Greek texts but has since developed into a research infrastructure incorporating texts in other ancient languages, as well as other Humanities disciplines. *Perseus'* data models have been developed with accessibility, sustainability, and openness at the forefront, allowing texts to be reused, analysed, and enhanced by external tools and/or projects.

Additionally, the *Scaife Viewer*<sup>101</sup>, launched in March 2018, provides an improved reading environment that integrates and aligns texts and translations from multiple sources, with the potential to be combined with external tools that allow annotation, for example (The Scaife Viewer Project, n.d.). Providing access to the *Scaife Viewer* before completion has encouraged users to contribute to further development via its *GitHub* repository<sup>102</sup>. *Perseus'* openness therefore fosters collaboration between its developers and user community, facilitating continuous innovation and improvement.

Of the subscription-based resources, *TLG* is a collection of digitised Greek texts and lexica, as well as a database of information about them, discoverable via powerful advanced search functionality (Farrington, 2017; Thesaurus Linguae Graecae, 2015). There are significant access restrictions to prevent copying or export of texts, which can cause user frustration (Bagnall & Heath, 2018, p. 176) and prevent researchers from consuming *TLG's* data with other tools or integrating it with other resources (Barker & Terras, 2016, p. 6). *L'Année Philologique* is a searchable bibliography of secondary literature in multiple languages relating to all aspects of the ancient Greek

---

<sup>101</sup> <https://scaife.perseus.org/>

<sup>102</sup> <https://github.com/scaife-viewer/scaife-viewer>

and Roman worlds (Brepolis, 2017b). From each record, links are provided to the relevant ancient texts held in other Brepolis databases, such as the *Library of Latin Texts*, as well as external resources such as *TLG* and *Perseus* (Brepolis, 2017a, p. 3). The *Loeb Classical Library* is a set of Greek and Latin texts, published with the aim of widening access to Classical literature by including the Greek or Latin version of a text on the left-hand page and an English translation on the right-hand page (Harvard University Press, n.d.). Its interface is visually appealing, in displaying pages as they would appear in the print versions. However, this close adherence to their physical counterparts can reduce scope for interaction with the digital texts. For example, texts must be explored by page number, with users unable to quickly access the chapter or section numbers usually provided in citations (Dik, 2015, pp. 495–496). The designers therefore likely compared the experience of using the digital *Loeb* to that of the printed versions, and in doing so failed to manage the expectations of users, who compare it to other digital tools and resources.

Two of the top six tools and resources (*Perseus* and *Papyri.info*) involve Linked Data, in combination with other technological approaches. Information about data models and standards used by the other four top tools and resources was scarce, perhaps indicating a greater drive for openness and transparency among Linked Data producers. Of those participants who mentioned *Perseus*, the majority were unfamiliar with the term ‘Linked Data’, whereas the majority of *Papyri.info* users had knowingly used a Linked Data tool or resource (Figure 4.3). This difference possibly reflects how explicit each of these resources is about their implementation of Linked Data technologies, or how intrinsic Linked Data is to their functionality and user expectations. Unlike *Papyri.info*, where Linked Data technologies are intrinsic to its functionality, *Perseus* has implemented Linked Data relatively recently. For example, *Perseus* provides a URI for each text by combining their own persistent identifiers with those provided by *Canonical Text Services (CTS)*<sup>103</sup> (Almas et al., 2014; Babeu, 2019). However, it is not primarily known for being a Linked Data resource.

---

<sup>103</sup> *CTS* (<http://cite-architecture.org/cts/>) provides persistent identifiers for specific passages of text, described using an RDF vocabulary (CITE Architecture, 2015). These standards can be used to align and compare different versions or translations of the same text, or to analyse text reuse and citation practices (Tiepmar & Heyer, 2017).

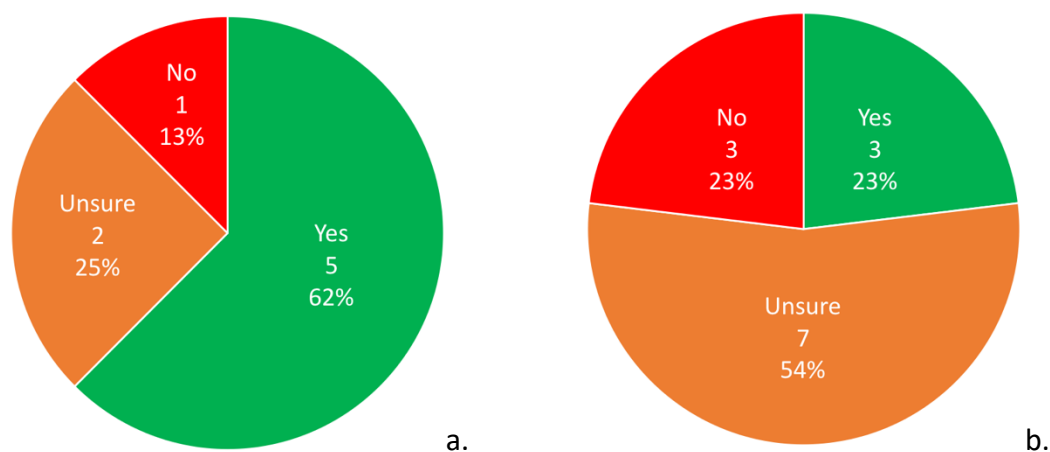


Figure 4.3 Pie charts showing whether users of particular resources (a. *Papyri.info*, including *Papyrological Navigator*; b. *Perseus*) had knowingly used Linked Data

Outside the top six tools and resources, two participants mentioned each of: *EAGLE* (2.2.4), *Pelagios'* (2.2.1) annotation platform *Recogito*, and *Papyri.info's Papyrological Navigator*, with none of the other Linked Data tools and resources from 2.2 included in Figure 4.2; however, one participant mentioned each of: *Pelagios* in general, *Pleiades* (2.2.1), and *Trismegistos* (2.2.4). These relatively low figures could indicate either that these tools and resources fulfil more specialised user needs (e.g., active interaction to perform a specific research task, rather than simply viewing content) and are therefore used less frequently, that barriers to their usability deter potential users, or that they remain less well-known than their more established counterparts.

It should be noted that this question was open-ended, which has led to some limitations. Responses would have been limited by participants' memories and the time they were willing to spend on their answers; therefore, including a pre-populated list might have ensured greater accuracy. However, it is significant that these responses reflect the tools and resources that immediately sprang to mind without prompting. Responses therefore imply regular usage, as well as identifying tools and resources that have most effectively permeated Ancient World researchers' consciousness.

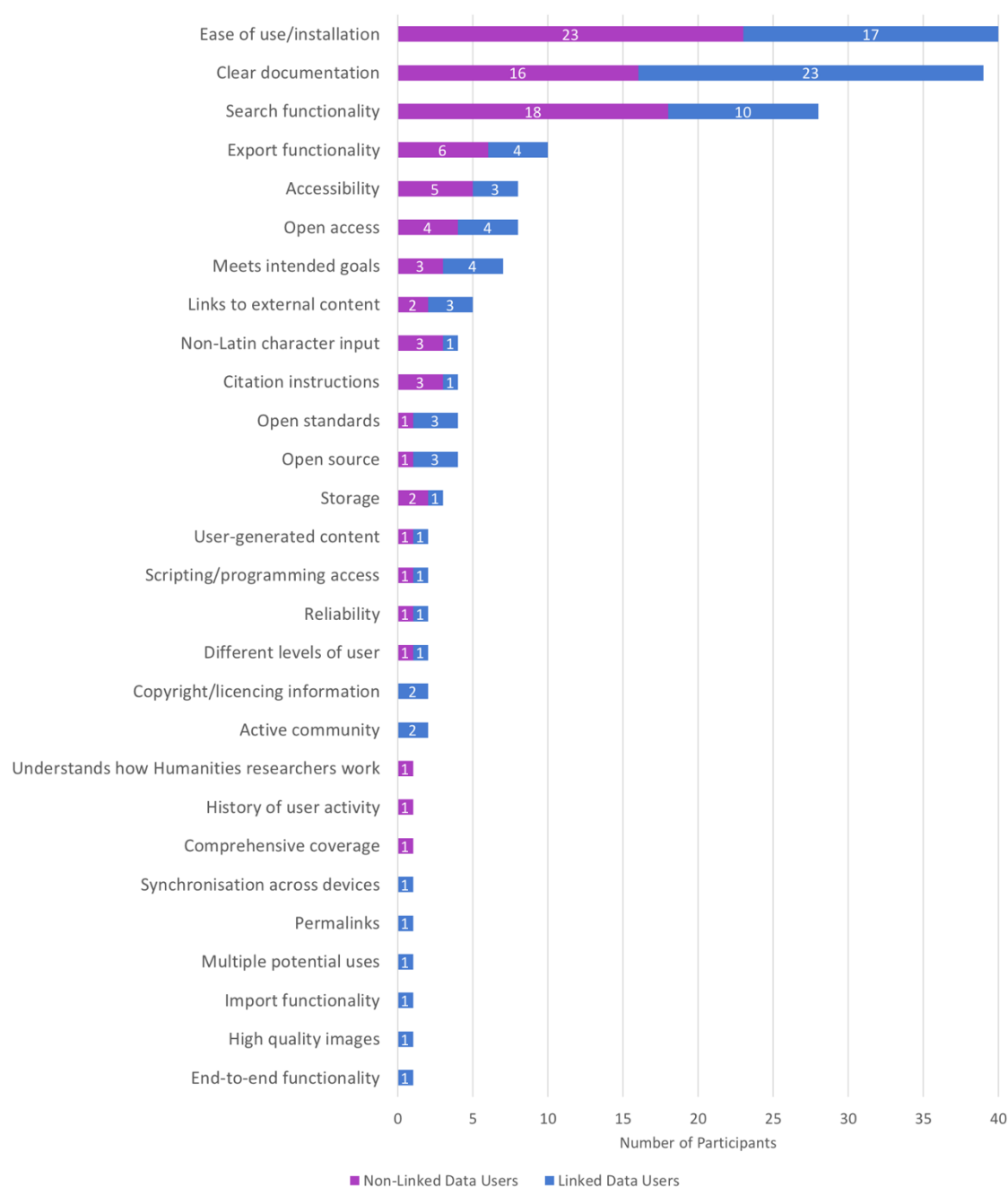
After naming digital tools and resources, participants were then asked to list what they considered to be features of a good digital tool or resource (Appendix 1, Q10), which I

classified using a series of ‘feature’ codes, shown in Figure 4.4<sup>104</sup>. The most frequently mentioned features were ease of use/installation, clear documentation, and search functionality, which are relatively general themes, compared with more specific requirements mentioned by fewer participants. Other features related to themes including accessibility, openness, interoperability, and reuse. Again, it should be noted that this was an open-ended question; no options or prompts were given, to avoid influencing participants’ choices. To assess how participants might have prioritised these features, interviewees were asked to fill out a brief follow-up survey, where they arranged ten features in order of importance. These ten features included those most frequently mentioned in Figure 4.4, as well as those mentioned by fewer participants that were particularly relevant to my research questions, such as ‘Reliability’ and ‘Understands how Humanities researchers work’.

Following completion of the follow-up survey, I reversed participants’ rankings to provide an importance score for each feature (i.e., a rank of 1 produced a score of 10, a rank of 2 produced a score of 9, etc.). I calculated the mean importance score for each feature and segmented the results by participants’ Linked Data usage, as shown in Figure 4.5. Although many features were universally popular (or unpopular) among all participants, others imply a difference in priority based on whether or not the participant had knowingly used Linked Data. Features that seemed particularly important to all participants included ease of use, clear documentation and search functionality. While open standards were not ranked highly by non-Linked Data users, their prioritisation of other features such as reliability and export functionality indicate areas where open standards could be advantageous, even if their application is not made explicit to the end user.

---

<sup>104</sup> A small number of additional features were highly specific, and were mentioned by one participant each, presumably relating to the research goals of the individuals concerned. These include ‘annotation functionality’, ‘bibliographies’, ‘bibliographical tools’, ‘cross-reference functionality’, ‘Linked Data’, and ‘relationship analysis functionality’.



*Figure 4.4 Bar chart showing features of a good digital tool/resource identified by survey participants, segmented by participants' Linked Data usage*



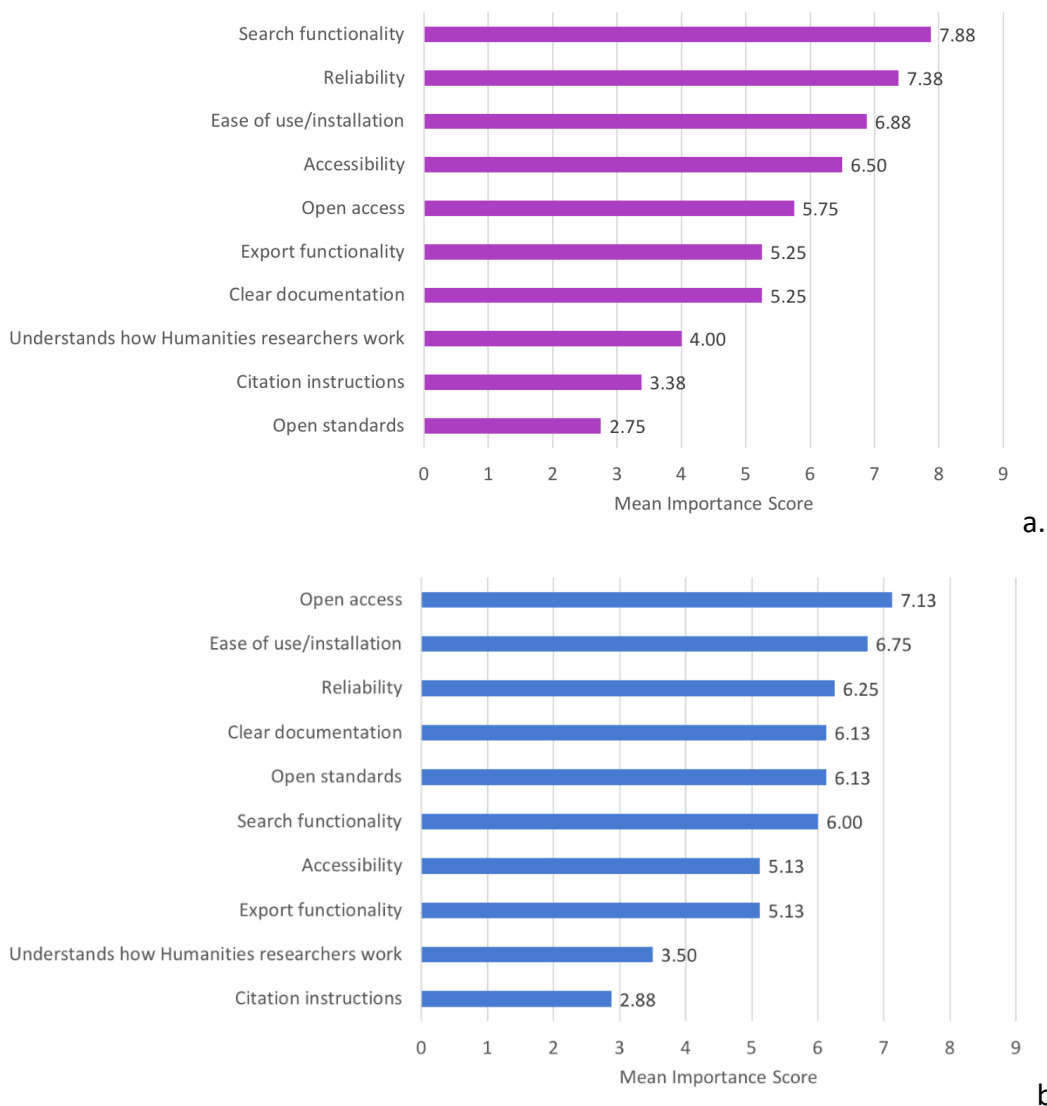


Figure 4.5 Features of digital tools/resources in order of importance as ranked by interview participants: a. non-Linked Data users; b. Linked Data users (8 participants in each group)

Many of the importance scores are very close (e.g., there is slightly more than one point between the top six features for Linked Data users in Figure 4.5b); therefore, a different sample of participants may have produced different results. Furthermore, several participants said they would have adjusted their rankings following our interview discussions, often due to asking for clarification on definitions (in particular, non-Linked Data users struggled with the term ‘Open standards’, hence its low position in Figure 4.5a). Some participants found the ranking part of the task unhelpful, e.g., PART008 saw the features as a sequence rather than discrete events and PART037 said their rankings had been fairly arbitrary, with the exception of their top and bottom two features. It must also be noted that the interview sample of 16 is considerably smaller than the 95 participants who answered the original survey question. However,

I did find that incorporating the follow-up survey provided a useful way of encouraging participants to think about features (and barriers), and their priorities, in advance of the interview.

In addition to identifying features of good digital tools and resources, the survey also asked participants to state any barriers they had experienced in their use (Appendix 1, Q11); again, I coded responses by barrier type, shown in Figure 4.6. It is clear from this diagram that the barriers affecting more researchers than any others are cost and the need for training. Other barriers referred to issues with maintenance and sustainability, usability, communication and transparency, data quality and reuse, scope, and lack of support.

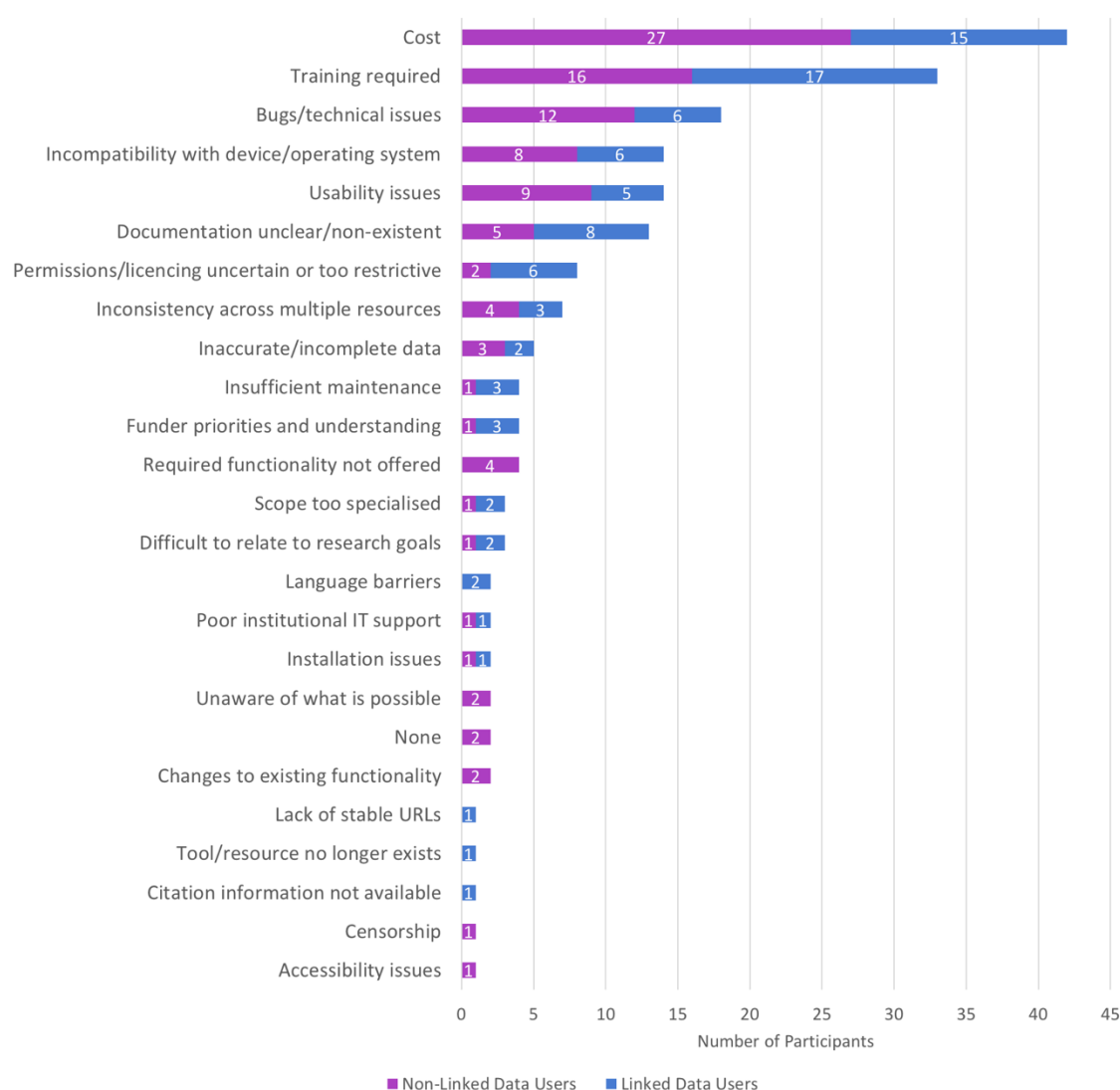


Figure 4.6 Bar chart showing barriers to digital tool/resource use identified by survey participants, segmented by participants' Linked Data usage

During the follow-up survey, I additionally asked interview participants to rank a selection of barriers in order of the extent to which they affected their use of digital tools and resources. Most barriers were selected due to the frequency with which they were mentioned in the original survey. I additionally chose 'Difficult to relate to research goals', 'Scope too specialised' and 'Unaware of what is possible', as these were particularly pertinent to my research questions. Using the same method described above, I assigned each barrier a mean importance score and used these to create the graphs in Figure 4.7, again segmented by participants' Linked Data usage.

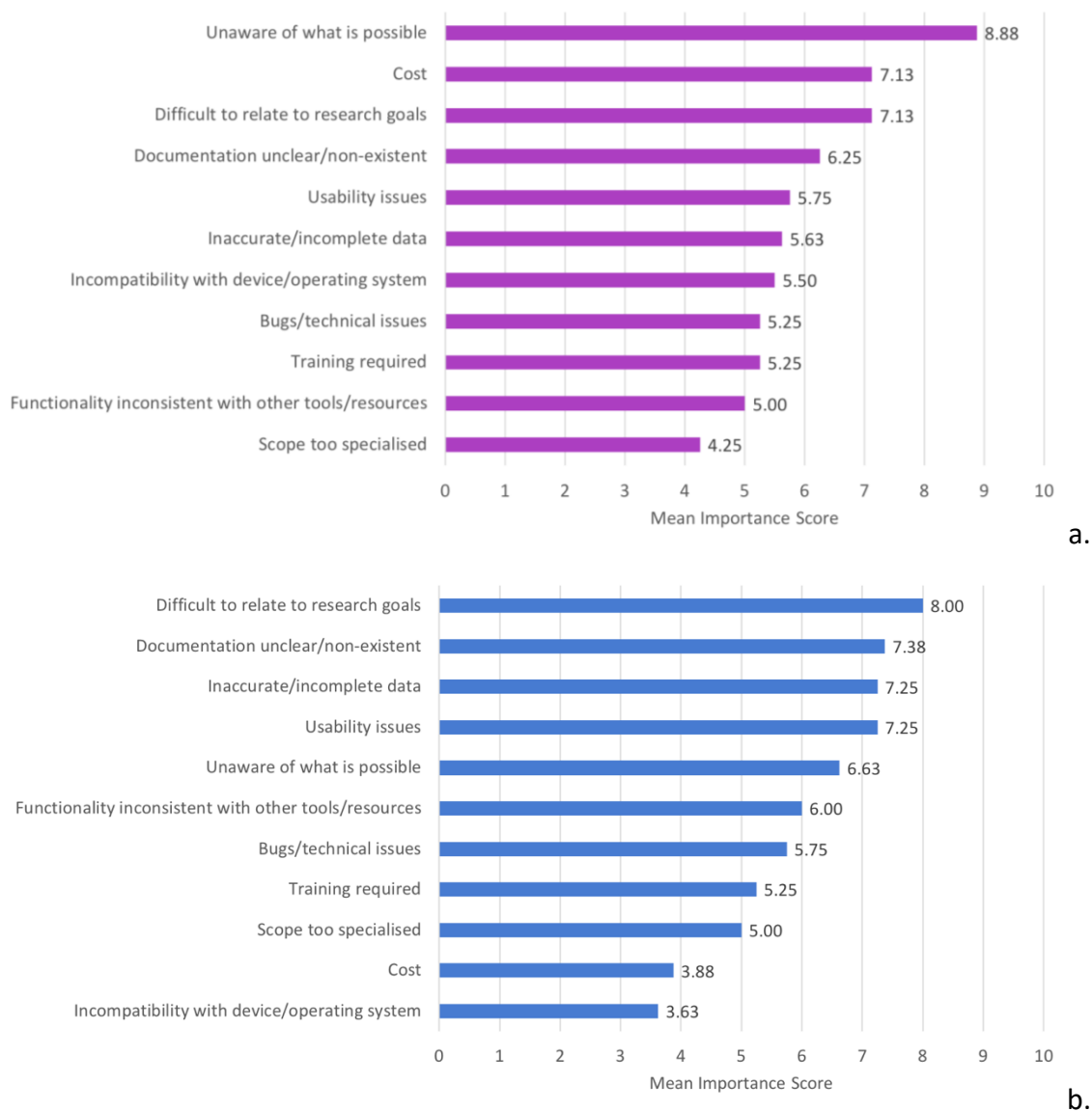


Figure 4.7 Barriers to digital tools/resources in order of the extent to which they affect interview participants: a. non-Linked Data users; b. Linked Data users

Although there were similar issues with the ranking system to those described in relation to features, above, comparing the rankings of these barriers indicates some difference in the priorities of participants who had or had not knowingly used Linked Data. As indicated by the original survey results in Figure 4.6, cost seemed to be a significant barrier to non-Linked Data users but had relatively little effect on Linked Data users. This discrepancy perhaps reflects many participants' reliance on subscription-based resources such as those discussed above, while many Linked Data resources are openly available. Furthermore, participants with higher levels of digital competence/confidence might be more aware of tools and resources available, including free alternatives to well-known subscription-based resources. Conversely, inaccurate or incomplete data seemed to have a greater effect on Linked Data users. All users, however, seemed to be affected most by finding tools or resources difficult to relate to research goals, being unaware of what is possible, unclear or non-existent documentation, and usability issues.

This section has comprised an overview of survey results relating to using digital tools and resources in general, including those most frequently used, features that particularly benefit participants, and barriers to their use. Overall, the most significant features and barriers related to usability, documentation, training, and cost, with aspects such as reliability, awareness and relevance increasing in importance in the follow-up surveys. There was some variation in responses between those who had knowingly used Linked Data and those who had not, indicating differing priorities. With this in mind, I will now discuss the survey results relating specifically to the use of Linked Data tools and resources.

#### 4.1.2 Using Linked Data Tools and Resources

As mentioned in 3.3, the 49 Linked Data users (37% of survey participants) were presented with a series of questions about their experiences. They were first asked which Linked Data tools or resources they had used, before stating the one with which they were most familiar and their perceived advantages of implementing a Linked Data approach for this tool or resource. Subsequently, participants were asked to compare their experience with that of using similar non-Linked Data tools or resources and suggest potential improvements.

The first question on Linked Data use (Appendix 1, Q14) asked participants to name Linked Data tools and resources they had used previously. I did not define ‘Linked Data’ at any point in the survey; therefore, responses to this question resulted from each participant’s own interpretation. As a result, on checking the documentation and/or publications for each tool and resource, I found a small number did not involve Linked Data (according to my broad definition in 1.3), largely because they used relational databases to connect data. I therefore removed them from my subsequent analysis.

Using the subset of responses confirmed as involving Linked Data, I classified the tools and resources into the categories used in 4.1.1. The majority (54%) related entirely or predominantly to the Ancient World, and the proportion of Specialist tools or resources (36%) was greater. There was a slightly lower proportion (10%) in the Academic category, with no tools or resources classed as Early Modern, Medieval or Generic, illustrating the small proportion of tools or resources obviously based on Linked Data, as well as their often specialised and technical nature. Responses relating specifically to Ancient World tools and resources are shown in Figure 4.8. The most frequently mentioned tools and resources, *Pleiades* and *Pelagios* (2.2.1), imply an interest in place, also demonstrated by the *Digital Atlas of the Roman Empire (DARE)*<sup>105</sup>, *Syriaca.org*<sup>106</sup> and the *Ancient World Mapping Center*<sup>107</sup>. As noted in 2.2.1, the concept of place is particularly amenable to Linked Data representation and related tools and resources are already relatively mature, explaining their popularity among survey participants. The remaining responses focus predominantly on objects or texts, and include *Trismegistos*, *Nomisma* (*CRRO*, *OCRE*, *MANTIS* and *CHRR* all use *Nomisma* data), the *British Museum* catalogue, *Papyri.info*, *EAGLE* and *Arachne* (all introduced in 2.2.4).

---

<sup>105</sup> <https://dh.gu.se/dare/>; provides persistent identifiers for ancient places (Bagnall & Heath, 2018, p. 183)

<sup>106</sup> <http://syriaca.org/>; provides persistent URIs for entities in the Syriac studies domain, such as places, and links to other Linked Data resources, including *Pleiades* (Michelson, 2016, pp. 75–77)

<sup>107</sup> <https://awmc.unc.edu/>; provides access to Linked Data resources, including *Pleiades*

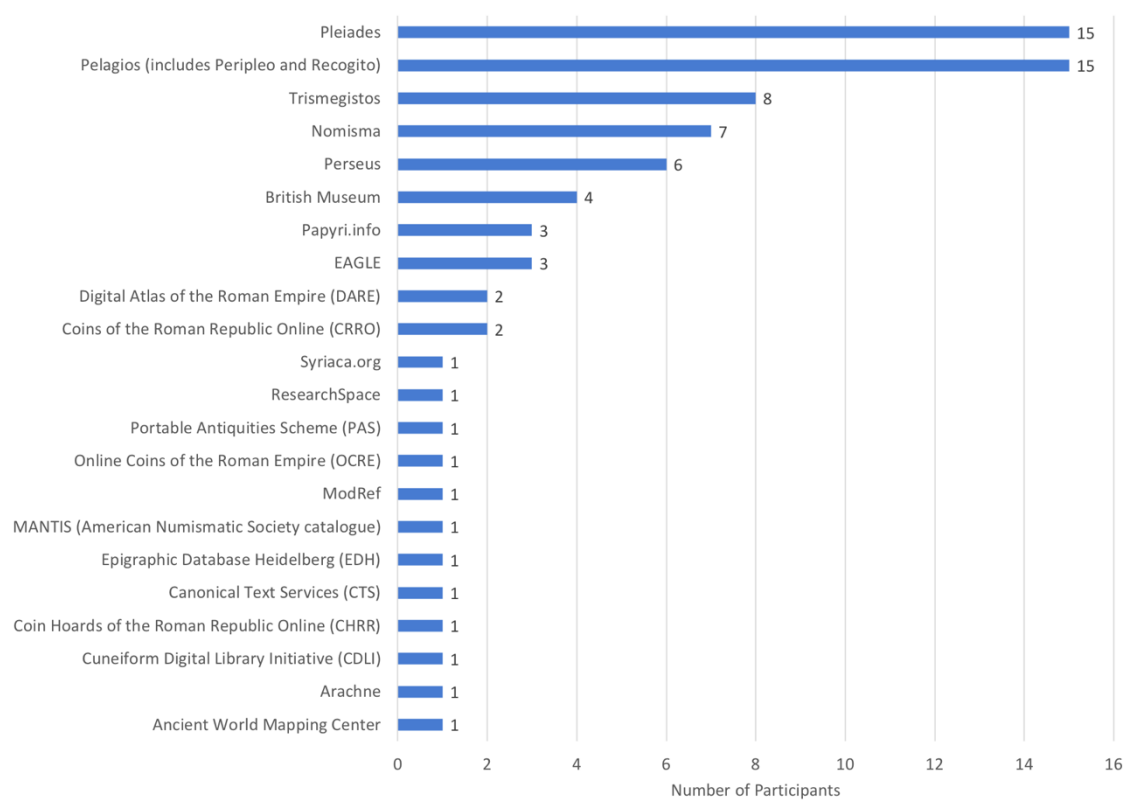


Figure 4.8 Bar chart showing all named Linked Ancient World Data tools/resources mentioned by Linked Data users

For the remaining questions in the Linked Data use section of the survey, participants were asked to specify the tool or resource with which they were most familiar (Appendix 1, Q14); results are shown in Figure 4.9. The vast majority were selected by one participant only, with *Pelagios*, *Pleiades* and *Papyri.info* being chosen more frequently, potentially indicating a relatively large user base, although the sample is too small to make any firm conclusions. Several of the most frequently mentioned tools and resources from Figure 4.8 do not appear (*Perseus* and *Nomisma* are particularly conspicuous by their absence) and others have moved further down the list (*Trismegistos* and *DARE*). This potentially indicates that while participants are familiar with these resources, they might use them relatively infrequently.

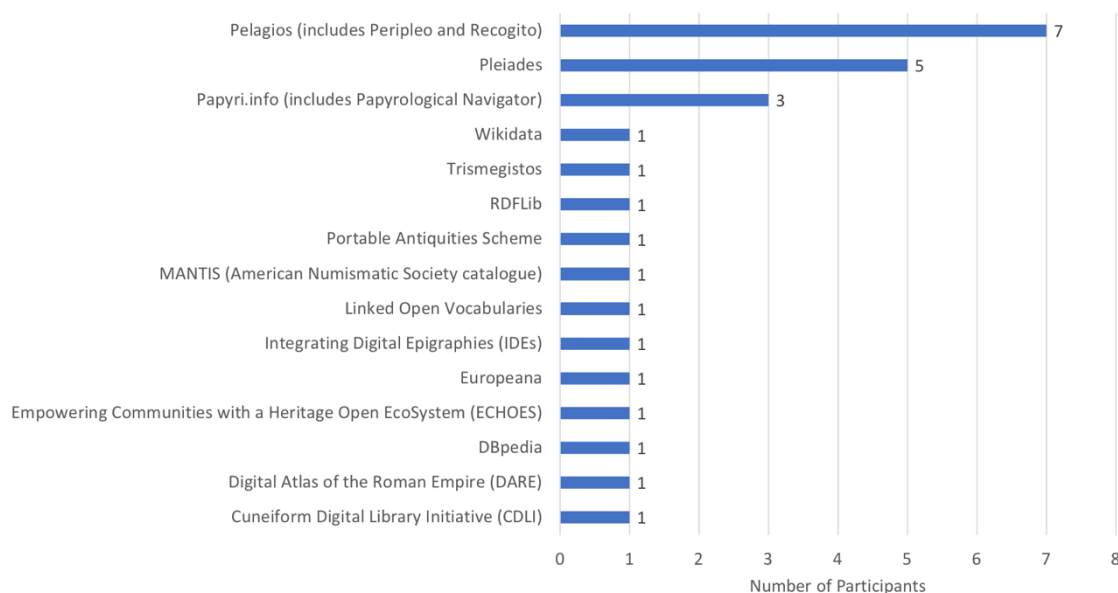


Figure 4.9 Bar chart showing Linked Data tools and resources used most frequently by Linked Data users

Other tools and resources mentioned in response to this question comprise specialist vocabularies and services for implementing and accessing Linked Data, with no specific disciplinary remit, such as *Wikidata* and *DBpedia* (both mentioned in 1.3). These responses demonstrate a variety of different approaches among participants, as well as a willingness to use Linked Data tools and resources outside of the Ancient World subject domain. Additionally, both Figure 4.8 and Figure 4.9 highlight the importance of *Pelagios* and *Pleiades* for research involving Linked Ancient World Data, as well as further confirming the significance of *Papyri.info* among the Ancient World research community more generally. The majority of tools and resources mentioned relate to places and objects, with no examples that focus on time (e.g. *PeriodO*, 2.2.2) or people (e.g. *SNAP:DRGN*, 2.2.3). Rather than indicating a lack of interest in these conceptual entities, this result is likely to suggest the relative lack of maturity of these resources, potentially resulting in barriers that deter their wider uptake, or less awareness of their existence among Ancient World researchers.

In relation to the Linked Data tool or resource with which they are most familiar, participants were asked whether the potential advantages of Linked Data were made clear to them (Appendix 1, Q17). Most participants (82%) indicated that this was indeed the case (5% responded in the negative, while 13% were unsure). Participants were then asked (Appendix 1, Q17a) what they understood the advantages of Linked Data to be, in relation to this specific tool or resource (Figure 4.10). While the sample

is too small to make any generalisations, this group of researchers found discoverability to be a particularly prominent advantage of implementing Linked Data for Ancient World research. Several of the comments categorised under 'discoverability' related more specifically to the ability to access multiple data sources from a central point, indicating that some of the appeal of Linked Data tools and resources might be their efficiency in condensing multiple searches into one. Other advantages identified by participants that facilitate discoverability include representing relationships between datasets, potentially via visualisation capabilities, and providing strong contextual information about objects and datasets.

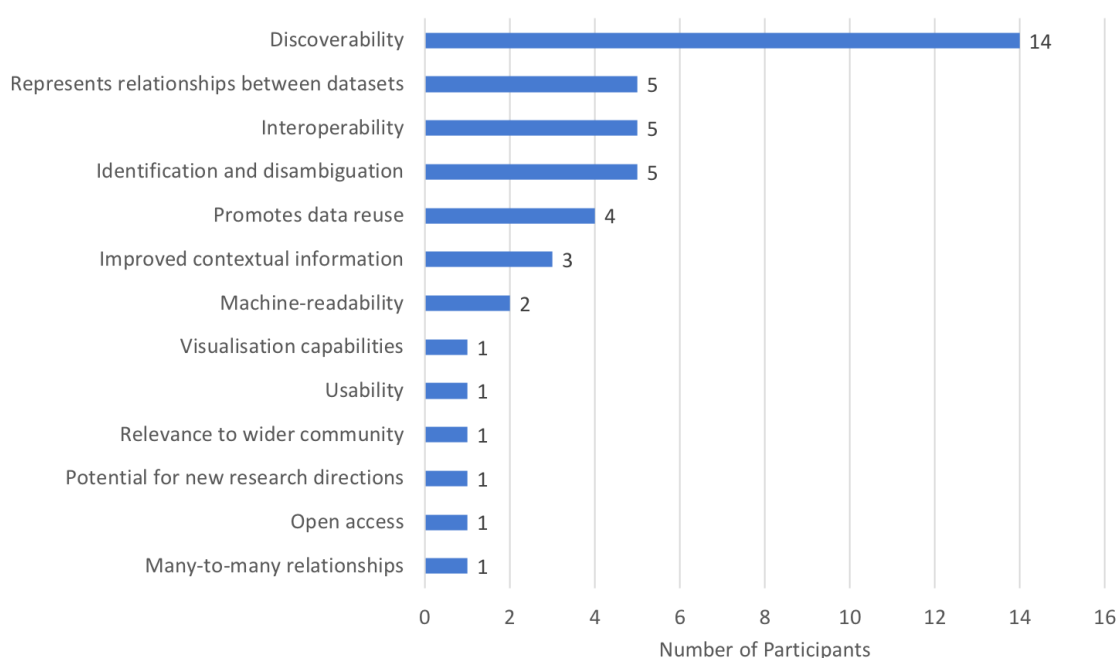


Figure 4.10 Bar chart showing advantages of Linked Data identified by participants from the information about specific tools/resources

To harness the potential for Linked Data discovery, tools and resources must additionally be usable, resulting in good search functionality, identified as a particularly important feature of digital tools and resources in 4.1.1. However, as Figure 4.10 shows, usability was mentioned by only one participant as a perceived advantage of Linked Data implementation for their chosen tool or resource (in this case, *Nomisma*). Having already discussed usability barriers to Linked Data tools and resources in 2.4, it is clear that usability is not an inherent advantage of implementing Linked Data technologies. Instead, it should be a key consideration when producing tools or resources with this approach.



Another group of advantages included in Figure 4.10 relate broadly to the theme of facilitating the research process. These include making tools and resources available open access, expanding their reach to a wider community, and increasing the potential for new research directions. Making these advantages clear is likely to appeal to a broad range of users and communicates the significance of structuring data in this way, from a research perspective. The remaining advantages apply to relatively complex operations and include interoperability, identification, and disambiguation, promoting data reuse, machine-readability, and many-to-many relationships. The fact that each was mentioned by a relatively small number of participants perhaps indicates that these advantages are not made immediately obvious but might be identified by consulting more detailed levels of the documentation.

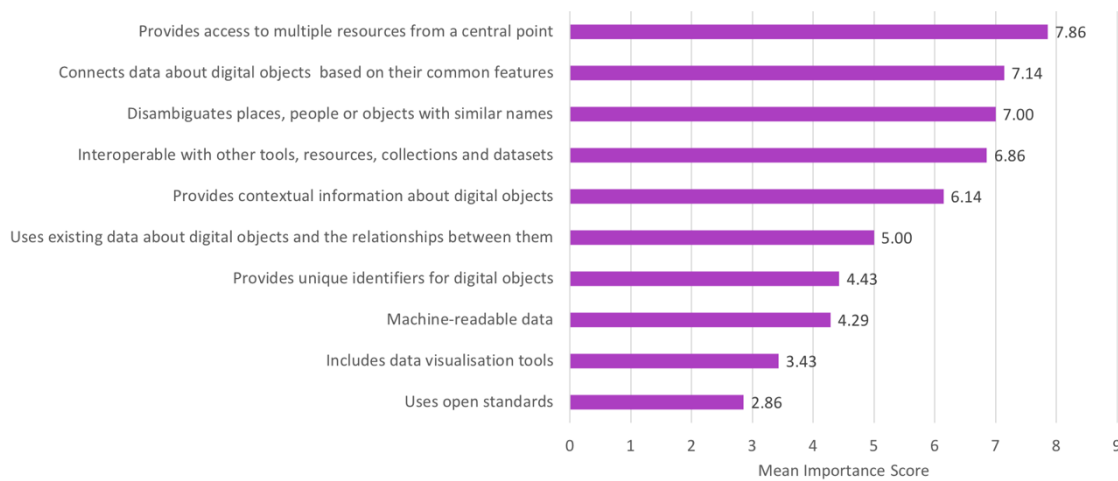
When comparing the responses to this question with the advantages I identified in 2.3, several were not mentioned by participants. For example, no one referred to the application of reasoning systems to Humanities data to make inferences; therefore, it is likely that the relevant information is either not included in tool or resource documentation, or that users do not consider it applicable to their needs. Although openness was mentioned, the related advantages of collaboration and sustainability were not made explicit in survey responses, potentially because users were more concerned with how a tool or resource met their immediate needs than its wider benefits for the research environment.

I used responses to this question to establish which advantages of Linked Data might be most appealing to those who have not knowingly used it, by including an extra question in the version of the follow-up survey sent to non-Linked Data user interview participants. Here, I provided ten "specific features of digital tools and resources" and asked participants to rank them in order of interest. Rather than copy the broad advantage categories included in Figure 4.10, I provided a greater level of specificity in their descriptions and split some categories into two, as shown in Table 4.1. Once I had received all responses, I converted participants' rankings to 'importance scores', as described in 4.1.1, and calculated the mean importance score for each feature; results are shown in Figure 4.11.

<b>"Specific Feature" Text from Follow-up Survey</b>	<b>Related Advantage from Original Survey</b>
Connects data about digital objects (e.g., texts, images, artefacts) based on their common features	Represents relationships between datasets
Disambiguates places, people or objects with similar names	Identification and disambiguation
Includes data visualisation tools	Visualisation capabilities
Interoperable with other tools, resources, collections and datasets	Interoperability
Machine-readable data	Machine-readability
Provides access to multiple resources (e.g., collections) from a central point	Discoverability
Provides contextual information about digital objects (e.g., texts, images, artefacts)	Improved contextual information
Provides unique identifiers for digital objects (e.g., texts, images, artefacts)	Identification and disambiguation
Uses existing data about digital objects (e.g., texts, images, artefacts) and the relationships between them	Promotes data reuse
Uses open standards	Interoperability

*Table 4.1 Linked Data "specific features" as included in the follow-up survey, with their relationships to advantages mentioned in the original survey*

Many advantages appear in a similar position on both Figure 4.11 and Figure 4.10. Both sets of participants focused primarily on discoverability, and are also clearly interested in representing relationships, interoperability, and disambiguation (non-Linked Data users were more interested in this application of unique identifiers than the identifiers themselves). Machine-readability and open standards appear towards the bottom of Figure 4.11; it became apparent during the interviews that participants struggled with the meaning of these terms, with several saying they might change their rankings following our discussion. The language and terminology used around Linked Data resources might therefore be considered another barrier to engaging fully with these technologies. If their advantages cannot be clearly understood, there is little incentive to use them. While it would be unwise to assume any definite conclusions from two small samples answering two slightly different questions, the advantages of Linked Data identified here provide some useful starting points for my discussion in subsequent chapters.



*Figure 4.11 Features of Linked Data tools/resources in order of importance as ranked by 7 non-Linked Data user interview participants*

After considering the perceived advantages of their chosen Linked Data tool or resource, survey participants were asked to compare it with other familiar digital tools and resources (Appendix 1, Q20). I divided the 24 responses into three categories:

**Positive:** participant gave entirely positive comments

**Negative:** participant gave entirely negative comments

**Neutral:** either a combination of the two, or the participant was noncommittal, or they provided information about the tool or resource rather than their opinion

Results indicate that although the positive comments (38%) clearly outweigh the negative (8%), most participants (54%) had neutral or mixed opinions. This result implies that there is more work to be done to improve the user experience, thereby illustrating the importance of RQ4, in determining how Linked Ancient World Data usability might be improved. Potential areas for improvement can be identified by looking in more detail at the comments themselves; I classified each comment type under a set of codes, with positive comments shown in Table 4.2 and negative comments in Table 4.3.

Comment	Number of Participants	Relevant Tool(s)/Resource(s)
Intuitive	6	Pelagios (2), Recogito (1), PAS (1), IDEs (1), Europeana (1)
Clear documentation	2	Pleiades (1), PAS (1)
Clear peer review system	1	Pleiades
Encourages user contributions	1	Wikidata
Good search functionality	1	PAS
High quality data	1	IDEs
Less complicated	1	Papyri.info

*Table 4.2 Positive comments from Linked Data users when comparing their chosen Linked Data tool/resource with other digital tools/resources*

Comment	Number of Participants	Relevant Tool(s)/Resource(s)
Takes time to understand	2	Peripleo (1), Europeana (1)
Difficult to see relevance	1	Pleiades
Few external links	1	IDEs
More filtering of search results required	1	Pelagios
Poor data quality	1	CRRO, CHRR*
Poor design	1	[tool/resource not specified]
Poor usability	1	[tool/resource not specified]
Slows down computer	1	Trismegistos
Unclear what to do	1	Pleiades

*Table 4.3 Negative comments from Linked Data users when comparing their chosen Linked Data tool/resource with other digital tools/resources (\*participant did not answer Q14 but included these resources in their response to Q20)*

The highest number of positive comments in Table 4.2 refer to the tool or resource being more intuitive than others they have used, which is contrary to the assumption in RQ4 that the usability of Linked Data tools or resources is relatively poor. These responses could therefore relate to particularly usable tools or resources or imply that participants' expectations of their usability were exceeded. Further comments relating to usability include improved search functionality and a lower level of complexity. Two

comments referred to the quality of the documentation, which correlates well with its importance in 4.1.1, as well as the greater degree of openness with which Linked Data tools and resources tend to be described. Other positive comments related to the provision and maintenance of high-quality data, while encouraging contributions from users.

Among the negative comments (Table 4.3), the only feature mentioned by more than one participant was the time required to understand the tool or resource and its structure, which (contrary to the positive comments about intuitiveness of some tools and resources) indicates a relatively poor level of usability. An explanation for why this process takes a relatively long time might be inferred from the barriers relating to training and documentation identified in 4.1.1, as well as the barriers to Linked Data implementation identified from previous research in 2.4. Further communication issues that could be addressed with sufficient documentation include 'Difficult to see relevance' and 'Unclear what to do'. Other negative comments relate to design, usability (including how the processing power required affects the speed of the user's device), and search functionality, as well as concerns about data quality. One participant commented on the lack of links to external sources, which could be due to an increased expectation that such content would be accessible via a Linked Data tool or resource.

The final question on Linked Data user experience (Appendix 1, Q22) asked participants to suggest what would improve their chosen tool or resource; 24 participants responded, with results shown in Table 4.4. While two participants were completely satisfied, the majority suggested at least one improvement, many of which are related to the themes identified at the end of the previous section. A particularly large number of comments related to improving the data behind the tool or resource, in terms of its quality and quantity, as well as its potential for integrating with and linking to external sources. Relatedly, two participants identified the benefits of encouraging contributions (e.g., data corrections or updates) from the user community. Four participants expressed a need for improved documentation, which includes two suggestions that this might take the form of use cases or tutorials aimed at less technically experienced users. Comments relating to usability and

discoverability included requesting better search functionality, simplified navigation, and curated content as a starting point for interaction with the tool or resource.

Improvement	Number of Participants	Tool(s)/Resource(s)
Better documentation	4†	Peripleo (1), Pleiades (1), Recogito (1)
Better quality data	4	MANTIS (1), Pleiades (1), PAS (1), Trismegistos (1)
Better search functionality	3	MANTIS (1), PAS (1), Wikidata (1)
More data	2	ECHOES (1), Europeana (1)
None	2	Linked Open Vocabularies (1), Papyri.info (1)
Better data integration	2†	Papyri.info (1)
Simplified navigation	2	Pelagios (1), PAS (1)
Better localisation	1	Recogito
Curated content	1	Pelagios
Easier to contribute	1	CRRO, CHRR*
Linked Data generation without coding	1	Peripleo
Larger community of contributors	1	Pleiades
More links to external data	1	IDEs

*Table 4.4 Improvements to Linked Data tools/resources suggested by Linked Data users (\*participant did not answer Q14 but included these resources in their response to Q20; †one of the participants who mentioned this improvement did not name a tool or resource in Q14)*

In this section, I have outlined the quantitative results relating to Linked Data usage, which has revealed topics to discuss further in later chapters. From using these tools and resources, participants understood that they were intended to facilitate the research process, by improving discoverability and interoperability, and providing disambiguation, all of which have the potential to open up new avenues for research. Comparisons with similar digital tools and resources were generally positive or neutral, with comments focusing on themes such as usability, discoverability, data quality, and documentation. Such topics were also at the forefront of participants' comments when asked to suggest improvements for these tools or resources.

The following sections will explore findings in relation to the production of digital tools and resources in general, and Linked Data in particular, which will provide further insight into potential usability improvements.

#### 4.1.3 Producing Digital Tools and Resources

The remainder of the survey focused on digital tool or resource production, with questions presented to the 31% of participants (3.4) who said they had been involved with the production of a digital tool or resource. Initial questions focused more generally on participants' preferred data structures and ascertained why non-Linked Data producers had not chosen this approach. Participants who had produced Linked Data then answered questions about their experiences, relating to their intended audiences, compliance with Berners-Lee's five-star model, and whether they would be willing to implement Linked Data again in future.

As part of the final section of the survey, producers were asked for their preferred data structures (Appendix 1, Q23a). Although this was an open-ended question, I did mention Linked Data, relational databases and text encoding as examples of possible responses. Responses from the 40 participants who answered are shown in Figure 4.12, segmented by whether or not the participant had produced a Linked Data tool or resource. Where a participant mentioned more than one approach (as was the case for 13 participants), all were included in the graph. By far the most popular responses were relational databases, text encoding and Linked Data, which suggests that including examples in the question text might have skewed the results.

Linked Data was mentioned by five participants who had not produced a tool or resource using this approach, which could indicate an increase in awareness and growing appreciation of its potential, or alternatively might reflect the survey's particular focus on Linked Data. Of those participants who selected Linked Data, one did so without having knowingly used a Linked Data resource, which (while surprising) is consistent with the finding in 3.4 that not all Linked Data producers were Linked Data users. It is unfortunate that no further information is available from the survey or interview phases to explain their circumstances, as a tool or resource's usability could be affected where the producer does not see themselves as a target user.

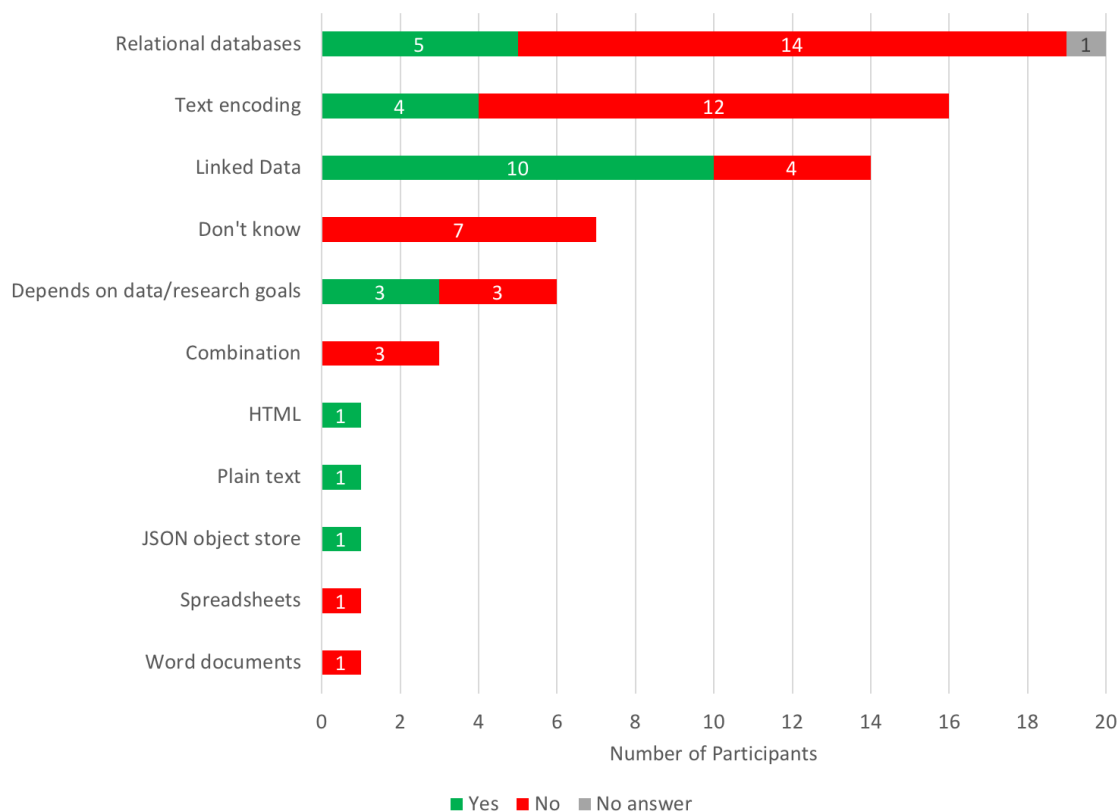


Figure 4.12 Bar chart showing producers' preferred data structures, segmented by their answer to the question of whether they have produced a Linked Data tool/resource

Some participants additionally provided reasons for their choice of data structure(s), which are shown in Table 4.5. The most frequently mentioned reason was 'Good integration with existing tools, resources and systems'. Half the participants who mentioned this reason did so in relation to Linked Data, which is likely due to its potential for interoperability, when implemented effectively, although differences in modelling approaches can pose a barrier to dataset alignment (2.4). A small number chose relational databases and text encoding, despite criticism that their inconsistent data models and terminology can restrict interoperability or prevent it entirely (1.2).



Reason	Number of Participants	Data Structure(s)
Good integration with existing tools, resources and systems	8	Linked Data (4), Plain text (1), Relational databases (2), Text encoding (1)
Established method for specified research processes	6	Relational databases (2), Text encoding (4)
Existing experience in this area	5	Relational databases (2), Text encoding (3)
Expressing relationships	5	HTML (1), Linked Data (4), Relational databases (2)
Expressing complexity	3	Linked Data (1), Relational databases (2), Text encoding (1)
Simplicity	3	Depends on data/research goals (2), Relational databases (1)
Discoverability	2	Linked Data (1), Relational databases (1)
Flexibility	2	Linked Data
Open standards	2	Linked Data (1), Plain text (1)
Sustainability	2	Plain text (1), Text encoding (1)
Usability by others	2	Depends on data/research goals (1), Plain text (1)
Advice from colleague	1	Relational databases
Contextualising capabilities	1	Linked Data
Insufficient understanding of Linked Data	1	Relational databases
Limited experience of other approaches	1	Relational databases, Spreadsheets, Word documents
Visualising relationships	1	Relational databases

*Table 4.5 Reasons participants provided for preferring particular ways of structuring data (NB: some reasons were provided in relation to more than one type of data structure, so the totals of the numbers in the 'Data Structure(s)' column sometimes exceed the 'Number of Participants')*

This choice is likely due to their familiarity with these methods, or because they designed their datasets with specific systems in mind. The reasons 'Established method for specified research processes' and 'Existing experience in this area' confirm this interpretation, indicating that some participants' choice of data structure may be based on the path of least resistance. Researchers want to use the solution that will be most efficient for them to implement, to achieve their research goals in the time

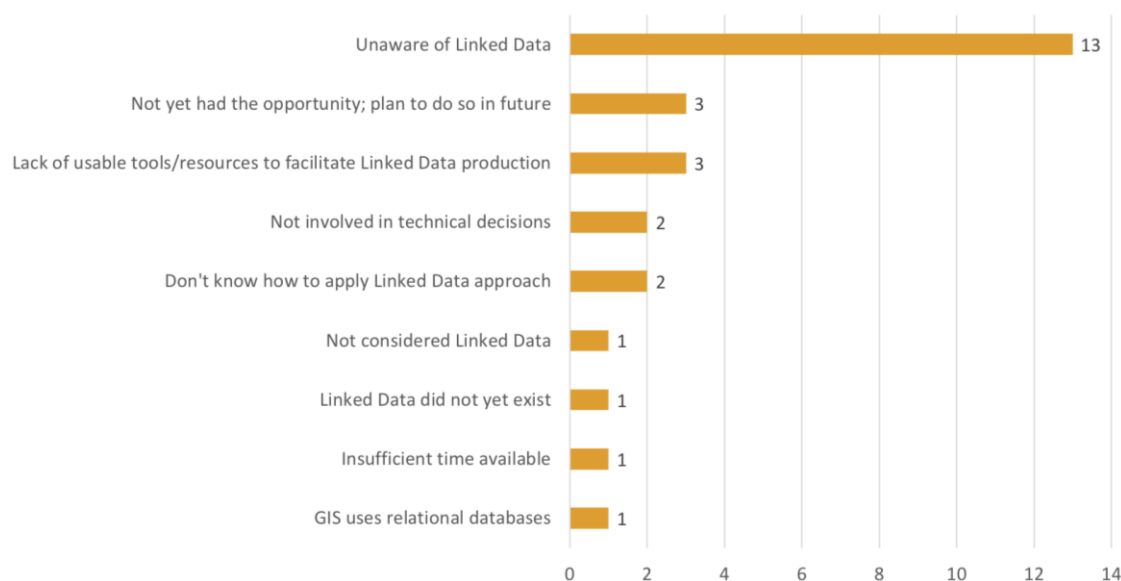
available, and in many cases, these types of data structure might be sufficient for their purposes. However, this could be at the expense of other options with greater potential for interoperability and reuse. These responses could additionally imply that participants are reluctant to consider new ways of working or are simply unaware that they exist.

Reasons mentioned by fewer participants concern specific features of their chosen data structure, including 'Expressing relationships', 'Expressing complexity', 'Simplicity', 'Discoverability', 'Flexibility', 'Open standards', 'Contextualising capabilities' and 'Visualising relationships'. Selecting a data structure based on features or functionality implies an informed decision. Linked Data appears frequently in these cases, with fewer mentions of relational databases or text encoding. All these reasons correspond to advantages of Linked Data discussed in 2.3; therefore, greater awareness of the benefits of Linked Data may cause a shift towards this approach.

Usability by others was only explicitly considered by two participants as a reason for selecting their preferred data structure (neither of whom mentioned Linked Data), although it was likely an implicit factor in some of the other reasons. I might infer from this that participants tended to prioritise the production of data for their specific purpose, with its reuse in other contexts being a secondary concern. Such a situation would suggest that there is little incentive to explicitly ensure usability in the production of digital tools or resources or their underlying data – a particular issue where Linked Data is concerned, due to its inherent complexity and lack of mature tools for its implementation, as discussed in 2.4.

Reasons for participants' preference of other approaches over Linked Data become clearer when non-Linked Data producers explained why they had decided not to use a Linked Data approach (Appendix 1, Q24). This was an open-ended question, answered by 26 participants, for which I categorised responses as shown in Figure 4.13. While many responses reflect the barriers discussed in 2.4, such as requiring training, time constraints and lack of usable resources for Linked Data production, by far the most frequent response was that the participant was unaware of its existence. In these

cases, therefore, participants made their decision on which data structure to use without being fully informed of all the options available.



*Figure 4.13 Bar chart showing reasons why producers had not chosen a Linked Data approach*

Three participants stated they had not used Linked Data previously but had plans to do so in future, suggesting that awareness of Linked Data technologies might be becoming increasingly common, but that this continues to be a gradual process. No participants answered this question with negative experiences of using Linked Data or by stating that any other data structure is inherently better, which supports the assertion that an increase in awareness, training, and usable tools and resources might encourage more researchers to produce data based on this approach.

Focusing on the 22 participants (35% of all digital tool and resource producers) who had been involved in producing Linked Data (3.4), they were presented with a series of extra survey questions about the tool or resource they had produced most recently. Of the 15 such tools or resources named by Linked Data producers (Appendix 1, Q25)<sup>108</sup>, there were two examples for which I could not find any clear evidence that Linked Data was involved, according to my definition in 1.3. One was a project that appears to have involved linking data using relational databases; the other was a participant's personal website that included no information about its data structures. Responses relating to these tools and resources have therefore been omitted from my analysis.

<sup>108</sup> To preserve participant anonymity, I will not name individual tools or resources in this section

To explore participants' plans for the usability of their tool or resource, I asked about its intended audience (Appendix 1, Q28). Results from the 14 participants<sup>109</sup> who answered this question are shown in Figure 4.14, which suggests at first glance that most tools and resources they produced were intended for an Ancient World audience, with some expectation that their users would possess technical skills. However, participants could select more than one option, and most did so (Figure 4.15), indicating their intention for these tools and resources to fulfil multiple user goals, thereby increasing their potential complexity.

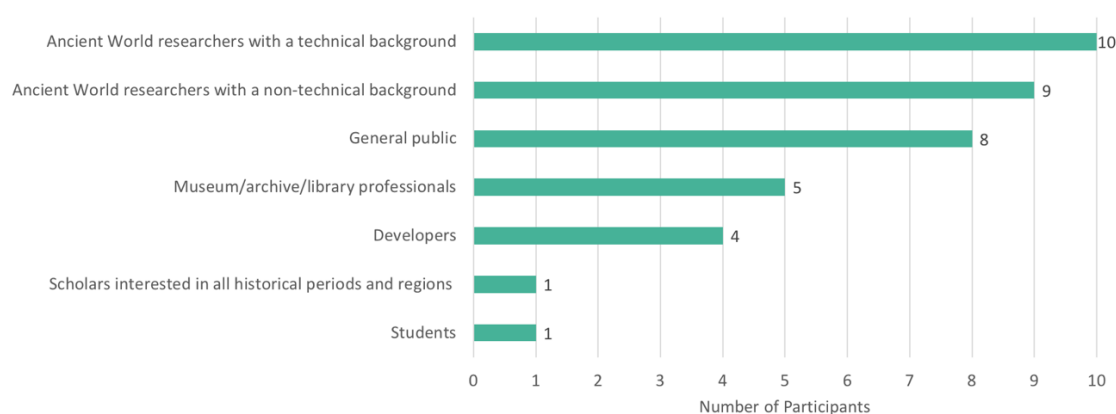


Figure 4.14 Bar chart showing the number of participants who selected each audience group in relation to the Linked Data tool/resource they most recently produced

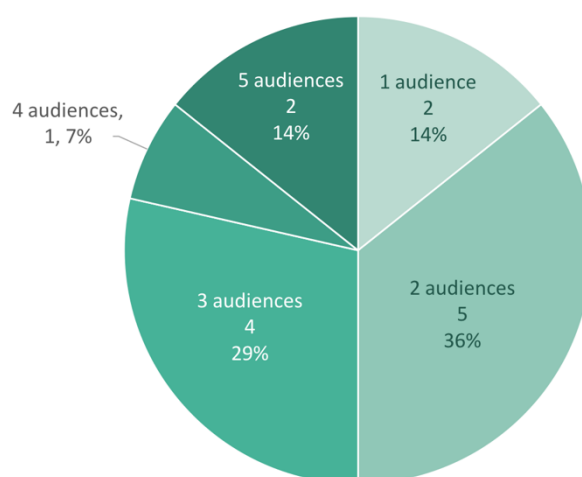


Figure 4.15 Pie chart showing the number of audience groups for which each tool/resource was intended

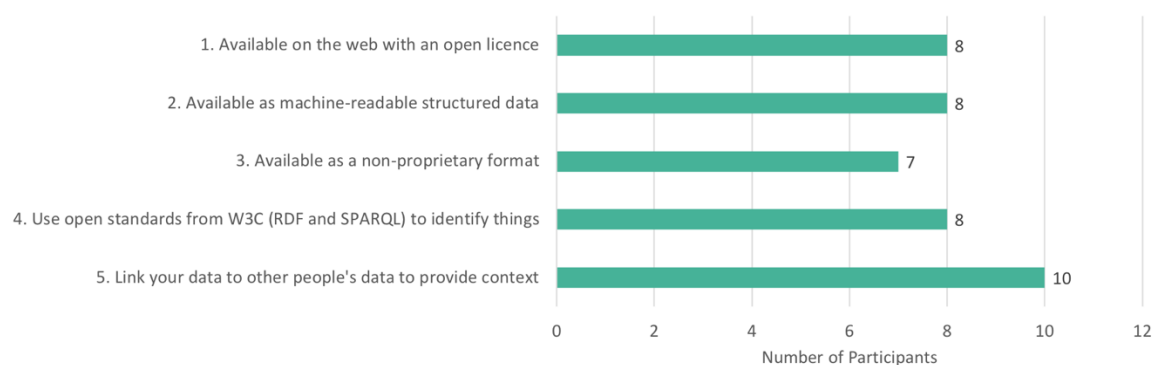
<sup>109</sup> One participant did not name a tool or resource in Q25

Combination	Audience Groups Included	Audience Groups Excluded	Number of Tools/Resources
All Ancient World researchers	<ul style="list-style-type: none"> <li>• Ancient World researchers with a technical background</li> <li>• Ancient World researchers with a non-technical background</li> </ul>		6
Non-technical audiences included	<ul style="list-style-type: none"> <li>• Ancient World researchers with a non-technical background</li> <li>• General public</li> </ul>		7
Technical audiences excluded		<ul style="list-style-type: none"> <li>• Ancient World researchers with a technical background</li> <li>• Developers</li> <li>• Museum/library/archive professionals</li> </ul>	3
Technical audiences only	<p>At least two of:</p> <ul style="list-style-type: none"> <li>• Ancient World researchers with a technical background</li> <li>• Developers</li> <li>• Museum/library/archive professionals</li> </ul>	<ul style="list-style-type: none"> <li>• Ancient World researchers with a non-technical background</li> <li>• General public</li> </ul>	3

*Table 4.6 Frequency of selected audience group combinations for which Linked Data tools/resources were intended*

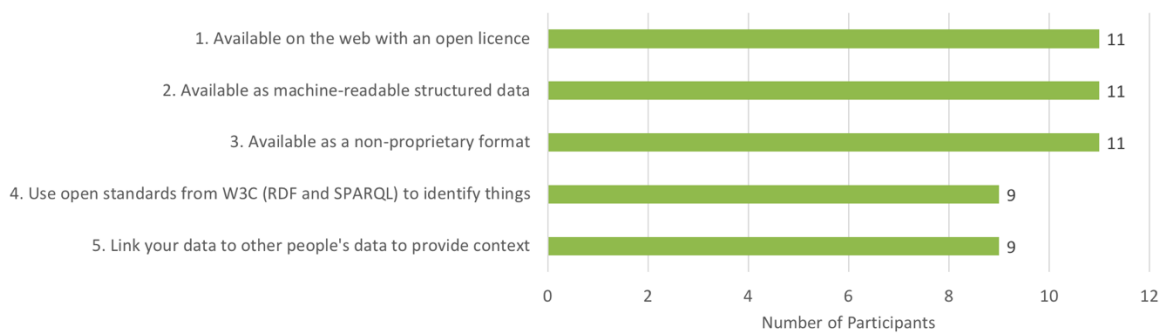
To gain an insight into the potential purposes of these tools and resources, I looked in more depth at the audience combinations included in participant responses (Table 4.6). Nearly half the tools and resources were intended for Ancient World researchers whose goals were likely to vary in levels of technical complexity, potentially requiring a wide range of functionality. Similarly, many tools and resources included audiences with few technical skills alongside more experienced users, again implying a need to balance complexity with usability. A small proportion of tools and resources were intended solely for less technical audiences, which could have led to a greater focus on producing a usable interface, albeit potentially at the expense of access to the underlying data. Finally, the same number of tools and resources were aimed specifically at audiences wishing to perform more technical operations, potentially excluding other user groups.

The following question (Appendix 1, Q29) asked Linked Data producers about the extent to which their data complied with Berners-Lee's five-star model (1.3). Of the 13 participants who responded, the number who selected each star is shown in Figure 4.16. In several cases, however, I found that the participant's assessment did not align with my own knowledge of the tool or resource in question, which led me to consult the documentation for each one, where possible. Unfortunately, one participant did not name a tool or resource, and another was no longer available; my assessment of the remaining 11 is shown in Figure 4.17. In all cases, my assessment either agreed with the participant's or added extra stars; there were no cases where I disagreed with a participant's selection of a star. It is also possible that more tools or resources complied with stars 4 or 5, but that this information was not readily available on their websites. Participants were therefore more likely to underestimate their compliance with the five-star model than overestimate. This underestimation was particularly apparent for stars 1-3, potentially due to a misunderstanding of the terminology, or because some participants assumed that selection of a star automatically implies compliance with those that precede it numerically (e.g., selecting star 4 would imply compliance with stars 1-3).



*Figure 4.16 Bar chart showing the number of participants who selected each of Berners-Lee's five stars in relation to the Linked Data tool/resource they produced most recently*

The survey results, supplemented by my own analysis, demonstrate that the majority of Linked Data tools and resources produced by participants meet all of the five-star criteria, with all meeting at least the first three stars. However, although compliance with the model implies that their data is accessible and interoperable, it provides no indication of usability, either of the data itself, or the tool or resource through which it is accessed.



*Figure 4.17 Bar chart showing the number of participants whose tool/resource actually meets the criteria for each of Berners-Lee's five stars*

Finally, Linked Data producers were asked whether they would choose Linked Data again if producing a similar resource in future (Appendix 1, Q30). Results indicate that most participants' experiences of Linked Data production were predominantly positive (78%), although some participants were unsure (22%). No one was completely deterred from further Linked Data production, which suggests that once they overcame the barriers discussed in 2.4, having invested their time in training and consideration of complex technical aspects, the majority were prepared to apply their experiences to future projects.

In this section, I have outlined the quantitative survey findings relating to digital tool and resource production, with a specific focus on Linked Data. Overall, the majority of participants favoured relational databases or text encoding over Linked Data, despite their aims aligning with known advantages of Linked Data, such as interoperability and the scope for expressing relationships. Rather than making an informed choice, in many cases participants did not choose to produce Linked Data because they were unaware of its existence. Others struggled with a lack of usable tools and resources to facilitate Linked Data production, or simply not knowing where to begin. Among those participants who had produced Linked Data, many intended wider usability of their tools and resources, and (intentionally or not) tended to comply with Berners-Lee's five-star model, although I noted that compliance does not necessarily lead to usability. The majority would be willing to apply this approach again, suggesting that investing more time in Linked Data training and production at the outset of a project could ultimately lead to multiple tools and resources that demonstrate its benefits.

Before moving on to explore survey findings relating to participants' research methods, I will conclude this section with a brief summary of my findings relating to digital tool and resource use and production.

#### 4.1.4 Summary: Tools and Resources

In this part of the chapter, I have explored initial quantitative findings from my surveys in relation to digital tool and resource use and production, with a specific focus on Linked Data. When investigating the most popular digital tools and resources used by participants in 4.1.1, the majority seemed to lack transparency around the data models, standards, and formats used; in many cases, such information was unavailable. This situation is concerning because it limits the scope for critical analysis by the end user, illustrated by the confusion among some survey participants when attempting to identify Linked Data tools and resources in 4.1.2. If the rationale behind data modelling processes is unclear, it is difficult to verify the authority or quality of a tool or resource's contents, thereby undermining its academic credibility.

My analysis has additionally facilitated the identification of several key themes, which represent aspects of Linked Data tools and resources where usability improvements could be most effective. Participants additionally discussed their use of these tools and resources in terms of features that were important to them, as well as barriers that would be prohibitive to their use. Combining these features and barriers demonstrates that the following aspects of digital tools and resources were most important to participants:

- usability;
- documentation;
- training;
- cost;
- reliability;
- awareness.

These aspects are not discrete entities, but are inextricably linked, all feeding into my main theme of usability, to varying extents. For example, *usability* implies *reliability* and potentially reduces the need for users to consult *documentation*. In turn, clear



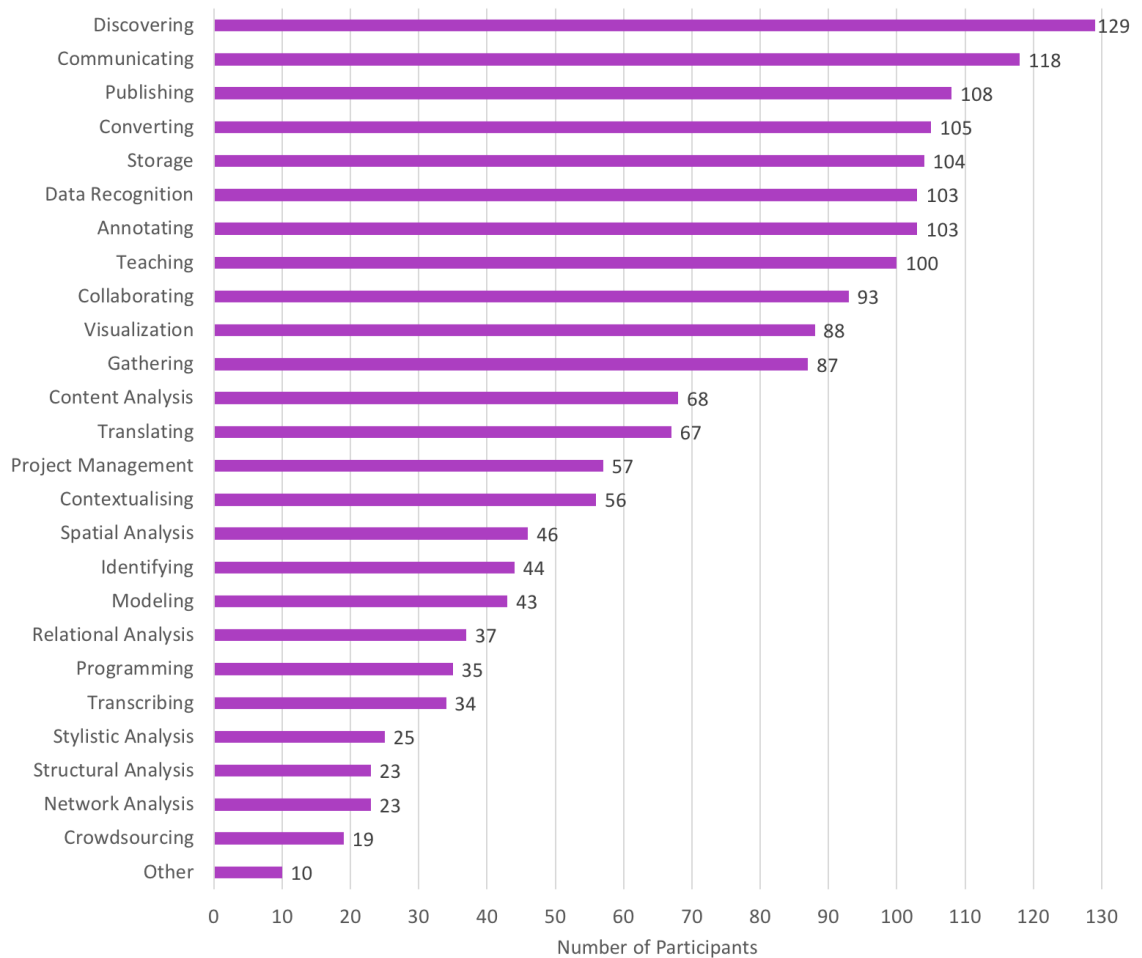
*documentation* could reduce the need for specific *training* to use the tool or resource, while additionally communicating its scope, thereby ensuring potential users are *aware* of how it might be able to help them. *Cost* and lack of *awareness* can act as gatekeepers to a tool or resource, either allowing users to incorporate it into their research process or preventing them from doing so. When participants were asked about their perceived advantages of Linked Data tools and resources, the majority focused on the themes of *usability*, *discoverability*, *data quality* and *documentation*. These themes align well with the more general responses above, indicating that Linked Data could facilitate existing research processes, even among less technically-inclined participants.

Many participants who had produced Linked Data demonstrated an interest in ensuring usability of their tool or resource by non-technical audiences, despite usability rarely factoring into a producer's choice of preferred data structure. The majority of Linked Data tools and resources complied with Berners-Lee's five-star model, indicating that they should meet the needs of users in terms of *data quality* and openness (i.e., lack of *cost*); however, such compliance does not necessarily imply that resulting tools and resources are *usable*, *reliable*, *discoverable* or *well-documented*. Most Linked Data producers would consider producing Linked Data again, demonstrating that this is a worthwhile approach for digital tools and resources relating to Ancient World research, with their numbers likely to increase in future. With this in mind, the initial findings from my survey indicate that now is the optimum time for investigating usability of existing resources and recommending measures that might be taken to optimise future fulfilment of user needs. Identifying ways in which Linked Ancient World Data usability might be improved, in response to RQ4, is therefore both pertinent and timely.

To explore the above issues fully, it was also crucial to identify existing research methods where Linked Data might most effectively be integrated, to address RQ3. The following section will therefore focus on quantitative findings relating to survey participants' research activities.

## 4.2 Research Methods

As discussed in 3.3, I asked survey participants to select activities from a list of TaDiRAH terms that applied to their use of digital tools and resources for Ancient World research, as well as their use and/or production of Linked Data more specifically (where applicable). Using an established framework instead of a free-text field provided consistent categorisation of research activities and improved efficiency of analysis. Additionally, it is highly likely that using a multiple-choice question resulted in more responses (163 responses on digital tools and resources in general) than a free-text question (the highest number of responses for a free-text question was 121). This section starts with an analysis of responses about methods relating to participants' use of digital tools and resources more generally, before focusing on the more specialised contexts of using and producing Linked Data tools and resources.



*Figure 4.18 Bar chart showing research activities participants associated with their use of digital tools/resources*

The first survey question involving TaDiRAH methods categories (Appendix 1, Q6) asked participants to select all activities they associated with their use of digital tools and resources for Ancient World research; results are shown in Figure 4.18. The most common digital research activity was *Discovering*, which aligns well with the identification of discoverability as a major advantage of Linked Data in 2.4 and 4.1.2. While many of the other most popular activities (e.g. *Communicating*, *Publishing*, *Converting*, and *Storage*) relate to more generic properties of digital tools and resources, the methods *Data Recognition*, *Annotating*, *Visualization*, and *Gathering* were all mentioned by more than half the participants who responded to this question, and might benefit more specifically from Linked Data integration; for example, these four activities align well with the tools and processes involved in *Pelagios* (2.2.1).

To gain more explicit information about which research methods might be most amenable to Linked Data integration, the survey subsequently (Appendix 1, Q18) asked Linked Data users about their intended activities when using the Linked Data tool or resource with which they are most familiar; results from the 33 participants who responded are shown in Figure 4.19. Here, *Discovering* has been overtaken by *Gathering* as the most common research activity, with *Data Recognition*, *Visualization*<sup>110</sup> and *Annotating* also ranked highly. The chart additionally demonstrates a rise in popularity among some more specialised activities when compared to Figure 4.18. These include *Contextualizing*, *Relational Analysis*, *Identifying*, and *Spatial Analysis*. While there may be less demand for these activities among Ancient World researchers overall, these could be areas in which Linked Data might be integrated more readily, to have a significant impact on specific user communities. Conversely, more generic activities (*Communicating*, *Publishing*, *Converting*, and *Storage*) have slipped further down the list, indicating that Linked Data technologies might be less relevant here.

---

<sup>110</sup> When referring to TaDiRAH terms by name, I use the same (-ize) spelling as included in TaDiRAH itself, e.g., *Visualization*; on all other occasions, I use the UK (-ise) spelling, e.g., *Visualisation*

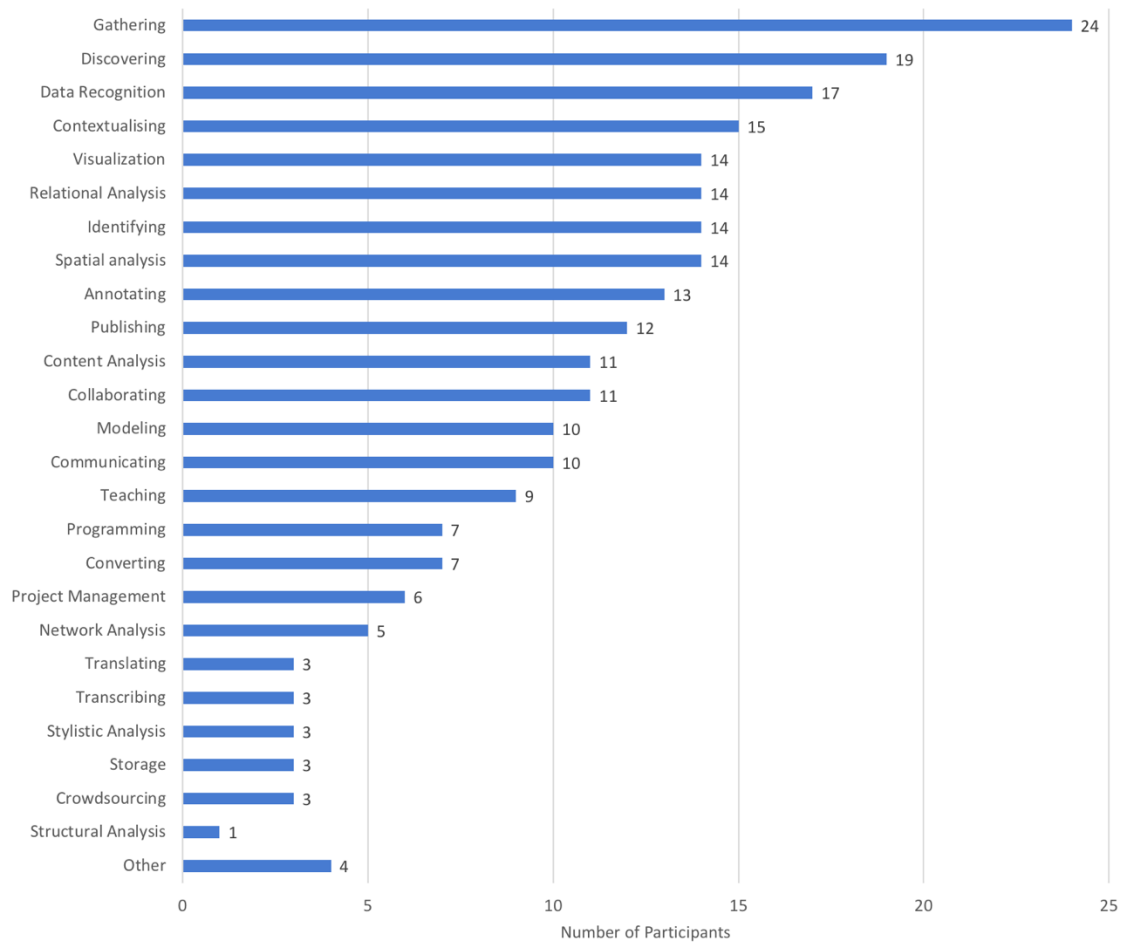


Figure 4.19 Bar chart showing research activities participants associated with their use of Linked Data tools/resources

Later in the survey (Appendix 1, Q27), Linked Data producers were asked to select research activities associated with the Linked Data tool or resource they most recently produced. A much smaller subset of 16 participants answered this question; results are shown in Figure 4.20. While many of the activities are in similar positions to those selected by Linked Data users (Figure 4.19), there are some noticeable differences. Most strikingly, *Discovering*, which was so prominent in responses to the previous questions, appears further down the list; *Data Recognition* also appears to be a lower priority for Linked Data producers. The relatively low position of *Discovering* in Figure 4.20 might indicate that (despite its obvious interest to potential users) discovery alone is no longer sufficient motivation to build a new tool or resource or enhance an existing one, potentially due to pressure from stakeholders to focus on more specialist activities. However, two generic activities, *Publishing* and *Collaborating*, featured relatively highly in relation to Linked Data production, but with relatively low interest

among Linked Data users, which perhaps implies a disconnect between actual user needs and those assumed by tool and resource producers.

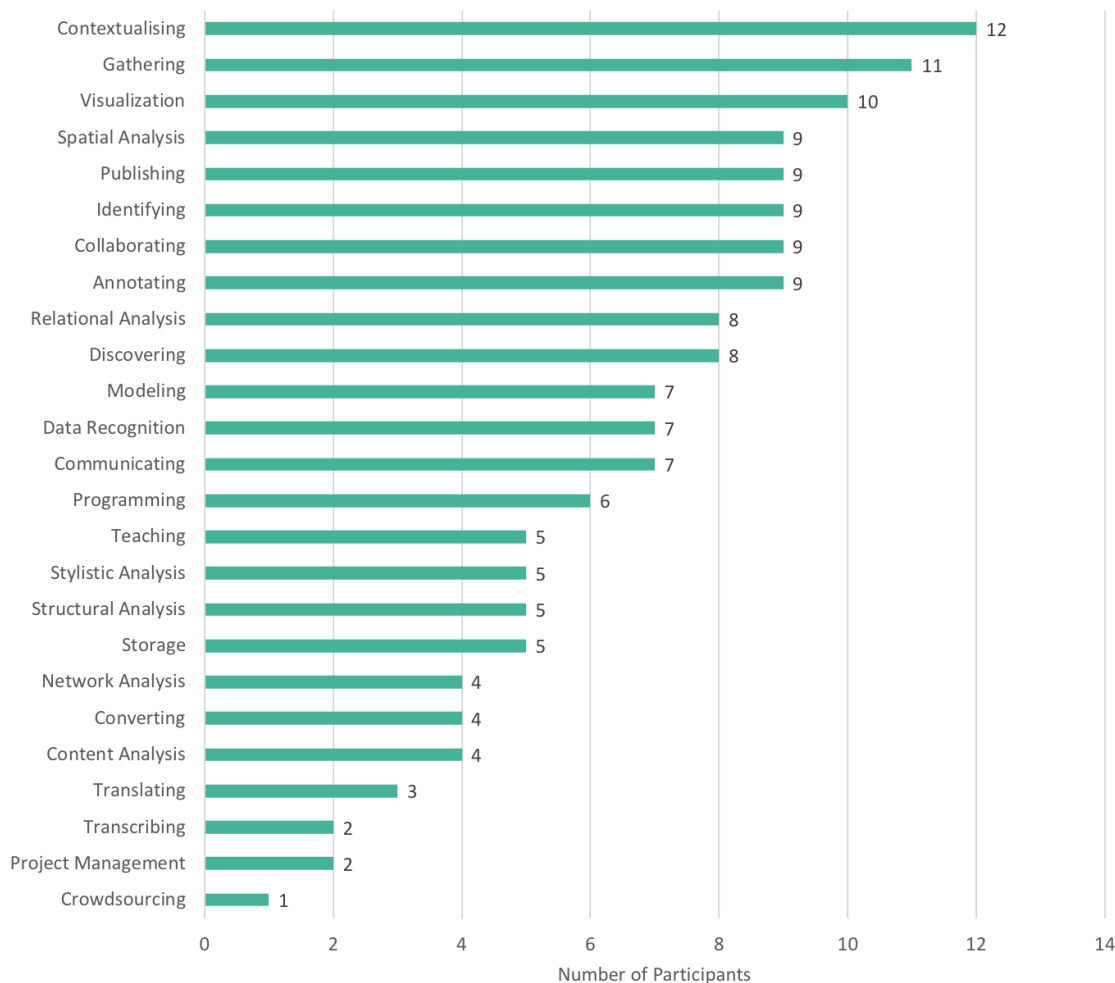


Figure 4.20 Bar chart showing research activities participants associated with production of Linked Data tools/resources

Several specialist activities were selected by a relatively high proportion of participants in relation to Linked Data use and/or production, but a relatively low proportion of participants associated them with general digital tool or resource use. These include *Contextualising*, *Spatial Analysis*, *Identifying*, *Modeling*, *Relational Analysis*, and *Programming*. All six activities were additionally selected by a higher proportion of Linked Data producers than users, with *Contextualising* having a particularly noticeable disparity between the two groups. These findings suggest a high level of interest in specialist activities among the most technically skilled participants, providing the motivation to produce tools or resources with such activities in mind. However, there appears to be less of a demand for such activities among many Ancient World researchers. Therefore, another potential barrier to Linked Data use (in addition to

those discussed in 2.4) might be that many such tools or resources are designed for research activities of interest to a relatively limited audience.

With all three of the research methods survey questions, there is a possibility that some or all participants were not sufficiently comprehensive in their responses. The variation in the number of methods selected was particularly pronounced in response to the Linked Data production question. While some participants were more selective, choosing only two or three methods, others chose ten or more. This discrepancy implies that some producers were thinking more broadly about all conceivable activities for which their tool or resource might be used, while others mentioned only the activities that aligned with their primary production goals. It is difficult to know how different people approached this question as they were not asked to rank the activities they selected.

In applying the TaDiRAH framework to my survey, I found that more specialised methods became more popular as the relative skill level of participants increased, from general digital tool or resource users, to Linked Data users, and finally Linked Data producers. However, *Discovering*, *Gathering*, *Data Recognition*, *Annotating* and *Visualization* appeared in the top half of the results for each of the three questions. These activities should therefore have a wide appeal to participants of all technical skill levels, while additionally being relevant to known advantages of Linked Data, as discussed in 2.3. I incorporated specific interview questions to explore these research methods in more detail; participants' responses form the basis for my discussion in the following chapter.

#### 4.3 Conclusions: Tools, Resources and Methods for Ancient World Research

This chapter has provided an overview of the quantitative survey results relating to digital tool and resource use and production (with specific reference to Linked Data) as well as associated research methods. To determine how Linked Data usability might be improved (RQ4), looking more widely at digital tools and resources for Ancient World research has already begun to provide important insights into the research processes and challenges experienced by potential users. Through analysing participants' responses to open-ended survey questions, I identified key components of usability,

which relate either to the user's interaction with a tool or resource, or its data (incorporating themes such as reliability and data quality), as well as various aspects that affect tool and resource production (including documentation and cost). As such, my discussion in response to RQ4, on improving the usability of Linked Ancient World Data tools and resources, will be split into two chapters: Chapter 6 will focus on specific components of the user experience and Chapter 7 on facilitating the production of a usable tool or resource.

Regarding the survey questions on research methods, *Discovering* was the most popular method overall. This finding corresponds to the popularity of 'Search functionality' among features of good tools and resources, as identified by participants in 4.1.1.1, as well as the discoverability advantages of Linked Data discussed in 2.3. A related feature was awareness, in that improving the discoverability of tools and resources ensures that more potential users are aware of their existence and therefore able to benefit from their use. Discoverability was also the most frequently mentioned advantage of Linked Data tools and resources, with comments indicating that this advantage often manifests itself in the ability to access multiple datasets from a central point. Responses in all sections of the survey demonstrated that participants were interested in using digital tools and resources to discover objects such as texts and artefacts, with places being of particular interest to Linked Data users. Applying Linked Data technologies to discovery tools and resources, as well as ensuring their usability by non-technical audiences, could therefore be a very effective means of demonstrating the benefits of this approach. Other methods on which I decided to focus during the interview phase were *Gathering*, *Data Recognition*, *Annotating* and *Visualization*, largely because of their relevance to participants with varying levels of technical skill.

While quantitative analysis has provided some interesting findings, I consider these a starting point for further investigation. As such, the following chapters will explore my findings in relation to more nuanced qualitative responses, which will begin in the following chapter by considering how Linked Data might be integrated with the above research methods.

## 5 Research Methods for Linked Data Integration

The previous chapter introduced my findings by focusing on elements of my survey that could be quantified, thereby providing a broad overview of the participant population. In this and the following two chapters, I will take a narrower, more detailed approach by focusing on qualitative elements, predominantly from my interviews, but also incorporating responses to open-ended survey questions. These findings will then form the basis for my recommendations in Chapter 8. The current chapter addresses the question of Linked Data integration with Ancient World research methods (RQ3). It builds on my findings from 4.2, in which I identified five research methods that were frequently selected in relation to digital tool and resource use in general, as well as Linked Data use and production. As such, I considered them potentially amenable to Linked Data integration.

This chapter will discuss each method in turn, in approximately the order in which they might be expected to occur during the research process (although, as I will demonstrate, this is not always the case). I will start with *Discovering* (5.1), then the related method *Gathering* (5.2), before exploring three methods associated with analysis, *Data Recognition* (5.3), *Annotating* (5.4), and *Visualization* (5.5). Within these sections, methods often manifest themselves as discrete tasks, with methods and tasks contributing to wider goals, or processes. The sections in this chapter are unequal in length and depth, with *Discovering* forming the most substantial part, due to the relative frequency with which it was mentioned by participants. Despite their popularity in the initial survey questions, I received fewer detailed comments about the other four methods, providing more limited scope for discussion. Having deconstructed responses to focus on individual *TaDiRAH* methods in these first five sections, I will then reconstruct the research process by providing two case studies (5.6) that illustrate how different methods are combined in practice.

### 5.1 Discovering

As demonstrated in the previous chapter (4.2), *Discovering* was the most popular research method among survey participants in relation to digital tool or resource use, and the second most popular in relation to Linked Data use, although it fell to joint ninth place for Linked Data production. In *TaDiRAH* (2014c), *Discovering* is defined as



"the activity of seeking out objects of research, research results, or other information which is useful in a given search perspective", incorporating basic searches, advanced queries and browsing. Comments relating to *Discovering* appeared regularly throughout my study; for example, from survey responses, I identified 'Search functionality' as the third most popular feature of good digital tools and resources (4.1.1).

I will start this section by introducing the various tools, tasks and features that participants associated with searching (5.1.1), before turning to the alternative discovery mode of exploration (5.1.2). I will end by discussing comments relating to the discoverability of relevant digital tools and resources (5.1.3) and their documentation (5.1.4), before summarising my findings in relation to this research method (5.1.5).

#### 5.1.1 Search

Most participant comments relating to *Discovering* focused on targeted searches of varying complexity, which usually involved inputting one or more terms into a search form, often with the option to add extra conditions. The advantages of using digital tools and resources for this purpose, as mentioned by participants, included convenience, speed, accuracy, and better availability than print materials at some institutions. Although *Discovering* appears as part of TaDiRAH's *Capture* category, implying it occurs in the initial stages of a project, participants' responses indicated that they search for new material throughout the research process; for example, PART089 spoke about performing bibliographic searches throughout the writing process. Participants often identified multiple routes to discovering digital materials about the Ancient World and sometimes tried different ways of obtaining a particular source if their original plans were derailed by access or technological barriers.

Some participants start the *Discovering* process by searching generic tools or resources to familiarise themselves with a topic, before moving on to discipline-specific sources, with PART078 remarking that image search can be at least as helpful as text search during this initial stage. Resources mentioned by participants included *Academia.edu*, *JSTOR*, *ResearchGate* and *Wikipedia*, although the majority achieved this goal using a

*Google* or *Google Scholar* search, with participants such as PART078 finding it more efficient to locate digital objects via *Google* than searching the resource in which they are held. PART089 praised the serendipity of the generic search engine, in potentially producing results that a more specific resource would filter out, including those that would not necessarily be available through academic tools and resources. When moving to tools and resources with a specific disciplinary remit, several participants indicated that they have their own individual 'hierarchies of usefulness', mentioning that they often start their search with the same tool or resource, while others were seen as a last resort. For example, PART078 always starts their searches using *Perseus*, due to its stability and familiarity, while PART089 described *L'Année Philologique* as a "wipe-up" once they had exhausted all other avenues. These hierarchies varied among participants with no obvious correlation between them.

In terms of *Discovering* tasks within specific tools and resources, most comments referred to full-text searching. For example, many survey participants (such as PART013 and PART038) found *TLG*, the most popular tool or resource in 4.1.1, to be important for accessing, searching, and analysing Ancient Greek texts, with PART005 using it for text mining. Other search tasks included querying metadata about digitised objects (with PART018 giving the example of *Trismegistos*, 2.2.4) and identifying bibliographic material for further reading (with PART035 giving the example of *L'Année Philologique*, 4.1.1). Several participants studied texts written in non-Latin scripts, and therefore found non-Latin (e.g., Greek) character input to be helpful, if not essential, when searching for relevant materials. PART043 stated that this functionality has already been achieved effectively in *Perseus*, through provision of an on-screen Greek keyboard, rather than requiring the user to adjust the keyboard settings on their device. Beyond the Ancient World domain, inclusion of non-Latin character input could have wider implications for Humanities research as a whole, potentially resulting in a more global approach to digital tool and resource development.

The search facilities used by participants had varying levels of complexity, suited to different types of query, as well as different levels of technical expertise; however, participants often found basic keyword search to be sufficient for their needs. PART061, who had themselves been involved in the production of digital tools and

resources, felt that a basic search box, allowing input of multiple terms in combination, is essential to meet user expectations. That said, most participants seemed to be confident about using advanced search options to narrow down their results. Even those with lower digital confidence were familiar with Boolean operators, such as 'AND', 'OR' and 'NOT', with PART078 criticising tools and resources that do not incorporate these terms. Filters and facets were mentioned by several participants, with PART018's survey response commenting on their effectiveness in *Papyri.info's Papyrological Navigator*, which allows users to filter their search results by date, collection or script. PART005 found the date facets in the *Beazley Archive*<sup>111</sup> particularly useful in narrowing their searches. PART078 and PART254 mentioned combining search terms to narrow down the results as they find a point of focus for the topic in question.

PART038 mentioned in the survey that providing both basic and advanced search options ensures that the tool or resource caters for different types of user and query. Similarly, PART041 recognised the importance of identifying different discovery goals and incorporated different types of search functionality in their own numismatic resource to ensure these goals were met. As well as a basic search, they included options that would allow users to identify an unfamiliar coin, explore images categorised using key terms (I will discuss exploration further in 5.1.2), and perform advanced queries on the underlying data. Segmenting potential audiences based on their end goals demonstrates a clear appreciation of different user journeys and exemplifies the user-centred design approach discussed in 7.3, below.

Some participants hoped for greater intelligence in the search systems they used, requesting more guidance that might lead to serendipitous discovery. For example, PART017 and PART255 felt the need to perform multiple searches using similar keywords, to be sure of finding all relevant materials, due to their knowledge of the resource (or prior experience) indicating that only exact matches will be returned, rather than synonyms or alternative spellings. Similarly, PART012 hoped for more efficiency in searching, stating that they would like to be able to easily access the most

---

<sup>111</sup> <https://www.beazley.ox.ac.uk/index.htm>

relevant results for them, while PART089 stated that they do not always know exactly what they are looking for, implying that they hoped the tool or resource in question would provide some direction.

Simple databases that produce literal results for user searches are therefore often insufficient for wider exploration of a research topic. As a potential solution to this issue, PART078 suggested that it would be helpful for digital tools and resources to suggest search terms that other users have input to yield similar results. The example they provided was online auction site *eBay*<sup>112</sup>, whose suggestions highlight terms they would not have previously considered, providing potentially more comprehensive results, as well as revealing connections between topics and terms. PART078 also stated that, if applied to academic tools and resources, such suggestions might help researchers to persist with their searches, providing new inspiration to “get you out of a rut”.

In a Linked Data system, such ‘intelligence’ might be achieved through descriptions and connections in metadata, particularly regarding disambiguation<sup>113</sup>. PART078 noted the negative impact on usability in cases where such an approach was not implemented. They gave the example that, if multiple localised instances of deities are not represented as discrete, albeit related, entities with their own identifiers, this can negatively impact discoverability. PART041 gave a similar example of searching for the Greek deity Aphrodite; sometimes they wished for Aphrodite’s Greek equivalent Venus to be included in results, while on other occasions they preferred to narrow their search to Aphrodite in particular, omitting any results for Venus. The ability to exert greater control over the results they receive could therefore considerably improve the efficiency of their discovery process. To take this example further, we might also consider a resource with a wider scope allowing users to specify that they are searching for the deity Venus rather than the planet. Indeed, PART001 found the lack of disambiguation between entities with similar names to be particularly problematic

---

<sup>112</sup> <https://www.ebay.co.uk/>

<sup>113</sup> I previously (2.3) discussed the advantages of disambiguation and alignment that Linked Data can provide to digital tools and resources and will return to these topics during my discussion of identifiers in 6.2.1, below.

when using resources with a broad disciplinary remit; for example, they often receive results relating to footballers when searching for Roman emperors.

The lack of ‘intelligent’ searching in many digital tools and resources caused several participants to express concern that their discovery processes are not sufficiently comprehensive. Once they had found the digital objects they required, they often identified the need for further searches to acquire more information: if there are gaps in the data provided by one resource, these might be filled by searching another. For example, PART005 often feels that they are “missing something”. To mitigate this issue, they start by performing a keyword search in *TLG*, then use *Perseus* to find contextual information about the results. Similarly, PART041 chose to use multiple discovery tools rather than a single source, providing increased breadth of material and greater confidence in the comprehensiveness of coverage. However, conducting the search process in this way can take significantly more time and often entails duplication of effort. To mitigate this issue, participants felt that consulting multiple resources from a central point would be beneficial<sup>114</sup>. Such functionality could serve to make the relationships between digital objects more apparent and introduce an element of serendipity. Crucially, a single search could reduce the time taken for the *Discovering* process, providing the researcher with more time to analyse and interpret their results.

Several participants spoke about how such a central point might work most effectively for their research purposes. PART041 mentioned that they would like the databases they use to be able to “talk to each other” but they are used to this not being the case and have developed their research processes to compensate for these limitations, e.g., by keeping multiple browser tabs open and moving between them. PART001 said that they would like to be able to search for a particular place in the Ancient World, then find “a linked network of artefacts” associated with that place at a particular time. They additionally stated that *Arachne* (2.2.4) should work well for this task, but that it did not provide the level of detail they required for places that are less well known.

---

<sup>114</sup> As well as including this functionality as an advantage of Linked Data technologies in 2.3, it was also mentioned in my survey by Linked Data users (4.1.2), with non-Linked Data user interview participants ranking it as the most important feature of Linked Data during the follow-up survey.

PART001 had also used the British Museum catalogue for this task; however, they commented that as it was restricted to internal links within the museum's own holdings, its search results could not be relied upon as a "representative sample of stuff from that particular place". Therefore, while such tools and resources are already in existence, their coverage was often not sufficient for (or compatible with) participants' requirements.

In discussing participants' experiences of searching, I found that many have developed processes that usually involve moving from generic to specific tools and resources and basic to advanced search. Within these processes, participants acknowledged that different types of search and levels of complexity are suited to different types of user and task. Participants wanted search tools and resources to be more 'intelligent' and provide them with more guidance about the next steps in their *Discovering* process, generate ideas for future searches, and incorporate synonyms and elements of disambiguation. All these aspects promote serendipity while maintaining a level of relevance to the original query. The above points additionally relate to participant concerns about a lack of comprehensiveness in the results produced by their queries. Such concerns might be mitigated by a central point through which to search multiple tools and resources, which aligns well with the advantages of Linked Data identified in 2.3 and 4.1.2.

Reliance on keyword search might stem from the ubiquity of generic search engines, such as *Google*, which influence how more specific academic tools and resources, such as those relating to the Ancient World, present their discovery functionality. However, my findings also suggest that keyword search might not always be the best means of *Discovering* new knowledge. An alternative approach is that of facilitating exploration by the user, the subject of the next section.

#### 5.1.2 Exploration

Although search seemed to be the most popular mode of *Discovering* conducted by participants, comments indicated that a more exploratory approach might integrate better with their ways of thinking. Such exploration tended to take place through following links between and within digital objects, either as a means of travelling from

one object to another, or to provide the user with contextual information without causing them to interrupt their research process.

Many participants wanted to see more connectivity between digital tools and resources, which was ranked as the second most important advantage of Linked Data by non-Linked Data user interview participants as part of my follow-up survey (4.1.2). Some of those participants commented that availability of more links within and between digital tools and resources could improve the accuracy and efficiency of discovery. For example, there was interest in connections between objects of different types, with PART012 giving the example of *LacusCurtius*<sup>115</sup>, which provides links between texts, maps and artefacts relating to the Roman world. The same participant also spoke about the effectiveness of *Pelagios*' links between texts, images, and maps, while suggesting that relatively few such resources are available for Ancient World research due to difficulties in production and publication. PART109 also found *Pelagios* to be "inspirational" in this regard, implying that they consider the provision of such connections between resources to be innovative, new, and rare. PART005 agreed that such features would be useful to them but had also encountered them relatively infrequently.

There were many specific comments about *Discovering* by means of following links in a text, either to other texts, to secondary scholarship, or to authority files. For example, providing links between digital resources can assist users in familiarising themselves with a topic about which they have insufficient knowledge to formulate effective combinations of search terms. Participant responses indicated that discovering new knowledge by making connections and following links can work more naturally with researchers' ways of thinking than constructing queries to enter into a search box. For example, PART078 commented that "*I love clicking through and just following that trail to see if there's anything I've missed that should be raised about an object*".

The exploratory process of following links might better emulate the experience of browsing a physical library than that of keyword searching. On the topic of library user

---

<sup>115</sup> <http://penelope.uchicago.edu/Thayer/E/Roman/home.html>

experience, PART078 spoke about how browsing printed materials in the physical library space facilitated their discovery of information, finding that this provided a quicker indication as to whether a particular item is relevant for their purposes. They also felt that discovery in a library environment, where publications have been classified and curated by librarians, leaves them open to the possibility of discovering material on ostensibly different topics, rather than restricting their search at an early stage. PART012 and PART089 mentioned that they prefer the user experience of the library; however, the financial and time costs of visiting distant locations to consult materials, as well as handling restrictions placed on rare and fragile objects, mean that digital resources often provide the only means of regular, guaranteed access. Since my study took place, this situation has of course been exacerbated by restrictions resulting from the COVID-19 pandemic, thereby increasing the importance of usable digital tools and resources for *Discovering* via exploration.

Connections can also facilitate discovery by allowing the user to trace the journey of physical objects, and their digital counterparts, through space and time. PART078 suggested that it would be helpful to be able to discover links between objects previously held by the same owner, or that were acceded into a collection in the same year, as such relationships can be used to demonstrate themes and ideas that have interested people over time. Making such connections more explicit could therefore enhance discovery of information, even for less confident researchers.

Links from relevant parts of texts or images can also be extremely helpful in providing contextual information about the subject in question and were identified as an advantage of Linked Data from survey responses (4.1.2). Among the benefits of these connections is the potential for pooling knowledge from different areas of expertise, as suggested by PART078. Some participants gave examples of tools and resources where these contextual links already exist, including *Trismegistos* (2.2.4), mentioned by PART089, which provides different grammatical forms of the same person's name, and the interactive map of Rome *Forma Urbis*<sup>116</sup>, mentioned by PART012, which provides helpful links between its catalogue, map, and bibliographic resources. Survey

---

<sup>116</sup> <https://formaurbis.stanford.edu/>



participant PART103 commented that *Papyri.info* was particularly helpful in providing links to related texts held in different collections, while PART254 gave *The Latin Library*<sup>117</sup> as a good example of where contextual links to information about places and people have been implemented well.

As a result of their experiences, PART043 and PART254 said that they would like for all digital texts to automatically provide contextual information when the user clicks on named entities within them, with PART254 saying that "*Connection of data is important, because sometimes you don't know what you're looking for, and that way you can find things more... easily*". Such functionality might take the form of a popup containing a person's birth and/or death dates, as well as related people, places, or images. This information might be obtained by linking to corresponding entries in external authority files, thereby reducing the number of searches the user would need to perform. Additionally, providing contextual information to a user while they are studying a text or image avoids interrupting their research process by causing them to perform additional searches.

The exploratory approach of following links is therefore not only extremely helpful for Ancient World research but has the potential for effective integration of Linked Data, to provide multiple directions for *Discovering* from a single digital object. Although *Discovering* is my focus for the current section, this provision of contextual information via links from a text or image additionally relates to other research methods. For example, these linked entities might first have been discovered by the tool or resource producer via some form of *Data Recognition* (5.3), then connected to the relevant URIs via *Annotating* (5.4). The relationship between these research methods highlights that they do not exist in isolation but are interlinked.

Having discussed two different *Discovering* pathways as applied to digital tools and resources, I will now turn to the topic of how (and whether) Ancient World researchers discover the tools and resources themselves.

---

<sup>117</sup> <http://www.thelatinlibrary.com/>

### 5.1.3 Discovering Digital Tools and Resources

While this section has focused thus far on *Discovering* digital objects within tools and resources (or via search engines), there were also many participant comments surrounding discovery of the tools and resources themselves, particularly where Linked Data is concerned. In my original survey, two participants mentioned that a significant usability barrier was being unaware of what is possible (4.1.1), due to a lack of knowledge about the types of tools, resources and technologies that are available. In the follow-up survey (completed by interview participants), this barrier became more prominent, achieving the highest mean importance score among non-Linked Data users. These findings indicate that discoverability issues should be considered a major obstacle to the uptake of digital tools and resources.

In terms of *Discovering* processes, some participants found that their usual starting point of a *Google* search (5.1.1) was often insufficient for finding tools and resources, rather than individual objects, because the names of these tools and resources are often not reflective of the content or functionality provided. PART078 suggested that this situation might be due to evolution of the tool or resource into something other than was originally envisaged, while pointing out that it has the effect of "excluding" the intended audience. Searching based on intended method or task might therefore be more fruitful. Most participants indicated that they discovered relevant tools and resources by serendipity, rather than conscious searching. Examples mentioned by PART005, PART008 and PART012 included conferences, mailing lists such as *Digital Classicist*, and social media (particularly *Twitter*). Other participants, such as PART254, found out about new tools and resources through conversations with colleagues, with PART005 finding that their MA tutor was particularly keen to normalise the use of digital tools and resources as part of Ancient World research. Other participants discovered new tools and resources via the more targeted means of attending training courses or workshops, with PART078 suggesting that a regular seminar to highlight new tools or resources would be a helpful addition to the current offering.

With specific regard to Linked Data, participants such as PART054 spoke about the need for interested researchers to be convinced about the benefits of this approach, then be presented with clear pathways for consuming and/or producing Linked Data.

PART061 agreed, advising that describing these potential benefits can be difficult without being able to demonstrate how Linked Data has been applied in practice, due to the steep learning curve for implementing this approach. PART054 additionally commented that, with the number of authority files and ontologies available, it can be complex and time-consuming to identify how Linked Data might best fit with a specific research project, particularly if the advantages of doing so are unclear. For example, it can take considerable time to find suitable tools, technologies, or ontologies, as there is such a vast array of resources across different disciplines, and it is impossible to gain familiarity with everything. PART061 suggested that the lack of time to address these issues, particularly on fixed-term projects, means that researchers are often deterred from producing or consuming Linked Data technologies, as it is unclear where to start, as well as what the limits of the work should be.

On a related note, several participants, such as PART063 in the survey and PART254 in interview, mentioned more generally the lack of a central catalogue for discovering not only which digital tools and resources might be relevant for their research, but which ones are likely to function reliably and contain trustworthy information. PART078 suggested that a "database of databases" would be helpful for researchers, to ensure that they have not missed any potentially useful resources, while PART012 suggested a curated list of the "top 25" digital Ancient World resources. PART008 emphasised the significant time efficiencies that such a resource would provide and suggested that it might be beneficial to focus on listing tools and resources that relate to specific themes or topics; the examples they gave were historical maps, gazetteers, and archives. However, they additionally acknowledged that the usefulness of any directory resource would diminish if it were not maintained. Although the *Digital Classicist* wiki (2.1.2) is a long-established resource that lists digital tools and projects relating to the Ancient World, it did not seem to be widely used among participants, with PART089 finding it difficult to navigate and PART254 being unaware of its existence.

Developing a platform to aid discovery or enhancing existing such resources (e.g., by incorporating some form of methods-based exploration), might go some way to filling

the perceived gaps in current provision, particularly where Linked Data is concerned. As such, I will return to this topic as part of my recommendations in 8.9.1.

#### 5.1.4 Documentation for Discovery

A topic that arose multiple times in relation to *Discovering*, and that was identified as one of the most popular features of digital tools and resources in 4.1.1, is that of documentation, i.e., how the information provided about tool or resource usage impacted participants' *Discovering* experiences. For example, PART078 advised that, for effective discovery, it is important to know both what is and is not included in the tool or resource. Similarly, instructions on how to perform searches were seen as crucial, although (as suggested above), participants found it preferable for digital tools and resources to allow standard Boolean syntax. Such documentation is particularly important when the user needs to perform more complex queries. For example, PART254 expected tutorials to demonstrate how to search effectively if specific terminology or dialects are required, or if input involves a non-Latin character set.

One way of producing documentation to assist with *Discovering* is by providing example queries that can be easily customised by users. PART109 found such examples to be particularly helpful when using *Nomisma's* (2.2.4) SPARQL endpoint, while also mentioning that their inclusion is relatively rare across Linked Ancient World Data tools and resources. Example queries ensure use of the correct ontologies and terms, while providing efficiency with a valid query for users to adapt. As mentioned in 2.4, this information can be particularly helpful for Linked Data tools and resources whose only search facility is a SPARQL endpoint, as different initiatives use existing ontologies in different ways, as well as incorporating their own vocabularies. SPARQL itself is an interesting means of discovery in that it bridges the gap between search and exploration, allowing queries to both retrieve information and explore datasets. However, the lack of comments relating to SPARQL from the majority of Linked Data users in my study indicates that few researchers take advantage of these capabilities. As suggested in 2.4, the primary reason for minimal SPARQL uptake is likely to be the lack of intuitive interfaces through which endpoints can be accessed. However, even more technically skilled researchers are likely to experience barriers imposed by

inconsistent application of ontologies and lack of documentation about how queries might be structured.

Key to the above points is the idea of transparency, i.e., effective, open communication about what the discovery functionality enables the user to do and how this is achieved by the system in question. If the tool or resource is unclear about how its search algorithms work or which query syntax/language to use, it might not produce any results at all. Alternatively, as PART001 remarked, if the user does not understand exactly what they are asking of the tool or resource, their results may be unreliable or incomplete. I will return to the topic of documentation, to facilitate usability more generally, in 7.4.

#### 5.1.5 Summary: *Discovering*

Having discussed participant comments relating to *Discovering*, I can infer that this tends to be the primary method that Ancient World researchers conduct using digital tools and resources, which occurs throughout the research process. Once researchers have access to the materials they require, subsequent methods might then be performed without any digital assistance. Such findings potentially indicate that *Discovering* has pervaded the researcher consciousness in a way that does not seem to be mirrored in other research methods. As search and discovery capabilities seem to be a core motivator for Ancient World researchers to engage with digital tools and resources in the first place, it is critical that the requisite functions are implemented effectively: an area in which Linked Data is likely to be beneficial.

To address RQ3, Linked Data might be integrated with *Discovering* by providing richer descriptions of digital objects that connect to external resources, providing disambiguation and alignment capabilities. Linked Data technologies might also be used to aggregate digital objects from multiple tools and resources, to ensure that searches are more efficient, streamlined, and comprehensive. In addition to facilitating searches, Linked Data technologies can also improve the user's experience in a more exploratory mode of *Discovering*: following connections between digital objects, revealing the relationships between them, and presenting contextual information to the user. However, a lack of awareness of digital tools and resources for Ancient World

research seems to be a problem, and one that particularly affects Linked Data technologies. Participant responses indicated demand for a centralised directory with information about the benefits of specific technologies, as well as links to relevant tools and resources; a topic to which I will return in my concluding chapter.

Having discussed *Discovering*, I will now explore the four other methods identified from my survey as being potentially amenable to Linked Data integration, starting with *Gathering*.

## 5.2 Gathering

*Gathering* is defined by TaDiRAH (2014d) as "*aggregating discovered resources, usually in some structured way*", giving the example of bringing together related papers for a literature review. In response to the methods questions in my survey, *Gathering* appeared in eleventh place in relation to the use of digital tools and resources in general but grew in popularity in response to the questions on the use and production of Linked Data, appearing in first and second places, respectively (4.2). Survey participants mentioned resources including *TLG* and *Perseus* as being particularly helpful for *Gathering*, with PART043 highlighting the speed with which *Perseus* allows users to "compare/collate information".

During the interviews, I found that several participants gathered information about their research topics into spreadsheets, structured to meet their individual needs, one of which forms the basis of Case Study 1 (5.6.1). PART017 is interested in Classical Receptions and manually populated their spreadsheet with details of films and television programmes that were relevant to their research, having identified that no such resource already existed. The spreadsheet includes both factual information and the participant's own critiques, providing a useful breakdown of all the data required for their project. PART017 stated that they found this approach invaluable for providing an overview of the topic, with contextual information, and they are continuously adding to the spreadsheet to ensure it is as comprehensive as possible. Another participant, PART109, initially used a spreadsheet to collate pertinent information about their research topic, which they then converted to Linked Data using Python scripts. Again, this was in response to a gap in current provision: even

though much of the data they gathered was already available as Linked Data, they had to combine it with new data they had created, to make it useful for their own research context.

Other participants also used online tools to perform digital *Gathering*. For example, PART041 created a personal website to collate and share relevant sources and ideas used in their research. They found that presentation of the information in this way allowed relevant resources to be linked together. In turn, these links facilitated search and exploration (i.e., *Discovering*, 5.1) by the participant, which integrated more effectively with their ways of working and thinking than navigating a collection of printed materials. Another participant, PART043, spoke about how *Twitter* hashtags are used to collate multiple people's experiences of the same location or event.

Further examples of *Gathering* involved more technologically complex approaches and software. For example, survey participant PART120 used databases and *RStudio*<sup>118</sup> to bring together different datasets for further analysis, while PART001 used *QGIS*<sup>119</sup> to combine and visualise geographical and artefactual information from publications and datasets. This latter participant found that these tasks were facilitated by clarity of information in the publications, as well as the provision of geographical coordinates in a consistent format. They additionally aimed to ensure that the data they had gathered was structured using open standards, to facilitate reuse by other researchers in future – a topic to which I will return in 6.2. At the other end of the spectrum, some participants mentioned non-digital *Gathering* processes, referring to bringing together their notes when reading source material. For example, survey participant PART106 prefers "printing the notes and combining them by hand", while PART249 uses "hand-written notebooks". PART089 mentioned that they found such non-digital *Gathering* to be laborious, and that using a digital tool such as *Scrivener*<sup>120</sup> provides considerably more potential for moving these notes around and reordering them into a different structure.

---

<sup>118</sup> <https://rstudio.com/>

<sup>119</sup> <https://www.qgis.org/en/site/>

<sup>120</sup> <https://www.literatureandlatte.com/scrivener/overview>

Several participants gave the same example as TaDiRAH, by using digital tools to gather materials for inclusion in a bibliography. In doing so, they created their own curated subsets of research materials, which could be searched from a single point. Having a central repository of information was particularly important for participants such as PART043 whose research spanned different subject areas or chronological periods. Generally, participants found reference management software to be helpful for collating and searching sources. However, PART017 commented that it can be difficult to find specific items once the system contains a large number of sources. These tools are therefore only effective at scale if accurate and well-structured metadata is applied throughout. Additionally, it can cause irreparable damage to the research process, and researcher trust, if the software fails and the data is lost. As a result of such an incident, PART078 adapted their *Gathering* processes to instead collate their source materials in themed *Word* documents, which integrates well with their existing ways of working.

Based on the above findings, Linked Data might facilitate *Gathering* by collating relevant resources to form an interconnected body of materials on the same topic, which is searchable/explorable from a central point, an example of which might be the directory suggested in 5.1.3. Additionally, as PART017 found, the similarities between research objects might be missed if each is the subject of a separate publication, with no means of bringing them together. On a larger scale, as I found in my survey (4.1.3), multiple producer participants selected data formats and structures due to their potential for integration with existing tools, resources, and systems, with half preferring to use Linked Data for this purpose. Linked Data could therefore be an ideal approach for forging these connections. PART061 advised that such interoperability is best achieved by implementing Linked Data technologies from the outset of a project, rather than attempting to integrate them with other data structures and formats at a later stage.

However, a greater degree of automation to facilitate *Gathering* might ultimately prove detrimental. Indeed, PART005 saw *Gathering* as fundamental to the acquisition, understanding and creation of knowledge, stating that they "*might gain an insight that I wouldn't if it had just been automatically done on some kind of algorithm*". However,



they did suggest that a balance might be achieved; the example they gave was the automatic alignment of references from ancient texts (e.g., those provided by *TLG*) with depictions of relevant scenes or entities on pottery (e.g., those described in the *Beazley Archive*). The researcher could then review the resulting combined dataset manually, with less time required for the initial *Gathering* stage. While such integration would not currently be possible between these two resources without developing a bespoke system, Linked Data implementation could facilitate this type of alignment across similar tools and resources in future, as discussed in 2.3.

In exploring experiences of digital *Gathering*, I found that, in addition to the use of reference management software, participants often collated data manually into their own research resources, such as spreadsheets or databases (which might subsequently be converted to Linked Data), or websites. Production of such resources might be facilitated by greater automation, which might comprise the use of Linked Data technologies to align datasets that refer to the same named entities. However, in the development of any such system, producers should be aware that *Gathering* is an important part of the research process. While there are efficiencies that could be introduced, too much automation might prevent researchers from making key connections and interpretations that can only be achieved through close familiarity with the data.

Such dataset alignment might include elements of *Data Recognition*, the subject of my next section.

### 5.3 Data Recognition

*Data Recognition* is defined by TaDiRAH (2014b) as "*the process of treating the immediate products of digital data capture... in a way to extract discrete, machine-readable units from them, such as plain text words, musical notes, or still or moving image elements*", giving the example of Optical Character Recognition (OCR). In my survey, I labelled Data Recognition as "Extracting pieces of information from a text (e.g. names, places)" (Appendix 1, Q6) to make its meaning more explicit to less

technically experienced participants<sup>121</sup>. I found it to be the sixth most popular method relating to the use of digital tools and resources, increasing to the third most popular for the use of Linked Data. Like *Discovering* (5.1), however, it appeared to be less popular among Linked Data producers, appearing in joint eleventh position (4.2).

None of the responses to open-ended survey questions referred specifically to *Data Recognition*. I therefore aimed to elicit more detail from participants by asking about this method during interviews, with responses indicating varied understandings of the phrase included in my survey. For example, PART017 spoke about using the ‘Find’ function in word processors, PDF readers or web browsers to identify relevant keywords within a text, with PART078 using this functionality to track changes in translation over time, or compare translations intended for different audiences. They found this much faster than reading the text in its entirety, giving them more time for analysis and interpretation. PART078 preferred performing this task within a single tool or resource, expressing a concern that they might introduce errors if they downloaded a text and worked on it outside of the resource in which it was originally contained.

Other participants discussed their use of digital tools and resources where some form of *Data Recognition* had been performed during production, i.e., those that facilitate the exploratory mode of *Discovering* by providing contextual information about specific entities (5.1.2). One such example, identified by PART089, was *Trismegistos*, which provides multiple linguistic forms for personal names, alongside information about their geographical context. There were far fewer instances of participants who had performed *Data Recognition* themselves, rather than using such pre-processed materials. Indeed, PART255 had manually structured the metadata of their own collection of digitised letters to produce a resource that provided contextual information about e.g., senders and addressees. However, they implied that automating *Data Recognition*, e.g., by using Named Entity Recognition (NER), might be a more efficient way of achieving this in future, as their resource increases in scale. Only PART005 spoke at length about performing *Data Recognition* themselves, using

---

<sup>121</sup> However, in hindsight, this narrower definition might have inadvertently restricted responses, e.g., from those researchers who work predominantly with images.

*Pelagios' Recogito* tool (2.2.1), as part of a project studying the language of an ancient text. Their work also incorporates three other research methods discussed in this chapter; as such, their experience forms the basis of my first case study on combining research methods (5.6.1).

Although I found relatively few examples of participants performing *Data Recognition* (when compared to the other research methods discussed in this chapter), this is likely due to the difficulty in separating it from other methods in the research process. As noted above, researchers might well perform *Data Recognition* as a stage in *Discovering* (5.1) or *Gathering* (5.2), without considering it to be a discrete method. Additionally, *Data Recognition* often takes place as a precursor to *Annotating*, the next method in my discussion.

## 5.4 Annotating

*Annotating* is defined by TaDiRAH (2014a) as "*the activity of making information about a digital object explicit by adding, e.g., comments, metadata or keywords to a digitized representation or to an annotation file associated with it*", incorporating linguistic commentary and links to external identifiers. In response to my survey questions on research methods, *Annotating* consistently appeared in the top 50%, ranking joint sixth (with *Data Recognition*, 5.2) for digital tool and resource use, ninth for Linked Data use, and joint fourth for Linked Data production (4.2). This latter position potentially indicates that producers are relatively interested in developing Linked Data tools with annotation functionality. Throughout my survey and interviews, *Annotating* was mentioned in relation to the use of digital tools and resources in general and the use and production of Linked Data tools and resources in particular. Participants predominantly referred to text as the object of their annotations; however, their responses could, in many cases, also apply to images or 3D models.

Participants' annotations tended to fall into two broad types:

1. Personal notes by the reader, e.g., thoughts, opinions or reminders;
2. Contextual information provided through internal or external links (semantic annotation).

Type 1 annotations are most useful to the annotator themselves, usually with very little incentive or desire to make them available to others. Due to the sometimes-personal nature of these annotations, and the long tradition of annotating written texts for this purpose, many participants tended to conduct this method using physical materials rather than digital tools. Accordingly, *Annotating* was often mentioned in relation to reading, an activity which many participants preferred not to perform digitally. I therefore inferred that *Annotating* might be seen as an extension of the reading process, i.e., if a participant is reading a printed text (due to either availability or personal preference), they would create any annotations on that same paper, meaning that their selection of annotation medium was likely based on their reading medium, rather than a conscious choice. PART081 in the survey and PART089 in interview confirmed that, for them, this is indeed the case, and that they are happy to adapt their ways of working to move between digital and non-digital *Annotating* based on whether they are reading a digital or physical text. Survey participants PART017 and PART027, however, found that using pen and paper assisted them to capture their thoughts more accurately. Both expressed general dissatisfaction with existing digital annotation tools largely due to their relative lack of efficiency.

For many researchers, pen-and-paper annotations might be sufficient for their needs, providing a means for them to record their thoughts at the point of reading a text and allowing them to return later to relevant passages. However, if a researcher makes multiple annotations on multiple documents over a long period of time, this results in a large amount of data that could become unmanageable, making specific annotations difficult to find and the entire body of work difficult to synthesise. Survey participant PART059 solved this issue by later adding their pen-and-paper annotations to a spreadsheet, clearly recognising the value provided by storing their annotations as structured data. Having annotations available in this format can be very useful, even if only to the individual researcher, and utilises familiar software as a *Gathering* (5.2) and *Discovering* (5.1) mechanism, for organising their own thoughts on a research topic and identifying key points.

In terms of digital *Annotating*, several participants spoke about using annotation functionality in software primarily intended for reading or writing, with specific examples including PART043 pasting images into *OneNote*<sup>122</sup> and making notes alongside them, and PART254 using *Preview*<sup>123</sup> to highlight and comment on PDFs. Like the spreadsheet comment above, this is another example of participants using familiar software as a solution to the issue of creating and storing annotation data (albeit without the same level of structure). However, even those participants who used such tools often expressed a preference for non-digital annotation. For example, PART254 found annotating PDFs in *Preview* to be much slower than using pen and paper, due to the time it took to change the cursor between commenting and highlighting modes.

In contrast with Type 1 annotations, semantic Type 2 annotations could potentially be helpful to others if made available as structured data (e.g., using Linked Data technologies) and could facilitate future research by building connections between and within sources. There is therefore a strong argument for conducting this type of *Annotating* digitally, then sharing the results with other researchers. Most participants who performed Type 2 *Annotating* used purpose-built tools. Their responses indicate the availability of a huge range of such tools, with no single one used by a particularly large number of participants.

Some annotation systems were not aimed at a specific subject domain, including *BRAT Rapid Annotation Tool*<sup>124</sup>, which is discipline-agnostic, as well as *TextGrid*<sup>125</sup> and *Pelagios' Recogito* platform, both of which have a broad Humanities remit. These tools tend to allow export in open formats that can easily be imported into other systems for further analysis. However, of the three, only *Recogito* provides RDF as an export format, alongside other open standards such as CSV, GeoJSON and TEI/XML (Simon et al., 2017). Other annotation tools were developed with Ancient World research in mind, including *Perseids' Arethusa* platform<sup>126</sup> and *Sematia*<sup>127</sup>. *Arethusa* uses texts encoded with TEI/XML, with *EpiDoc* for papyri and manuscripts (Almas et al., 2015;

---

<sup>122</sup> <https://www.microsoft.com/en-gb/microsoft-365/onenote/>

<sup>123</sup> <https://support.apple.com/en-gb/guide/preview/>

<sup>124</sup> <https://brat.nlplab.org/>

<sup>125</sup> <https://textgrid.de/en/>

<sup>126</sup> <https://www.perseids.org/tools/arethusa/app/#/>

<sup>127</sup> <https://sematia.hum.helsinki.fi/user/>

Almas & Beaulieu, 2016, p. 180), while *Sematia* imports and pre-processes texts from *Papyri.info*, before exporting them to *Arethusa* for annotation via an API (Vierros, 2018, p. 108). Unfortunately, other than *Recogito* (used in Case Study 1, 5.6.1, below), most of the above tools were only mentioned briefly in survey responses and were not used by interview participants; therefore, I do not have details of these users' experiences.

As I will demonstrate in 5.6.1, *Recogito* incorporates *Data Recognition* in addition to *Annotating*; it also includes *Visualization* capabilities. Some of the other annotation tools mentioned above similarly combine multiple research methods. *Arethusa* incorporates *Data Recognition* for morphological analysis, by identifying potential matches for commonly used words and providing information about different grammatical forms. The relationships between annotated terms can then be visualised (5.5) as a tree structure (Perseids, n.d.). It therefore additionally incorporates TaDiRAH methods not included in this chapter such as *Structural* and *Stylistic Analysis*. Similarly, *BRAT* automatically identifies named entities, such as people, organisations and places, as well as parts of speech and relationships between terms (Stenetorp et al., n.d.). Based on these three examples, it seems that *Annotating* often serves as an intersection between multiple research methods.

Several participants had been involved in the production of digital tools and resources that involved *Annotating*. PART255 worked on a project to annotate a series of letters, which included manually identifying entities such as people, places, and topics (mentioned briefly in 5.3). As part of this project, the participant worked with external developers to produce a tool that would allow researchers with minimal coding knowledge to perform this annotation, which could then be converted to XML. The tool's remit was initially broad; however, over the course of the project, various customisations were applied that would now make it difficult to use for any other document types. PART255 did not seem particularly concerned by this development, as the tool now provides the ideal system to use for this project. However, without the flexibility to be used in other contexts, it is difficult to envisage how sustainable such a tool would be in the longer term.

In my discussion of *Annotating*, I have discovered its intersections with other research methods, and its potential for integration with Linked Data technologies. Using this approach, connections between semantic (Type 2) annotations could potentially build up a vast data source containing different researchers' interpretations of the same topic (and the relationships between them), which could then be visualised or used as the basis for another analysis method. However, Type 1 and 2 annotations do not necessarily occur as discrete tasks and can be difficult to separate. Some researchers might not consider a Linked Data solution to be relevant to the types of annotation they tend to create, or they might find the use of digital technologies for *Annotating* to be too disruptive to their existing workflows. There might additionally be privacy concerns unless there is the functionality for selected (Type 1) annotations to remain private. I will therefore take such considerations into account when making recommendations for further research and development in this area (8.9.2).

I will now turn to discussing *Visualization*, one of the methods with which *Annotating* intersects, having previously mentioned it in relation to *Discovering* (5.1).

## 5.5 Visualization

The final method, *Visualization*, is defined by TaDiRAH (2014f) as "*activities which serve to summarise and present in a graphical form, and to use such graphical forms analytically, that is to detect patterns, structures, or points of interest in the underlying data*", incorporating images, maps, timelines, graphs and tables. The term therefore encompasses a wide variety of visual forms that can be used both to facilitate research and present outputs. In my survey, *Visualization* increased in popularity from tenth position for using digital tools and resources in general, to joint fifth position for using Linked Data, to third position for producing Linked Data (4.2). These results indicate that Linked Data can indeed facilitate *Visualization*, as well as potentially suggesting that it tends to be more popular among technically experienced researchers, likely due to the skills required to create and interpret more complex data visualisations. Additionally, as PART109 indicated, *Visualization* can be crucial to facilitating user understanding of a dataset and identifying new research directions, thereby potentially demonstrating the value of a Linked Data approach. PART005 advised that there are many ways of visualising data, each of which appeals to different people,

depending both on their individual learning styles and the type(s) of data with which they are working.

Although the popularity of *Visualization* seemed to increase with technical ability, many participants with lower confidence in their digital skills also conducted this research method. The most widely accessible modes of *Visualization* tended to be tables, graphs, and charts, with their creation largely facilitated by *Microsoft Excel*. For example, PART005 said that even producing a simple table can be a helpful way to visualise data, enabling the user to identify patterns and connections without requiring significant technical knowledge. Both PART017 and PART055 (in the survey) mentioned their use of *Voyant Tools*<sup>128</sup> to create visualisations that highlight patterns, themes, and keywords in text-based materials. In particular, PART017 described how their goals in using this tool were primarily to organise their research findings and identify key themes or directions for exploration, rather than inclusion in their final outputs. Both participants commented positively on their user experience of *Voyant Tools* in terms of its simplicity. *Voyant Tools* therefore provides a strong example of a simple interface that allows users to create Humanities data visualisations quickly and effectively.

Many participants spoke about their use of spatial visualisations, relating strongly to the TaDiRAH method *Spatial Analysis*. Several Ancient World-specific *Visualization* tools were mentioned by participants, including *ORBIS: The Stanford Geospatial Network Model of the Roman World*<sup>129</sup>, the *Digital Atlas of the Roman Empire (DARE)*<sup>130</sup> and *Pelagios' Periplo* tool (2.2.1). PART254 praised *ORBIS* for its ability to visualise distances and journey times depending on different modes of transport, which facilitated their teaching of related topics. They also commented positively on *DARE*, due to its comprehensive coverage of places mentioned in ancient literature, as well as the functionality to follow journeys made in specific campaigns. PART012 spoke about *Pelagios*, although it was clear from their comments that they were referring specifically to *Periplo*. They found that its interactivity provides a better user

---

<sup>128</sup> <https://voyant-tools.org/>

<sup>129</sup> <https://orbis.stanford.edu/>; uses location data from *Pleiades*, 2.2.1 (Meeks & Grossner, 2012)

<sup>130</sup> <https://imperium.ahlfeldt.se/>



experience than a static image, as the user can click parts of the map or perform searches to discover relevant information. Their comments highlight the use of *Visualization* to facilitate exploration, as discussed in 5.1.2. All three tools and resources incorporate Linked Data technologies; I might therefore infer from participants' positive experiences that Linked Data integration can facilitate the effectiveness of spatial visualisations.

While tools and resources that participants used for exploring spatial visualisations tended to focus on the Ancient World, those for creating such visualisations had often been developed with a non-discipline-specific remit, like some of the *Annotating* tools mentioned above (5.4). For example, PART043 used *Google Maps* to enhance their understanding of the geographical elements of their research, by visualising locations and journeys in relation to each other. PART109 used *Carto*<sup>131</sup>, praising the ease with which a map can be generated from an Excel spreadsheet, then customised based on user requirements. Others, such as survey participant PART095, used Geographic Information Systems (GIS) software to visualise their data, in addition to tasks such as data management and statistical analysis. Apart from *Recogito* (discussed in relation to visualising annotation data in 5.4 and 5.6.1), no Linked Data tools were mentioned in this context, potentially indicating an absence of well-known, usable examples for creating spatial visualisations.

Multiple participants described their use of *Visualization* tools for representation of objects, buildings, and places. One such participant, PART012, spoke at length on 3D modelling of ancient places, for which they use *SketchUp*<sup>132</sup>, as well as other software primarily intended for architects and game designers, assisted by digital texts and maps. *SketchUp* targets a broad audience, incorporating primary through higher education, as well as industries such as architecture and construction. Its website emphasises its ease of use, which was also highlighted by PART012. Their experiences exemplify my assertion in 4.1.1 that many tools used in archaeological research are specialist in nature, but non-discipline specific. PART012 did, however, reflect that their model might be improved by embedding contextual information to assist

---

<sup>131</sup> <https://carto.com/>

<sup>132</sup> <https://www.sketchup.com/>

exploration (5.1.2). Such data might be added via the application of Linked Data using e.g. the *Semantic Collaborative Ontology for Three-dimensional visualisation of Cultural Heritage (SCOTCH)* (Vitale, 2016). *SCOTCH* would provide the additional benefit of allowing the participant to document their modelling processes but might be difficult to integrate with proprietary formats.

Another type of *Visualization* mentioned in survey and interview responses was that of networks, which additionally relates to other TaDiRAH methods, *Network* and *Relational Analysis*. In addition to Case Study 2 (5.6.2, below), PART255 intended to create network visualisations from data produced as part of a collaborative research project. Their preferred software was *Nodegoat*, which stores data in a relational database with some Linked Data integration, including dynamically populating data models by querying SPARQL endpoints (Bree & Kessels, 2013). However, the participant encountered an access barrier in that there is a cost associated with setting up a licence for multiple users on the same project. Their comments exemplify the barriers that cost can pose to using digital tools and resources (as identified in 4.1.1), but also indicate an expectation that these tools and resources should be free to use, even when there is increased complexity in user requirements (and their sustainability needs to be ensured).

I have already mentioned links between *Visualization* and *Discovering* (5.1). PART119 took these connections further when they were involved in producing a digital resource, by presenting search results as visualisations, such as maps, timelines and pie charts. They also stated the importance of providing visualisations in different formats, incorporating a table view and CSV export, to facilitate transparency and reusability. This idea of transparency introduces another issue discussed by several participants. They felt that some producers use visualisations to control the degree of access to their data, i.e., providing access to the visualisation but not the data itself. PART001 found that this level of control by external organisations made it difficult to reuse data or connect it to other datasets. Data (re)usability is a key theme that arose from my research, with findings that apply to multiple research methods; as such, I will discuss this topic further in 6.2.

Like *Annotating*, I can identify *Visualization* as occurring at the intersection of multiple research methods, largely because it requires other analysis methods to have taken place as a preliminary step. Such methods include *Network*, *Relational* or *Spatial Analysis*, as well as *Data Recognition* (5.3) and *Annotating*, all of which currently benefit from Linked Data integration. In turn, effective visualisations can be used to facilitate discovery via exploration (5.1.2). In this section, I have additionally found multiple diverse examples of *Visualization* types that are used and/or produced by Ancient World researchers. Those where Linked Data has already been applied particularly effectively include spatial and network visualisations, with additional potential for implementation in 3D models. While participants generally found visualisations to be an intuitive and helpful means of interacting with data, some also recognised the importance of providing the data itself, accompanied by information on how it has been processed.

Having discussed each of the five research methods individually, I will now provide examples of how participants have combined them.

## 5.6 Case Studies: Combining Research Methods

Thus far, I have considered each research method separately. However, as noted multiple times in the above discussion, there are numerous intersections between methods and, as such, they are often conducted in combination. The case studies in this section therefore demonstrate the reality of researchers' interactions with digital tools and resources, by illustrating how two participants combined several research methods in different ways. Firstly, Case Study 1 (5.6.1) describes PART005's experiences of combining *Discovering*, *Data Recognition*, *Annotating* and *Gathering*, using various tools and resources. Secondly, Case Study 2 (5.6.2) outlines PART061's experiences of producing a tool that combines *Annotating*, *Visualization* and *Discovering*.

#### 5.6.1 Case Study 1: Discovering – Data Recognition – Annotating –Gathering

PART005 conducted four research methods in combination to facilitate their study of the type of language used in an ancient text, using *Perseus*, *Recogito* and the *Beazley Archive*. Their sequence of tasks associated with each method was as follows:

**Discovering:** searching *Perseus* for the relevant text and exporting it.

**Data Recognition:** using *Recogito* to identify and "map in links between vocabulary to look for patterns", finding that using a digital tool made this task considerably more efficient than working through the text manually.

The participant acknowledged that this task might not explicitly have been the purpose for which *Recogito* was intended, "adapting it away from that core functionality", thereby demonstrating the importance of flexibility in the design of tools and resources, to widen the scope for their applicability.

**Annotating:** annotating relevant parts of the text, then exporting these annotations from *Recogito* in CSV format to create a spreadsheet.

**Gathering:** enhancing the spreadsheet by manually adding information and links from *Perseus* and the *Beazley Archive*.

During our interview, this participant said that they would ideally have liked to be able to use a tool or resource that links references in the text to visual representations of the relevant entity, as well as connecting authority files for entities such as places and people. As discussed in 5.2, they additionally implied that that creating their own dataset might not have been necessary if they were better able to explore connected resources. I might therefore infer that enhancing the *Discovering* process by increasing connections between existing resources (as discussed in 5.1) could make the *Gathering* process more efficient.

### 5.6.2 Case Study 2: Annotating – Visualization – Discovering

In our second example, PART061 is working on a project to produce a tool for a researcher to annotate philosophical texts, using Linked Data technologies; it is accompanied by a discovery tool to allow other users to access annotation data via queries and visualisations. PART061 has designed its functionality as follows:

**Annotating:** the researcher makes annotations on a digitised text, "to represent those topics and arguments in a Linked Data graph".

**Visualization:** the user visualises these annotations as a network, for use in research or teaching. The eventual aim is that users should be able to make use of the tool without a detailed understanding of the underlying technologies.

**Discovering:** the user explores the visualisations and/or performs SPARQL queries on the dataset.

In allowing user interaction via these different research methods, Linked Data technologies have facilitated sharing the annotations with other researchers, as well as people who might use them in a teaching context. *Visualization* (5.5) and *Discovering* might therefore be natural progressions from the *Annotating* (5.4) stage of the research process.

\*       \*       \*

Rather than limiting this chapter to a somewhat artificial separation between research methods, these case studies have demonstrated how such methods have been combined in practice, from both a user and producer perspective. Furthermore, rather than confirming the trajectory anticipated by my structure for this chapter, of discovery through to analysis and presentation of results, the case studies (as well as the previous sections) demonstrate that each method could be conducted at any point in the research process. For example, *Visualization* might be applied to facilitate exploratory *Discovering*, or *Data Recognition* might be performed throughout the *Annotating* process.

## 5.7 Conclusions: Research Methods for Linked Data Integration

In this chapter, I have discussed the five research methods of *Discovering* (5.1), *Gathering* (5.2), *Data Recognition* (5.3), *Annotating* (5.4), and *Visualization* (5.5), while acknowledging that the level of detail across sections was uneven. *Discovering* was mentioned by almost all participants, across the digital competence/confidence spectrum, producing a wealth of material from which to derive my findings. The other four methods, however, seemed far more specialised, with each discussed at length by relatively few participants. This chapter has also demonstrated that none of these methods occurs in isolation and that they are, in many cases, interlinked, as exemplified by the two case studies.

Furthermore, based on the above discussion, *Discovering* seems to be a constant throughout the research process, and is often a researcher's ultimate goal when employing other methods. Similarly, I have identified that *Annotating* and *Visualization* form intersections with multiple research methods, which might explain their relative popularity among survey participants in 4.2. *Discovering* and *Gathering* were also strongly connected with each other, with participants' responses indicating that some of their *Gathering* tasks were conducted in response to the lack of integration between existing discovery tools. Therefore, while identifying distinct methods can facilitate a greater understanding of the research process, imposing too much division between them can have the opposite effect.

All five research methods have the potential to benefit from Linked Data integration, either individually, or in combination. Examples outlined already include:

1. Searching multiple resources from a central point;
2. Providing contextual information:
  - a) Recognising and annotating entities (within a text, image, or visualisation) with identifying and contextualising information, to facilitate discoverability and exploration;
  - b) Enhancing visualisations with documentation about their production;

3. Gathering and connecting tools, datasets, and bibliographic materials:

- a) Facilitating an individual's research process;
- b) Collating and curating lists of digital tools and resources, particularly those relating to Linked Data production and consumption.

Many of these examples already exist but were not always familiar to participants. Notable by its absence in our discussions was *ResearchSpace* (2.2.4), which combines several of the research methods from this chapter, possibly due to its limited availability at the time of my study. My findings will form the basis for recommendations in Chapter 8, to suggest how the integration of Linked Data with these research methods might best be achieved in future, thereby addressing RQ3.

Having explored my findings in relation to RQ3, the next chapter will commence my investigation of RQ4, on the broader topic of improving Linked Ancient World Data usability.

## 6 User Engagement with Linked Data

In Chapter 4, I provided a broad overview of my survey population by presenting findings based on quantitative data produced from their responses, in relation to digital tool and resource use and production, as well as associated research methods. I used the findings on the latter topic as the foundation for Chapter 5, where I used qualitative responses to identify where Linked Data might most effectively be integrated with existing methods for Ancient World research (RQ3). The current and following chapters will take a similar approach to discussing qualitative findings about the usability of digital tools and resources and their underlying data, with specific reference to Linked Data (RQ4). Here, I will focus on the user's perspective, while Chapter 7 will concentrate on issues affecting producers.

As explained in 1.4, my definition of usability is based on that provided by ISO (2018). In the context of this thesis, the term "usability" refers broadly to the extent to which Linked Data (and the tools and resources used to interact with it) can be used by Ancient World researchers to achieve their "specified goals" with "effectiveness, efficiency and satisfaction". As most participants accessed data via a purpose-built tool or resource, I will start this chapter with a brief section on responses relating to tool and resource usability (6.1). However, most responses related more to the usability of the data itself; therefore, I will then provide a more detailed discussion on the multiple factors that affect data usability (6.2), before summarising my findings in 6.3. Throughout this chapter, I will identify a series of themes (training, collaboration, user-centred design, documentation, access, and sustainability) that will form the basis for further discussion in Chapter 7, representing potential areas where Linked Data producers might ensure or improve usability in future.

### 6.1 Tool and Resource Usability

As implied in the examples discussed in 2.2, Linked Ancient World Datasets are often accessed via a digital tool or resource, whose usability is key to the usability of the underlying data, particularly for those users with minimal technical experience. Conversely, poor tool or resource usability forms a barrier to accessing Linked Data and deters the user from future interaction. The theme of tool and resource usability additionally relates to 'Ease of use/installation', the most popular feature of digital



tools and resources mentioned by participants in 4.1.1. The majority who mentioned this feature were non-Linked Data users, indicating that this is a significant concern for users without an advanced level of digital competence; for example, survey participant PART256 stated that digital tools and resources "should be... easy to use for computer illiterates like me". To be more specific, participants wanted digital tools and resources to have interfaces that allow them to easily perform their desired tasks or methods, such as those discussed in the previous chapter. This point was summarised effectively by PART061:

*"It is, as I see it, one of the big [challenges] to... make tools and resources which are naturally or easily focused on technology to become usable by people who do not necessarily want to care about technology but want to achieve a particular task."*

Users who are focused on a particular task, rather than driven by technology, are therefore unlikely to have the time or patience to use a digital tool or resource that does not allow them to complete this task easily and efficiently. For example, survey participant PART258 stated that encountering usability issues related to their desired task usually results in their abandoning the tool or resource. The risk of producing tools and resources with such limited usability is particularly pertinent to approaches such as Linked Data, where the complexity of the underlying data could affect the usability of the interface.

Participant comments relating to tool and resource usability can be grouped under the themes of user interfaces (6.1.1) and reliability (6.1.2), which I will discuss below. Although my focus remains Linked Ancient World Data, many of the comments made in this section apply to digital tools and resources of any type, as well as Humanities research more generally.

#### 6.1.1 User Interfaces

I will start my discussion of tool and resource usability by exploring participant comments relating to user interfaces, defined by the International Organization for Standardization (2020) as the "set of all the components of an interactive system that

provide information and controls for the user to accomplish specific tasks with the interactive system". The interface is the first aspect of a tool or resource that users will usually encounter, and its effectiveness is crucial for access to functionality and data by the widest possible audience. In this section, I will discuss key features that were important to participants, the incorporation of visual elements and instructions, the merits of implementing an existing system, and accessibility features, as well as why some participants chose not to develop a user interface for access to their data.

Participants expected the interfaces they encountered to be intuitive, i.e., that it should be clear how to perform their desired task(s). The interface should render the tool or resource easy to use, with clear information and visual cues, without the user needing to consult documentation to perform basic functions. Comments indicated that intuitiveness can best be achieved by ensuring that interfaces are designed for a generally interested user with minimal technical experience. For example, PART109 stated that requiring coding or programming experience for any interaction with Linked Data to take place is an immediate barrier to a significant proportion of Ancient World researchers. Key components to engage this wider audience, as suggested by PART005 (in relation to the *Beazley Archive*), included strong visual elements, combined with clear information, while incorporating effective search functionality. PART001 also appreciated simple navigation of a tool or resource, avoiding complex multi-level hierarchies or ambiguous terminology used in menus. Participants tended to prefer tools and resources that provide a clear indication of expected user journeys, rather than presenting too many options at the start, with PART037 and PART001 criticising software such as Oxygen<sup>133</sup> and QGIS<sup>134</sup> (respectively) for overwhelming them. On a similar note, PART017 praised their university library catalogue, which is easy to search and produces multiple helpful results with clear filtering options, to avoid overwhelming the user with too much information at the outset. These responses therefore demonstrate an interest in simplifying the user experience, even for those who wish to perform complex tasks.

---

<sup>133</sup> <https://www.oxygenxml.com/>

<sup>134</sup> <https://qgis.org/en/site/>

As I discovered in 4.1.1, participants often used digital tools and resources to access digitised versions of texts that had previously been made available in print. Some interface designers have found it tempting to incorporate skeuomorphism, emulating these physical items in their digital presentation, as in the case of the *Loeb Classical Library*. While this might initially appear to create a visually appealing, familiar environment, user interaction with a screen is fundamentally different from that with a printed book and adhering too closely to the appearance of a physical object can be at the expense of expected digital functionality. In my brief introduction to this resource, I cited Dik's (2015) review, which highlights the discovery and navigation issues caused by these design choices. These comments were echoed by my participants, with PART012 speaking of their frustration with the *Loeb* content being restricted to single, numbered pages. Attempting to reassure users by replicating elements of the physical experience can therefore do more harm than good.

Several participants, including PART012, found that commercial tools and resources, intended for a wider remit and user base, were more intuitive than those produced for niche academic audiences. While it is reasonable to expect that it might not be possible for relatively complex, specialist operations performed by academic users to be presented as intuitively as more general tasks, such as basic keyword searches, there is still much that can be learned from commercial interface design by academic tool and resource producers. As PART041 suggested, when planning the design of a digital tool or resource, serious consideration should be made to producers' own experiences (as users) of both positive and negative aspects of familiar digital tools and resources.

Several comments mentioned combining the visual elements of interface design with clear language that guides the user on their journey, while being explicit about the purpose, remit and structure of the tool or resource. For example, PART061 recommended that Linked Data producers ensure the necessary information to perform a task (including transparency about the system they are using and the provenance of the data) is easily accessible via the user interface, while being mindful not to overload the less technically experienced user. Such a balance might be

achieved by providing minimal information upfront but supplementing this with clear links to a more detailed explanation. I will discuss such documentation further in 7.4.

An important consideration in interface design is accessibility, the fifth most popular feature of digital tools and resources mentioned by survey participants in 4.1.1. I had initially assumed participants were referring to measures taken to ensure a tool or resource is usable by disabled people, based on the definition of "accessibility" by the W3C's Web Accessibility Initiative (2019). Some participants did indeed intend their use of the term "accessibility" to be interpreted in this context. For example, PART012 acknowledged that funded tools or resources (in particular) should be accessible to disabled users; PART078 suggested incorporating resizable text, and PART054 advised that such features could be readily incorporated into a design by using existing frameworks or code libraries. However, during the interviews it became apparent that other participants had intended a definition similar to that included in the *FAIR Guiding Principles* for research data (1.2)<sup>135</sup>. Broadly, these participants wanted to ensure that they were able to access the functionality and data of a tool or resource, an issue often related to subscription charges (which I will discuss in 7.5).

While the usability benefits of providing an interface are clear, not all producer participants planned on developing one for their own data. PART109 stated that in their case, while recognising the importance of providing a user interface to access their data, its absence was due to a lack of time and technical expertise. This issue was particularly apparent in the case of solo projects, such as PhDs. Responses therefore indicated that it would be advantageous for data producers to have a simple and intuitive means of making data available via a user interface, thereby enhancing its usability and potential for reuse. Such a system would be particularly beneficial for providing a usable mechanism to explore complex data structures, such as that of RDF, to increase uptake of Linked Data tools and resources among less technically experienced researchers; as such, I will discuss possible solutions as part of my recommendations in 8.3.

---

<sup>135</sup> "To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol  
A1.1 the protocol is open, free, and universally implementable" (Wilkinson et al., 2016)

This section demonstrated that the potential audience for a Linked Ancient World Dataset can be maximised by enabling interaction via a tool or resource that incorporates a simple, intuitive, and accessible user interface. Strong visual elements can enhance the user's experience and interaction with content and data, provided aesthetics does not take priority over functionality. User interfaces should minimise the requirement for specific skills or knowledge, providing access to key information as required, while overall usability could be improved by considering how disabled users might experience the tool or resource. Not all producers had the time, resources, or skills to build an interface through which potential users could access their data; it could therefore be beneficial to use an existing system through which this could be achieved with minimal time, effort, and training. In all cases, producers of new tools and resources for Ancient World research, particularly those involving Linked Data, would benefit from establishing the needs of their potential audiences and taking these into account when designing the user interface.

I will return to discussing user-centred design approaches in 7.3, where I will examine the themes revealed in this chapter from a producer perspective. For now, I will continue the topic of tool and resource usability by discussing the impact of reliability issues.

#### 6.1.2 Reliability

Reliability was mentioned by only two participants during my original survey but was ranked highly by interview participants in the follow-up survey (4.1.1). Fundamentally, as stated by PART008 and PART119, participants needed to be able to trust that a tool or resource would continuously work as expected, with PART061 specifically highlighting the relationship between reliability and usability. Here, I will discuss aspects of tool and resource reliability that were important to participants, including availability, stability, and consistency, as well as how bugs and technical issues might be resolved.

A fundamental characteristic of reliability, as suggested by PART037, is whether the user can trust that a tool or resource will be available when required. Unavailability

was a particular problem for participants whose time was limited, such as PART005, and deterred them from continuing to use some tools and resources. As mentioned in 2.4, lack of availability is a particular issue for Linked Data tools and resources due to dependence on external data sources, as well as SPARQL endpoints being notoriously temperamental. Several participants found that Linked Data tools and resources were particularly prone to reliability issues in general, which survey participant PART002 suggested might be due to their relative lack of maturity compared with tools and resources based on other technologies. Although this should improve over time with more training, experience and sharing of best practice (and funding), PART061 speculated that it could be a potential deterrent to Ancient World researchers using or producing Linked Data in the immediate future.

A similar reliability issue is stability, which applies particularly to tools or resources with complex and dynamic visual elements that require a strong internet connection and/or substantial memory on the user's device. For example, PART041 found that both *Digital Augustan Rome*<sup>136</sup> and *Antiquity À-la-carte*<sup>137</sup> "end up either crashing browsers or systems". Similarly, PART254 found that *ORBIS: The Stanford Geospatial Network Model of the Roman World*<sup>138</sup> often crashed their web browser, occasionally rendering its use impossible. PART078 often experienced issues where tools or resources crashed after they had input a query, meaning they had to start their search again. Providing a low-bandwidth alternative to complex visual interfaces might therefore be desirable to avoid frustration that could deter future use.

Another key component of reliability is consistency, which manifests itself in various ways, such as securing the user's trust that performing the same action multiple times will produce the same outcome. For example, PART043 was suspicious of the Beinecke Library's digital collections<sup>139</sup> because they found that inputting the same query on different occasions produced varying results. In addition to causing a frustrating experience for the user, this inconsistency highlights a lack of transparency

---

<sup>136</sup> <https://www.digitalaugustanrome.org/>

<sup>137</sup> <http://awmc.unc.edu/wordpress/alacarte/>

<sup>138</sup> <https://orbis.stanford.edu/>

<sup>139</sup> <https://beinecke.library.yale.edu/digital-collections/digital-collections-beinecke-library>

surrounding the search process, which can undermine the user's trust that future queries will produce reliably comprehensive results.

The concept of consistency similarly applies to the behaviour of a tool or resource across different environments. In many cases, consistency can be achieved by ensuring that users can interact with a tool or resource in a web browser rather than requiring download and installation of specific software (another barrier to interaction, as identified by PART254). However, compatibility issues still remain with web-based tools and resources, e.g., PART041 and PART254 mentioned that some browsers interpret code in different ways, with PART012 finding that some tools and resources do not display correctly on mobile devices. If a specific browser or device is required for effective interaction with a tool or resource, this immediately and significantly reduces the potential audience. A related issue is that of functionality across different tools and resources; as PART078 stated, participants tend to base their expectation of how a tool or resource works based on their previous experiences elsewhere. Any inconsistency in this regard imposes a usability barrier, causing the user to spend more time learning how to use the tool or resource and less time performing their intended task.

Reliability also relates to the user's ability to interact with a tool or resource without being impeded by bugs or technical issues. Different participants were affected to varying extents, with responses suggesting that those using specialised academic tools and resources are affected more than those who predominantly use commercial software, likely due to their increased time and budget available for identifying and fixing issues. Participants performing more complex operations, such as PART001, were particularly affected by a lack of feedback returned from a system when their input contained an error, rendering it difficult or impossible to fix the issue. Providing information that allows users to easily identify and resolve such errors results in a more efficient interaction with the tool or resource.

In the above section, I have used survey and interview responses to demonstrate that Linked Data tools and resources are particularly prone to reliability issues, which can cause mistrust and deter repeated usage. Availability and stability were identified as

key issues by several participants, which apply particularly to Linked Data tools and resources. Consistency is another key aspect of reliability, which involves ensuring that the same action conducted at different times will produce consistent results, that the tool or resource will perform consistently across multiple browsers, devices and platforms, and that design elements behave consistently with those of similar tools and resources. A related issue was the degree to which users experience bugs and technical errors. As suggested in 6.1.1, such issues might be resolved, or at least mitigated, by conducting user research to inform development, providing clear information to the user, and encouraging feedback about potential errors. Reliability issues might also be addressed by regular maintenance, which will form part of my discussion on the theme of sustainability in relation to Linked Data production in 7.6.

\* \* \*

Through considering participant responses on the topics of user interfaces and reliability, key aspects of tool and resource usability, several themes are beginning to emerge. Firstly, there is a clear need for Linked Data producers to receive appropriate training – both to develop a usable interface, allowing others to interact with their data, and to ensure that interface functions in a reliable manner. However, I also noted that implementing an existing system with required functionality can often be a suitable (and more efficient) alternative. Secondly, many of the issues identified in this section might be avoided if the user is placed at the heart of the development process. Implementing aspects of user-centred design is likely to result in more intuitive interfaces, while also facilitating the early identification of potential reliability issues.

Thirdly, users' experiences can be improved, and their expectations managed, with appropriate documentation, providing information about the scope of the tool or resource and how it functions. Although the need for extensive documentation might be decreased by developing an intuitive interface (as above), making relevant information available at key points can be extremely effective, particularly in reducing the need for the user to obtain specific training to use the tool or resource. Finally, the potential audience can be maximised by minimising access barriers. Such barriers



might relate to disabilities or arise due to compatibility issues with the user's browser or device.

Having discussed participant comments relating to tools and resources through which they access data about the Ancient World, I will now discuss the usability of the data itself.

## 6.2 Data Usability

Tool and resource usability, discussed in the previous section, is underpinned by usability of the underlying data. As a result, many more participant comments related to the data itself. Some spoke about key considerations to make when modelling data (6.2.1), such as incorporating usable standards and vocabularies. Others commented more on their concerns about data quality (6.2.2), including topics such as trust and accuracy. Additional responses related to aspects of data sharing (6.2.3) to facilitate reuse by others.

### 6.2.1 Data Modelling

Decisions made from the outset of the modelling process, and their documentation, can have a significant impact on usability of the resulting data. Factors discussed in this section include interoperability, open standards, identifiers, and ontologies, as well as ensuring that the data incorporates sufficient granularity to meet the initial project goals, without precluding future reuse in different contexts. Although some of these measures are particularly applicable to a Linked Data approach, much of the below discussion could apply more broadly to any datasets produced for Ancient World (or indeed Humanities) research.

Key to interoperability and, indeed, sustainability of data is the facility to export it from one tool or resource and import it into another. Export functionality was the fourth most popular feature of digital tools and resources in 4.1.1, mentioned by both Linked Data and non-Linked Data users. Related comments reflected a broad spectrum of user needs. PART089 and survey participant PART251 were specifically interested in having the ability to copy and paste text into a word processor or spreadsheet software, or to download ready-made documents in formats such as PDF. Others, however, had more complex requirements. For example, in my survey, PART059 mentioned the ability to

conduct queries, then download the results in a usable format that maximises potential compatibility with external tools or resources' with PART040 specifically referring to Linked Data in this context. Participants such as PART032 and PART089 criticised *TLG* (4.1.1) for actively preventing users from exporting data. Ensuring that export functionality is available and usable improves the efficiency of the research process and increases the potential scope for the dataset's reuse.

Types of reuse that are possible when data is made available using open standards (1.2) include text analysis, spatial analysis, and visualisation, as mentioned by PART054, with relevant tools and resources having different data import requirements. While, in some contexts, it is not unreasonable to expect the user to perform some data conversion themselves, in others this is more difficult. For example, PART119 wished to explore cuneiform texts using *Perseus' Scaife Viewer* (4.1.1) but was prevented from doing so because these texts are not available in TEI/EpiDoc and cannot readily be converted. In this example, restriction of data formats not only limits the scope for users to take advantage of new tools and techniques, but it places the onus for all interaction with that data on the people who originally produced it, which is not sustainable in the long term and is likely to lead to user frustration.

To mitigate such compatibility issues, it can be helpful to provide data exports in multiple formats. Many tools and resources make their data available as comma separated values (CSV) files, an accepted (and, as mentioned by PART008, expected) method for sharing data produced in more complex formats. Although RDF provides a richer representation of a dataset, CSV provides significantly more options for future reuse and increases accessibility to a wider audience. Minimising the need for conversion has the additional effect of preserving data integrity by reducing the potential for user error. For example, PART001 uses Light Detection and Ranging (Lidar) data, which requires considerable additional processing that could pose a technical barrier and lead to errors from inexperienced users. PART119 mentioned that they initially publish data in the format most helpful for them, but then add different formats in response to user requirements, taking a similar approach to *Recogito* (Simon et al., 2019, p. 157).

Where data is not made available using open standards, this limits the potential scope for reuse, or can make it completely impossible. For this reason, PART001 recommended converting any data originally produced in a proprietary format before making it available for export, provided this is permitted within copyright restrictions. Such data conversion not only increases the scope for reuse but also the potential sustainability of the dataset, as it is no longer attached to a specific piece of software or specific producers and can continue to be usable after that software becomes deprecated.

With regard to Linked Data, RDF and SPARQL are themselves open standards. Tools and resources using these technologies should therefore be particularly amenable to data export, with inbuilt interoperability and extensibility. For example, PART054 made their data available for export as RDF via a repository, in addition to providing access through their online resource. However, as I found in my discussion of barriers to Linked Data use and production in 2.4, implementing open standards does not always lead to usability. For example, PART061 provides access to their data via a SPARQL endpoint but acknowledges that this restricts access to users with specific technical knowledge and skills. The participant did say that they would like to support use of their data by less technically experienced users but have not yet been able to achieve this. It might, therefore, be advisable to additionally release such data in a simplified form, using more familiar formats.

As mentioned in 1.3, open standards are a key component of Berners-Lee's five-star model, with which most Linked Data producer participants had complied (4.1.3). While I noted that compliance does not necessarily imply usability, the resulting openness can have a positive impact in this area. PART109 noticed a marked difference between those resources that provided five-star Linked Data and those that did not, finding that non-compliant datasets tended to be more difficult to consume, with issues ranging from unpredictability of unstable URIs to a total inability to export data at all. Indeed, PART041 suggested that the more people who use and produce Linked Data, the more potential for interoperability between datasets, increasing the potential richness of descriptions and links.

Another consideration when modelling data is providing identifiers for digital objects, such as URIs (1.3). In my survey, a small number of participants mentioned features such as permalinks and links to external content (4.1.1), both of which can be facilitated through the implementation of persistent identifiers (1.2), i.e., where an identifier can be reliably associated with a digital object over time. Persistent, unique identifiers are intrinsic to Linked Data technologies, which indicates the suitability of this approach for publishing research data of any discipline, including topics relating to the Ancient World. Providing explicit information about the persistence of any identifiers reassures users that they will still exist in future and can therefore be safely shared with others, with the same confidence as more traditional publication citations. Persistent identifiers also ensure disambiguation, as discussed in 2.3 and 5.1.1; for example, PART255 applied a numbering system to person entities, to ensure that each had a unique identifier even when their names were identical.

PART005 found persistent identifiers particularly helpful from a public engagement perspective, as they provide an easy and reliable way to direct interested people to relevant content from resources such as the *Beazley Archive*. Similarly, PART078 stated that applying a unique and persistent identifier to a digital object ensures that users will continue to be able to find, access and share it in future. Indeed, PART041 mentioned that the lack of persistent identifiers caused usability problems when interacting with the British Museum's online collections (2.2.4). They suggested that this might be addressed by inputting accession numbers, but the different cataloguing systems used by different departments mean that such numbers are not always unique. As well as identifiers for individual objects, some tools and resources enable users to generate identifiers for specific queries. PART119 was involved in producing a resource that incorporated this feature, to avoid the user having to input the same query multiple times (which can be a complex process) and allowing them to share a persistent URL with others, e.g., via a more traditional publication.

URL obsolescence was a significant issue for both survey and interview participants. PART254 said that this deterred them from citing online resources because the URL would not reliably continue to exist in the long term, or that the content to which it

refers might be significantly updated, whereas print is reliably static over time.

PART041 had a negative experience when the American Numismatic Society (ANS) restructured their database, changing all the URLs in the process, many of which the participant had stored for future reference. Now PART041 takes screenshots of any online materials they use, which they store on their own website, linking back to the original source (while it continues to exist). This behaviour indicates distrust in the stability of digital resources and their identifiers and demonstrates the importance of ensuring that identifiers are persistent, e.g., by resolving to a new location if a resource is restructured.

In addition to affecting citation practices, link instability can cause problems with the functionality, and therefore usability, of other tools and resources that depend on use of these identifiers, an issue that particularly affects those using a Linked Data approach. There are likely many cases where identifiers are being used without the knowledge of the original data producers or publishers. If these identifiers are not persistent, this might render a third-party tool or resource unusable and act as a deterrent to implementing Linked Data in future. Therefore, updates to Linked Data tools and resources have much wider usability implications than changes to tools and resources based on other technologies.

Several participants spoke about their use and production of ontologies for identifying classes and properties that describe digital objects. For example, PART037 particularly enjoyed using the *Ontologies of Linguistic Annotation (OLiA)*<sup>140</sup>, which reconciles the tags used for parts of speech in different languages, stating that it achieves what should be the ultimate aim of creating an ontology: saving the time and effort of future users. PART109 agreed with this sentiment and was considering how to ensure that their own ontology would be as reusable as possible, e.g., by separating it into a non-discipline-specific ‘core’ and an Ancient World ‘extension’; I will discuss this type of extensible approach to development further as part of my recommendations in 8.3. The same participant acknowledged that the best way to optimise reusability would be to incorporate existing ontologies rather than creating a new one but found that their

---

<sup>140</sup> <http://www.acoli.informatik.uni-frankfurt.de/resources/olia/>

requirements were too specific to rely solely on such resources (they specifically mentioned *Pleiades*, 2.2.1, and the *Digital Atlas of the Roman Empire (DARE)*), as they were built with a different scope in mind.

Serious consideration is required with regard to what to include in an ontology, as deciding on the appropriate term to use in each situation can be very difficult. As found by PART037, in some cases, the time taken to make such decisions can pose a significant barrier to ontology production. However, if the process is rushed, the end result may be incomplete or poorly conceived, rendering alignment with other vocabularies difficult or impossible, particularly if definitions are unclear. Another concern of PART037 was that, even where definitions are clear, it is not always possible to know if two people have exactly the same concept in mind when using the same term. However, if the definitions are too specific, this can limit the scope for reuse in different contexts; for example, PART109 found that, in many cases, they were unable to use *CIDOC CRM* (1.3) because its definitions did not exactly align with theirs.

Like ontologies, dataset structures should also achieve a balance between sufficient specificity to answer the original research question and sufficient generality to be used for other purposes. Regarding this latter point, PART255 expressed a concern about the potential for datasets to be too rich for reliable reuse, if they incorporate too much interpretation from their producer(s). The participant felt that such excessively detailed datasets could undermine the research process by causing future users to engage with the data differently than if they were categorising the data themselves. Similarly, PART001 advised that, when considering whether to use an existing dataset, a researcher must first establish what has happened to that dataset already, before attempting to make any interpretations based on its contents. Accompanying documentation or metadata can assist the user in making this decision; however, there will likely be some degree of unconscious assumptions and biases that would be impossible to record.

This section has explored various decisions relating to data modelling that facilitate usability by others. Firstly, open standards increase the potential scope for data reuse

in different contexts, and usually facilitate transparency of information about data structures, as well as long-term sustainability. For those users with experience interacting with Linked Data (including the open standards RDF and SPARQL), compliance with Berners-Lee's five-star model tended to be a reliable indicator of interoperability and reuse potential. Secondly, persistent identifiers facilitate accurate, unambiguous and reliable discovery and citation of digital objects, while maintaining the functionality of third-party tools and resources that rely on linking to this content. Thirdly, while implementing established ontologies is usually preferable from a reusability perspective, new ontologies can be beneficial provided that sufficient consideration is given to definitions of terms and the relationships between them. Finally, ontologies and datasets should achieve an appropriate level of granularity, incorporating enough detail to be useful and usable in the relevant context, while aiming to avoid influencing future users with excessive interpretation.

Throughout this discussion, I have found that the process of creating a dataset or ontology should be accompanied by the production of documentation containing information about its structure, as well as key decisions made, and theoretical approaches taken. Effective, transparent, communication about data modelling is therefore key to Linked Data usability. However, usability considerations do not stop when modelling decisions have been made. As I will discuss in the following section, there are many other factors affecting data usability, which fall under the broad theme of data quality.

#### 6.2.2 Data Quality

Although data quality was not mentioned in the survey as a feature of good tools and resources in 4.1.1, it was implied in several comments about reliability. Additionally, the opposite (inaccurate/incomplete data) appeared in the top half of the barriers mentioned by participants in the same section. The theme of data quality was much more prominent during the interview phase, where it became increasingly clear that data quality (or lack thereof) can have a significant impact on usability. In this section, I will start by discussing user trust and assessing a dataset's authority, before exploring comments relating to data accuracy and completeness. I will continue by looking at how such issues might be addressed by allowing and facilitating user contributions,

with the final section relating to the difficulties in representing uncertainty, a key concern for Humanities data in general, and Linked Ancient World Data in particular.

For data to be usable by academic researchers of any discipline, users need to be confident that it originates from a trustworthy source. In several cases, I found that participants viewed digital information sources with more suspicion than their print counterparts, with survey participant PART140 particularly sceptical due to the amount of incorrect information propagated online. As mentioned in 6.2.1, information about the data's provenance and authority can be communicated via documentation or metadata about its production and modelling, thereby enabling users to perform an accurate critical assessment (and avoiding the issues I encountered in 4.1 when attempting to ascertain whether various tools and resources met my definition of Linked Data). Several participants, such as PART001, therefore recommended that users have some understanding of data structures and vocabularies. In particular, Linked Data producers such as PART054 felt that users should have a basic conceptual understanding of Linked Data to use tools or resources based on this approach, but that this should not necessarily require comprehensive knowledge of the technical details. Producers should therefore achieve an appropriate level of transparency about a tool or resource, while ensuring that users with lower levels of digital competence/confidence are not overwhelmed by superfluous technical information.

Once a user has determined that the data source is sufficiently authoritative and trustworthy for it to be used in scholarly research, they must also be able to assess the data's accuracy and completeness. PART001 considered inaccurate and incomplete data to be "a fact of... digital life" that all dataset users experience to some degree: every researcher has different goals that affect their approach and the tasks they intend to perform, and some level of processing or enhancement is likely to be required to adapt datasets to their specific needs. Any gaps or inaccuracies might not have been relevant to the producer in the same way that they are to the user and can often be overcome if the data itself is applicable to their purposes. Therefore, while varying levels of accuracy and completeness can cause data usability issues, in many cases these issues should not become barriers, provided that the user is well-informed



about the dataset's extent and limitations, and that they are prepared to address any issues they discover.

For the often-complex subjects of Humanities datasets, some participants felt that Linked Data provides scope for more accurate representation than tabular formats (which I also found in 2.3). Survey participant PART002 felt that Linked Data was the best way of connecting structured data where representing relationships to other sources is important, while PART008 found it to be the most accurate approach for representing many-to-many relationships. PART054 even observed that, in some cases, Linked Data might actually provide a simpler approach than relational databases, which can become very complicated as more relationships are added. However, there remain some common issues experienced by participants, including encountering gaps or errors in the data, or outdated information.

Participants found gaps in datasets to be a relatively minor issue, with PART078 commenting that once a user has access to key information about an entity, they then have the tools to discover more about that entity from other sources. No one dataset can ever claim to be truly comprehensive; for example, as PART255 found, major generic authority files often contain only the most prominent entities that relate to a particular subject area, whereas a more specialised dataset would contain less well-known entities, within a much narrower remit. Cases such as these led to participants creating their own ontologies to fill the gaps (6.2.1) or using print resources to supplement existing digital information, PART255's eventual solution.

Generally, participants found that any concerns about incomplete data could be mitigated by data producers adopting a transparent approach, where documentation containing information about what is, and what is not included in the dataset, is clear and readily available. Many datasets can never be considered complete as they are continuously updated based on new knowledge, or new data input and digitisation efforts. In some cases, as PART012 suggested, it can often be beneficial to divide the work into manageable tasks, releasing a series of smaller datasets (e.g., based around specific geographic regions), rather than waiting until one major dataset is complete. Early and partial release of data ensures that users can derive benefit from it more

quickly, and their feedback used to improve subsequent additions. Furthermore, PART037 advised that availability of these partial datasets might also attract the attention of potential collaborators who can already fill in some of the gaps, allowing future work to be more effectively prioritised.

Accuracy issues tended to be a more serious concern; examples included PART254's disagreements with translations from ancient languages, as well as PART041 encountering incorrect associations between object metadata and images. PART041 also found transcription errors in *Nomisma* (2.2.4) data, which had the effect of important objects not appearing in search results because they were not associated with the correct keywords. Identification of such errors can be achieved relatively easily by domain experts but could go unnoticed by those new to the subject area, which could lead to flaws in any resulting scholarship. In some cases, such errors may result in the researcher deeming the data unusable (as PART078 said, "If you're researching and the data's inaccurate, why bother?"), with them instead seeking another, more reliable, source.

Often, errors or inaccuracies arise from information that was correct when a dataset was first created but has since become outdated; for example, some datasets were originally produced from digitised print publications. Although such 'traditional' publications have the advantage of conveying academic authority, their information is dependent on interpretation of the available evidence, potentially requiring reconsideration if new discoveries are made. In one example, PART089 lacked trust in *Perseus* (4.1.1), because subscription-based resources such as the *Oxford Latin Dictionary (OLD)*<sup>141</sup> and *TLG's Liddell-Scott-Jones Lexicon (LSJ)*<sup>142</sup> contained more up to date information<sup>143</sup>. PART012 expressed similar sentiments about *LacusCurtius*<sup>144</sup>, which provides primary and secondary texts relating to the Roman world, suggesting that it could potentially be improved with the addition of more recent critical works - although this would be unviable due to current copyright laws. Issues surrounding

---

<sup>141</sup> <https://www.oxfordscholarlyeditions.com/page/the-oxford-latin-dictionary>

<sup>142</sup> <http://stephanus.tlg.uci.edu/lcj/>

<sup>143</sup> *Perseus* hosts an older version of the *LSJ*, while *TLG* hosts an updated version (Pantelia, 2011).

<sup>144</sup> <http://penelope.uchicago.edu/Thayer/E/Roman/home.html>

openness and availability of digital objects relate to the wider theme of open access, which I will discuss in 7.5.

A similar example is that of *Nomisma*, whose *Coinage of the Roman Republic Online* (*CRRO*) dataset is based on information from a 1974 publication<sup>145</sup>, without reflecting updates from subsequent scholarship (American Numismatic Society, n.d.). Although this information is made clear on the *CRRO* homepage<sup>146</sup>, it is not acknowledged on individual object records. PART041 felt that such an omission could risk users unfamiliar with the subject area making inaccurate interpretations, as well as perpetuating outdated chronologies and typologies.

Due to accuracy and currency issues in existing Linked Datasets, PART109 urged caution in the production of new datasets, advising that quality should be prioritised over quantity. PART041 agreed, suggesting that more time and funding should be invested in data cleaning and maintenance than in production. There are, however, sustainability implications; employing a member of staff responsible for maintaining data quality would be difficult to maintain in the long term as the amount of data increases and would be impossible for a fixed-term academic project.

Various tools and resources manage this issue by inviting user contributions, i.e., encouraging users to let producers know about any errors in their data (and how to address them), as well as providing a mechanism to add new data. PART041 had been involved in the production of a resource that incorporated both types of user contribution and expressed frustration that not all tools or resources incorporate these features, while acknowledging the importance (and potential financial implications) of human moderation. PART054 mentioned their generally positive experiences when correcting errors or enhancing contextual information in *Pleiades* (2.2.1).

In other cases, participants had created their own datasets and hoped to add them to a relevant resource once completed. For example, PART255 hoped that integrating their dataset with an established database held by a national cultural heritage

---

<sup>145</sup> Crawford, M. H. (1974). *Roman Republican coinage*. Cambridge University Press.

<sup>146</sup> <http://numismatics.org/crro/>

institution could be a way of ensuring that dataset's sustainability or increasing its potential audience. However, achieving such integration at scale can be difficult. PART008 found that some Linked Data resources, such as *GeoNames* (2.2.1) and *Pleiades*, do not currently support batch data uploads, and was (understandably) unwilling to input hundreds or thousands of individual records; therefore, these resources (and their audiences) do not benefit from this kind of data enrichment.

Allowing and, indeed, encouraging and facilitating user contributions means that a dataset benefits from a much wider range of expertise than that of the people who originally produced it, thereby harnessing knowledge from an expert user community rather than simply imparting information to them in a unidirectional manner, without opportunity for dialogue and enrichment. Comments from participants such as PART054 indicated that user communities feel valued if their contributions are encouraged, facilitated, and integrated, and are willing to spend the time to improve a tool or resource in which they feel invested. Conversely, as experienced by PART041, they feel frustrated if their offers to assist in improving a tool or resource are ignored. Users are therefore more likely to continue using a tool or resource to which they can contribute, which should support arguments in favour of its long-term sustainability.

To facilitate user contributions, participants found that tools and resources should include clear and current documentation, providing information and instructions. In contrast, PART008 found *Pleiades'* documentation to be confusing and out of date. PART054 found it helpful when such documentation was complemented by feedback from data producers or moderators, to ensure that contributions meet quality and format requirements. For tools or resources where there might be some debate about differing interpretations, PART041 suggested incorporating a wiki, to contain discussion and peer-review of suggested changes. However, while user contributions are advantageous to data usability for the reasons discussed above, they do not completely address issues of data quality and should not be considered as a substitute for more formal long-term maintenance and sustainability arrangements, issues which I will discuss in 7.6.

Another issue that particularly affects Humanities data is the representation of uncertainty, as discussed in 2.4. If careful consideration is not given to the terms and ontologies used, a Linked Data approach has the potential to exacerbate this problem, because the triple structure can convey a misleading impression that it represents known facts or overstate the level of connection between entities. For example, PART109 mentioned that they frequently encountered datasets where properties such as 'skos:exactMatch' were used<sup>147</sup>, only to find that the entities in question were similar but not equivalent. This approach can result in incorrect alignment and inferencing, which could serve to undermine users' trust in a dataset or perpetuate these issues in publications.

As a result of these potentially misleading statements, PART001 said that they would prefer for Linked Data to be used solely for modelling known facts, rather than interpretations. In particular, they found that attempting to ascertain the provenance of these interpretations, as well as the data models and ontologies used, was extremely time-consuming, thereby negating the potential efficiency advantages of a Linked Data approach. Although restricting the application of Linked Data technologies to known facts could be one solution, this would be extremely difficult to apply in practice as the line between fact and interpretation in Humanities datasets is often blurred. Furthermore, the advantages of a Linked Data approach could be preserved by effective communication with users, to manage their expectations. PART037 asserted that, if users are prepared to consider RDF triples as claims rather than facts, they are then able to assess these claims, based on the claimant and context, in a similar way to other sources of evidence they encounter during their research. They recommended that pertinent information to incorporate within a Linked Data structure might include associating each claim with a claimant, any degree of doubt, and the temporal and/or geographic scope to which it applies; however, they additionally cautioned that it can be difficult to achieve this effectively using RDF.

PART008 wanted to allow users to contribute to their dataset, while avoiding potential inconsistencies between multiple interpretations. They therefore developed a scheme

---

<sup>147</sup> As opposed to the broader 'skos:closeMatch', for example

for claimants to quantify their confidence in each claim and explained the system fully in the accompanying documentation. However, PART008 also made it clear that they felt there should be an accepted standard that will persist over time and apply across different contexts and datasets, to ensure consistency and manage user expectations. Such a standard must be readily usable, without requiring the user to learn extensive new concepts or terminology.

This section has explored various aspects of data quality, and their relationship to usability, as discussed by participants during the survey and interviews. Key components included trust in and authority of Linked Ancient World Data, its accuracy and completeness, how these might be addressed by facilitating user contributions, and the difficulty in representing uncertainty. While some participants found it difficult to trust digital tools and resources at all, others perhaps placed too much trust in the datasets they used, without questioning potential issues of uncertainty in Ancient World data. More digitally experienced participants tended to occupy a 'middle ground', where they critically assessed datasets before use and kept any limitations in mind during their research.

A clear theme has emerged, relating to the importance of keeping users informed about how the data is structured, as well as any characteristics that might have implications for its quality and reliability. For data to be usable, its quality need not be perfect, but any potential issues relating to e.g., accuracy, completeness or uncertainty should be communicated effectively to the user. Such information might be included in documentation or, ideally, as part of the dataset itself. These measures should ensure that more researchers reach the 'middle ground' mentioned above, where they can place sufficient trust in Linked Ancient World Data for it to form a key part of their research, while being mindful that it cannot be used uncritically and must be treated as any other literary, historical, or archaeological source.

Once a producer has assessed their dataset's quality and ensured that any potential issues are communicated to users, they must then consider another set of factors that impact that data's usability: those relating to how the data is shared.

### 6.2.3 Data Sharing

Although the short-term goals of a digital project are often to answer specific research questions, several participants, such as PART008 and PART017, expressed that a longer-term goal should be to produce a resource that endures beyond the project. In this way, the data can be used, and potentially enhanced, to answer other research questions without duplication of effort. Indeed, PART008 produced their resource with reuse in mind, to help a particular research community by providing relevant geographic information at a central point, to act as a foundation for future projects. The potential for reusability also acted as a catalyst for PART001 to produce digital data at all. They acknowledged that pen and paper would suffice for their purposes but chose to produce digital data to facilitate reproducibility of their work and maximise the scope for reuse in different contexts, including combining datasets. Additionally, as demonstrated in 6.2.2, sharing data can ultimately result in quality improvements via user contributions. Data sharing is usually an intrinsic aspect of applying Linked Data technologies, with the first star of Berners-Lee's five-star model (1.3) requiring data to be openly available online. Indeed, PART061 said that there is no reason to use a Linked Data approach for producers who do not intend to share their data, as it relies upon a fully open infrastructure for making connections and analysing relationships.

Some participants were less enthusiastic about sharing their data. PART005 had not previously considered sharing their dataset but, during the interview discussion, thought they might be interested. However, they acknowledged that there is a difference between producing a dataset for personal use and producing one that is useful and usable for other people. As such, data cleaning would be required, which can take significant time that might not be available towards the end of their project. Considering issues of data sharing would therefore be most worthwhile at the outset of a project, to ensure data usability is kept in mind throughout development.

Other participants were even more cautious. For example, PART012 chose to make selected components of their resource explorable within a web browser, without the ability to download the underlying data, ensuring continued authority and control over their work. However, such restrictions could impede extensive critical analysis of their outputs and render it impossible for other researchers to reuse the data in different

contexts. PART037 expressed concerns about making their data public before completion, due to the potential for it to undergo substantial changes in the near future, although they have made it available to interested colleagues and implemented their feedback. While it is important for data to be in a useful and usable state before sharing, as I found in 6.2.2 it can be beneficial to publish it while incomplete, as this can help encourage collaborations or user contributions to enhance its quality.

Further barriers to sharing data include the cost of hosting and the technical support required, both of which might be mitigated by placing the dataset in a trusted repository. However, from a usability perspective, having an intuitive interface (6.1.1) available to facilitate exploration of the data by less technical researchers is usually preferable, and (as PART255 remarked) this requires continued technical support, necessitating additional funding.

In addition to the above considerations, participants also spoke about the importance of ascertaining conditions of a dataset's reuse. In my survey, 'copyright/licencing' was mentioned by two participants as a feature of good tools and resources, with six mentioning 'permissions/licencing uncertain or too restrictive' as a barrier to tool or resource use (4.1.1). Such aspects were more of a concern to Linked Data users, who tended to be more technically experienced. During the interviews, PART012 mentioned that it can be difficult to locate information about copyright and licencing on the digital tools and resources they use, with PART041 stating that, in this situation, they tended to operate on the principle that it is "better to ask forgiveness than permission". Therefore, while a lack of information about copyright and licencing can act as a deterrent to some, there will always be users who choose to reuse data unless they are explicitly told not to. Making data available under an open licence, such as those provided by Creative Commons<sup>148</sup>, provides an explicit indication that reuse is permitted (and any conditions that should be applied), thereby increasing reusability.

---

<sup>148</sup> <https://creativecommons.org/licenses/>



Evidence of reuse following data sharing is often difficult to obtain but could be crucial to enhancing that dataset's usability and facilitating its connection with similar resources. PART054 mentioned that, having shared their dataset via an established repository, it is relatively easy to see how many times it has been downloaded, but implied that it is impossible to know how (or if) the data was consumed, without the user actively making contact. Facilitating two-way communication between the user and producer could lead to future collaborations, or incorporation of new information to improve the original dataset.

This section has explored issues surrounding data sharing, as experienced by participants. Many appreciated the benefits of sharing research data, such as assisting future researchers and encouraging collaboration, and therefore took steps to share their own data effectively. I found that it was often more conducive to take such steps at the start of a project, considering how future users might interact with the resulting data, rather than waiting until the end to discover that significant data cleaning is required. Participants suggested that data can be preserved at low (or no) cost by sharing it via an established repository, but its usability can be optimised by making it available via an intuitive interface, if funding permits. Users require information about digital data or content to enable them to use it with confidence. Such information includes the licence under which it has been made available, while making explicit the conditions for its reuse.

\* \* \*

My findings relating to data usability complement those on tool and resource usability from the previous section, reinforcing the importance of some of the themes identified earlier, as well as introducing new ones. Of the previous themes, training is key to producing usable data, in terms of selecting and designing appropriate models and ontologies, as well as ensuring that the data is accompanied by sufficient descriptive information when it is shared. On this latter point, such documentation is critical to the usability of a dataset by communicating information about structures and formats used, as well as licencing and reuse conditions. Documentation also manages users' expectations around data accuracy, completeness and uncertainty, and forges user

trust by demonstrating the authority behind any assertions a dataset contains. Access is another theme that reappears, in terms of making data freely available, by using open standards and disseminating via trusted repositories.

The first new theme to emerge is that of collaboration, in the form of inviting user contributions and assisting future reuse. Both activities can also act as a catalyst for more formalised collaborations on new projects or initiatives, as well as helping to build communities around technical and disciplinary areas of mutual interest; such communities might ultimately improve the usability of future tools and resources. Appropriate management of these collaborations is key to sustainability, the second theme arising from this section. Sustainability can be facilitated through the implementation of open standards, as well as considering the needs of future users at the outset of the data production process.

### 6.3 Conclusions: User Engagement with Linked Data

Using the definition of usability summarised at the outset of this chapter, and based on the above discussion, I can conclude that the effectiveness and efficiency of Linked Ancient World Data tools and resources might be improved with simple, intuitive interfaces to the data, which function in a reliable and consistent manner. Steps to improve the usability of the underlying data should start from the modelling process, with the application of open standards, persistent identifiers, and established ontologies, while ensuring an appropriate level of granularity. Furthermore, data quality can be enhanced by gaining users' trust, via communication about any potential issues, as well as inviting their feedback and contributions. Finally, sharing this data openly and publicly, with clear licencing information, maximises its potential audience.

In discussing participants' comments relating to both tool/resource usability and data usability, several themes have emerged, each of which draws on findings from one or more of the topics explored in this chapter:

- Training
- Collaboration
- User-centred design

- Documentation
- Access
- Sustainability

Each of these themes represents a key consideration during the production process, ensuring that the resulting Linked Data, and the tool or resource through which it can be accessed, are usable by Ancient World researchers. As such, the following chapter will explore participant responses relating to the above six themes and their application to Linked Data production.

## 7 Producing Usable Tools and Resources

In the previous chapter, I focused on specific measures that might be taken to ensure, or improve, the usability of Linked Ancient World Data tools and resources. However, in doing so, it became clear that their usability was also significantly affected by decisions made during, and even before, their production process. As a result, I identified six factors to explore in more detail. These factors are based on themes that emerged throughout my discussion, and which relate both to the usability of digital tools and resources and to their underlying data. By the end of Chapter 6, it had become clear that exploring these six themes further, in terms of their application to the production process, could be crucial in developing recommendations to improve Linked Ancient World Data usability (RQ4).

This chapter is structured according to the order in which these six themes might be expected to occur during the production process. I will start by discussing factors that affect the inception of digital projects, including participants' experiences of digital skills training (7.1), as well as collaboration with other researchers and technical experts (7.2). To continue, I will turn to the production process, by looking at key aspects of user-centred design (7.3) and documentation (7.4). Finally, I will consider the publication and continuation of the resulting tools and resources, by exploring issues such as the relationship between cost and access (7.5), and long-term sustainability (7.6).

### 7.1 Training

Throughout my study, many participants commented on the training they had (or had not) received and how this affected their production of digital tools and resources. Researchers' training and skills can have a significant and lasting impact on any digital tools or resources that they produce. However, participants generally felt that they lacked digital skills training, with those more digitally competent/confident participants being predominantly self-taught. PART005 additionally observed that it is easier to build upon existing skills than to develop new ones, acknowledging that their previous experiences affected their choices in terms of digital tool and resource use and production.

As I mentioned in 2.4, lack of training is already a known barrier to Linked Data production, an area where participants considered training to be vital to achieve an effective outcome. For example, in my survey, PART008 specifically mentioned their difficulty in understanding RDF, while those participants who had received appropriate training found that it increased their confidence to produce Linked Data. Training they had attended included the Digital Humanities Summer Institute (DHSI)<sup>149</sup> (PART119), Ontotext<sup>150</sup> and Sunoikisis DC<sup>151</sup> (PART109), as well as ‘hack days’ organised during the early stages of the *Pelagios* (2.2.1) project (PART054). PART054 and PART109 additionally spoke about their experiences at the Digital Humanities at Oxford Summer School (DHOxSS)<sup>152</sup>. Their comments were predominantly positive; however, both felt that there was insufficient time dedicated to the practical implementation of Linked Data. PART109 suggested that building up a single case study over the course of the week might have been more effective training than conducting a series of distinct, unrelated tasks. Receiving training based around a real-world use case should therefore encourage researchers to consider a Linked Data approach when planning a new digital tool or resource, while ensuring that the resulting output is usable by others.

However, there are several barriers to attending such courses. Although participants did not comment on the matter, the most effective, comprehensive, training tends to be delivered via intensive residential courses (e.g., DHOxSS), which are likely to be prohibitively expensive to those researchers or institutions with less funding. As an alternative, PART054 spoke about using the free version of a Massive Open Online Course (MOOC) but found it difficult to maintain momentum as it largely involved setting aside time to watch videos, with no interaction from the course leaders. Time itself was another barrier to training, mentioned by several participants in the survey as well as PART012 in interview. They often felt compelled to prioritise their immediate needs over potential long-term benefits, i.e., using their existing skills rather than developing new ones. Even PART109, who had received training in Linked

---

<sup>149</sup> <https://dhsi.org/>

<sup>150</sup> <https://www.ontotext.com/services/semantic-technology-trainings/>

<sup>151</sup> <https://www.dh.uni-leipzig.de/wo/sunoikisisdc/>

<sup>152</sup> <https://digital.humanities.ox.ac.uk/digital-humanities-oxford-summer-school>

Data production, appreciated that the time investment might deter other researchers from doing the same.

Other participants commented on the lack of training opportunities available to them, often implying that the onus should be on institutions to provide more digital training, which would make it easier for them to identify and access relevant courses. However, such training was often provided on an ad hoc basis, rather than as a central part of their professional development. In several cases, participants (such as PART254) felt that their only option was to teach themselves, while appreciating that more formalised training would ultimately have been more beneficial to them. Where training was available, it tended to focus on the familiar formats of tabular or relational data, which (as PART061 remarked) requires a relatively low investment of time for participants to develop sufficient understanding to apply what they have learned. As a result, researchers tend to favour these types of data structure when developing their own tools or resources, as found in 4.1.3.

To address this situation, participants including PART008, PART012 and PART109 recommended that more comprehensive digital training should become an integral part of postgraduate, or even undergraduate, Humanities degree courses. They suggested that incorporating training in this way would increase future researchers' digital confidence and introduce different approaches to data modelling, including Linked Data, as well as basic programming skills. However, PART012 also acknowledged the potential difficulty in implementing such initiatives because Ancient World researchers with sufficient digital skills are still quite rare. To mitigate this issue in the short term, they suggested that openly available training resources should be created, although they were unclear about who might be responsible for their development and maintenance. Such resources do indeed exist, in the form of *The Programming Historian*<sup>153</sup> and *SunoikisisDC*<sup>154</sup>; however, the former was not mentioned at all and the latter was only mentioned by one participant, indicating a lack of awareness of these resources among the wider Ancient World research community.

---

<sup>153</sup> <https://programminghistorian.org/>

<sup>154</sup> <https://www.dh.uni-leipzig.de/wo/sunoikisisdc/>

In summary, usability of digital tools and resources, particularly those involving Linked Data, can be affected by the level of digital training that producers have received. Although various training courses and initiatives are available, participants generally felt ill-equipped in this area due to a lack of awareness of existing such resources, and with limited time to seek out new opportunities. The need for training is particularly apparent when considering possible data structures, resulting in participants choosing to use more familiar approaches rather than learning about new ones, such as Linked Data. One possible strategy to avoid this situation in future is by introducing digital skills training at an earlier stage, e.g., as part of the undergraduate curriculum. Such training might most effectively take the format of a continuing project, to allow practical application of new skills developed. I will return to the topic of digital training for Humanities researchers as part of my recommendations, in 8.1.

Another potential solution is by forming collaborations between researchers with complementary skillsets, which I will discuss in the following section.

## 7.2 Collaboration

Throughout my literature review in Chapter 2, I found that collaboration tends to be a common (or even intrinsic) feature in the production of effective and usable Linked Data tools and resources. Working collaboratively provides more opportunities for researchers to learn from each other and fosters sustainability (7.6, below) by avoiding a single point of failure. Such collaboration need not always necessitate a formalised arrangement as part of a large project but might simply consist of researchers and practitioners providing each other with mutual support and advice to ensure that the tools and resources they produce are as effective, usable, and sustainable as possible.

In my study, each producer participant was involved in a project of differing scale, ranging from individual initiatives, to small voluntary teams, to major international collaborations. Survey responses showed that producer demographics (Figure 7.1) were broadly similar to those of the entire survey population (3.4), albeit with some slight discrepancies. However, there was a noticeable increase in the proportion of male participants (from 45% to 57%) and a decrease in the proportion of female

participants (from 50% to 38%) (Figure 7.1d). Additionally, male participants generally had a higher degree of confidence in their digital abilities, with a median digital competence/confidence score<sup>155</sup> of 4.14 out of a possible 5, compared to 3.57 for female participants. While these results might appear to suggest a greater aptitude or interest in digital production among men, they might instead reflect that men are more likely to demonstrate greater confidence in their abilities by overestimating their competence. With women more likely to take the opposite approach, fewer opportunities are made available to them, potentially due to their abilities being underestimated.

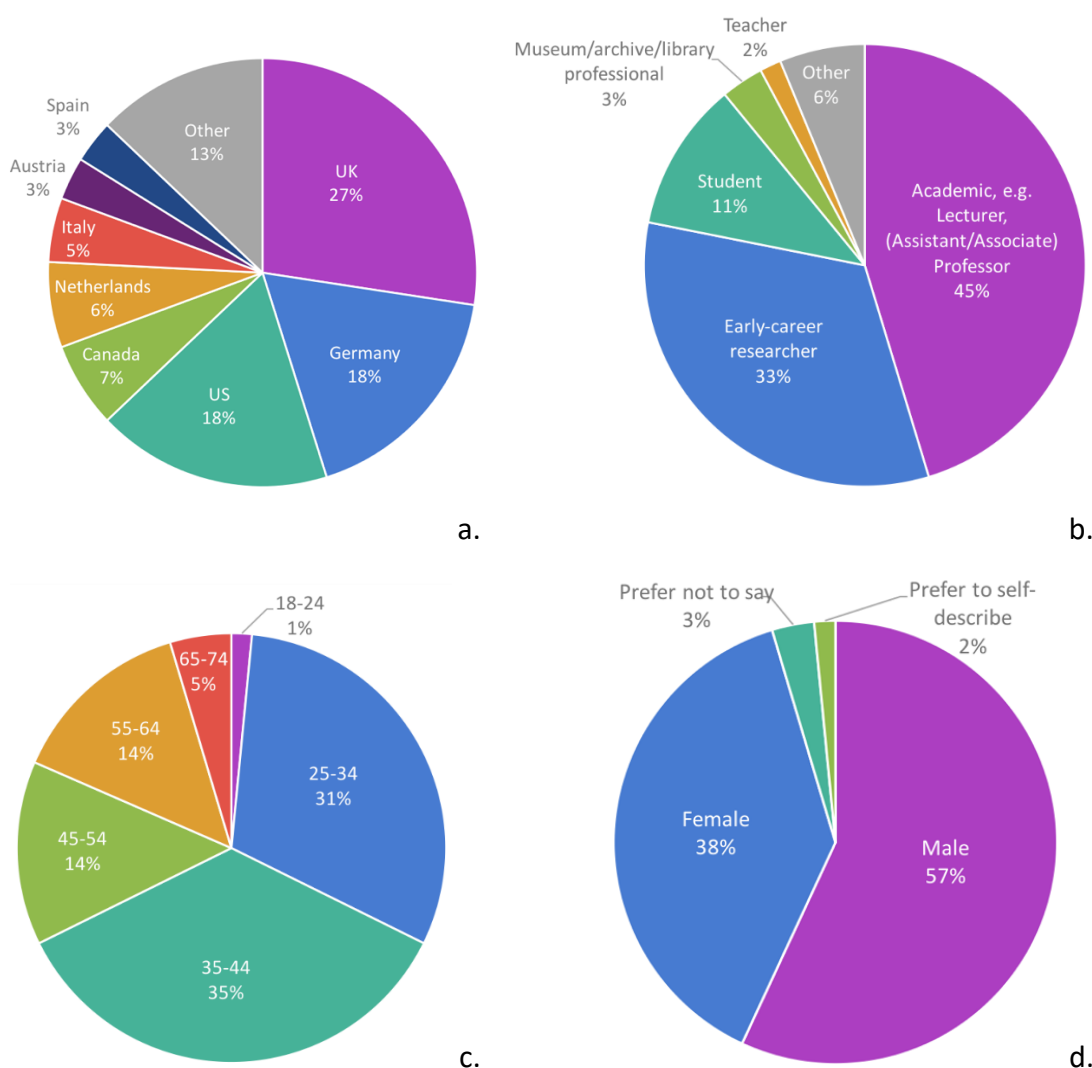


Figure 7.1 Pie charts showing producer demographics (a. location, b. role, c. age, d. gender)

<sup>155</sup> In 3.4, I calculated a digital competence/confidence score for each participant by converting their level of agreement with each of seven statements to a score between 1 and 5, then taking the mean of these scores. The medians referred to in this section are the medians of these mean scores.



Several participant comments indicated that this is often, indeed, the case. Of the female producers I interviewed, the majority had produced their digital tools and resources by working alone, with only PART260 having been involved in a large-scale collaborative project. She spoke about how her technical skills were persistently underestimated on this project due to her gender, an experience that tarnished her views on the Digital Humanities as a whole:

*"I assumed that I would have lead researcher and a great deal of ownership over the project... and I feel like I was extremely underutilised... I think this was incredibly gendered. I told them as soon as I came on the project that I could do more work than they were having me do. They would not let me edit the main database, they gave me a shadow database that [colleague] was required to check before it was integrated... I am shocked and horrified by the gendered politics of how I was treated on this project and how little my expertise was valued."*

*"It can be a gendered space and there can be politics... I don't assume that male... Digital Humanists will be my ally."*

Experiences like that of PART260 can act as a significant deterrent to becoming involved in future digital projects, potentially contributing to the above gender imbalance in tool and resource production. Such a lack of representation deprives teams of important insights from members of the research community that might have led to usability improvements for a wider audience.

Although some participants worked with their collaborators face-to-face, much work also took place remotely, a trend made possible through the use of digital technologies, which (since my research took place) has been accelerated by the COVID-19 pandemic. Such remote collaboration expands the scope for working across organisational and geographical boundaries, and facilitates the development of volunteer-run projects, such as the *Classical Language Toolkit (CLTK)*<sup>156</sup>, to which

---

<sup>156</sup> <http://cltk.org/>

PART119 is a contributor. They found *CLTK* to be particularly helpful in bringing developers together by implementing *GitHub*<sup>157</sup>, which facilitates code sharing and version control. However, PART255 found that, even when using such platforms, remote development must be managed effectively to avoid versioning issues or file conflicts, while PART078 felt that working remotely had the effect of isolating them from their team. Both these latter participants have therefore identified areas where remote collaborations might require particular attention to ensure their effectiveness.

Several participants spoke about working with developers, who were usually either technical professionals or Computer Science researchers. Some projects included a combination of domain experts and developers, with participants highlighting the importance of multidisciplinary teams that value both subject and technical expertise. For example, PART061 is a computer scientist who finds it most effective to work as part of an equal partnership with Humanities researchers, in a process of co-development. Similarly, PART041 found that allowing considerable time for meetings and discussion between all parties was extremely beneficial to the resulting resource. These findings echo those of Barker et al. (2011, p. 16, 2012, p. 188) that such projects often work most effectively when Humanities and digital experts are embedded in the same team structure, rather than the investigative team consisting solely of Humanities researchers, with the technical experts considered as service providers who might not be sufficiently informed or equipped to offer appropriate support.

For researchers and developers to form a successful collaboration, it is crucial for all members of a team to understand and appreciate each other's goals, ways of working and how potential solutions might best be realised. PART119 spoke at length about their experiences of working with a web development firm who implemented user-centred design (7.3, below), an approach they were keen to apply to future projects. Although the company was commissioned to carry out the work, rather than being fully integrated with the project team, their way of working was collaborative and responsive: they worked with the researchers to establish the most effective way of meeting potential user needs, rather than simply delivering on the original

---

<sup>157</sup> <https://github.com/cltk/cltk>

specification without further consultation. PART061 found that there can be a steep learning curve associated with this way of working, sometimes with considerable time required to build a sufficient level of understanding between all parties, which can lead to a period of uncertainty about the resources required. This additional time commitment might be off-putting to some researchers, or might cause others to assume that it will result in unnecessary delay (even if it ultimately saves time over the course of the project), particularly when justifying the time and resources required in a funding application.

Not all participants worked collaboratively, however, and expressed differing views about the impacts of working alone. PART012 felt that a lack of collaborators and funding provided them with greater freedom and control over their project, as well as more opportunity to enjoy the work, although they did suggest that a greater level of support would be preferable. In particular, they found that the IT professionals in their institution did not have the specialist knowledge to support complex Digital Humanities projects.

The need for support was particularly highlighted in relation to projects aiming to implement Linked Data in a usable way, with PART109 speaking about the importance of receiving such support from their supervisor. However, in most cases, such support seemed to be lacking, which particularly affected participants working alone or in small teams. For example, PART037 mentioned the potential benefits of working with a Linked Data specialist, to ensure that their resource is robust and to advise on possible improvements. However, they found it difficult to secure any formal collaborators without funding, because everyone they approached was already very busy with their own projects.

In PART008's case, despite knowing about Linked Data technologies and recognising their potential in the context of their project, the production team were pragmatic about their own capabilities and, being unable to secure external support, reluctantly chose a relational database approach. Lack of support can therefore affect decisions about which data structure(s) and format(s) to implement, potentially limiting researchers' choices to more familiar tabular or relational structures, as found in 4.1.3.

To assist in such cases, PART061 suggested projects that had already implemented Linked Data (or other digital technologies) in a Humanities context, should share best practices and lessons learned. Publishing these experiences openly and transparently might therefore be a sustainable way to share knowledge and facilitate the production of more usable resources in future. As such, I will return to the idea of knowledge sharing as part of my recommendations for communities of practice in 8.7.

In this section, I have discussed participants' experiences of a collaborative approach and the associated benefits it provides to digital projects, particularly regarding sustainability. Although some advantages to lone working were identified, responses generally indicated that to produce a usable digital tool or resource, some element of collaboration was preferable, even on an informal basis of providing support. Such support was felt to be particularly important where Linked Data was concerned, due to unfamiliarity with the relevant technologies. I also found that the most effective collaborations are inclusive and diverse, ensuring that contributions from all members are valued equally. In addition to the under-representation of female producers, this point additionally relates to the integration of technical experts within a project team. Although many such collaborations are formally associated with particular projects, tools, or resources, collaboration also manifests itself in informal support arrangements and knowledge sharing, which might subsequently form the basis for more formal arrangements. Collaboration additionally links to the theme of training (7.1), in that it can bring together researchers with complementary skills from taking different training pathways. As such, I will return to both topics in my recommendation for effective teams (8.1).

As identified by PART119 above, collaboration can facilitate the implementation of a user-centred design process, which forms the subject of my next section.

### 7.3 User-Centred Design

As suggested in 6.1, implementing user-centred design can be an extremely effective strategy for ensuring the usability of a digital tool or resource. User-centred design is also sometimes called 'human-centred design', defined by the International Organization for Standardization (2019) as an *"approach to systems design and*

*development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques".* In practice, a user-centred design approach seeks to identify what users need from a tool or resource, as well as testing their responses to it. Several participants had positive experiences of incorporating elements of this approach during production of their tool or resource, with others expressing the intention of doing so in future.

At the outset of the development process, one user-centred design method involves identifying pathways that different users might take through a tool or resource to achieve different goals. For example, in 7.2, I mentioned PART119, who had worked with a web development company that employed a user-centred design approach. This participant found that implementing such techniques from the start of a project can result in significant differences to the user experience than if the tool or resource is based solely on the production team's personal opinions and experiences, ensuring wider usefulness and usability. For example, their initial approach to presenting content and functionality was to "put everything in one place", until the web developers prompted them "to think more about how a diversity of people would go through the interface". These user journeys can be identified in different ways; during my research, both audience-focused and task-focused methods were mentioned.

My initial approach when constructing the survey and interview questions was to focus on different audiences, asking Linked Data producers to select the intended audience group(s) for their tool or resource from options including other researchers, students, cultural heritage practitioners, teachers, and the "general public". However, I found that the latter term tends to be very loosely defined, with little clarity on specific user needs. As demonstrated in 4.1.3, above, Linked Data producer participants usually intended for their tools and resources to be used by multiple audience groups, although there appeared to be a general consensus that academic tools and resources would predominantly be intended for academic audiences.

When discussing this topic in more detail during the interview phase, some participants found it most effective to start with one primary audience, then consider

others if there is sufficient time and/or funding available. For example, PART061, whose annotation, visualisation and discovery tool is described in Research Methods Case Study 2 (5.6.2), primarily focused on academic researchers, although the eventual aim for usability of their tool by a wider audience influenced some of their development and data modelling decisions. PART037, on the other hand, took the approach of focusing on a specific (academic, technically skilled) demographic, while attempting to source collaborations that would facilitate future developments aimed at a wider audience.

While the definition of specific audience groups before consideration of their potential goals can be very effective, PART041 suggested that more usable results might be achieved by the reverse approach, i.e., first identifying user goals, then suggesting potential audiences to which they might apply (I gave the example in 5.1.1 of how they applied this approach to the search functionality of their own digital resource). The participant went on to state that this resource is still relevant and useful for these different purposes, years after its development. Similarly, PART119 established user goals by including the content editors for their resource in discussions about defining the tasks they would need to carry out, as well as features they would expect to be available to them.

Several participants spoke about conducting user testing during development of their tools and resources. PART255 planned to incorporate workshops into their development process, where their "media team" would film and observe user interactions and gain feedback. PART037 found that obtaining such feedback at an early stage can help to identify barriers that affect tool or resource usability, as well as identify any implicit assumptions they might have made about user understanding, which could be addressed by providing clearer information. For tools or resources aimed at multiple audiences, participants found it helpful to ensure that representatives from each audience group took part in the user testing; for example, PART054 tested their resource with both academics and members of the public. In other cases, participants gained high-level feedback through informal conversations with colleagues, with PART061 finding that many were interested in their resource's capabilities for demonstrating the benefits of Linked Data.

Where user testing was not included in the project plan from the outset, some participants found that it became an afterthought, with little opportunity to implement the resulting feedback before the end of the project. For example, PART119 focused on ensuring their resource was ready for presentation at a conference, with insufficient time to conduct user testing, or scope within the project plan. Similarly, PART054 expressed regret that their user testing had been rushed, as the positive feedback they initially received gave way to complaints about usability issues and gaps in the accompanying documentation. They subsequently suggested that a more thorough user testing process might have identified and resolved these issues during development.

The above user-centred design techniques discussed by participants included identifying user journeys and conducting user testing. While some took the approach of speculating on potential user demographics before anticipating the tasks they might perform, others suggested that a task-centred process might be more effective; as such, I will explore this idea further when making recommendations about user consultation in 8.4. Several participants found the process of user testing to be extremely beneficial, while highlighting the need for sufficient time to implement improvements.

Although a user-centred design approach is a reliable way to ensure the usability of a tool or resource for its intended purposes, it can be complemented well with appropriate documentation, the topic of my next section.

## 7.4 Documentation

Documentation, comprising information about and instructions for the use of digital tools and resources, is often key to their usability. In my original survey, clear documentation was the second most popular feature of good tools and resources, after ease of use/installation, while the lack thereof was the sixth most frequently mentioned barrier (4.1.1). Documentation generally tended to be mentioned by more participants who had knowingly used Linked Data than those who had not and appeared to be more important to Linked Data users than ease of use/installation. In

the follow-up survey, completed by interview participants, clear documentation seemed to be mid-priority for Linked Data users and low priority for non-Linked Data users, suggesting that its presence was considered less important than that of other factors, such as ease of use, reliability, or open access. However, unclear or non-existent documentation remained a significant barrier for both groups, which indicates that its absence is noticed, particularly if the tool or resource is unreliable or difficult to use.

Documentation exists in different forms. Firstly, it can include information about the tool or resource and its underlying data, making explicit its remit and any limitations. Including such information enhances users' trust in the data, as well as managing their expectations in terms of its quality, as demonstrated in 6.2.2. Secondly, documentation can include usage instructions, user guides, or tutorials, which might contain details about how advanced searches can be performed, e.g., by providing example queries, or explaining which Boolean operators are accepted, as detailed in 5.1.1. Thirdly, use cases can serve to demonstrate the potential of a tool or resource, providing examples of how it has been used by others, as discussed in 6.2.3. Including these types of documentation might negate the need for specialist training in order to use the tool or resource, the second most frequently mentioned barrier in 4.1.1.

When encountering a digital tool or resource for the first time, most interview participants preferred to try using it straight away, before looking at the documentation. There were differing levels of expectation among participants about how successful this strategy was likely to be; however, most expected to be able to access and use sufficient documentation to identify and, ideally, resolve any issues they encountered. PART008, however, was an exception to this rule. They stated that they check the documentation first, then read it to assess comprehensiveness, trustworthiness, and currency. If the documentation is lacking in these areas and there is no clear way of obtaining clarification from its producer(s), the participant does not then attempt to use the tool or resource as they do not consider it to be a worthwhile investment of their time. Similarly, PART037 said that they could sometimes work out how to complete a task without documentation if they spent enough time, but that they often do not have this time available as other tasks take priority. Several



responses to this question implied that some participants assumed that consulting the documentation first was the 'correct' thing to do and seemed almost guilty when admitting that they operate differently. However, if basic functionality can be accessed easily via an intuitive user interface, this should reduce the need to consult documentation in the first instance, until the user requires detailed information or seeks to perform more complex operations.

Furthermore, responses revealed an unwillingness among several participants to consult documentation at all. Those with less technical experience, such as PART017, PART089 and PART254, often assumed that documentation would be too long and difficult to understand, suggesting previous confusion resulting from overly technical language. It is possible that these responses arose from a misunderstanding of what I meant by my use of the term 'documentation' during the interview phase. However, even participants with higher levels of digital competence/confidence, such as PART037, sometimes found that tool and resource producers had made incorrect assumptions about users' understanding and computing environments. PART109 and PART254 additionally commented that they did not have sufficient time available to consult documentation in detail, and that if it took too long to find the answer to their question, they would stop using the tool or resource.

For those participants who had been involved in producing digital tools and resources themselves, some had produced associated documentation, while others had not. The former participants, such as PART008 and PART061, intended for their documentation to assist less technically experienced users, as well as to encourage contributions from the wider community (discussed in 6.2.2). To ensure that documentation makes sense to others, PART109 recommended a focus on communicating to the user, with PART037 advising that it should be written collaboratively where possible (or to at least ask other people for feedback before publication). PART061 suggested that such documentation might most effectively be written by researchers whose interests and skills align closely with the potential users of the resource, rather than technical experts, to ensure effective communication with the target audience. PART037 additionally mentioned the importance of including comments when writing code, ensuring that all variables are defined. To improve the speed and accuracy of

documentation production, as well as managing its maintenance over time, PART037 also found it helpful to automate some elements that were reproduced in multiple places. Such an approach avoids the confusion that arises from outdated documentation. However, the participant acknowledged that human verification and clarity checking were still required.

Participants such as PART037 and PART109 found that creating documentation as an integral part of the production process tended to be most effective, as it can prompt changes that result in a more usable tool or resource. However, several said that they did not enjoy the process of producing documentation and that, in practice, it was often left until the end of a project or was not produced at all. For example, PART054 rushed their documentation towards the end of a fixed term project, while under pressure to upload more data to the resource, producing substandard, incomplete results. While managing projects in this way can have negative implications for usability (as in the above example), PART119 emphasised that their development approach minimised the need for documentation, by incorporating instructions within the interface and generally ensuring ease of use throughout. However, they did acknowledge the importance of providing some separate documentation, such as provenance information or instructions for more complex operations.

Clear documentation can mitigate usability and reliability issues, while managing user expectations, as well as reducing the time taken to complete specific tasks, or the need for specialist training, thereby breaking down real or perceived skill barriers and widening the potential audience. As such, I found documentation was considered important by participants of all levels of technical experience. However, it was valued particularly by those who tend to perform more complex operations, with participants usually consulting documentation only when they had a specific need. The extent of documentation might therefore be reduced by ensuring ease of use via an intuitive interface, particularly for relatively basic operations. Some participants had a perception that documentation would be difficult to understand, which could indicate a lack of clarity in existing tools and resources. Producers suggested that this issue might be mitigated by writing documentation as an integral part of the development process and ensuring it is comprehensible by the intended users before deployment. I

will continue discussing documentation as part of my wider recommendation for openness and transparency in 8.2.

Having explored several factors relating to digital tool and resource planning and production, the next consideration I will discuss is that of access restrictions and associated costs.

## 7.5 Access and Cost

Several usability issues identified in the preceding chapter relate to the theme of access to digital tools and resources. As mentioned in 6.1.1, the word ‘accessibility’ often appeared in survey and interview responses, usually with the implication of the FAIR data (1.2) definition of the word, specifically principles A1 and A1.1, i.e. data should be available via a "standardized protocol" that is "open, free, and universally implementable" (Wilkinson et al., 2016, p. 4). Accordingly, for many participants, access to a tool or resource depends largely on whether there is an associated cost. Cost was the barrier identified by the largest number of participants in my original survey, the majority of whom were not Linked Data users, as supported by the follow-up survey findings (4.1.1). Although several participants referred to software licences, hardware and storage, most responses mentioning cost related to online journal or database subscriptions, with PART049 and PART090 referring specifically to *TLG* (4.1.1). These responses indicate that subscription-based resources such as *TLG* are widely used by those who can access them, while those who do not have access to a subscription feel disadvantaged.

Participants usually expected their institutions to pay for such subscriptions, with PART078 and PART089 stating explicitly that they would be unwilling to pay themselves; many tools and resources that adopt the subscription model are therefore likely to be missed by researchers who do not have access via an institution. Several mentioned the difficulties they had experienced in requesting institutional access to a new resource, appreciating that subscriptions must necessarily be limited due to budgetary constraints. For example, PART041 explained how their user experience with the *Jacoby Online*<sup>158</sup> database (of fragmentary Ancient Greek historians) was

---

<sup>158</sup> <https://brill.com/view/db/bnjo>

compromised as they were only able to access selected articles in PDF via interlibrary loan. These limitations had a direct impact on the type of research they were able to conduct, as they were unable to discover materials via keyword searches on the resource itself. PART041 went on to speak very positively about the open access nature of many digital numismatic tools and resources. The benefits that this openness has provided, including the ability to make new connections and discover new research areas, as well as to contribute to a community, has led their research to focus increasingly in this area. Openness (or, conversely, cost) of digital tools and resources can therefore have a significant impact on which topics a researcher is able to study.

Those whose institutions have more generous budgets for online resources often did not consider cost to be an issue because they had not experienced such barriers, although PART043 expressed concern about losing their access if they had to leave their job. Some participants talked about paying for subscriptions themselves but both PART008 and PART043 noted that unless a tool or resource is free of charge, costs tend to be prohibitively expensive for an individual researcher; there does not seem to be an affordable 'middle ground'. PART119 considered paying for access to a tool or resource as an investment that they were not prepared to make without first being able to test its suitability via a free trial.

Reliance on subscription-based tools and resources tended to be more problematic for participants whose work was not heavily computational, e.g., those whose main digital activities involved searching primary and secondary literature databases. Conversely, cost tended to be a lesser concern for Linked Data users due to the proliferation of freely available tools and resources. Several of these participants, including PART001, PART061 and PART109, said that most tools and resources they use have no cost associated with them, or that there are usually free of charge alternatives available to allow them to avoid using paid tools and resources altogether. PART078 found that the availability of open access resources is increasing, resulting in usability improvements. The participant suggested that this shift might be due to producers recognising that their tools and resources might be useful to a wider audience, rather than restricted to a niche group of researchers.

As discussed in 6.2.3, the first star of Berners-Lee's five-star model (1.3) requires Linked Datasets to be openly available, a principle advocated by most producer participants. However, development and maintenance of these resources does come at a cost to the producers. At the outset of a project, initial development costs are often covered by funding, but (as I will discuss in more detail in the following section), when this funding expires, alternative sources of income must be sought to cover the cost of hosting, maintenance, and further improvements. As a result, licence or subscription fees may be introduced. One example is *Trismegistos* (2.2.4), whose data continue to be openly available, but whose visualisation features can now only be accessed via a subscription. Although the producers had considered every available alternative (Depauw, 2019), it remains the case that an access barrier has been imposed on the very functionality that was intended to make the underlying data more usable to those with minimal technical skill. If similar measures are taken elsewhere, this is likely to restrict the usability of Linked Ancient World Data to a relatively small audience of technically experienced researchers. The decision to charge for access to these features is associated with the related issue of sustainability, which I will discuss further in 7.6.

Survey and interview comments demonstrate that cost posed a significant barrier to many participants, resulting in inequality of access to key digital tools and resources, which particularly affected those with less interest in performing complex technical operations. The inability of individuals and institutions to pay for access has the effect of limiting tool and resource usability, which in some cases can restrict the research directions that an individual is able to pursue. Conversely, tools and resources intended to be openly available to a wider audience often tended to be more usable.

The interlinked factors of access and cost will feature as part of my openness (8.2) and sustainability (8.6) recommendations, with sustainability itself being the final theme for discussion in this chapter.

## 7.6 Sustainability

I have mentioned sustainability at various points throughout my discussion, due to its impact on the long-term usability of a tool or resource. Similarly, ensuring usability

increases uptake, thereby supporting the case for sustainability into the future. Sustainability (or lack thereof) affected many participants from both a user and producer perspective, with responses demonstrating that while multiple potential ways of sustaining digital tools and resources are available, there is no clear solution for how this might be achieved in a consistent and reliable manner.

In terms of the underlying data, use of open standards, such as those associated with Linked Data, promotes sustainability through continued interoperability with other tools and resources, as discussed in 6.2.1. Some participants, such as PART054 and PART109, therefore chose to archive this data in trusted repositories for future consumption by others, a practice that has increased following funder mandates to preserve research data. While such measures are critical for ensuring long-term data availability and preservation, they do not ensure sustainability of a usable tool or resource through which that data can be accessed and explored by researchers with a range of technical skill levels. The remainder of this section will therefore focus on participant responses relating to tool and resource sustainability, rather than data preservation.

Several responses mentioned the ephemerality and precarity of digital materials. As a result of concerns that their preferred tools or resources would cease to exist, PART017 and PART041 made their own backups by printing, taking screenshots, or copying datasets to their own storage. However, existence does not in itself denote sustainability. For example, PART008 and PART041 identified usability barriers that occur due to digital tools and resources seemingly being abandoned by their creators, without being kept up to date with new technological developments or providing a reliable point of contact. As PART012 stated, ongoing maintenance is required to promote continued usability, e.g., upgrading to ensure compatibility with newer browsers or devices. PART061 recommended planning tool or resource development to ensure that minimal maintenance will be required, thereby reducing the potential need for external technical support.

In some cases where digital tools or resources continued to be maintained, their producers gave little consideration to the impact of any changes on external sources

that made use of their data. Survey participant PART104 stated that this issue particularly affects the usability of Linked Data tools and resources: if the URIs from one dataset are not stable or persistent (6.2.1), any links to them from other sources become broken, and usability suffers as a result. Even when there is some warning that such changes are going to happen, this can significantly increase the maintenance required to update these third-party tools and resources, which might not be practical within their budgets or timescales.

To consider how such issues might be addressed, I asked producer participants about their strategies for ensuring sustainability of their own digital tools and resources. Responses indicated that, like the other factors discussed in this chapter, sustainability should be planned from the outset of a digital project. Several participants spoke about the need for balance between developing a tool or resource to address specific research questions, while also ensuring that its scope is sufficiently broad to encourage further uptake beyond the original project. As demonstrated in 5.4 above, PART255's project became increasingly specific over time, resulting in a resource with limited usefulness in other contexts. PART008 suggested that, to truly ensure this more general usability, the implementation of Linked Data is crucial – an approach used by all participants in the following examples. Of these, PART061 ensured that potential for wider usability was incorporated into their project specification from the start, while PART037 and PART109 recognised that their resources might be applied in different contexts later on, making the necessary alterations accordingly. PART054, however, cautioned that their project had been too concerned with such flexibility, with the result that their resource had become so generic as to obscure its original topic and purpose.

The main form of financial support received (or sought) by producer participants was that of research grants. These grants tend to be fixed term and are usually geared towards the production of a digital tool or resource alongside more traditional publications, with little consideration of what might happen to it after the funding ends. Indeed, PART061 expressed concern that funding bodies were apparently focusing more on preservation than sustainability: "Most of what I have seen so far is that preservation means archiving and archiving is... obviously not a way to... keep a

tool sustainably running". PART041 suggested that funding bodies will eventually need to introduce specific grants for updating these "dead" or "dying" digital tools and resources. However, both they and PART061 acknowledged that decision-makers are likely to consider this a less attractive prospect than the more explicitly innovative work of building new tools and resources, particularly if existing uptake is low. As a result, securing maintenance funding is rare and is often only a short-term solution. Subsequently, as PART012 observed, many digital tools and resources fall into obsolescence as their producers move on to the next funded project, with some continuing to be maintained on a voluntary basis, as discussed below.

The time pressures of fixed-term projects additionally mean that if any part of the process takes longer than anticipated, there is an impact on the final output. Often, as indicated above, the aspects that are most affected are those that would have most effectively facilitated its usability, such as interface design (6.1.1), user testing (7.3) and documentation (7.4), which are often considered secondary to data completeness and functionality. These issues lead to tools and resources that are difficult to use by anyone outside the project team, a situation that (as mentioned above) contributes to lack of uptake, reducing the likelihood of securing further funding.

These experiences indicate a need for funding bodies, institutions, and researchers to move away from the idea of a 'complete' digital tool or resource. Indeed, participants such as PART008, PART037, and PART109 described how they never considered the development process to be over, but instead identified new ways in which their tools or resources could be improved. New digital tools and resources should therefore be considered as investments that can be reused by other researchers, and which must be maintained and developed to ensure their continued usability. Together, these tools and resources could form an ecosystem upon which future developments might build. A Linked Data approach has the potential to be particularly effective in this context; I will therefore discuss this in more detail as part of my recommendations on extensibility in 8.3.

On a related note, some participants spoke about ensuring a smooth transition between research projects and established tools or resources, by specifying a clear



pathway from grant funding to institutional support. PART061 recommended that funders stipulate that digital tools and resources must be maintained by their institutions for a set amount of time following the end of a project, which should, in turn, result in those institutions investing in more robust technological infrastructure. PART012 agreed, while acknowledging that it is unclear how institutions might secure sufficient funds to provide such a service to all digital projects.

Some participants had indeed been successful in securing institutional support to maintain their digital tools and resources. PART054 had worked with a Digital Humanities lab, who have continued to host their resource beyond the five years originally agreed because it shares the same technical infrastructure as other resources that they manage. PART119 had an agreement from their own institution to provide hosting in perpetuity but would require more funding for any further development work. PART061 had secured fixed-term hosting from their institution, with potential for further, targeted funds for development, based on uptake of the tool. Therefore, rather than committing the institution to hosting such tools and resources in perpetuity, which would gradually consume more and more funds, they are taking a pragmatic approach based on usability and usefulness, which could encourage future producers to prioritise these aspects in subsequent projects.

When discussing the question of ongoing maintenance and hosting, some participants thought that a national initiative might be a more appropriate solution. For example, PART012 suggested that a national body, such as Jisc<sup>159</sup>, might provide these services for digital tools or resources that can no longer be supported elsewhere, on the condition that they meet certain technical criteria. The participant likened these criteria to those stipulated by online app stores, but they might also include recommendations that such tools and resources should utilise technologies that will facilitate their maintenance, such as open standards. However, PART001 mentioned that even national initiatives can be precarious, giving the UK's *Archaeology Data Service (ADS)*<sup>160</sup> as an example of an infrastructure that used to be state-funded, but this is now no longer the case. Instead, the *ADS* now charge organisations to deposit

---

<sup>159</sup> <https://www.jisc.ac.uk/>

<sup>160</sup> <https://archaeologydataservice.ac.uk/>

their data, with the University of York committing to preserving it (Archaeology Data Service, 2019).

Where funding is not available, several participants spoke about alternative ways through which the maintenance of a digital tool or resource might be achieved. PART037 gave the example of *TEI* (1.2) as a community-driven initiative that is independent of individual funders or institutions. Setting up such a community requires considerable engagement from a wide user base, as well as a clear focus, but could be the ultimate sustainability goal. As such, this approach has been taken by several of the initiatives discussed in earlier chapters, including *Pelagios* (2.2.1) (Kahn et al., 2021), as well as *CIDOC CRM* (Bruseker et al., 2017, p. 111) (1.3) and *EAGLE* (Santucci et al., 2016) (2.2.4). Subsequently, PART037 had made their own attempts at community-building by asking researchers from other projects to incorporate their format and become involved in its development. Similarly, PART109 spoke of their intention to raise awareness of their resource by promoting it to other projects and institutions.

Other means of ensuring long-term sustainability of a digital tool or resource, not mentioned by participants, include raising funds through donations or subscriptions. In terms of the former, *Papyri.info* (2.2.4) opened a call for donations to cover the costs of appointing a permanent member of staff for its maintenance (Torallas Tovar & Schubert, 2019). As regards the latter, such subscriptions might be required to use the tool or resource at all, or they could form a two-tier system where basic access is free and a fee is charged for additional functionality, as in the case of *Trismegistos* (7.5). However, as demonstrated above, cost was considered by participants as a significant barrier to use; therefore, introducing a subscription is likely to impede usability by those with limited financial means and cause them to instead seek free alternatives.

Many tools and resources, however, ultimately rely on individuals or small teams working on a voluntary basis. Two participants spoke about conducting digital projects outside their 'day jobs', rather than as an integral part of an academic/research role. PART012 found that this way of working can affect the pace of development and maintenance, particularly as their career developed and they gained more

administrative responsibilities. They also found it difficult to raise the funds required to cover basic hosting and software costs, which often relied upon applying for multiple small-scale grants. PART037 mentioned that their ideal scenario would be securing a position at a research institution where they would be paid to develop and maintain their resource, and where they might be in a better position to start building a community around it. Finally, both participants acknowledged that if such projects remain exclusively in the hands of individuals, they have an inherent single point of failure. Indeed, PART037 expressed concern about what would happen to their resource if they died.

In this section, I have shown that the current research environment prioritises preservation of static datasets over sustainability of working tools and resources. Ensuring the usability of a tool or resource beyond the original project can help to increase its uptake once it is published, potentially leading to further funding. However, key factors that facilitate usability often tend to be overlooked due to time constraints during development. Many participants expected or hoped for their institutions to provide long-term maintenance or hosting, but such arrangements are currently applied inconsistently due to lack of appropriate infrastructure and, ultimately, funding. As a result, many tools and resources persist on a voluntary basis, with the most effective endeavours becoming self-sustaining community initiatives. Several of my recommendations relate to sustainability; as such, I will return to this topic in 8.6.

## 7.7 Conclusions: Producing Usable Tools and Resources

In this chapter, I have built upon my findings from Chapter 6 to explore the wider context of digital tool and resource production and identify factors that impact usability. Although I have presented these factors in an approximate order in which they might occur during the production process, I found that, in practice, usability can most effectively be achieved by considering each one from the outset. Such considerations include:

1. Consistent and effective training in digital tool and resource production:

- a) Ensuring informed decisions about appropriate data models and technologies.
2. Formal or informal collaboration:
- a) Bringing together researchers and technical experts with complementary skillsets;
  - b) Managing teams in an inclusive and equitable way.
3. User-centred design principles:
- a) Promoting the development of tools and resources that researchers can use intuitively;
  - b) Identifying 'user journeys' based on the tasks they intend to perform;
  - c) Performing user testing with sufficient time to implement the results;
  - d) Reducing the extent of documentation required.
4. Clear and effective documentation:
- a) Fostering transparency and user trust;
  - b) Minimising the need for specialist training to use a digital tool or resource.
5. Long-term sustainability:
- a) Institutional support;
  - b) Community engagement;
  - c) (Inter)national infrastructures.
6. Cost to users/institutions:
- a) Charging for full or partial access introduces inequalities and compromises wider usability.

Training (1) and collaboration (2) ensure that people with sufficient expertise are involved from the start of the project, potentially forming the foundations for user communities. Such collaborations of appropriately skilled individuals might effectively incorporate user-centred design (3) and documentation production (4) as integral

components of the development process resulting in a transparent and usable tool or resource. This usability should encourage wider uptake, which in turn increases interest in ensuring long-term sustainability (5), without passing on the cost to end users (6). In the above discussion, I have found that many Digital Humanities tools and resources could benefit from consideration of these fundamental issues. However, this process might be particularly advantageous for the usability of tools and resources involving Linked Data, due to its relative complexity and unfamiliarity when compared with tabular, relational, or hierarchical data models.

I will return to these and my previous findings in the following chapter, where I will sum up and make recommendations to improve the usability of future Linked Ancient World Data tools and resources.

## 8 Conclusions and Recommendations

Discussion of my findings in the previous four chapters has demonstrated that Linked Ancient World Data tools and resources can be made more usable, and integrate more effectively with existing research methods, if several key factors are considered in the initial project planning stages. Fundamentally, these tools and resources should be developed with future users and producers in mind from the outset, ensuring openness, transparency, extensibility, and sustainability. However, usability is not only a technological issue, but is often more dependent on the people involved: from selecting an appropriate team of collaborators, to ensuring that potential users are consulted throughout development and beyond, while assisting the formation of communities. Crucially, the usability of Linked Ancient World Data tools and resources depends on a series of decisions made by their producers, the institution(s) with which they are affiliated, and the organisation(s) providing the funding. The majority of this chapter will present recommendations for how each of these stakeholders might approach various aspects of a Linked Ancient World Data project; however, I will start by summarising each chapter of my thesis so far.

Chapter 1 introduced the issue that my thesis aims to address: how the usability of Linked Data might be improved in a Humanities research context, using the Ancient World as a case study. I subsequently presented my research questions, which sought to assess the current state of Linked Ancient World Data implementation, before conducting research into how such technologies might be integrated with existing research methods, and how their usability might be improved in this context. I also introduced several approaches to data modelling and sharing in Humanities research, before providing a more detailed explanation of my key topics, Linked Data and usability.

In Chapter 2, I found that Linked Data projects are still rare in Humanities research, although a relatively high proportion were developed to facilitate study of the Ancient World (RQ1a). On exploring some key Linked Ancient World Data tools, resources, and initiatives, I found that they tend to focus on one of four key concepts (place, time, people, or objects), while promoting connections between them. The examples included in this discussion illustrated how Linked Data technologies have been

implemented in an Ancient World research context thus far (RQ1b), adding disciplinary nuance to the generic five-star model. Advantages to the application of a Linked Data approach in this context (RQ1c) included the role of persistent identifiers in aligning and disambiguating entities, as well as rich descriptions for more accurate representation of the inherent complexities in digital and physical objects. Linked Data also provides potential for interoperability between disparate datasets, connecting collections for more efficient, holistic, and serendipitous discovery. Furthermore, its open standards promote data sharing and reuse, leading to collaboration and facilitating preservation. However, I found that barriers (RQ1d) were encountered when attempting to align different vocabularies, communicate uncertainty, and represent complex objects in a consistent way. Additionally, I found that many researchers were not aware of Linked Data technologies, did not have sufficient training to implement them, or encountered usability issues – the primary focus of my thesis.

As a result of these findings, and following recommendations from previous usability research, I designed a study based on a survey and interviews, aimed at Ancient World researchers. In addition to my methodology, Chapter 3 discussed the definition and classification of digital research methods in the Humanities (RQ2), by exploring several different initiatives. These frameworks ranged from the generalised Scholarly Primitives, which distils the research process to seven fundamental activities, to the detailed Taxonomy of Digital Research Activities in the Humanities (TaDiRAH) that provides a deeper level of granularity and differentiation between methods. Due to the digital nature of my topic and its balance between generality and specificity, I decided to use TaDiRAH to classify and analyse participant responses to my survey. In this chapter, I also explained the rationale behind my study design and presented demographic information about my survey participant population, which I used in selecting my interview sample.

Chapter 4 analysed the quantitative findings from my survey, beginning to address the integration of a Linked Data approach with existing Ancient World research methods (RQ3), and improving usability of Linked Ancient World Data (RQ4). I divided my findings into those relating to participants' experiences of using and producing digital

tools and resources, and those relating to associated research methods. In terms of the former, I identified six key factors that influence the usability of Linked Ancient World Data tools and resources: usability, documentation, training, cost, reliability, and awareness. Regarding the latter, I highlighted five TaDiRAH methods of particular interest to Ancient World researchers, with potential for Linked Data integration:

*Discovering, Gathering, Data Recognition, Annotating, and Visualization.*

I discussed these methods further in Chapter 5, where I gave particular attention to *Discovering*, the most popular research method described by participants. I also noted that the boundaries between research methods are almost invariably blurred, with many interconnections between them, particularly where *Annotating* and *Visualization* are concerned. This chapter additionally identified that key areas for integrating Linked Data with these methods might include providing centralised search and exploration tools, enhancing existing resources with contextual information (potentially via annotation), and increasing connectivity between resources, e.g., to produce curated directories.

In the following chapters, I expanded on the user experience topics identified in Chapter 4, dividing my discussion into findings regarding tools, resources, and their underlying data (Chapter 6), and those relating to the production process (Chapter 7). In Chapter 6, I found that usability can be greatly improved by ensuring that tools and resources are designed in an intuitive way and that they function reliably and consistently; participants also highlighted the importance of data openness and quality. More fundamentally, however, I found that the usability of Linked Ancient World Data tools and resources is significantly influenced by decisions made before and during the production process. As such, I identified six key themes to explore in the following chapter: training, collaboration, user-centred design, documentation, access, and sustainability.

In Chapter 7, I identified the importance of Linked Ancient World Data producers having sufficient knowledge and experience to make informed decisions at all stages of the production process, ideally leading to effective, inclusive collaborations and, ultimately, communities of practice. Such collaboration might also extend to the



process of tool or resource design, by consulting with users and ensuring that the resulting output will meet their needs. Similarly, effective communication with users can lead to transparent, comprehensive documentation. All these factors impact the sustainability of, and future access to, the tool or resource and, therefore, its continued usability.

The above discussions regarding the human and technological factors that affect Linked Ancient World Data usability have formed the foundation for a series of recommendations as to how these factors might be addressed in future developments, both within and beyond this disciplinary context. Recommendations are aimed broadly at researchers involved in Linked Ancient World Data production, with some directed more specifically at roles such as project leaders or developers, as well as stakeholders from the wider research community, such as funders and institutions. I have ordered them based on the usual chronology of when they might be considered during the preparation and execution of a Humanities research project, while acknowledging that each project will take an individual approach.

The first group of recommendations concern actions that might be taken and issues that might be considered to improve Linked Ancient World Data usability in the short to medium term. Topics within this group include effective teams (8.1), openness and transparency (8.2), extensible development (8.3), user consultation (8.4), and facilitating discovery (8.5).

## 8.1 Effective Teams

As I found in 7.2, working collaboratively to produce Linked Ancient World Data tools and resources is likely to produce more usable results than individual projects. Effective collaboration potentially leads to more efficient ways of working, by bringing together researchers and developers with complementary areas of expertise. It also promotes sustainability from the outset of a project, rather than having a single point of failure. Effective teams are therefore the foundation of usable Linked Ancient World Data projects, facilitating transparency (8.2), extensibility (8.3), user consultation (8.4), and sustainability (8.6), topics that I will discuss later in this chapter. In this section, I will outline recommendations for how such teams might be assembled, and how

institutions and funders might facilitate this process by improving access to, and awareness of, appropriate training.

- **Project Leaders: *Build diverse, inclusive teams, with appropriate technical and research experience***

When building such teams, project leaders should be mindful of ensuring inclusivity, in terms of demographic characteristics. In doing so, it is important to reflect on how the composition of the team might affect usability of the resulting output by different audience groups. This latter point reflects my findings in 7.2 that women appear to be underrepresented among digital tool and resource producers in the Ancient World domain, with one participant feeling that poor treatment by colleagues on a past project was largely due to her gender.

These findings exemplify a wider phenomenon. Despite the popular perception of gender equality in the Digital Humanities (suggested by the number of women in prominent roles), recent analysis of conference paper acceptances (Eichmann-Kalwara et al., 2018) and citation practices (Earhart et al., 2021) demonstrates that female researchers continue to be underrepresented. While gender is the characteristic that came to the fore over the course of my research, many other intersecting aspects come into play, such as race, language, sexuality, and ability. Indeed, responses to Bordalejo's (2018) survey on diversity in the Digital Humanities indicated that most participants were "binary, white, affluent, and Anglophone". If teams of people producing digital tools or resources are not representative of the people who use them (as indicated by the demographics of my wider survey population), their usability by wider audiences is likely to be restricted.

Furthermore, in any team intending to produce a Linked Ancient World Data tool or resource (or in the Digital Humanities more generally), it is crucial that technical experts, information professionals, and subject specialists are considered equal partners, as found in 7.2. In particular, the project team should ideally include at least one person with experience of producing one or more sustainable, useful, usable digital resources, with other members of the team having received appropriate training (see below). Sufficient time should be allocated for discussions between all

parties at various points throughout the project, to ensure mutual understanding and to allow all team members to share their views.

- **Institutions: *Greater investment in Digital Humanities training across career stages***

My findings demonstrate that consistent availability of Digital Humanities training across institutions, disciplines, and career stages would be of benefit to all researchers from undergraduate level upwards<sup>161</sup> (7.1), with such training being particularly crucial for producing usable, reliable, Linked Data tools and resources (6.1). Having received appropriate training, researchers are able to make more informed decisions about how they can most effectively model their data and maximise its usability by others. They are also better able to manage digital projects and communicate key information to technical and domain specialists, as well as project stakeholders. However, the current situation often results in digitally interested researchers seeking training elsewhere, while those who are less digitally confident remain constrained by the limitations of their institution.

In terms of content, such courses should initially aim to provide a basic understanding of how data can be modelled, incorporating tabular, relational, text encoding and Linked Data approaches. However, rather than considering only the technical aspects, researchers should also be encouraged to consider how each approach affects aspects such as data richness, accuracy of representation, and potential for connection with external sources. In fact, as suggested in 6.2, this latter aspect might be more important for those researchers interested in consuming, rather than producing, digital data, to promote understanding and critical evaluation of tools, resources, and associated documentation. Other key topics might include aspects of managing digital tool and resource production, such as user-centred design, documentation production and sustainability.

---

<sup>161</sup> Some institutions already provide such training as part of undergraduate modules, or even as a larger component of degree courses (e.g., King's College London, n.d.; Loughborough University, 2021; UCL Centre for Digital Humanities, 2020; University of Edinburgh, 2019).

Following my findings in 7.1, such training might be most effective if it takes the form of a project that is built on over the course of the training. For example, the course might start by considering a set of physical objects in a cultural heritage collection, then proceed through aspects such as digitisation, data modelling and representation, and online presentation. Mahony and Pierazzo (2012, pp. 221–222) found a similar approach to be extremely effective when teaching King’s College London’s ‘Medieval Manuscripts in the Digital Age’ course, even for less digitally confident students. They found that attendees were better able (and more willing) to engage with digital technologies when they appeared as part of a wider research methodology, applied to familiar objects, rather than as a series of abstract, unrelated exercises.

For undergraduate and postgraduate students, such training might take the form of a module included as part of their degree, with similar courses offered to staff as part of the institution’s library or IT training programme. Training materials might be produced and delivered collaboratively by teams including Humanities researchers, library staff and technical specialists, taking advantage of existing Digital Humanities networks where they exist, or forming them where they do not. General sessions on digital methods, data modelling, and project management might be supplemented by subject-specific sessions organised by departments, to raise awareness of relevant tools, resources, and techniques. As this would require significant investment, it might alternatively be achieved by partnering with other institutions to share knowledge and resources. This process might be facilitated by utilising existing links such as doctoral training partnerships or long-term collaborations.

- **Institutions/Funders: *Promote the use of existing freely available Digital Humanities training resources***

For institutions with limited budgets for Digital Humanities training (or in the interim before new courses can be developed), another solution would be to raise awareness of freely available tutorials and workshops by initiatives such as the *Programming Historian*<sup>162</sup> and *Sunoikisis DC*<sup>163</sup>. Although each of these initiatives has a disciplinary focus, both include content applicable to other Humanities subject areas. Similarly, it

---

<sup>162</sup> <https://programminghistorian.org/>

<sup>163</sup> <https://www.dh.uni-leipzig.de/wo/sunoikisisdc/>

could be helpful for funders to direct researchers to these materials if they are considering a proposal for a digital project. However, relying on such resources as the sole source of Digital Humanities training would not be advisable in the long term, because only researchers who are already digitally confident are likely to engage with them. A potential middle ground is for institutions to arrange training courses that directly use these materials; for example, many universities have incorporated *Programming Historian* tutorials into their syllabi (Crymble, 2018). Organising training in this way reduces the burden of creating new content, while allowing staff to focus on teaching, using high quality peer-reviewed materials. Its provision might further be facilitated by the development of overarching, community-led directories, which I will discuss in 8.9.1, below.

## 8.2 Openness and Transparency

Many of my findings in Chapters 6 and 7 related to the themes of openness and transparency, from the implementation of open standards and data sharing (6.2), to provision of clear documentation (7.4). These characteristics ensure that users can understand how Linked Ancient World Data has been modelled. Furthermore, honesty about any limitations of that dataset helps gain users' trust and provides the tools they need to make their own critical evaluations and interpretations. Reuse of the data is facilitated and encouraged, with clear instructions for how this might best be achieved. In this section, I will start by making recommendations about data openness, before focusing in more detail on transparency, recommending how effective, usable, documentation might be produced.

- **Project leaders/Developers: *Implement open, FAIR, data***

As Linked Data requires the implementation of open standards, such as RDF, and Linked *Open* Data is inherently open, *FAIR*, data (1.2), Linked Data should automatically be findable, accessible, interoperable, and reusable. However, there are some additional considerations that should be made to maximise its usability. Data export, in particular, is one area in which Linked Ancient World Data producers might provide more openness and transparency. As found in 5.5, data export options should be made clearly available to accompany any visualisations, to provide users the opportunity to gain in-depth understanding of a dataset, rather than a superficial overview.

Additionally, it can be beneficial to allow users to export data in familiar formats, such as CSV, which might more readily be consumed by other tools in their research ‘pipeline’ than its original format of RDF (6.2.1). Indeed, Simon et al. (2019) emphasise that using *Recogito* is just one step in a researcher’s process; their data import and export options should allow seamless transitions between *Recogito* and the previous and subsequent steps.

Another way in which producers can enhance the openness and transparency (as well as the reusability) of Linked Ancient World Data is by ensuring that their URIs are persistent. Assuring users that URIs will be persistent in the long term increases their trust in the tool or resource, as well as the likelihood that they will share those URIs with others (6.2.1), either in citations, or by integrating the dataset with their own tool or resource. Implementing persistent URIs avoids usability and reliability issues in these third-party tools and resources, and potentially facilitates their sustainability, in reducing the amount of maintenance they require (7.6).

- **Project leaders/Developers: *Produce clear, discoverable documentation***

Transparency can also be achieved by incorporating clear documentation, which should enhance tool or resource usability by providing instructions, examples and use cases, potentially negating the need for resource-specific training (as shown in 7.4). As such, while it should not be considered a long-term substitute for an intuitive user interface, extending, or improving existing documentation could be an effective first step in increasing usability, particularly where substantial changes to the interface would be prohibitively expensive or time-consuming. When producing documentation, it is advisable to consider the subject/technical knowledge of anticipated users (or lack thereof) by explaining terminology and using clear, accessible language, to avoid limiting the potential audience. In terms of content, effective documentation should include information from the following broad categories (alongside tutorials and publications, such as use cases):

**General information:** brief overview of tool or resource aims, institution(s) and individual(s) responsible for its production, and funding source(s).

**Scope:** what is and what is not included within the tool or resource (5.1.4).

**Access:** any restrictions or technical requirements that affect access to the tool or resource (6.1, 7.5).

**Functionality:** for example, explaining how to search effectively using Boolean operators or wildcards, or providing example queries for a SPARQL endpoint (5.1.4).

**Data provenance:** the original source of the data, and how it has been altered or processed before publication as part of this tool or resource (6.2.2).

**Data model:** how the data has been structured, including standards and formats used, as well as the rationale behind these decisions (6.2.1).

**Reuse:** copyright and licencing information, including any restrictions on future reuse of either the tool/resource itself, or the data contained within it (6.2.3).

**User contributions:** instructions for how to add new data or contribute to future development (6.2.2).

**Maintenance:** who is responsible for hosting and maintaining the tool or resource (7.6).

I have built upon these categories to develop a detailed checklist for producing Linked Ancient World Data project documentation, provided in Appendix 5.

- **Developers: *Integrate basic information and instructions with the user interface***

As participants were unlikely to interrupt their workflow to consult documentation, some elements might most effectively be integrated within the user interface itself, in addition to inclusion within the main documentation pages (6.1.1, 7.4). Examples of such integration could include providing basic instructions in situ, such as displaying the phrase “Enter search keywords”, or including example terms, within a search box.

Similarly, in cases where the user has made an error in their query, a clear message should be displayed, providing feedback to help them resolve the issue (6.1.2). Where users require further assistance, this might be provided by more detailed instructions in a popup accessed by clicking or hovering over a question mark icon. Any data quality issues should be communicated within relevant object metadata rather than only the main documentation pages (6.2.2).

For those users who require more extensive information, such as a diagram of the data model, a description of the modelling process, or a step-by-step tutorial, these should remain as separate pages in the resource, easily accessible via direct links included at relevant points in the user journey. This kind of layered approach to structuring the information about a tool or resource would assist users by making the most widely applicable advantages more apparent, while providing a clear entry point to more technical information if required.

- **Project leaders: *Consider documentation a key component of the production process***

Planning effective integration with a tool or resource also ensures that documentation is considered from the start of the project and throughout the development process (7.4). Fully integrating documentation into the production process, as well as the tool or resource itself, enhances usability, while also mitigating against people leaving the production team; if the documentation is up to date, their knowledge will not leave with them. Furthermore, collaborating on the documentation itself, either by having multiple authors or seeking feedback from colleagues, should aid clarity and assist user comprehension.

### 8.3 Extensibility

The previous section included recommendations for how Linked Ancient World Data producers might implement openness and transparency measures to ensure (re)usability by future users. I will now take these ideas further by recommending how they might also consider future producers, facilitating usability as a result. Learning from my participants' experiences, incremental development is often the most effective approach. Producers might consider building upon tools and resources that



already exist (where possible) and/or ensuring that new tools and resources can themselves be readily expanded and enhanced. Such extensible development can avoid duplication of effort, as well as reducing time and costs required. There is also a greater incentive to ensure usability, by envisioning a tool or resource from the outset as a long-term prospect, which might be expanded or enhanced over time by people both inside and outside the original project team. In some ways, this approach mirrors that of Linked Data as a relatively small-scale, extensible implementation of the wider Semantic Web. In this section, I will start by discussing the possibilities for building on existing tools and resources, before recommending how Linked Ancient World Data producers might consider developing a tool or resource in an extensible way.

- **Project leaders/Developers: *Build upon existing tools and resources (where possible)***

As an initial step, I recommend that potential producers review existing digital tools and resources that incorporate the intended research methods, applied to a similar or related subject area. These tools or resources might be discovered via mailing lists targeted at the relevant research community, or resource directories such as those discussed in 8.9.1, below, in addition to social media or internet searches. Project leaders and developers should work collaboratively to identify potentially compatible tools or resources and assess their suitability for involvement in the project, in consultation with the tool or resource producers. If such a system has implemented openness and transparency measures, as recommended in 8.2, alongside open source code, there might be scope to adapt it to provide appropriate functionality that allows users to interact with a new dataset. For example, the Linked Data Greek pottery resource *Kerameikos.org*<sup>164</sup> is based on the same software architecture as *Nomisma* (2.2.4) (Gruber & Smith, 2015, p. 209). In some cases, such as that of *Pelagios'* original *Peripleo* tool (Simon et al., 2016b), it might be possible to create a new instance of that system and customise it accordingly. In others, new datasets and/or functionality might be integrated with the original. The latter option would be ideal for researchers who have produced useful and compatible datasets but lack the time or technical knowledge to develop a tool or resource through which it might be accessed.

---

<sup>164</sup> <https://kerameikos.org/>

- **Data producers: Use an RDF-compatible CMS to provide an interface for Linked Data access**

For Linked Data producers with limited technical skills and/or insufficient funding to design a new interface, an alternative option is to implement an open-source content management system (CMS) that supports RDF data structures, as suggested in 6.1.1. Although implementing a CMS will not be an option for every Linked Ancient World Data project, they provide more flexibility than existing subject- or task-specific platforms. Widely adopted systems that have previously been used in a similar context are likely to already be usable by Humanities researchers and are often readily customisable if changes are recommended following user feedback. They also tend to be updated regularly by active communities, thereby reducing the amount of time and resources required for maintenance. Two such systems that have been applied to Digital Humanities projects include *Omeka S*<sup>165</sup> and *WissKI*<sup>166</sup>.

*Omeka S* is a new version of the *Omeka* CMS, intended primarily for cultural heritage collections. It uses Linked Data technologies, provides each object with a URI, and allows integration of new or existing vocabularies (Omeka, n.d.). Previous implementations of *Omeka S* in Humanities research projects (Bruneau et al., 2021; Jones & Muftic, 2020; Lombardo et al., 2020) indicate its reliability, as well as demonstrating its potential to facilitate user exploration of a Linked Dataset without requiring substantial technical skills on behalf of the producer. However, it does not include an integrated SPARQL endpoint, thereby limiting the potential for Linked Data implementation.

*WissKI* is a Virtual Research Environment (VRE) that acts as a plugin to the larger, more generic, *Drupal*<sup>167</sup> CMS, providing integration between a *Drupal* site and an RDF triplestore (WissKI, n.d.). Like *Omeka S*, previous Humanities research projects (Scholz et al., 2014; Vlachidis et al., 2018; Wettlaufer et al., 2015) have demonstrated *WissKI*'s potential for increasing the usability of Linked Datasets, particularly as (unlike *Omeka*

---

<sup>165</sup> <https://omeka.org/>

<sup>166</sup> <http://wiss-ki.eu/>

<sup>167</sup> <https://www.drupal.org/>

5) it incorporates a SPARQL endpoint. However, it lacks the flexibility of *Omeka S* from a data modelling perspective, requiring all data to be structured using *CIDOC CRM* (1.3), which some of the above producers found to be quite restrictive. *WissKI* might therefore be a good solution for publishing Linked Data that already complies with *CIDOC CRM*; however, it could be difficult to implement for producers who have already used a different approach.

While both *Omeka* and *WissKI* provide the possibility for producing usable resources from Linked Ancient World Datasets, both have some shortcomings. An ideal system would combine both flexibility of data models and a SPARQL endpoint; currently, it seems that researchers must choose one or the other, or customise their chosen system, with the latter option being problematic for time-poor researchers with minimal technical skills.

- **Developers: Take an extensible, modular approach to development**

In practice, many existing tools and resources are built using project-specific databases that preclude integration. To avoid this issue in future, projects identifying the need for a new system should keep openness and transparency (8.2) in mind throughout development, to facilitate further development by others. These steps can potentially extend the life of the tool or resource, as well as its wider usefulness. For example, when developing the *Folioscope* text viewer, Hedges et al. (2017, p. 10) aimed to be as flexible as possible in terms of permitted data structures.

Extensibility also applies to the scope of the tool or resource. In many cases, it can be beneficial to start with simple expectations that incorporate potential for expansion (as recommended in 6.2.2), rather than taking an over-ambitious approach that precludes further integration or never reaches 'completion'. For example, the producers of *PeriodO* (2.2.2) describe their initial aims as being too "optimistic", which resulted in reprioritisation of their development goals following consultation with their advisory board. In doing so, they divided the remaining work into two separate phases, ensuring that a usable output could be completed by the end of the first phase, which could then be enhanced in the second (Buchanan et al., 2016, p. 8). Such an approach

also provides the opportunity to gain user feedback at an early stage, which might then be applied to later versions or future developments.

This modular approach was also recently undertaken when upgrading *Pelagios'* (2.2.1) *Recogito* platform, which involved separating its functionality into separate components, to facilitate integration and further development by other members of the community (Kahn et al., 2021, p. 95; Simon et al., 2019, p. 162). Geser (2016, p. 24) particularly recommends this approach in a Linked Data context, stating that it has the additional benefit of familiarising less experienced producers with Linked Data technologies at a smaller scale, allowing them to build their knowledge and skills in this area for application to further developments. For the data itself, such simplification might involve prioritising the production of smaller datasets, which can be more rigorously checked for quality and academic credibility, and could be more viable to update, rather than making large quantities of data available without the capacity to verify and maintain it.

Another potential solution is to consider how a tool or resource with a more specific remit might be generalised for application in another context or integrated with another tool or resource. For example, the principles behind PART109's idea of dividing their ontology into a generic 'core' and specialised 'extension' (6.2.1) might be applied to Linked Ancient World Data tool or resource development more generally. A similar approach, implemented by Antonini et al. (under review), is to apply an overarching model that can be used to align multiple datasets. The authors used an "Experimentation and Observation" approach, via the *Crowdsourcing of Evidences Ontology*<sup>168</sup>, to unite the *Reading Experience Database*<sup>169</sup> and *Listening Experience Database*<sup>170</sup>, two comparable RDF datasets with differing structures and foci. As a result, both datasets are now more widely usable in different contexts, rather than being bound to their original research questions.

---

<sup>168</sup> <https://github.com/eureadit/crowdsourcing-ontology>

<sup>169</sup> <https://www.open.ac.uk/Arts/reading/UK/>

<sup>170</sup> <https://led.kmi.open.ac.uk/>

## 8.4 User Consultation

In 6.1.1, participants spoke about intuitive design, i.e., tools or resources should be easy to use, with clear information and visual cues. The user should not need to consult documentation to perform basic functions, or to have any coding or programming experience. While this is true for digital tools and resources in general, particular attention must be given to those based on a Linked Data structure, to ensure that the complexity of the underlying data structure does not lead to an unnecessarily complicated user interface. Intuitive interfaces can most effectively be achieved by consulting users from the outset of a digital project, and throughout the development process.

- **Project leaders/Developers: *Incorporate user-centred design techniques***

As found in my survey and interviews, interface design is crucial for ensuring that researchers outside the immediate project team can engage with a Linked Data tool or resource, with several participants recommending a user-centred approach (6.1.1). By consulting users about their expectations and observing their behaviour with existing tools and resources, producers can map out user journeys. These journeys can then be used as the basis for establishing pathways that might be taken through the new tool or resource, to achieve a particular goal. For design to be intuitive, such pathways should guide the user through the tool or resource, presenting their options clearly at each stage, without requiring them to check documentation (7.3, 8.2).

- **Project leaders/Developers: *Consider a task-centred approach***

Although, in many cases, user goals are likely to relate to their demographics, defining potential audiences based on their intended tasks might produce more effective results, as described in the Task-Centred User Interface Design approach introduced by Lewis and Rieman (1993). After all, ISO's (2018) definition of usability, used throughout this thesis, explicitly relates to the user's ability to perform their desired task, rather than general ease of use (6.1). It would therefore be worthwhile for producers of future Linked Data resources to consider carefully which broad activities most closely align with the primary aims of their project, then define possible tasks associated with these activities. Once these tasks are established, it would then be advantageous to consider the probable levels of technical skills and experience of the users who might

carry out each one, then design their journeys through the tool or resource accordingly (as exemplified by PART041 in 5.1.1).

This process might be effectively achieved in a Digital Humanities context via the alignment of user goals to research activities, e.g., by using TaDiRAH terms. Clear definitions of potential tasks, associated with key terms that indicate related research methods, can additionally facilitate discovery, by enabling researchers to search or browse based on their research aims. Shifting the focus from demographics to tasks might additionally help to break down the somewhat nebulous idea of the “general public”, ensuring that development decisions are made with a clear outcome in mind, rather than an assumed idea of who might be carrying out unspecified activities (7.3). Hedges and Dunn (2017, pp. 27–50) took a similar approach when developing their typology to describe activities conducted via online crowdsourcing platforms. In doing so, they related tools and processes developed from McCarty and Short’s *Methodological Commons* (3.1) to the objects of these actions (assets) and the outputs produced as a result. Several of the processes included in their typology directly align with TaDiRAH terms, such as *Contextualizing*, *Transcribing* and *Translating*.

However, as found throughout Chapter 5, and demonstrated in 5.6, boundaries between these activities are often blurred and many occur in combination; therefore, focusing on individual tasks out of context might be counterproductive. A similar finding was identified during the development of Project Bamboo (Dombrowski, 2014), where an approach involving strict definition of discrete methods was felt to be reducing researchers’ experiences to a lowest common denominator, resulting in mistrust among the community it intended to serve. A balance must therefore be achieved by anticipating specific activities, while being receptive to the idea of different use motivations among individual researchers and avoiding attempts to restrict users to rigid pathways. It is, of course, impossible for producers to predict every conceivable way in which users might engage with their tool or resource; therefore, it is advisable for tools and resources to be developed in an extensible way (as discussed in 8.3) that does not limit the potential for current or future user engagement.

- **Project leaders/Developers: *Perform user testing throughout the development process***

User consultation should not be seen as an isolated act or discrete aspect of a project, but an integral component that is revisited throughout development, by inviting users to test the design and updating it based on their feedback, in a process of co-creation. Such testing might facilitate the identification of potential reliability issues (6.1.2), such as bugs or technical errors, or areas where the functionality is inconsistent, either with the behaviour of other tools or resources, or within the tool or resource itself.

Resolving these issues before a Linked Data tool or resource is made publicly available will improve usability and secure users' trust that it can be relied upon to work as they expect. More generally, user testing might help to identify usability barriers, or assumptions the producers have made about users' technical or domain knowledge.

As found in 7.3, user testing should be planned from the outset of a project to ensure there is sufficient time to implement the feedback received before the tool or resource is made available to a wider audience.

Ideally, user testing should then continue to occur over the lifetime of the tool or resource, provided there is sufficient funding to perform maintenance work (see 8.6, below). As an example, existing Linked Data tools and resources could seek to increase their user base by learning more about the types of query different users would be interested in investigating. Framing these use cases as a series of user journeys of varying levels of complexity could indicate key areas where users enter a resource, or functionality that might be highlighted on the landing page. The process might also help to identify areas where it would be helpful to lead more technically advanced users to detailed information about the resource structure, or how they might access the data via an API or SPARQL endpoint. If users overestimate the complexity of a tool or resource, user testing would similarly reveal how best to pare down the information on the interface and direct users to the relevant area of the resource for their research needs.

## 8.5 Facilitating Discovery

In Chapter 5, I discussed my findings regarding digital research methods conducted by participants, while considering their integration with Linked Data technologies. I particularly focused on *Discovering* due to the wealth of responses from both my survey and interviews, which led to my recommendation for implementing an exploratory mode of discovery. The theme of discovery continued into 6.1.2, where I mentioned the impact of SPARQL endpoint unavailability on usability. I will discuss both discovery topics further below, in relation to recommendations for Linked Ancient World Data producers.

- **Project leaders/Developers: *Facilitate discovery through exploration***

In 5.1, I used findings from my study to discuss how Linked Data technologies might facilitate *Discovering*. Participants indicated that their discovery processes could be enhanced by searching multiple resources from a central point, as well as having greater opportunities for exploration (5.1.2). Such exploration might involve using *Visualization* (5.5) techniques, such as maps or networks, to illustrate the connections between digital objects, potentially providing inspiration for new research pathways. These visualisations can be combined with links to contextual information (a process that could itself be facilitated by *Data Recognition* and *Annotating* tools; see 8.9.2, below). One way in which such exploration might be achieved without developing a completely new interface (following principles of extensibility, 8.3) is by using *Pelagios'* (2.2.1) new version of *Peripleo*, *Peripleo Lite*<sup>171</sup>, which allows the user to select the Linked Datasets they require for geographical exploration and visualisation (Barker, 2021). Another option could involve installing an instance of *ResearchSpace* (2.2.4). However, no participants mentioned how usable this system is in practice, likely because it was only available as a demonstrator at the time of my study.

A similar principle is advocated by Whitelaw's (2015) "generous interfaces", where visual elements provide users with a broad overview of entire collections, while facilitating discovery of relevant objects and the relationships between them. Additionally, referring particularly to an Ancient World context, Barker and Terras

---

<sup>171</sup> <http://pelagios.org/peripleo-lite/>



(2016) state that engagement with the wider context of digitised objects allows users to gain a greater appreciation of their significance and increases the potential for "serendipitous discovery". This latter point about serendipity also corresponds with my findings in 5.1.1, where users wanted the opportunity to discover new connections from seemingly unrelated resources, rather than being presented with exactly what they have asked for. As semantic search becomes more sophisticated and complex queries become more achievable, this is a key aspect for producers to bear in mind when developing Linked Data tools to support Ancient World research, or Humanities research more generally.

- **Developers: *Optimise SPARQL endpoint reliability***

To further facilitate discovery in Linked Ancient World Data tools and resources, SPARQL endpoint availability should be secured as far as possible (6.1.2). Specific measures might include restricting users from inputting particularly complex queries that overload the server, as recommended by Shaw et al. (2016, pp. 58–59). Other solutions include dividing datasets into 'Linked Data Fragments', containing all triples relating to a particular entity, which can be queried on the client-side (the user's machine) (Verborgh et al., 2014, 2016), as well as caching query results on the user's machine to facilitate future retrieval (Janevska et al., 2014). However, the former increases query times and struggles to support more complex queries, such as filtering, and the latter's benefits are only apparent in situations where the user needs to repeat the same query patterns over time. Optimising the availability of SPARQL endpoints therefore continues to be an area where further research is required.

\* \* \*

Through the above five sections, I have presented a series of short to medium term recommendations for improving the usability of Linked Ancient World Data, with the theme of collaboration occurring throughout. Most explicitly, such collaboration involves Linked Data producers working together, combining their complementary skills developed through training in the application of digital methods and tools to Humanities research. However, collaboration can also take the form of anticipating the needs of future users and developers, through producing open, transparent tools and

resources that facilitate extension, enhancement, and customisation. Additionally, collaboration with existing and potential users during the development process can ensure usability for a wider audience, rather than restricting uptake to those with advanced technical skills. This user consultation might lead to new ways of approaching the discovery process, by prioritising exploration, while also being mindful of technical issues that can particularly affect Linked Data tools and resources.

Further measures are required for Linked Ancient World Data usability to persist into the long term, which leads to my second group of recommendations. I will first discuss various strategies to enable and promote sustainability (8.6), before recommending how producers and users can come together to form ongoing communities of practice (8.7).

#### 8.6 Sustainability, Preservation and Funding

With sustainability being a theme that appeared throughout discussion of my findings, I ended the previous chapter by focusing on participant comments that specifically related to this topic (7.6). Planning for sustainability ensures continued usability and usefulness of a Linked Ancient World Data resource by researchers outside the team who produced it. Sustained uptake (facilitated by active user communities, 8.7) builds trust, increasing the potential for data reuse. Similarly, considering future producers by implementing extensibility measures (8.3) further promotes sustainability by reducing the amount of development time required to create new tools and resources, and ensuring data can be preserved in a usable format. All these factors demonstrate that ensuring sustainability of Linked Ancient World Data tools and resources will provide funders with a greater 'return' (in the form of research outputs and wider impact) on their original investment. However, I also found that current funding models tend to focus on data preservation, often at the expense of tool and resource sustainability. This section will start with a brief discussion on data preservation, before recommending several potential sustainability strategies for Linked Ancient World Data producers, as well as how these processes might be facilitated by funding bodies, institutions and policymakers.

- **Project leaders/Developers: *Preserve research data***

As discussed in 7.6, there are multiple options available for preserving research data, facilitated by using open standards (an inherent quality of Linked Data), effective documentation, and deposition in a trusted repository, accompanied by copyright and licencing information. Such preservation measures are crucial to ensure that the data remains available for future use, even if it can no longer be accessed via a specific tool or resource. As such, data preservation is often required by funding bodies and tends to be the main sustainability focus of Digital Humanities projects, rather than sustainability of the tool or resource itself. It should therefore be considered the minimum required to ensure some level of Linked Ancient World Data sustainability.

- **Project leaders/Developers: *Maximise potential for uptake by ensuring usability***

While it is not possible, or indeed desirable, to sustain all Linked Ancient World Data tools or resources, it is advisable to give the time and attention required to encourage and assess uptake by the wider community, rather than abandoning them once the original funding comes to an end. To maximise usability, and therefore, the likelihood of uptake by other researchers, development of such tools and resources should include consideration of my above recommendations. Implementing open standards and providing clear documentation will broaden the tool or resource's potential for use in other research contexts (8.2), while extensibility should minimise the maintenance required and increase potential for integration with other tools, resources, or datasets (8.3). Ultimately, however, it should have an intuitive user interface that enables less digitally confident researchers to access the data and perform their desired tasks (6.2.3, 7.6), which can most effectively be achieved through user consultation (8.4).

Taking the above measures to increase uptake of the tool or resource, alongside consideration of my below recommendations in 8.7, could lead to the establishment of a user community, which in itself provides a strong case for long-term sustainability. Adopting a community-led model, where users become contributors who take an active role in supporting and maintaining the tool or resource, can sometimes reduce ongoing maintenance costs.

- **Funders: *Require usability and sustainability measures in project plans***

Following the above recommendation for maximising uptake of Linked Ancient World Data tools and resources through ensuring usability, it would be prudent for funding bodies to stipulate these aspects as part of their application requirements, alongside a sustainability plan, to ensure that they are making a sound investment. Although a project plan is a crucial part of any funding application, criteria that relate specifically to digital tool and resource production are currently often absent or assumed to be implicit in the more general guidelines. For example, the AHRC asks applicants to consider the users of their outputs more broadly, but does not explicitly relate this point to digital tools or resources, or provide guidance on how it might be achieved in practice (Arts and Humanities Research Council, 2020, pp. 76, 81).

Elsewhere, the situation is more positive. The United States' National Endowment for the Humanities (NEH) highlights the importance of usability in digital tool or resource production, specifically mentioning that applicants should consider their potential audiences, consult accessibility guidelines, and conduct user testing (Office of Digital Humanities, 2021, pp. 1–2, 7–8, 27). Unlike many other funders, the NEH also explicitly requires documentation to be produced to accompany digital tools and resources (in addition to their underlying data) (Office of Digital Humanities, 2021, pp. 1–4).

Documentation was previously required in the AHRC's Technical Plan (Arts and Humanities Research Council, 2018), which has since been replaced by a Data Management Plan with no explicit mention of documentation (Arts and Humanities Research Council, 2020). As identified in this thesis, user-centred design and documentation are crucial for both usability and sustainability; as such, they would be extremely worthwhile additions to future funding application requirements.

Although data openness and preservation are now common funder requirements, sustainability of functional tools or resources through which those datasets are accessed is mentioned less consistently. While the NEH require a sustainability plan as part of their application process for high-budget 'Level III' projects in their Digital Humanities Advancement Grants programme (Office of Digital Humanities, 2021, p. 23) and the Andrew W. Mellon Foundation (2019a, p. 3, 2019b, p. 3, 2020, p. 3) ask

grant applicants to state how digital outputs would be sustained in the long term, the issue is not mentioned at all by the European Research Council (ERC) (ERC Scientific Council, 2019a; European Research Council, 2021). Sustainability is similarly absent from the AHRC's current research funding guide (Arts and Humanities Research Council, 2020), despite the lengthy sustainability guidelines previously provided for their Technical Plan (Arts and Humanities Research Council, 2018). The AHRC's move from a Technical Plan to a Data Management Plan therefore appears to diminish the importance of tool or resource sustainability, focusing instead on data preservation.

- **Funders: *Reframe development of digital tools and resources as a long-term investment that benefits the research community***

In some cases, unforeseen obstacles (or adjustments based on recent technological developments) might affect project timescales, meaning that the intended goals cannot be realised before the funding end date. Where this occurs, extension funding to allow the project team to complete aspects such as interface design and documentation, without compromising on their quality, should ideally be seen as a strong investment to ensure lasting impact of the resulting tool or resource (and thereby reducing the need for near-identical projects in future). However, in practice, additional funds are rarely made available for this purpose (7.6). For example, the AHRC and ERC allow previous grant recipients to apply for 'Follow-on Funding' and 'Proof-of-Concept Grants', respectively; however, their guidelines (Arts and Humanities Research Council, 2020; ERC Scientific Council, 2019b) clearly state that the purpose of these grants is to fund innovative new ideas that have arisen from the previous projects, rather than extending or improving existing work. In fact, the AHRC (2020, p. 40) explicitly states that 'Follow-on Funding' "cannot be used to... develop or extend an existing website or resource".

The NEH, however, does allow applications for 'Level I' (low-budget) grants for "the revitalization and/or recovery of existing digital projects", while emphasising the benefits of integration with more established systems (Office of Digital Humanities, 2021, p. 1). Most encouragingly, the Andrew W. Mellon Foundation (2020, p. 2) indicate that existing recipients are eligible to apply for renewal of support, with both

*Pelagios*<sup>172</sup> (2.2.1) and *ResearchSpace*<sup>173</sup> (2.2.4) having successfully secured additional funding. Applications for such funding should be encouraged, or even ring-fenced, and not overlooked in favour of new projects whose proposals promise ground-breaking innovation, but which might be left in a similar position at the end of the funded period. The above measures will serve to reframe the development of digital tools and resources as a long-term investment that benefits the research community rather than a one-off cost to produce expendable outputs.

- **Institutions: *Greater investment in staff to manage sustainability of Digital Humanities tools and resources***

Ultimately, however, research grants are not a sustainable means of long-term funding for digital tools and resources, particularly once they have reached the stage where the main costs are for hosting and routine maintenance, rather than active development. At this point, responsibility for such support largely falls to the institutions with which their producers are affiliated. Such support can take several different forms, described by Maron and Pickle (2014): either an existing department, usually the library, provides Digital Humanities support as one of their services; a specific Digital Humanities lab or centre takes on this role alongside their own research and development work, or a network of units or departments work together, each providing small-scale specialised support.

Whichever approach is taken, it is the individuals involved who have the most impact on the effectiveness of institutional support. Smithies et al. (2019, paras 21–22) argue that the most fundamental means of ensuring sustainability of Digital Humanities tools and resources is to employ permanent staff responsible for their management, who are supported in regularly updating their knowledge and skills. Similarly, the Centre for Information Modelling at the University of Graz recognises the importance of investing in people as well as technology (Neuefeind et al., 2020, p. 5). Maron and Pickle (2014, p. 48) additionally highlight the importance of gaining the long-term support and understanding of senior administrators responsible for allocating budgets. Effective institutional support therefore requires the provision of secure roles with clear career

---

<sup>172</sup> <https://mellon.org/grants/grants-database/?grantee=&q=pelagios>

<sup>173</sup> <https://mellon.org/grants/grants-database/?grantee=&q=researchspace>

pathways, alongside communication and advocacy to the wider university to demonstrate why this work matters.

- **Policymakers: *Develop a national, international, or discipline-based Digital Humanities infrastructure***

Where ongoing institutional support is not possible, another possibility is to secure hosting (and, potentially, maintenance) from a national or international infrastructure. Indeed, survey results from the nascent UK-Ireland Digital Humanities Network suggest that inconsistent support provision across universities might be addressed with a shared Digital Humanities infrastructure between the UK and Ireland (Romanova et al., 2021, p. 16). In the UK, some support is already available from Jisc<sup>174</sup>, who provides technological advice and infrastructure to UK Higher and Further Education institutions. Although it previously funded Digital Humanities projects, such as the first two phases of *Pelagios* (2.2.1) (Jisc, 2014; The Digital Classicist Wiki, 2020), it now works with institutions to develop tools and services that meet collective needs, such as a shared repository and preservation facilities (Jisc, n.d.-a, n.d.-b) and might therefore be in a position to support future sustainability of Digital Humanities tools and resources, given sufficient funding.

A more extensive infrastructure that provides Digital Humanities-specific support throughout Europe is DARIAH<sup>175</sup>, which brings together researchers and practitioners, and provides support, tools, and training. DARIAH is keen to support sustainability through interoperability and ensure that the tools and services<sup>176</sup> developed by member organisations are compatible with each other and openly available, in a similar way to the modular approach described in 8.3 (Edmond et al., 2020, pp. 218–219). DARIAH’s members include multiple national Digital Humanities organisations, who might be able to assist with sustainability arrangements for future projects, although these organisations themselves appear to be based at research institutions (and therefore subject to the issues highlighted above).

---

<sup>174</sup> <https://www.jisc.ac.uk/>

<sup>175</sup> <https://www.dariah.eu/>

<sup>176</sup> <https://www.dariah.eu/tools-services/tools-and-services/>

Even these major infrastructures cannot guarantee their continued existence, however. Indeed, Kansa et al. (2018, p. 502) caution that any national, publicly funded, services are at risk of closure due to governments realigning their priorities. In other cases, research cultures and approaches might shift. For example, the UK's *Arts and Humanities Data Service (AHDS)* was closed in 2008, following a withdrawal of support from the AHRC (Arts and Humanities Data Service, 2010). Rather than aiming for broad multidisciplinary support, producers might benefit more from a discipline-specific solution. For Archaeology, the UK's *ADS* and the international, US-based, service *OpenContext*<sup>177</sup> provide extremely valuable services. However, as I mentioned in 7.6, the *ADS* has moved from being state funded to requiring subscriptions from depositors; *OpenContext* also supports itself through subscription fees. Additionally, although both organisations provide publication options, as well as connectivity with related outputs, they currently only provide data preservation facilities, rather than full tool or resource sustainability. However, given sufficient funding, this could potentially present a future direction for such platforms.

- **Institutions: *build a global Linked Humanities Data Consortium***

Another model that takes elements from both institutional and (inter)national support is that adopted by the *International Image Interoperability Framework (IIIF)*<sup>178</sup>, which provides technologies that enhance the experience of interacting with online images, and the *Open Library of Humanities (OLH)*<sup>179</sup>, an open access journal publishing platform. Each of these initiatives is financially supported by a global consortium of institutions. While *IIIF* offers two tiers of membership based on the amount an institution is willing to pay (International Image Interoperability Framework, n.d.), each member institution of *OLH* pays an affordable annual subscription fee, based on their size and country (Open Library of Humanities, n.d.), thereby taking a more inclusive and equitable approach. These contributions fund the required technological and human infrastructure to provide free, public access to *IIIF* and *OLH* resources. This type of consortium model could provide a consistent means of supporting and maintaining

---

<sup>177</sup> <https://opencontext.org/>

<sup>178</sup> <https://iiif.io/>

<sup>179</sup> <https://www.openlibhums.org/>



Linked Humanities Data tools and resources, without the risks inherent in relying on a single institution, or even a single country.

## 8.7 Communities of Practice

My findings demonstrate that communities of practice are often at the heart of usable, sustainable, Linked Ancient World Data tools and resources (6.2). Encouraging and nurturing user communities from an early stage increases uptake, as well as the likelihood that users will want to provide feedback on tool or resource usability or make their own contributions. Development might therefore continue beyond the funded period of the project, and data quality might be improved over time. Such communities can also benefit the users themselves, in identifying other researchers with similar interests, with the potential for forming collaborations or mutual support arrangements. In this section, I will recommend how user contributions might be encouraged, and how communication between interested parties might be facilitated, particularly with regard to knowledge sharing.

- **Project leaders/Developers: *Invite user contributions***

An effective means of building communities is to invite user input into a Linked Ancient World data tool or resource. Such contributions are often at the data or content level; for example, *Pleiades* (2.2.1) encourages users to submit place data, while *Papyri.info* (2.2.4) provides a means for users to edit translations via its *Papyrological Editor* tool. Other modes of contribution can have a significant impact on how the tool or resource works; for example, PART119 spoke about contributing to the development of the *Classical Language Toolkit* (7.2). Elsewhere, users are given the opportunity to become involved in the governance of tools, resources, and related initiatives, with a notable example being *Pelagios'* (2.2.1) move from funded project to community network. All three types of contribution provide benefit to tools and resources (and their producers) by improving data accuracy and comprehensiveness, as well as identifying and, potentially, resolving current and future issues (6.2.2), a key point also noted by Tupman (2021). However, they also benefit contributors themselves in facilitating greater ownership over their user experience. Demonstration of these benefits can form the roots of community building.

To encourage the above contributions, a Linked Ancient World Data tool or resource should fundamentally be something that researchers want to use and are able to trust. Project leaders and developers aiming for their tool or resource to evolve into a community of practice would therefore be advised to consider earlier recommendations in this chapter. Of particular relevance in this context are the provision of data using open standards, as well as open source code, accompanied by clear documentation (8.2), as well as considering future developers by implementing extensibility (8.3). Finally, there should be a simple means of contributing data or content by less technically experienced researchers. This mechanism might consist of an intuitive interface, produced through user-centred design (8.4), or might use technologies with which the contributor is likely to be familiar, such as *Nomisma's* (2.2.4) batch upload system that uses *Google* spreadsheets. Crucially, the tool or resource should provide a way of contributing data on both a small and large scale.

User contribution mechanisms are effective in distributing labour and allowing community ownership of tools and resources, while facilitating continuous improvement of content at minimal cost. However, they often rely upon researchers and technical specialists volunteering their own time, which can be problematic for those not currently employed in secure jobs where it is acceptable to spend some of their paid hours engaged in this work. Additionally, further financial or practical support, either through institutions or (inter)national organisations, is often required for continued hosting of usable tools and resources rather than simply the underlying data, as discussed in 8.6.

- **Project leaders/Developers: *Facilitate communication between users, contributors, and producers***

At the heart of communities is communication. Linked Ancient World Data tools and resources should therefore provide a means of opening dialogue between users, contributors, and producers. For example, feedback should be provided to contributors, to assure them that their input is valued and to give advice about improving the quality of any further contributions they might choose to make. On a larger scale, introducing a broader communication mechanism can serve to foster community development by allowing users, contributors, and producers to discover

and interact with other people that have similar research and technological interests. Such communication might result in collaborations (7.2, 8.1), ranging from mutual support arrangements through to funded projects. It would also serve to inform producers about how their data is being used, where any usability issues might be, and how it might be improved. Familiar and easily accessible communication mechanisms include web forums, mailing lists, or social media.

- **Project leaders: *Encourage reflection and knowledge sharing***

Communication can also take the form of knowledge sharing, following reflection on the development of a Linked Ancient World Data tool or resource. Such reflection is a useful exercise in itself, to inform and improve producers' future work. However, when their experiences, best practices, and lessons learned are shared with the research community, e.g., via a blog, others can also benefit from this knowledge (7.2). Future producers might, therefore, apply their learning to avoid or mitigate any issues encountered during production and improve the efficiency of the development process. Reflection and knowledge sharing are particularly important when working with new and unfamiliar technologies, such as Linked Data, to move from a mindset of short-term experimentation to one of long-term usability.

\* \* \*

The above recommendations have presented ways in which Linked Data tools and resources might be sustained in the long term, through financial and infrastructural support, as well as developing communities of practice. I found that significant improvements might be made to usability with shifts in funding models and requirements, with institutions working together and investing in staff to ensure that usability persists over time. Again, my findings demonstrate that social factors involving collaboration, sharing, and mutual support are equally, if not more, important than technological factors in this regard, with community forming the lynchpin of a usable Linked Ancient World Data tool or resource throughout its lifespan.

Following the above discussion, I will now take a broader view than my Ancient World case study, by drawing together my recommendations into a *Five-Star Model for Linked Humanities Data Usability*.

## 8.8 Five-Star Model for Linked Humanities Data Usability

The previous sections have demonstrated various measures that can be taken to improve the usability of Linked Ancient World Data tools and resources. However, it can also be argued that they might equally be applied to Linked *Humanities* Data more broadly, particularly because my use of *TaDiRAH* ensured a focus on Humanities research methods, rather than techniques specific to study of the Ancient World. Although these recommendations might be implemented to improve existing tools and resources, usability can most effectively be ensured by decisions made at the outset of a project. Additionally, as I discovered in 4.1.3, while the majority of Linked Datasets produced by participants complied with Berners-Lee's Five-Star Model (1.3), such compliance does not necessarily result in usability of the tool or resource through which this data can be accessed. Further guidelines are therefore required to promote usability of Linked Humanities Data tools and resources, in addition to their openness and interoperability. Based on my key findings and recommendations, I propose the following *Five-Star Model for Linked Humanities Data Usability*, aimed at project leaders:

- ★ **Transparent:** provide clear documentation about data structures and functionality
- ★★ **Extensible:** encourage integration of new data, while allowing extension of the system using modular components
- ★★★ **Intuitive:** develop clear user journeys to facilitate completion of desired tasks, without interrupting workflows
- ★★★★ **Reliable:** ensure consistent functionality, while minimising downtime
- ★★★★★ **Sustainable:** incorporate a clear strategy to support continued functionality for (at least) a fixed period, alongside long-term data preservation

Stars 1 and 2 require open standards and documentation, as recommended in 8.2, as well as the extensible approach discussed in 8.3, while stars 3 and 4 each relate to different aspects of the user consultation process, recommended in 8.4. Finally, star 5 involves implementing recommendations on sustainability, from 8.6. Throughout this process, it is crucial for people with requisite skills and knowledge (or sufficient interest to acquire them through training) to work collaboratively from the outset (8.1). Eventually, these collaborations might lead to the development of communities of practice (8.7) that can assist in maintaining the tool or resource in the long term.

As mentioned in 1.4, my research was conducted in parallel with work on 'Linked Open Usable Data' (LOUD), led by Robert Sanderson and the *Linked Art* initiative, which, similarly, resulted in the definition of five 'Design Principles' (Linked Art Contributors, n.d.-b). Although there are some commonalities between my five-star model and the LOUD design principles, they have different audiences and goals in mind. While my model provides guidance for project leaders about to embark on production of a Linked Ancient World Data tool or resource, to ensure it is usable by the research community, the LOUD principles are aimed at data scientists, to ensure the resulting data is usable by developers. The main similarity between the two is the recommendation to produce "Documentation with working examples", to facilitate understanding of a data model. Elsewhere, the LOUD principles recommend that data scientists provide data that is "The right Abstraction for the audience", that has "Few Barriers to entry", that is "Comprehensible by introspection", and that has "Few Exceptions, instead many consistent patterns". Although the more detailed explanations for the LOUD principles differ from mine, the principles themselves could equally be applied in the context with which this thesis is concerned. The two models might therefore be used in conjunction with each other, by different people involved in the same Linked Humanities Data project.

In the above sections, I have discussed measures that might be taken by producers, institutions, and funders to improve the usability of current and future Linked Humanities Data tools and resources. I will now discuss potential areas for future research and development to support the integration of Linked Data with existing research methods, before making my final conclusions.

## 8.9 Future Work

I will now build upon my recommendations, and address gaps identified during my study, by presenting two ideas for new developments and associated research. Both are intended for effective integration with existing research methods, with the first focusing on *Discovering* and the second combining *Annotating* with *Data Recognition* and *Visualization*.

### 8.9.1 Resource Directories

Several of my findings, particularly in the areas of awareness (5.1.3) and training (7.1, 8.1) suggest the potential value of curated, trustworthy resource directories that would provide researchers with a single, discoverable access point to multiple tools, resources, and tutorials. Such directories would break down barriers to accessing these materials and acquiring relevant skills, while minimising the time taken to discover suitable resources. However, there are some considerations that must be made before their production, particularly regarding sustainability. After providing an example of the directory sustainability issue, I will outline two potential future resource directory types, alongside suggestions for their effective implementation.

Resource directories need sustained input from the research community for their content to remain current and useful. However, Grant et al. (2020) caution that, in practice, the work in maintaining such directories largely falls to an individual or small group of contributors. They describe how the director of the *DiRT Directory* was unable to continue managing the resource when it no longer fell within the remit of their paid employment, which led to its eventual closure and incorporation within the *Text Analysis Portal for Research (TAPoR)*<sup>180</sup>. Strategies that future such directories might take to facilitate long-term sustainability include minimising the scope, for ease of maintenance (8.3), adopting a collaborative approach to directory management (8.1), engaging contributors from the user community (8.7), and attempting to secure institutional or infrastructural support (8.6) from the outset of their development.

One area where a resource directory could be helpful is by listing appropriate tools and resources for training in digital methods. Many such directories in fact already

---

<sup>180</sup> <http://tapor.ca/home>

exist, as part of institutional *LibGuides*<sup>181</sup>; however, there are various issues with this approach. Firstly, provision is inconsistent: some institutions' *LibGuides* are more comprehensive than others, while yet more have no Digital Humanities section at all. Secondly, there is significant overlap in their content, with any inconsistencies likely being due to a lack of awareness by the maintainer or lack of time for updating on a regular basis. Thirdly, many contain links to resources that are no longer available, which could be frustrating or off-putting for users. Essentially, therefore, multiple individuals or small teams (who are already overstretched and underfunded) are duplicating each other's work to produce multiple similar, yet incomplete, resources. Alongside these efforts is an attempt at a more centralised resource directory in the form of Alan Liu's *DH Toychest*<sup>182</sup>, which suffers from similar issues with comprehensiveness and currency, likely due to its reliance on an individual curator. Additionally, it is less likely to be discoverable by researchers new to Digital Humanities, who are more likely to consult resources provided by their own institutions.

A solution to these issues could be for librarians currently responsible for updating their institutions' Digital Humanities *LibGuides* to collaborate in creating a centralised directory, either using the *LibGuides* platform itself, or an open-source CMS, such as *Drupal* or *Wordpress*. The resulting directory could then be linked from individual *LibGuides*, thereby maximising discoverability by researchers, minimising the maintenance required, and allowing librarians to focus their efforts on updating information and links to institution-specific provisions. Once created, librarians, resource producers, or users might then be able to contribute links to relevant training resources, moderated by an editorial team. While hosting could be provided by a single institution, it might more effectively be arranged through a consortium model, as described in 8.6.

---

<sup>181</sup> <https://www.springshare.com/libguides/>; Digital Humanities examples can be found from institutions such as the University of Exeter (<https://libguides.exeter.ac.uk/digitalhumanities/resources>), MIT (<https://libguides.mit.edu/c.php?g=176357&p=4205334>), and the University of North Carolina (<https://guides.lib.unc.edu/dhatcarolina/skills>)

<sup>182</sup> <http://dhresourcesforprojectbuilding.pbworks.com/w/page/69244243/FrontPage>

Another potential future resource directory would facilitate Linked Ancient World Data production. As demonstrated throughout my discussion, decisions made during the production process can have a significant impact on Linked Data usability, as well as ensuring consistency and facilitating connectivity to the wider ecosystem. Producers therefore need to know about resources that are already available, relevant, and appropriate for application to their particular context. In turn, these resources (and the information about them) must be readily discoverable. One way of achieving this is via a Linked Ancient World Data resource directory.

To initiate the process of building such a directory, I have been involved in a collaboration with Elton Barker, Gabriel Bodard and Paula Granados García, to create the *LAWD Catalogue*<sup>183</sup>. In the interests of keeping the resource small and manageable, our remit is currently fairly strict (datasets contained in the catalogue must fit the scope of the *Digital Classicist* wiki and must be available in a Linked Data format, such as RDF). There is a publicly available form<sup>184</sup> for users to contribute new datasets, which we moderate prior to inclusion. We presented the *LAWD Catalogue* at the Linked Pasts symposium in December 2020 and received helpful suggestions about datasets and metadata fields that might be included, as well as increasing awareness among the research community.

As expected, several Linked Pasts attendees raised concerns about potential sustainability issues; to address this, we hope to build a user community while the *LAWD Catalogue* remains narrow in scope, to ensure there would be sufficient support to effectively manage its expansion (Granados García & Middle, 2021). Currency and sustainability might additionally be assisted by structuring catalogue metadata using Linked Data itself. In particular, the *Data Catalog Vocabulary (DCAT)*<sup>185</sup> promotes the use of a standard data model to enable integration between directories. For example, the *LAWD Catalogue* is now a partner in the *Pelagios Registry Activity*<sup>186</sup>, which is in the process of developing a registry of place-related Linked Data resources (therefore,

---

<sup>183</sup> <https://tinyurl.com/LAWDcatalogue>

<sup>184</sup> <https://tinyurl.com/LAWDform>

<sup>185</sup> <https://www.w3.org/TR/vocab-dcat-2/>

<sup>186</sup> <https://pelagios.org/activities/registry/>



an overlapping but not identical scope). Using *DCAT* to structure both directories would allow them to make use of each other's information.

Future expansion of *LAWD Catalogue* scope could involve broadening our definition of Linked Data or disciplinary scope, or including tools, resources, and ontologies alongside datasets. Additional elements might include practical guidance on producing Linked Data, including key considerations and steps in the modelling process, as well as a conceptual overview of how it works. Written content should be clear, to ensure understanding by researchers without significant technical experience. Ideally, members of the user community would contribute new tools and resources, provide updates, and share their experiences, use cases, and lessons learned. The community could also facilitate collaboration, allowing other users to find potential partners for future projects. As the number of experienced users increases, so too would the potential for community support, rather than reliance on a particular individual or group.

At the time of writing, the *LAWD Catalogue* is currently available as a Google Spreadsheet, although we hope to present it via a more dynamic interface or visualisation tool in future. In addition to allowing users to search this list by tool or resource name, or type, it might also be helpful to incorporate the facility to search or browse by task, e.g., associating each resource with the relevant TaDiRAH method(s). This task-based approach would facilitate discovery for researchers completely new to Linked Data, as well as introducing more experienced producers to previously unfamiliar materials.

We would also seek to promote the *LAWD Catalogue* to researchers outside the Linked Ancient World Data community, for example through relevant mailing lists, conferences, training courses and workshops – both those targeted at Ancient World researchers with existing digital interests and experience, as well as those who might pursue such avenues in future. While such a resource could never fully replace training courses or tailored support, it could help to break down the barriers to Linked Ancient World Data production and make the prospect of a graph (as opposed to tabular) data

model seem more achievable and less daunting, particularly for those researchers with existing technical experience.

This section has introduced two possible future resource directories that could benefit the Ancient World (or, indeed, Humanities) research community, with regard to digital training and Linked Data production, while considering potential sustainability and management issues. In particular, both examples emphasise the importance of collaboration and community-building, as discussed in 8.1 and 8.7, to ensure high-quality, up-to-date directories, as a result of distributed effort across institutions.

#### 8.9.2 Digital Annotation Research and Development

Section 5.4 demonstrated the benefits of using digital annotation tools in general, as well as the particular advantages of Linked Data, in enabling structured, semantic annotations that might be shared and used by other researchers. I also showed that combining *Annotating* with other activities such as *Data Recognition* and *Visualization* can be extremely powerful. However, while those participants who already use digital annotation tools tended to be very digitally engaged, others preferred the experience of reading and annotating on paper, particularly when those annotations comprise personal opinions and notes that they would not wish to become publicly available. To harness the potential wider usefulness of semantic annotations, while being mindful of the continued need for private annotations, further research will be required into how existing tools might work together as part of an intuitive process that is useful both to the annotator and the research community.

Based on my findings so far, there are various annotation features that might be of interest to Ancient World researchers. Annotations should be stored privately by default, although the option to make them public should be clearly available. Named Entity Recognition (NER) can greatly facilitate the process of identifying corresponding authority files (e.g., for people or places) from external sources, allowing semantic annotation, while providing users with the option to review and confirm the connection. If all publicly available annotations can be visualised on a network or map (for example), this can reveal links between researchers' work, potentially forming a

virtual collaboration by filling each other's gaps, as well as making connections to external resources.

Most of these features are already available in tools such as *Recogito* or *BRAT*; however, my research showed that less digitally confident researchers found digital annotation tools to be too disruptive to their workflow and tended to prefer pen and paper as a result. Standard input methods include highlighting and typing, whereas the option to 'handwrite' annotations on a tablet or phone, which can then be converted to text, might integrate more effectively with researchers' existing ways of working (provided they already own a compatible device). Further research would demonstrate whether connecting such functionality to existing annotation tools would be of benefit to those researchers who prefer manual annotation methods, i.e., would it be helpful to them, or the wider community, for their annotations to become digital and, where appropriate, semantically linked and publicly available?

Since my study took place, there seems to have been an increase in popularity of digital annotation tools for text- and image-based research in the Humanities and cultural heritage. In addition to *Recogito* and its image annotation counterpart *Annotorious*<sup>187</sup>, *Hypothes.is*<sup>188</sup> and *IIIF* viewers such as *Mirador*<sup>189</sup> have been widely adopted. *Hypothes.is* allows annotations to be made public or private, with the additional option to share only with a specified group of collaborators (Lemay, n.d.), although its primary aim is to facilitate research communication, rather than create semantically structured data. *IIIF* takes the approach of conceptualising all content as annotations; for example, an image is considered to be an annotation on a virtual canvas, a transcription of any text it contains is represented as an annotation superimposed at the relevant points, and comments or links are presented separately from the object to which they refer (T. Crane, 2017b). Like *Recogito*, annotations are structured using the *Web Annotation Data Model*<sup>190</sup> (T. Crane, 2017a), thereby producing semantic, interoperable outputs. *IIIF* technologies have been implemented

---

<sup>187</sup> <https://recogito.github.io/annotorious/>

<sup>188</sup> <https://web.hypothes.is/>

<sup>189</sup> <https://projectmirador.org/>

<sup>190</sup> <https://www.w3.org/TR/annotation-model/>

by multiple institutions including libraries, museums and archives, as well as the cultural heritage data management platform *Arches*<sup>191</sup> (Devolder, 2021).

As a result of my findings, and following the above developments in the area of semantic annotation, further research would be extremely helpful to identify how these tools, and their associated communities of practice, might most effectively work together to facilitate the integration of Linked Data with *Annotating*, alongside *Data Recognition* and *Visualization*. Although such research into digital annotation practices might take a similar approach to my survey and interview study, other possible formats might include user observations with existing tools or focus groups/workshops that bring multiple Ancient World researchers together to provoke discussion between them.

## 8.10 Conclusions

In this thesis, I have presented the rationale, methodology and findings from my survey and interview study into the usability of Linked Ancient World Data and its integration with existing research methods. This final chapter has provided a summary and discussed how these findings might be implemented in future Humanities projects more broadly, as well as identifying areas deserving of further research. As is evident in the above discussion, the primary factors that facilitate Linked Data usability could equally apply to digital tools and resources based on different data structures. However, because Linked Data production can be more complex from the outset and because the potential for reuse is so great, more care must be taken in ensuring its usability by the research community. As a result, more researchers will be able to benefit from its advantages, apply a similar approach to their existing research methods, and thereby contribute to the Linked Data ecosystem.

Usability is affected by decisions made during the planning stages of tool or resource development, or even during the grant application process. In particular, taking a collaborative, user-centred approach will produce tools and resources that are more likely to integrate with existing research methods. With regard to Linked Ancient World Data, I found that such methods might include *Discovering*, *Gathering*, *Data*

---

<sup>191</sup> <https://www.archesproject.org/>

*Recognition, Annotating* or *Visualization*, either individually or in combination. The user community would benefit greatly from further research into how these methods might most effectively be conducted in the digital space, as well as how a Linked Data approach might facilitate the tasks associated with them.

In a similar vein, the interlinked issues of awareness and training have a broader impact on Linked Data usability. While these might partially be mitigated by the provision of community-maintained resource directories, steps must be taken at an institutional or national level to ensure appropriate provision of digital skills training to researchers and students in Humanities disciplines. Improved institutional and national investment in Digital Humanities activities will also have the most significant impact on long-term tool and resource sustainability, ensuring usability and access to Linked Ancient World Data for all, now and into the future.

## References

- Adams, A., & Cox, A. L. (2008). Questionnaires, in-depth interviews and focus groups. In P. Cairns & A. L. Cox (Eds.), *Research Methods for Human Computer Interaction* (pp. 17–34). Cambridge University Press.  
<http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521870122&ss=toc>
- AHRC ICT Methods Network. (2008, March 31). *Resources by Method*. AHRC ICT Methods Network.  
<https://www.methodsnetwork.ac.uk/resources/method.html>
- Alexander, K., Cyganiak, R., Hausenblas, M., & Zhao, J. (2011, March 3). *Describing Linked Datasets with the VoID Vocabulary*. W3C. <https://www.w3.org/TR/void/>
- Almas, B., Babeu, A., & Krohn, A. (2014). Linked Data in the Perseus Digital Library. *ISAW Papers*, 7(3). <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/almas-babeu-krohn/>
- Almas, B., & Beaulieu, M.-C. (2016). The Perseids Platform: Scholarship for all! In G. Bodard & M. Romanello (Eds.), *Digital Classics Outside the Echo-Chamber* (pp. 171–186). Ubiquity Press. <https://doi.org/10.5334/bat>
- Almas, B., Beaulieu, M.-C., & Höflechner, G. (2015). *Perseids And Arethusa: Building Tools That Build Digital Humanists*. Digital Humanities 2015: Global Digital Humanities, Sydney, Australia.  
[http://dh2015.org/abstracts/xml/ALMAS\\_Bridget\\_May\\_Perseids\\_And\\_Arethusa\\_Building/ALMAS\\_Bridget\\_May\\_Perseids\\_And\\_Arethusa\\_Building\\_Tools.html](http://dh2015.org/abstracts/xml/ALMAS_Bridget_May_Perseids_And_Arethusa_Building/ALMAS_Bridget_May_Perseids_And_Arethusa_Building_Tools.html)
- Aloia, N., Binding, C., Cuy, S., Doerr, M., Fanini, B., Felicetti, A., Fihn, J., Gavrilis, D., Geser, G., Hollander, H., Meghini, C., Niccolucci, F., Nurra, F., Papatheodorou, C., Richards, J., Ronzino, P., Scopigno, R., Theodoridou, M., Tudhope, D., ... Wright, H. (2017). Enabling European Archaeological Research: The ARIADNE E-Infrastructure. *Internet Archaeology*, 43. <https://doi.org/10.11141/ia.43.11>
- American Numismatic Society. (n.d.). *Coinage of the Roman Republic Online*. Coinage of the Roman Republic Online. Retrieved 2 November 2020, from <http://numismatics.org/crro/>
- Angelis, S., Benardou, A., Chatzidiakou, N., Constantopoulos, P., Dallas, C., Dunning, A., Edmond, J., Ekman, S., Garnett, V., Gavrilis, D., Jordan, C., Kalogeropoulou, K.,

- Karozos, K., Lace, I., Latham, R., Lefferts, M., Makri, D.-N., Mantes, A., Papaki, E., ... Waterman, K. (2015). *Europeana Cloud Deliverable 1.3 User Requirements Analysis and Case Studies Report. Content Strategy Report*. Europeana.  
[http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_Cloud/Deliverables/D1.3%20D1.6%20User%20Requirements%20Analysis%20and%20Case%20Studies%20Report%20Content%20Strategy%20Report.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.3%20D1.6%20User%20Requirements%20Analysis%20and%20Case%20Studies%20Report%20Content%20Strategy%20Report.pdf)
- Angjeli, A., Mac Ewan, A., & Boulet, V. (2014). *ISNI and VIAF – Transforming ways of trustfully consolidating identities*. IFLA WLIC 2014, Lyon, France.  
<http://library.ifla.org/985/>
- Antonini, A., Adamou, A., Suárez-Figueroa, M. C., & Benatti, F. (under review).  
 Experiential Observations: An Ontology Pattern-based Study on Capturing the Potential Content within Evidences of Experiences. *Semantic Web Journal*.  
<http://www.semantic-web-journal.net/system/files/swj2761.pdf>
- Archaeology Data Service. (2019). *Finance Model*. Archaeology Data Service; re3data.org - Registry of Research Data Repositories.  
<https://archaeologydataservice.ac.uk/about/finance.xhtml>
- Arts and Humanities Data Service. (2010, April 20). *Arts and Humanities Data Service*.  
 Arts and Humanities Data Service.  
<https://web.archive.org/web/20100420183308/http://ahds.ac.uk/news/futureAHDS.htm>
- Arts and Humanities Research Council. (2015). *Technical plan*. Arts and Humanities Research Council.  
<http://www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan/>
- Arts and Humanities Research Council. (2018, March 4). *Technical plan*. Arts and Humanities Research Council.  
<https://web.archive.org/web/20180304084451/http://www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan/>
- Arts and Humanities Research Council. (2020). *Research Funding Guide: Version 5.1*.  
<https://ahrc.ukri.org/documents/guides/research-funding-guide1/>
- Ast, R., & Essler, H. (2018). Anagnosis, Herculaneum, and the Digital Corpus of Literary Papyri. In N. Reggiani (Ed.), *Digital Papyrology II: Case Studies on the Digital*

- Edition of Ancient Greek Papyri* (pp. 63–74). De Gruyter.  
<https://doi.org/10.1515/9783110547450-003>
- 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* (Vol. 10, pp. 53–72). De Gruyter Saur.  
<https://doi.org/10.1515/9783110599572-005>
- Bagnall, R. S., & Heath, S. (2018). Roman Studies and Digital Resources. *The Journal of Roman Studies*, 108, 171–189. <https://doi.org/10.1017/S0075435818000874>
- Baker, J. (2014, April 30). *Preserving Your Research Data*. Programming Historian.  
<https://programminghistorian.org/en/lessons/preserving-your-research-data>
- Bambaci, L., Boschetti, F., & Gratta, R. D. (2019). Qohelet Euporia: A Domain specific Language for the Encoding of the critical Apparatus. *International Journal of Information Science and Technology*, 3(5), 26–37.
- Barats, C., Schafer, V., & Fickers, A. (2020). Fading Away... The challenge of sustainability in digital studies. *Digital Humanities Quarterly*, 14(3).
- Barbera, M. (2013). Linked (open) data at web scale: Research, social and engineering challenges in the digital humanities. *JLIS.It*, 4(1), 91–101.
- Barker, E. (2020). What would Indy do? Resisting post-truth through the practice of annotation. In C. Roosevelt (Ed.), *Spatial Webs: Mapping Anatolian Pasts for Research and the Public* (pp. 5–31). Koç University Research Centre for Anatolian Civilizations. <http://oro.open.ac.uk/69161/>
- Barker, E. (2021, October 6). Early steps in digitally mapping Pausanias’s Description of Greece. *Pelagios*. <https://medium.com/pelagios/early-steps-in-digitally-mapping-pausaniass-description-of-greece-548301b2a54d>
- Barker, E., Bissell, C., Hardwick, L., Jones, A., Ridge, M., & Wolffe, J. (2012). Colloquium: Digital technologies: Help or hindrance for the humanities? *Arts and Humanities in Higher Education*, 11(1–2), 185–200.  
<https://doi.org/10.1177/1474022211428311>
- Barker, E., Bouzarovski, S., Pelling, C., & Isaksen, L. (2011). HESTIA (the Herodotus Encoded Space-Text-Imaging Archive): An interdisciplinary project. In European Science Foundation (Ed.), *Research Infrastructures in the Digital Humanities* (pp. 16–17). European Science Foundation.



[http://www.esf.org/fileadmin/Public\\_documents/Publications/spb42\\_RI\\_DigitalHumanities.pdf](http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf)

- Barker, E., & Terras, M. (2016). Greek Literature, the Digital Humanities, and the Shifting Technologies of Reading. *Oxford Handbooks Online*.  
<https://doi.org/10.1093/oxfordhb/9780199935390.013.45>
- Baumann, R. (2013). The Son of Suda On-Line. *The Digital Classicist 2013*, 91–106.
- Benardou, A., Constantopoulos, P., & Dallas, C. (2013). An Approach to Analyzing Working Practices of Research Communities in the Humanities. *Journal of Humanities & Arts Computing: A Journal of Digital Humanities*, 7(1/2), 105–127.  
<https://doi.org/10.3366/ijhac.2013.0084>
- Benardou, A., & Dunning, A. (2018). From Europeana Cloud to Europeana Research: Tools, users and methods. In A. Benardou, E. Champion, C. Dallas, & L. Hughes (Eds.), *Cultural Heritage Infrastructures in Digital Humanities* (pp. 136–152). Routledge.
- Berman, M. L., Åhlfeldt, J., & Wick, M. (2016). Historical Gazetteer System Integration: CHGIS, Regnum francorum Online, and GeoNames. In M. L. Berman, R. Mostern, & H. Southall (Eds.), *Placing Names: Enriching and Integrating Gazetteers* (pp. 110–125). Indiana University Press.  
<http://lup.lub.lu.se/record/f279fef4-4d0b-4242-9b23-328a8eb9d816>
- Berman, M. L., Mostern, R., & Southall, H. (2016). Introduction. In M. L. Berman, R. Mostern, & H. Southall (Eds.), *Placing Names: Enriching and Integrating Gazetteers* (pp. 1–11). Indiana University Press.
- Berners-Lee, T. (1998). *Semantic Web Road map*. W3C - Design Issues.  
<https://www.w3.org/DesignIssues/Semantic.html>
- Berners-Lee, T. (2010). *Linked Data—Design Issues*.  
<https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34.
- Binding, C., Tudhope, D., & Vlachidis, A. (2019). A study of semantic integration across archaeological data and reports in different languages. *Journal of Information Science*, 45(3), 364–386. <https://doi.org/10.1177/0165551518789874>
- Birnbaum, D. J. (2020, December 22). *Project checklist*. Obduradon.Org.  
<https://dh.obdurodon.org/project-checklist.xhtml>

- Biston-Moulin, S., & Thiers, C. (2018). The Karnak Project: A Comprehensive Edition of the Largest Ancient Egyptian Temple. In A. De Santis & I. Rossi (Eds.), *Crossing Experiences in Digital Epigraphy. From Practice to Discipline* (pp. 155–164). De Gruyter Open. <https://doi.org/10.1515/9783110607208-013>
- Bodard, G., Cayless, H., Depauw, M., Isaksen, L., Lawrence, F., & Rahtz, S. (2017). Standards for Networking Ancient Person data: Digital approaches to problems in prosopographical space. *Digital Classics Online*, 3(2), 28–43. <https://doi.org/10.11588/dco.2017.0.37975>
- Bordalejo, B. (2018). Minority Report: The Myth of Equality in the Digital Humanities. In E. Losh & J. Wernimont (Eds.), *Bodies of Information: Intersectional Feminism and Digital Humanities*. University of Minnesota Press. <https://dhdebates.gc.cuny.edu/read/untitled-4e08b137-aec5-49a4-83c0-38258425f145/section/46cde2dd-e524-4cbf-8c70-e76e2c98f015#ch18>
- Borek, L., Dombrowski, Q., Perkins, J., & Schöch, C. (2016). TaDiRAH: A Case Study in Pragmatic Classification. *Digital Humanities Quarterly*, 10(1). <http://www.digitalhumanities.org/dhq/vol/10/1/000235/000235.html>
- Bree, P. van, & Kessels, G. (2013). *About nodegoat*. Nodegoat. <https://nodegoat.net/about>
- Brepolis. (2017a). *L'Année philologique: Bibliography of the Classical World*. Brepolis. [https://aboutbrepolis.files.wordpress.com/2018/09/brepolis\\_apl\\_en.pdf](https://aboutbrepolis.files.wordpress.com/2018/09/brepolis_apl_en.pdf)
- Brepolis. (2017b, June 20). *L'Année philologique (APh)*. Brepolis. <https://about.brepolis.net/lannee-philologique-apl/>
- Brickley, D., & Miller, L. (2014, January 14). *FOAF Vocabulary Specification*. <http://xmlns.com/foaf/spec/>
- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. Taylor & Francis.
- Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, 8(2), 29–40.
- Broux, Y. (2017a). Ancient Profiles Exploited: First Results of Named Entity Recognition applied to Latin Inscriptions. *U Schyłku Starożytności - Studia Źródłoznawcze*, XVI, 11–33.

- Broux, Y. (2017b). Egyptian names and networks in Trismegistos (800 BC – AD 800). In G. Rosati & M. C. Guidotti (Eds.), *Proceedings of the XI International Congress of Egyptologists: Florence, Italy 23-30 August 2015* (pp. 64–68). Archaeopress.
- Broux, Y. (2017c). Towards a Universal Facebook of the Ancient World. In S. Orlandi, R. Santucci, F. Mambrini, & P. M. Liuzzo (Eds.), *Digital and Traditional Epigraphy in Context: Proceedings of the EAGLE 2016 International Conference* (pp. 343–352). Sapienza Università Editrice. <http://www.editricesapienza.it/node/7610>
- Broux, Y., & Depauw, M. (2015). Developing Onomastic Gazetteers and Prosopographies for the Ancient World Through Named Entity Recognition and Graph Visualization: Some Examples from Trismegistos People. In L. M. Aiello & D. McFarland (Eds.), *Social Informatics* (pp. 304–313). Springer International Publishing. [https://doi.org/10.1007/978-3-319-15168-7\\_38](https://doi.org/10.1007/978-3-319-15168-7_38)
- Bruneau, O., Lasolle, N., Lieber, J., Nauer, E., Pavlova, S., & Rollet, L. (2021). Applying and developing semantic web technologies for exploiting a corpus in history of science: The case study of the Henri Poincaré correspondence. *Semantic Web*, 12(2), 359–378. <https://doi.org/10.3233/SW-200400>
- Bruseker, G., Carboni, N., & Guillem, A. (2017). Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM. In M. L. Vincent, V. M. López-Mencherro Bendicho, M. Ioannides, & T. E. Levy (Eds.), *Heritage and Archaeology in the Digital Age: Acquisition, Curation, and Dissemination of Spatial Cultural Heritage Data* (pp. 93–131). Springer International Publishing. [https://doi.org/10.1007/978-3-319-65370-9\\_6](https://doi.org/10.1007/978-3-319-65370-9_6)
- Buchanan, S., Golden, P., Kansa, E. C., Rabinowitz, A., & Shaw, R. (2016). *Periods, Organized (PeriodO): A gazetteer of period assertions for linking and visualizing periodized data*. <https://hcommons.org/deposits/item/hc:12295/>
- Buil-Aranda, C., Hogan, A., Umbrich, J., & Vandenbussche, P.-Y. (2013). SPARQL Web-Querying Infrastructure: Ready for Action? In H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. Noy, C. Welty, & K. Janowicz (Eds.), *The Semantic Web – ISWC 2013* (pp. 277–293). Springer. [https://doi.org/10.1007/978-3-642-41338-4\\_18](https://doi.org/10.1007/978-3-642-41338-4_18)
- Buzi, P., Bogdani, J., & Berno, F. (2018). The ‘PATHs’ Project: An Effort to Represent the Physical Dimension of Coptic Literary Production (Third–Eleventh centuries). *Comparative Oriental Manuscript Studies Bulletin*, 4(1), 39–58.

- Calvanese, D., Liuzzo, P., Mosca, A., Remesal, J., Rezk, M., & Rull, G. (2016). Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire. *Engineering Applications of Artificial Intelligence*, 51, 212–229. <https://doi.org/10.1016/j.engappai.2016.01.005>
- Cayless, H. A. (2011). *The Papyrological Navigator: Project Integration with RDF*. 31–34. [https://dh2011.stanford.edu/wp-content/uploads/2011/05/DH2011\\_BookOfAbs.pdf](https://dh2011.stanford.edu/wp-content/uploads/2011/05/DH2011_BookOfAbs.pdf)
- Cayless, H. A. (2019). Sustaining Linked Ancient World Data. In M. Berti (Ed.), *Digital Classical Philology* (pp. 35–50). De Gruyter. <https://doi.org/10.1515/9783110599572-004>
- Celano, G. G. A. (2018). An Automatic Morphological Annotation and Lemmatization for the IDP Papyri. In N. Reggiani (Ed.), *Digital Papyrology II: Case Studies on the Digital Edition of Ancient Greek Papyri* (pp. 139–148). De Gruyter. <https://doi.org/10.1515/9783110547450-008>
- CITE Architecture. (2015, March 23). *Generating RDF*. CITE Architecture. <http://cite-architecture.github.io/hocuspocus/specs/hocuspocus/rdf/Rdf.html>
- Cobb, J. (2015). The Journey to Linked Open Data: The Getty Vocabularies. *Journal of Library Metadata*, 15(3–4), 142–156. <https://doi.org/10.1080/19386389.2015.1103081>
- Crane, G. R. (2014). *The Hellespont Project: Integrating Arachne and Perseus* (No. 100476). Tufts University. <http://dx.doi.org/10.17613/M6HH2Q>
- Crane, T. (2017a, March). *An Introduction to IIIF*. Digirati. <https://resources.digirati.com/iiif/an-introduction-to-iiif/index.html>
- Crane, T. (2017b, May). *Annotations: How IIIF resources get their content*. Digirati. <https://resources.digirati.com/iiif/an-introduction-to-iiif/annotations.html>
- Crymble, A. (2018, May 22). *Uses of the Programming Historian*. Programming Historian. <https://programminghistorian.org/posts/Uses-Of-The-Programming-Historian>
- Dallas, C., Chatzidiakou, N., Benardou, A., Bender, M., Berra, A., Clivaz, C., Cunningham, J., Dabek, M., Garrido, P., Gonzalez-Blanco, E., Hadalin, J., Hughes, L., Immenhauser, B., Joly, A., Kelpšienė, I., Kozak, M., Kuzman, K., Lukin, M., Marinski, I., ... Zebec, T. (2017). European survey on scholarly practices and

digital needs in the arts and humanities—Highlights Report. *Zenodo*.

<https://doi.org/10.5281/zenodo.260101>

della Porta, D. (2014). In-Depth Interviews. In D. della Porta (Ed.), *Methodological Practices in Social Movement Research* (pp. 228–261). Oxford Scholarship Online.

<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198719571.001.0001/acprof-9780198719571-chapter-10>

Delmas-Glass, E., & Sanderson, R. (2020). Fostering a community of PHAROS scholars through the adoption of open standards. *Art Libraries Journal*, 45(1), 19–23.

<https://doi.org/10.1017/alj.2019.32>

Depauw, M. (2018). Trismegistos: Optimizing Interoperability for Texts from the Ancient World. In A. De Santis & I. Rossi (Eds.), *Crossing Experiences in Digital Epigraphy: From Practice to Discipline* (pp. 193–201). De Gruyter.

<https://doi.org/10.1515/9783110607208-016>

Depauw, M. (2019, July 22). *Keep TM alive!* Trismegistos.

<https://www.trismegistos.org/keeptrismegistosalive.php>

Depauw, M., & Gheldof, T. (2014). Trismegistos: An Interdisciplinary Platform for Ancient World Texts and Related Information. In Ł. Bolikowski, V. Casarosa, P. Goodale, N. Houssos, P. Manghi, & J. Schirrwagen (Eds.), *Theory and Practice of Digital Libraries—TPDL 2013 Selected Workshops* (Vol. 416, pp. 40–52).

Springer International Publishing. [https://doi.org/10.1007/978-3-319-08425-1\\_5](https://doi.org/10.1007/978-3-319-08425-1_5)

Depauw, M., & Stolk, J. (2015). Linguistic Variation in Greek Papyri: Towards a New Tool for Quantitative Study. *Greek, Roman, and Byzantine Studies*, 55, 196–220.

Deutsches Archäologisches Institut. (n.d.). *About iDAI.chronontology*.

IDAi.Chronontology. Retrieved 23 April 2020, from

<https://chronontology.dainst.org/info/about>

Devolder, N. (2021, March). *Roadmap and Releases*. Arches Project.

<https://www.archesproject.org/roadmap/>

Digital Classicist. (2014, September 2). *Category:Tools*. The Digital Classicist Wiki.

<https://wiki.digitalclassicist.org/Category:Tools>

- Digital Classicist. (2016a, February 5). *Linked Ancient World Data Institute*. Digital Classicist Wiki.  
[https://wiki.digitalclassicist.org/Linked\\_Ancient\\_World\\_Data\\_Institute](https://wiki.digitalclassicist.org/Linked_Ancient_World_Data_Institute)
- Digital Classicist. (2016b, May 31). *Category:Projects*. The Digital Classicist Wiki.  
<https://wiki.digitalclassicist.org/Category:Projects>
- Digital Classicist. (2019a, August 13). *Category:Events*. The Digital Classicist Wiki.  
<https://wiki.digitalclassicist.org/Category:Events>
- Digital Classicist. (2019b, August 13). *Category:Linked open data*. The Digital Classicist Wiki. [https://wiki.digitalclassicist.org/Category:Linked\\_open\\_data](https://wiki.digitalclassicist.org/Category:Linked_open_data)
- Digital Classicist. (2019c, December 16). *Category:FAQ*. The Digital Classicist Wiki.  
<https://wiki.digitalclassicist.org/Category:FAQ>
- Dik, H. (2015). Review Essay: On First Looking into the Digital Loeb Classical Library. *The Classical Journal*, 110(4), 493–500. JSTOR.  
<https://doi.org/10.5184/classicalj.110.4.0493>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons, Incorporated.  
<http://ebookcentral.proquest.com/lib/open/detail.action?docID=1762797>
- Dombrowski, Q. (2014). What Ever Happened to Project Bamboo? *Literary and Linguistic Computing*, 29(3), 326–339. <https://doi.org/10.1093/lc/fqu026>
- Dombrowski, Q., & Perkins, J. (2014). TaDiRAH: Building Capacity for Integrated Access. *Dh+lib*. <http://acrl.ala.org/dh/2014/05/21/tadirah-building-capacity-integrated-access/>
- Earhart, A. E., Risam, R., & Bruno, M. (2021). Citational politics: Quantifying the influence of gender on citation in Digital Scholarship in the Humanities. *Digital Scholarship in the Humanities*, 36(3), 581–594.  
<https://doi.org/10.1093/lc/fqaa011>
- Edmond, J., Fischer, F., Romary, L., & Tasovac, T. (2020). Springing the Floor for a Different Kind of Dance: Building DARIAH as a Twenty-First-Century Research Infrastructure for the Arts and Humanities. In J. Edmond (Ed.), *Digital Technology and the Practices of Humanities Research* (pp. 207–234). Open Book Publishers. <https://doi.org/10.11647/OBP.0192>
- Eichmann-Kalwara, N., Jorgensen, J., & Weingart, S. B. (2018). Representation at Digital Humanities Conferences (2000–2015). In E. Losh & J. Wernimont (Eds.), *Bodies*

*of Information: Intersectional Feminism and Digital Humanities*. University of Minnesota Press. <https://dhdebates.gc.cuny.edu/read/untitled-4e08b137-aec5-49a4-83c0-38258425f145/section/5dcc1fee-caef-4c10-aa3c-08a9bbf0b68b#ch06>

Elliott, T., & Gillies, S. (2009). Digital Geography and Classics. *Digital Humanities Quarterly*, 3(1).

Elliott, T., Heath, S., & Muccigrosso, J. (2012). Report on the Linked Ancient World Data Institute. *Information Standards Quarterly*, 24(2/3), 43–45.  
<https://doi.org/10.3789/isqv24n2-3.2012.08>

ERC Scientific Council. (2019a). *Information for Applicants to the Synergy Grant 2020 Call Version 1.1*. European Research Council.  
[https://ec.europa.eu/research/participants/data/ref/h2020/other/guides\\_for\\_applicants/h2020-guide20-erc-syg\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/h2020-guide20-erc-syg_en.pdf)

ERC Scientific Council. (2019b). *Information for Applicants to the Proof of Concept Grant 2020 Call Version 1.1*. European Research Council.  
[https://ec.europa.eu/research/participants/data/ref/h2020/other/guides\\_for\\_applicants/h2020-guide20-erc-poc\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/h2020-guide20-erc-poc_en.pdf)

Esterberg, K. G. (2002). *Qualitative methods in social research*. McGraw-Hill.

European Research Council. (2021). *Information for Applicants to the Starting and Consolidator Grant Calls Version 2.0*. European Research Council.  
[https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/information-for-applicants\\_erc-stg\\_erc-cog\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/information-for-applicants_erc-stg_erc-cog_en.pdf)

Farrington, S. (2017, January 2). Review: Thesaurus Linguae Graecae. *Society for Classical Studies*. <https://classicalstudies.org/scs-blog/scott-farrington/review-thesaurus-linguae-graecae>

Faught, K. S., Whitten, D., & Green, K. W. (Jr). (2004). Doing Survey Research on the Internet: Yes, Timing Does Matter. *Journal of Computer Information Systems*, 44(3), 26–34. <https://doi.org/10.1080/08874417.2004.11647579>

Flanders, J., & Jannidis, F. (2016). Data Modeling. In S. Schreibman, R. G. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 229–237). John Wiley & Sons, Ltd.

Fostano, K., & Morreale, L. K. (2019a, January 31). PART 1 Active Projects Digital Documentation Process (DDP) Archiving Wizard. *The Digital Documentation*

- Process. <https://digitalhumanitiesddp.com/part-1-active-projects-digital-documentation-process-ddp-archiving-wizard/>
- Fostano, K., & Morreale, L. K. (2019b, January 31). PART 2 Active Projects Digital Documentation Process (DDP) Archiving Wizard. *The Digital Documentation Process*. <https://digitalhumanitiesddp.com/part-2-active-projects-digital-documentation-process-ddp-archiving-wizard/>
- Frunzeanu, E., Robineau, R., & MacDonald, E. (2016). Biblissima's Choices of Tools and Methodology for Interoperability Purposes = Biblissima: Selección de herramientas y de metodología para fomentar la interoperabilidad. *CIAN-Revista de Historia de Las Universidades*, 19, 115–132.  
<https://doi.org/10.20318/cian.2016.3146>
- Geonames. (n.d.). *About GeoNames*. Geonames. Retrieved 3 January 2019, from <https://www.geonames.org/about.html>
- Gerth, P. (2016). *ARIADNE D14.2: Pilot deployment experiments*. ARIADNE.  
[http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/D14.2\\_Pilot\\_deployment\\_experiments\\_Final.pdf](http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/D14.2_Pilot_deployment_experiments_Final.pdf)
- Gerth, P., Beck, D. M., Schmidle, W., & Cuy, S. (2016). Sculptures in the Semantic Web Using Semantic Technologies for the Deep Integration of Research Items in ARIADNE. In M. Matsumoto & E. Uleberg (Eds.), *Proceedings of the 44th Annual Conference on Computer Applications and Quantitative Methods in Archaeology CAA 2016—Oceans of data* (pp. 11–22).
- Geser, G. (2016). *ARIADNE WP15 Study: Towards a Web of Archaeological Linked Open Data*. ARIADNE. [http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE\\_archaeological\\_LOD\\_study\\_10-2016-1.pdf](http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE_archaeological_LOD_study_10-2016-1.pdf)
- Gibson, L., & Kahn, R. (2016). Digital Museums in the 21st Century: Global Microphones or Universal Mufflers? *Museological Review*, 20, 39–51.
- Gillies, S. (2015, April 21). *Technical Introduction to Places*. Pleiades: A Gazetteer of Past Places. <https://pleiades.stoa.org/help/technical-intro-places>
- Golden, P., & Shaw, R. (2016). Nanopublication beyond the sciences: The PeriodO period gazetteer. *PeerJ Computer Science*, 2. <https://doi.org/10.7717/peerj-cs.44>



- Granados García, P., & Middle, S. (2021, March 15). Collection and Promotion of Real-World LOD Resources: Report on Activity 5 for the LP6 Symposium. *The Stoa: A Review for Digital Classics*. <https://blog.stoa.org/archives/4011>
- Granados-García, P. L. (2020). *Cultural Contact in Early Roman Spain through Linked Open Data* [PhD, The Open University]. <http://oro.open.ac.uk/73887/>
- Grant, K., Dombrowski, Q., Ranaweera, K., Rodriguez-Arenas, O., Sinclair, S., & Rockwell, G. (2020). Absorbing DiRT: Tool Directories in the Digital Age. *Digital Studies/Le Champ Numérique*, 10(1), 1–18. <https://doi.org/10.16995/dscn.325>
- Green, H. E. (2014). Under the Workbench: An analysis of the use and preservation of MONK text mining research software. *Literary and Linguistic Computing*, 29(1), 23–40. <https://doi.org/10.1093/lc/fqt014>
- Grieshaber, F. (2016). Graph of Dated Objects and Texts. In M. Berti (Ed.), *Epigraphy Edit-a-thon: Editing chronological and geographic data in ancient inscriptions*. <https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa-221532>
- Grossner, K. (2019, October 28). Beta Release v0.2. *World-Historical Gazetteer*. <http://whgazetteer.org/2019/10/28/beta-release-v0-2/>
- Grossner, K., & Hill, T. (2017). *From Linking Places to a Linked Pasts Network*. [http://kgeographer.com/pubs/LinkedPastsNetwork\\_7Dec.pdf](http://kgeographer.com/pubs/LinkedPastsNetwork_7Dec.pdf)
- Grossner, K., Janowicz, K., & Kessler, C. (2016). Place, Period, and Setting for Linked Data Gazetteers. In M. L. Berman, R. Mostern, & H. Southall (Eds.), *Placing Names: Enriching and Integrating Gazetteers* (pp. 80–96). Indiana University Press. <https://vbn.aau.dk/en/publications/place-period-and-setting-for-linked-data-gazetteers>
- Grossner, K., & Meeks, E. (2014). *Topotime: Representing Historical Temporality*. Digital Humanities 2014, Lausanne, Switzerland. <http://dharchive.org/paper/DH2014/Paper-892.xml>
- Gruber, E. (2016). Using XForms to Create, Publish, and Manage Linked Open Data. *XML London 2016 Conference Proceedings*, 99–107. <https://xmllondon.com/2016/xmllondon-2016-proceedings.pdf>
- Gruber, E. (2018). Linked Open Data for Numismatic Library, Archive and Museum Integration. In M. Matsumoto & E. Uleberg (Eds.), *Oceans of Data: Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in*

- Archaeology* (pp. 55–62). Archaeopress. <https://caa-international.org/2019/01/28/caa-2016-proceedings-published/>
- Gruber, E., & Meadows, A. (2021). Numismatics and Linked Open Data. *ISAW Papers*, 20. <http://dlib.nyu.edu/awdl/isaw/isaw-papers/20-6/>
- Gruber, E., & Smith, T. J. (2015). Linked Open Greek Pottery. In F. Giligny, F. Djindjian, L. Costa, P. Moscati, & S. Robert (Eds.), *Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* (pp. 205–214). Archaeopress. <http://www.archaeopress.com/Public/displayProductDetail.asp?id=%7BE35F9954-5653-493D-884B-4A7D2DE66610%7D>
- Gunter, V., Matthias, S., & Vogeler, G. (2019, September). *Data exchange in practice: Towards a prosopographical API (Preprint)*. Biographical Data in a Digital World 2019, Varna, Bulgaria. <https://hcommons.org/deposits/item/hc:29017/>
- Hannemann, J., & Kett, J. (2010, August). *Linked Data for Libraries*. World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden. <https://www.ifla.org/past-wlic/2010/149-hannemann-en.pdf>
- Harvard University Press. (n.d.). *The Digital Loeb Classical Library*. Harvard University Press. Retrieved 30 July 2021, from <https://www.hup.harvard.edu/features/loeb/digital.html>
- Hawes, G. (2020, May 22). *User testers needed!* MYTHLAB. <https://www.manto-myth.org/blog/user-testers-needed>
- Hawes, G., & Smith, S. (2020, April 23). *A web of myth*. MYTHLAB. <https://www.manto-myth.org/blog/building-a-web-of-myth>
- Hawes, G., & Smith, S. (2021, July 1). *A dataset of mythical people with stable URIs*. MYTHLAB. <https://www.manto-myth.org/blog/a-dataset-of-mythical-names-with-stable-uris>
- Heath, S. (2018). SPARQL as a first step for querying and transforming numismatic data: Examples from Nomisma.org. In A. Meadows, F. Duryat, & S. Glenn (Eds.), *Alexander the Great. A Linked Open World* (pp. 35–52). Ausonius Éditions.
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1–136. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>

- Hedges, M., & Dunn, S. (2017). *Academic Crowdsourcing in the Humanities: Crowds, Communities and Co-Production*. Elsevier Science & Technology.  
<http://ebookcentral.proquest.com/lib/open/detail.action?docID=5123094>
- Hedges, M., Jordanous, A., Lawrence, K. F., Roueché, C., & Tupman, C. (2017). Computer-Assisted Processing of Intertextuality in Ancient Languages. *Journal of Data Mining & Digital Humanities*. <https://doi.org/10.46298/jdmdh.1375>
- Hedrick, C. (2018, October 14). *Review: Searching EAGLE (The Electronic Archive of Greek and Latin Epigraphy)*. Society for Classical Studies; Society for Classical Studies. <https://classicalstudies.org/scs-blog/charles-hedrick/review-searching-eagle-electronic-archive-greek-and-latin-epigraphy>
- Hering, K. (2014). Provenance Meets Source Criticism. *Journal of Digital Humanities*, 3(2). <http://journalofdigitalhumanities.org/3-2/provenance-meets-source-criticism/>
- Hickey, T. B., & Toves, J. A. (2014). Managing Ambiguity In VIAF. *D-Lib Magazine*, 20(7/8). <https://doi.org/10.1045/july2014-hickey>
- Holterhoff, K. (2017). From Disclaimer to Critique: Race and the Digital Image Archivist. *Digital Humanities Quarterly*, 11(3).
- Horne, R. (2020a). Mapping Power: Using HGIS and Linked Open Data to Study Ancient Greek Garrison Communities. In C. Travis, F. Ludlow, & F. Gyuris (Eds.), *Historical Geography, GIScience and Textual Analysis: Landscapes of Time and Place* (pp. 213–227). Springer International Publishing.  
[https://doi.org/10.1007/978-3-030-37569-0\\_13](https://doi.org/10.1007/978-3-030-37569-0_13)
- Horne, R. (2020b). Beyond lists: Digital gazetteers and digital history. *The Historian*, 82(1), 37–50. <https://doi.org/10.1080/00182370.2020.1725992>
- Hughes, L., Constantopoulos, P., & Dallas, C. (2016). Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines. In S. Schreibman, R. G. Siemens, & J. Unsworth (Eds.), *A new companion to digital humanities / edited by Susan Schreibman, Ray Siemens, and John Unsworth*. (pp. 150–170). John Wiley & Sons, Ltd.
- International Image Interoperability Framework. (n.d.). *How to Join the IIIF Consortium and Membership Benefits*. IIIF | International Image Interoperability Framework. Retrieved 26 October 2021, from <https://iiif.io/community/consortium/join/#benefits>

- International Organization for Standardization. (2011). *ISO 25964-1:2011(en), Information and documentation—Thesauri and interoperability with other vocabularies—Part 1: Thesauri for information retrieval*. International Organization for Standardization.  
<https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en>
- International Organization for Standardization. (2018). *ISO 9241-11:2018(en), Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:en>
- International Organization for Standardization. (2019). *ISO 9241-210:2019(en), Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems*. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en>
- International Organization for Standardization. (2020). *ISO 9241-110:2020(en), Ergonomics of human-system interaction—Part 110: Interaction principles*. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-110:ed-2:v1:en>
- Isaksen, L. (2011). *Archaeology and the semantic web* [PhD, University of Southampton]. <http://eprints.soton.ac.uk/206421>
- Isaksen, L., del Rio Riande, G., De León, R., & Hernández, N. (2018, June). *Getting to Grips with Semantic and Geo-annotation using Recogito 2*. Digital Humanities 2018, Mexico City. <https://dh2018.adho.org/en/getting-to-grips-with-semantic-and-geo-annotation-using-recogito-2/>
- Janevska, M., Jovanovik, M., & Trajanov, D. (2014, April). *HTML5 based Facet Browser for SPARQL Endpoints*. 11th International Conference for Informatics and Information Technology, Bitola, Macedonia.
- Jisc. (n.d.-a). *Preservation*. Jisc. Retrieved 4 May 2021, from <https://www.jisc.ac.uk/preservation>
- Jisc. (n.d.-b). *Research repository*. Jisc. Retrieved 4 May 2021, from <https://www.jisc.ac.uk/research-repository>
- Jisc. (2014, July 12). *Pelagios2*. Jisc.  
[https://web.archive.org/web/20140712144338/http://www.jisc.ac.uk/whatwedo/programmes/di\\_informationandlibraries/resourcediscovery/pelagios2.aspx](https://web.archive.org/web/20140712144338/http://www.jisc.ac.uk/whatwedo/programmes/di_informationandlibraries/resourcediscovery/pelagios2.aspx)
- Jones, K., & Muftic, S. (2020). Endangered African Languages Featured in a Digital Collection: The Case of the †Khomani San, Hugh Brody Collection. *Proceedings*

*of the First Workshop on Resources for African Indigenous Languages*, 1–8.

<https://www.aclweb.org/anthology/2020.rail-1.1>

- Jordanous, A., Lawrence, K. F., Hedges, M., & Tupman, C. (2012). Exploring manuscripts: Sharing ancient wisdoms across the semantic web. *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, 1–12. <https://doi.org/10.1145/2254129.2254184>
- Kahn, R., Isaksen, L., Barker, E., Simon, R., de Soto, P., & Vitale, V. (2021). Pelagios—Connecting Histories of Place. Part II: From Community to Association. *International Journal of Humanities and Arts Computing*, 15(1–2), 85–100. <https://doi.org/10.3366/ijhac.2021.0263>
- Kansa, E. C., Kansa, S. W., Wells, J. J., Yerka, S. J., Myers, K. N., DeMuth, R. C., Bissett, T. G., & Anderson, D. G. (2018). The Digital Index of North American Archaeology: Networking government data to navigate an uncertain future for the past. *Antiquity*, 92(362), 490–506. <https://doi.org/10.15184/aqy.2018.32>
- Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, 15(3), 261–266. <https://doi.org/10.1093/intqhc/mzg031>
- Kemman, M., Kleppe, M., & Scagliola, S. (2014). Just Google It—Digital Research Practices of Humanities Scholars. In C. Mills, M. Pidd, & E. Ward (Eds.), *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities*. HRI Online Publications. <https://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman>
- Keusch, F. (2015). Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. *Management Review Quarterly*, 65(3), 183–216. <https://doi.org/10.1007/s11301-014-0111-y>
- King's College London. (n.d.). *Undergraduate courses | Department of Digital Humanities*. King's College London. Retrieved 19 May 2021, from <https://www.kcl.ac.uk/ddh/undergraduate/undergraduate>
- Klump, J., & Huber, R. (2017). 20 Years of Persistent Identifiers – Which Systems are Here to Stay? *Data Science Journal*, 16(0), 9. <https://doi.org/10.5334/dsj-2017-009>

- Koch, W., & Koch, G. (2017). Aggregation and management of metadata in the context of Europeana. *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare*, 70(2), 170–178.
- Kuczera, A. (2017). Graphentechnologien in den Digitalen Geisteswissenschaften. *ABI Technik*, 37(3), 179–196. <https://doi.org/10.1515/abitech-2017-0042>
- Lauer, C., McLeod, M., & Blythe, S. (2013). Online Survey Design and Development: A Janus-Faced Approach. *Written Communication*, 30(3), 330–357. <https://doi.org/10.1177/0741088313488075>
- Lawrence, K. F., & Bodard, G. (2015). Prosopography is Greek for Facebook: The SNAP:DRGN Project. *Proceedings of the ACM Web Science Conference*, 44:1-44:2. <https://doi.org/10.1145/2786451.2786496>
- Lee, C. A. (Cal). (2011). A framework for contextual information in digital collections. *Journal of Documentation*, 67(1), 95–143. <https://doi.org/10.1108/002204111111105470>
- Lemay, K. (n.d.). *Annotating with Groups*. Hypothes.is. Retrieved 17 May 2021, from <https://web.hypothes.is/help/annotating-with-groups/>
- Lewis, C., & Rieman, J. (1993). *Task-Centered User Interface Design: A Practical Introduction*. HCI Bibliography. <http://hcibib.org/tcuid/>
- Limp, W. F. (2011). Web 2.0 and Beyond, or On the Web, Nobody Knows You're an Archaeologist. In E. C. Kansa, S. W. Kansa, & E. Watrall (Eds.), *Archaeology 2.0: New Approaches to Communication and Collaboration* (pp. 265–280). Cotsen Institute of Archaeology. <https://escholarship.org/uc/item/1r6137tb>
- Linked Art Contributors. (n.d.-a). *CIDOC-CRM Class Analysis*. Linked Art. Retrieved 26 February 2020, from [https://linked.art/model/profile/class\\_analysis.html](https://linked.art/model/profile/class_analysis.html)
- Linked Art Contributors. (n.d.-b). *LOUD: Linked Open Usable Data*. Linked Art. Retrieved 21 July 2021, from <https://linked.art/loud/>
- Liu, D., Bikakis, A., & Vlachidis, A. (2017). Evaluation of Semantic Web Ontologies for Modelling Art Collections. In M. Kirikova, K. Nørsvåg, G. A. Papadopoulos, J. Gamper, R. Wrembel, J. Darmont, & S. Rizzi (Eds.), *New Trends in Databases and Information Systems* (pp. 343–352). Springer International Publishing. [https://doi.org/10.1007/978-3-319-67162-8\\_34](https://doi.org/10.1007/978-3-319-67162-8_34)
- Liuzzo, P., Mambrini, F., & Franck, P. (2017). Storytelling and Digital Epigraphy-Based Narratives in Linked Open Data. In M. Ioannides, N. Magnenat-Thalmann, & G.

- Papagiannakis (Eds.), *Mixed Reality and Gamification for Cultural Heritage* (pp. 507–523). Springer International Publishing. [https://doi.org/10.1007/978-3-319-49607-8\\_20](https://doi.org/10.1007/978-3-319-49607-8_20)
- Lombardo, V., Karatas, T., Damiano, R., Mattutino, C., & Sasakura, M. (2020, September 29). Bringing Digital Curation to Archaeological Projects: Evidence from the BeArchaeo Project. *Proceedings of the AVI2CH Workshop on Advanced Visual Interfaces and Interactions in Cultural Heritage Co-Located with 2020 International Conference on Advanced Visual Interfaces (AVI 2020)*. AVI2CH 2020: Advanced Visual Interfaces and Interactions in Cultural Heritage, Ischia, Italy. <http://ceur-ws.org/Vol-2687/paper3.pdf>
- Loughborough University. (2021). *English with Digital Humanities BA*. Loughborough University. <https://www.lboro.ac.uk/study/undergraduate/courses/a-z/english-with-digital-humanities/>
- Lucas, S. R. (2014). Beyond the existence proof: Ontological conditions, epistemological implications, and in-depth interview research. *Quality & Quantity*, 48(1), 387–408. <https://doi.org/10.1007/s11135-012-9775-3>
- Mahony, S. (2011). Research communities and open collaboration: The example of the Digital Classicist wiki. *Digital Medievalist*, 6, Article 0. <https://doi.org/10.16995/dm.26>
- Mahony, S. (2017). The Digital Classicist: Building a Digital Humanities Community. *Digital Humanities Quarterly*, 11(3), Article 3. <http://www.digitalhumanities.org/dhq/vol/11/3/000335/000335.html>
- Mahony, S., & Bodard, G. (2010). Introduction. In G. Bodard & S. Mahony (Eds.), *Digital Research in the Study of Classical Antiquity* (pp. 1–11). Routledge. <https://pdfs.semanticscholar.org/a181/452b411d9233782bd2e76a8bfba30b815b4e.pdf>
- Mahony, S., & Pierazzo, E. (2012). Teaching Skills or Teaching Methodology? In B. D. Hirsch (Ed.), *Digital Humanities Pedagogy: Practices, Principles and Politics* (pp. 215–225). Open Book Publishers.
- Manning, P., & Mostern, R. (2015). *World-Historical Gazetteer*. <https://hcommons.org/deposits/item/hc:12283/>
- Mannocci, A., Casarosa, V., Manghi, P., & Zoppi, F. (2014). The Europeana Network of Ancient Greek and Latin Epigraphy Data Infrastructure. In S. Closs, R. Studer, E.

Garoufallou, & M.-A. Sicilia (Eds.), *Metadata and Semantics Research* (pp. 286–300). Springer International Publishing. [https://doi.org/10.1007/978-3-319-13674-5\\_27](https://doi.org/10.1007/978-3-319-13674-5_27)

MANTO. (2020). *Part 1: Principles*.

<https://static1.squarespace.com/static/5e8189e975b8de5d6147ef4b/t/5ef1bd0d28e7676403df6711/1592900882962/The+Bible+Part+1+23+June+2020.pdf>

Maron, N. L., & Pickle, S. (2014). *Sustaining the Digital Humanities: Host Institution Support Beyond the Start-up Phase*. Ithaca S+R.

<https://doi.org/10.18665/sr.22548>

Mayr, P., Tudhope, D., Clarke, S. D., Zeng, M. L., & Lin, X. (2016). Recent applications of Knowledge Organization Systems: Introduction to a special issue. *International Journal on Digital Libraries*, 17(1), 1–4. <https://doi.org/10.1007/s00799-015-0167-x>

McCarty, W., & Short, H. (2002). *Mapping the Field*. EADH - The European Association for Digital Humanities. <http://eadh.org/publications/mapping-field>

Meeks, E., & Grossner, K. (2012). ORBIS: An Interactive Scholarly Work on the Roman World. *Journal of Digital Humanities*, 1(3).

<http://journalofdigitalhumanities.org/1-3/orbis-an-interactive-scholarly-work-on-the-roman-world-by-elijah-meeks-and-karl-grossner/>

Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, C., Gavrilis, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C., & Vlachidis, A. (2017). ARIADNE: A Research Infrastructure for Archaeology. *Journal on Computing and Cultural Heritage*, 10(3), 18:1-18:27.

<https://doi.org/10.1145/3064527>

Meroño-Peñuela, A., Ashkpour, A., van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., & van Harmelen, F. (2014). Semantic technologies for historical research: A survey. *Semantic Web*, 6(6), 539–564. <https://doi.org/10.3233/SW-140158>

Michelson, D. A. (2016). Syriaca.org as a Test Case for Digitally Re-Sorting the Ancient World. In C. Clivaz, P. Dilley, & D. Hamidović (Eds.), *Ancient Worlds in Digital Culture* (pp. 59–85). Brill. [https://doi.org/10.1163/9789004325234\\_005](https://doi.org/10.1163/9789004325234_005)



- Miller, W. L., & Crabtree, B. F. (2004). Depth Interviewing. In S. N. Hesse-Biber & P. Leavy (Eds.), *Approaches to qualitative research: A reader on theory and practice* (pp. 185–202). Oxford University Press.
- Modest, W., & Lelijveld, R. (Eds.). (2018). *Words Matter: An Unfinished Guide to Word Choices in the Cultural Sector*. Tropenmuseum, Afrika Museum, Museum Volkenkunde, Wereldmuseum.  
[https://www.materialculture.nl/sites/default/files/2018-08/words\\_matter.pdf.pdf](https://www.materialculture.nl/sites/default/files/2018-08/words_matter.pdf.pdf)
- Morreale, L. (2019). *Medieval Digital Humanities and The Rite of Spring: Thoughts on Performance and Preservation*. BodoArXiv.  
<https://doi.org/10.34055/osf.io/7p2t6>
- Neuefeind, C., Schildkamp, P., Mathiak, B., Karadkar, U., Stigler, J., Steiner, E., Vasold, G., Tosques, F., Ciula, A., Maher, B., Newton, G., Arneil, S., & Holmes, M. (2020, July 21). *Sustainability Strategies for Digital Humanities Systems*. Digital Humanities 2020. <https://hcommons.org/deposits/item/hc:32079/>
- Niccolucci, F., & Hermon, S. (2017). Expressing reliability with CIDOC CRM. *International Journal on Digital Libraries*, 18(4), 281–287.  
<https://doi.org/10.1007/s00799-016-0195-1>
- Oates, B. J. (2005). *Researching Information Systems and Computing* (1 edition). SAGE Publications Ltd.
- OCLC. (2019, May 23). *Scholars' Contributions to VIAF*. OCLC Research.  
<https://www.oclc.org/research/themes/data-science/viaf-scholars.html>
- O'Connell, D. C., & Kowal, S. (1995). Basic Principles of Transcription. In J. A. Smith, R. Harré, & L. Van Langenhove (Eds.), *Rethinking Methods in Psychology* (pp. 93–105). SAGE Publications Ltd. <https://doi.org/10.4135/9781446221792>
- Office of Digital Humanities. (2021). *Notice of Funding Opportunity, Funding Opportunity Title: Digital Humanities Advancement Grants*. National Endowment for the Humanities.  
<https://www.neh.gov/sites/default/files/inline-files/Digital%20Humanities%20Advancement%20Grants%20NOFO%20June%202021%2020210115-HAA.pdf>
- Oldman, D., & Tanase, D. (2018). Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace. In D. Vrandečić, K.

- Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L.-A. Kaffee, & E. Simperl (Eds.), *The Semantic Web – ISWC 2018* (Vol. 11137, pp. 325–340). Springer International Publishing. [https://doi.org/10.1007/978-3-030-00668-6\\_20](https://doi.org/10.1007/978-3-030-00668-6_20)
- Omeka. (n.d.). *Project*. Omeka. Retrieved 12 March 2021, from <https://omeka.org/about/project/>
- Open Data Institute. (2018). *Open Standards for Data*. Open Standards for Data Guidebook. <https://standards.theodi.org/>
- Open Library of Humanities. (n.d.). *Open Library of Humanities: A better path to open access for the humanities*. Open Library of Humanities. Retrieved 6 May 2021, from [https://www.openlibhums.org/media/files/olh\\_prospectus.pdf](https://www.openlibhums.org/media/files/olh_prospectus.pdf)
- Orlandi, S., Giberti, L. M. C., & Santucci, R. (2014). EAGLE: Europeana Network of Ancient Greek and Latin Epigraphy. Making the Ancient Inscriptions Accessible. *Lexicon Philosophicum: International Journal for the History of Texts and Ideas*, 2, 315–326.
- Packard Humanities Institute. (n.d.). *Digital Text Archives*. Packard Humanities Institute. Retrieved 7 May 2019, from <https://packhum.org/texts.html>
- Packard Humanities Institute. (2017, November 1). *Searchable Greek Inscriptions*. PHI Greek Inscriptions. <https://inscriptions.packhum.org/>
- Pagé-Perron, É. (2017, February). *Expanding Digital Assyriology With Open Access and Machine Learning: A Cuneiform Digital Library Initiative White Paper for the Global Philology Project*. Global Philology Open Conference. <https://doi.org/10.17613/M6MP5H>
- Palmer, C. L., Teffeau, L. C., & Pirmann, C. M. (2009). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. OCLC Research. <http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf>
- Pantelia, M. (2011, February 24). *The Liddell-Scott Jones Lexicon*. Thesaurus Linguae Graecae. <http://stephanus.tlg.uci.edu/lsg/about.html>
- Papaki, E., Benardou, A., Chatzidiakou, N., Dallas, C., Hadalin, J., Tasovac, T., & Vosyliute, I. (2015, January 22). DARIAH Survey on Digital Practices in the Arts

- and Humanities. *Community Engagement: A DARIAH-EU VCC2 Working Group*.  
<http://dariahre.hypotheses.org/285>
- Papyri.info. (n.d.). *About Papyri.info*. Papyri.Info. Retrieved 7 May 2019, from  
<http://papyri.info/docs/about>
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Sage Publications.
- Perkins, J., Dombrowski, Q., Borek, L., & Schöch, C. (2014). Building Bridges to the Future of a Distributed Network: From DiRT Categories to TaDiRAH, a Methods Taxonomy for Digital Humanities. *International Conference on Dublin Core and Metadata Applications*, 181–183.
- Perseids. (n.d.). *Arethusa*. Arethusa. Retrieved 18 January 2021, from  
<https://www.perseids.org/tools/arethusa/app/#/>
- Pierazzo, E. (2016). Textual Scholarship and Text Encoding. In S. Schreibman, R. G. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 307–321). John Wiley & Sons, Ltd.
- Polis, S., & Razanajao, V. (2016). Ancient Egyptian Texts in Context. Towards a Conceptual Data Model (the Thot Data Model—Tdm). *Bulletin of the Institute of Classical Studies*, 59(2), 24–41. <https://doi.org/10.1111/j.2041-5370.2016.12036.x>
- Portigal, S. (2013). *Interviewing Users: How to Uncover Compelling Insights* (1st edition). Rosenfeld Media.
- Prag, J. R. W., & Chartrand, J. (2018). I. Sicily: Building a Digital Corpus of the Inscriptions of Ancient Sicily. In A. De Santis & I. Rossi (Eds.), *Crossing Experiences in Digital Epigraphy* (pp. 240–252). De Gruyter.  
<https://doi.org/10.1515/9783110607208-020>
- Prandoni, C., Fresa, A., Alfarano, N., Amato, G., Zoppi, F., Mannocci, A., Liuzzo, P., Mambrini, F., Orlandi, S., & Santucci, R. (2017). Searching inscriptions through the EAGLE Portal. In S. Orlandi, R. Santucci, F. Mambrini, & P. M. Liuzzo (Eds.), *Digital and Traditional Epigraphy in Context: Proceedings of the EAGLE 2016 International Conference* (pp. 173–185). Sapienza Università Editrice.  
<http://www.editricesapienza.it/node/7610>

- Presner, T. (2012). How to Evaluate Digital Scholarship. *Journal of Digital Humanities*, 1(4). <http://journalofdigitalhumanities.org/1-4/how-to-evaluate-digital-scholarship-by-todd-presner/>
- Rabinowitz, A. (2014). It's about time: Historical periodization and Linked Ancient World Data. *ISAW Papers*, 7. <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/rabinowitz/>
- Rabinowitz, A., Shaw, R., Buchanan, S., Golden, P., & Kansa, E. (2016). Making Sense of the Ways We Make Sense of the Past: The PeriodO Project. *Bulletin of the Institute of Classical Studies*, 59(2), 42–55. <https://doi.org/10.1111/j.2041-5370.2016.12037.x>
- Rabinowitz, A., Shaw, R., & Golden, P. (2018). Making up for Lost Time: Digital Epigraphy, Chronology, and the PeriodO Project. In A. De Santis & I. Rossi (Eds.), *Crossing Experiences in Digital Epigraphy* (pp. 202–215). De Gruyter. <https://doi.org/10.1515/9783110607208-017>
- Ramsay, S. (2004). Databases. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A Companion to Digital Humanities* (pp. 177–197). Blackwell Publishing Ltd. <https://doi.org/10.1111/b.9781405103213.2004.x>
- Reggiani, N. (2017). Digital Papyrology I: Methods, Tools and Trends. In *Digital Papyrology I*. De Gruyter. <https://doi.org/10.1515/9783110547474>
- Renear, A. H. (2004). Text Encoding. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A Companion to Digital Humanities* (pp. 218–239). Blackwell Publishing Ltd.
- Ridge, M. (2016). *Making digital history: The impact of digitality on public participation and scholarly practices in historical research* [PhD, Open University]. <http://oro.open.ac.uk/45519/>
- Rocco, A. (2017). EDB 2.0. How Eagle Europeana project improved the Epigraphic Database Bari. In S. Orlandi, R. Santucci, F. Mambrini, & P. M. Liuzzo (Eds.), *Digital and Traditional Epigraphy in Context: Proceedings of the EAGLE 2016 International Conference* (pp. 115–130). Sapienza Università Editrice. <http://www.editricesapienza.it/node/7610>
- Rockwell, G. (2012). Short Guide To Evaluation Of Digital Work. *Journal of Digital Humanities*, 1(4). <http://journalofdigitalhumanities.org/1-4/short-guide-to-evaluation-of-digital-work-by-geoffrey-rockwell/>

- Romanova, N., Ciula, A., Smithies, J., Jakeman, N., Murphy, O., Tupman, C., & Winters, J. (2021). *Capacity enhancement in Digital Humanities in the United Kingdom and Ireland: Training and beyond*. OSF Preprints.  
<https://doi.org/10.31219/osf.io/9e4az>
- Ross, S., Ballsun-Stanton, B., Sobotkova, A., & Crook, P. (2015). Building the Bazaar: Enhancing Archaeological Field Recording Through an Open Source Approach. In A. T. Wilson & B. Edwards (Eds.), *Open Source Archaeology: Ethics and Practice* (pp. 111–129). De Gruyter. <https://doi.org/10.1515/9783110440171-009>
- Santucci, R., Orlandi, S., Ferraro, A., Ragone, G., Giberti, L. M. C., Prandoni, C., & Fresa, A. (2016). *EAGLE Sustainability Plan: De 6.6*. Europeana. [https://www.eagle-network.eu/wp-content/uploads/2016/05/EAGLE\\_D6.6\\_Sustainability\\_Plan\\_V1.0.pdf](https://www.eagle-network.eu/wp-content/uploads/2016/05/EAGLE_D6.6_Sustainability_Plan_V1.0.pdf)
- Sauro, J. (2011). *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. CreateSpace Independent Publishing Platform.
- Schneider, P., Jones, J., Hiltmann, T., & Kauppinen, T. (2018). Challenge-derived design practices for a semantic gazetteer for medieval and early modern places. *Semantic Web*, 1, 1–20.
- Scholz, M., Holdenried, M., Dreyer, B., & Goerz, G. (2014). *EDEN – An Epigraphic Web Database of Ancient Inscriptions*. CIDOC 2014.  
[http://network.icom.museum/fileadmin/user\\_upload/minisites/cidoc/ConferencePapers/2014/J-2\\_Scholz\\_Holdenried\\_Dreyer\\_Goerz\\_paper.pdf](http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2014/J-2_Scholz_Holdenried_Dreyer_Goerz_paper.pdf)
- Schweizer, T., & Geer, B. (2019). Gravsearch: Transforming SPARQL to query humanities data. *Semantic Web Journal, Special Issue Cultural Heritage*.  
<http://www.semantic-web-journal.net/system/files/swj2412.pdf>
- Scriba, G., & Stockinger, U. (2016). Timeo interretem et dona ferentem. On ARACHNE and the Potential and Limits of Publishing Archaeological Catalogues Online. In I. Pinto Coelho, J. Bento Torres, L. Serrão Gil, & T. Ramos (Eds.), *Entre ciência e cultura: Da interdisciplinaridade à transversalidade da arqueologia. Actas das VIII Jornadas de Jovens em Investigação Arqueológica* (No. 4; Issue 4, pp. 497–502). CHAM. <https://edoc.unibas.ch/74530/>

- Shaw, R., Rabinowitz, A., & Golden, P. (2018). *A Deep Gazetteer of Time Periods*. Digital Humanities 2018, Mexico City. <https://dh2018.adho.org/en/a-deep-gazetteer-of-time-periods/>
- Shaw, R., Rabinowitz, A., Golden, P., & Kansa, E. (2016). A sharing-oriented design strategy for networked knowledge organization systems. *International Journal on Digital Libraries*, 17(1), 49–61. <https://doi.org/10.1007/s00799-015-0164-0>
- Simon, R., Barker, E., & Isaksen, L. (2012). Exploring Pelagios: A visual browser for geo-tagged datasets. *International Workshop on Supporting Users' Exploration of Digital Libraries, 23-27 September 2012, Paphos, Cyprus*. International Workshop on Supporting Users' Exploration of Digital Libraries, Paphos, Cyprus. <http://oro.open.ac.uk/34496/>
- Simon, R., Barker, E., Isaksen, L., & Cañamares, P. de S. (2017). Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2. *Journal of Map & Geography Libraries*, 13(1), 111–132. <https://doi.org/10.1080/15420353.2017.1307303>
- Simon, R., Isaksen, L., Barker, E., & de Soto Cañamares, P. (2016a). The Pleiades Gazetteer and the Pelagios Project. In M. L. Berman, R. Mostern, & H. Southall (Eds.), *Placing Names: Enriching and Integrating Gazetteers* (pp. 97–109). Indiana University Press.
- Simon, R., Isaksen, L., Barker, E., & de Soto Cañamares, P. (2016b). Peripleo: A Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time. *The Code4Lib Journal*, 31. <http://journal.code4lib.org/articles/11144>
- Simon, R., Pilgerstorfer, P., Isaksen, L., & Barker, E. (2014). Towards semi-automatic annotation of toponyms on old maps. *E-Perimetron*, 9(3), 105–112.
- Simon, R., Vitale, V., Kahn, R., Barker, E., & Isaksen, L. (2019). Revisiting Linking Early Geospatial Documents with Recogito. *E-Perimetron*, 14(3), 150–163.
- Simou, N., Chortaras, A., Stamou, G., & Kollias, S. (2017). Enriching and Publishing Cultural Heritage as Linked Open Data. In M. Ioannides, N. Magnenat-Thalmann, & G. Papagiannakis (Eds.), *Mixed Reality and Gamification for Cultural Heritage* (pp. 201–223). Springer International Publishing. [https://doi.org/10.1007/978-3-319-49607-8\\_7](https://doi.org/10.1007/978-3-319-49607-8_7)

- Smithies, J., Westling, C., Sichani, A.-M., Mellen, P., & Ciula, A. (2019). Managing 100 Digital Humanities Projects: Digital Scholarship & Archiving in King's Digital Lab. *Digital Humanities Quarterly*, 13(1).
- Smith-Yoshimura, K. (2013, November 25). First Scholars' Contributions to VIAF: Greek! *Hanging Together*. <https://hangingtogether.org/?p=3455>
- Smith-Yoshimura, K. (2014, March 11). New Scholars' Contributions to VIAF: Syriac! *Hanging Together*. <https://hangingtogether.org/?p=3634>
- Smith-Yoshimura, K. (2018). Analysis of 2018 International Linked Data Survey for Implementers. *The Code4Lib Journal*, 42. <https://journal.code4lib.org/articles/13867>
- Stenetorp, P., Pyysalo, S., & Topić, G. (n.d.). *Annotation examples*. Brat Rapid Annotation Tool. Retrieved 18 January 2021, from <https://brat.nlplab.org/examples.html>
- TaDiRAH. (2014a, July 18). *Annotating*. TaDiRAH - Taxonomy of Digital Research Activities in the Humanities. <http://tadirah.dariah.eu/vocab/index.php?tema=22&/annotating>
- TaDiRAH. (2014b, July 18). *Data Recognition*. TaDiRAH - Taxonomy of Digital Research Activities in the Humanities. <http://tadirah.dariah.eu/vocab/index.php?tema=9&/data-recognition>
- TaDiRAH. (2014c, July 18). *Discovering*. TaDiRAH - Taxonomy of Digital Research Activities in the Humanities. <http://tadirah.dariah.eu/vocab/index.php?tema=10&/discovering>
- TaDiRAH. (2014d, July 18). *Gathering*. TaDiRAH - Taxonomy of Digital Research Activities in the Humanities. <http://tadirah.dariah.eu/vocab/?tema=11>
- TaDiRAH. (2014e, July 18). *TaDiRAH - Taxonomy of Digital Research Activities in the Humanities*. <http://tadirah.dariah.eu/vocab/index.php>
- TaDiRAH. (2014f, July 18). *Visualization*. TaDiRAH - Taxonomy of Digital Research Activities in the Humanities. <http://tadirah.dariah.eu/vocab/index.php?tema=32&/visualization>
- Thaller, M. (2020, July 8). On vagueness and uncertainty in historical data. *A Digital Ivory Tower*. <https://ivorytower.hypotheses.org/88>
- The Andrew W. Mellon Foundation. (2019a). *Grant Proposal Guidelines: Arts and Culture*. The Andrew W. Mellon Foundation.

- [https://mellon.org/media/filer\\_public/7b/4a/7b4a57b0-7b75-444f-8ac4-b452ac2638ca/grant\\_proposal\\_guidelines\\_9-2019\\_-\\_arts\\_and\\_culture\\_1.pdf](https://mellon.org/media/filer_public/7b/4a/7b4a57b0-7b75-444f-8ac4-b452ac2638ca/grant_proposal_guidelines_9-2019_-_arts_and_culture_1.pdf)
- The Andrew W. Mellon Foundation. (2019b). *Grant Proposal Guidelines: Public Knowledge*. The Andrew W. Mellon Foundation.
- [https://mellon.org/media/filer\\_public/cf/94/cf9466f7-605d-4acb-b9d2-7f4e3374626f/grant\\_proposal\\_guidelines\\_9-2019\\_-\\_public\\_knowledge\\_2.pdf](https://mellon.org/media/filer_public/cf/94/cf9466f7-605d-4acb-b9d2-7f4e3374626f/grant_proposal_guidelines_9-2019_-_public_knowledge_2.pdf)
- The Andrew W. Mellon Foundation. (2020). *Grant Proposal Guidelines: Higher Learning, Public Affairs*. The Andrew W. Mellon Foundation.
- [https://mellon.org/media/filer\\_public/3d/b6/3db6c74a-6966-45f3-a40f-6549178a940b/grant\\_proposal\\_guidelines\\_9-2019\\_-\\_higher\\_learning\\_public\\_affairs\\_1.pdf](https://mellon.org/media/filer_public/3d/b6/3db6c74a-6966-45f3-a40f-6549178a940b/grant_proposal_guidelines_9-2019_-_higher_learning_public_affairs_1.pdf)
- The Digital Classicist Wiki. (2020, September 1). *Pelagios*. The Digital Classicist Wiki.
- <https://wiki.digitalclassicist.org/Pelagios>
- The Scaife Viewer Project. (n.d.). *The Scaife Viewer Project*. The Scaife Viewer Project. Retrieved 29 July 2021, from <https://scaife-viewer.org/>
- Thesaurus Linguae Graecae. (2015). *The History of the TLG*. Thesaurus Linguae Graecae. <http://stephanus.tlg.uci.edu/history.php>
- Thiery, F., & Engel, T. (2016). The Labelling System: A Bottom-up Approach for Enriched Vocabularies in the Humanities. In S. Campana, R. Scopigno, G. Carpentiero, & M. Cirillo (Eds.), *CAA2015. Keep the Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* (pp. 259–268). Archaeopress.
- <http://archaeopress.com/ArchaeopressShop/Public/displayProductDetail.asp?id={27710CAF-069B-43D3-A53B-4BB1A293F456}>
- Tiepmar, J., & Heyer, G. (2017). An Overview of Canonical Text Services. *Linguistics and Literature Studies*, 5(2), 132–148. <https://doi.org/10.13189/lis.2017.050209>
- Tolle, K., & Wigg-Wolf, D. (2016). How To Move from Relational to 5 Star Linked Open Data – A Numismatic Example. In S. Campana, R. Scopigno, G. Carpentiero, & M. Cirillo (Eds.), *Keep the Revolution Going: Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* (Vol. 1, pp. 275–281). Archaeopress.
- <http://archaeopress.com/Public/displayProductDetail.asp?id=%7B27710CAF-069B-43D3-A53B-4BB1A293F456%7D>



- Torallas Tovar, S., & Schubert, P. (2019, November 18). *The Campaign for Papyri.info*. Securing the Future of Papyrology. <https://www.supportpapyri.info/campaign/>
- Tóth-Czifra, E. (2021, February 10). OpenMethods Spotlights #2: Interview with Luise Borek and Canan Hastik about TaDiRAH. *OpenMethods*. <https://openmethods.dariah.eu/2021/02/10/openmethods-spotlights-2-interview-with-luise-borek-and-canan-hastik-about-tadirah/>
- Trismegistos. (n.d.). *Database Structure*. Trismegistos. Retrieved 2 June 2020, from [https://www.trismegistos.org/about\\_databasestructure.php](https://www.trismegistos.org/about_databasestructure.php)
- Trismegistos. (2019, July). *What changes on 1 January 2020?* Trismegistos. [https://www.trismegistos.org/keeptrismegistosalive\\_whatchanges.php](https://www.trismegistos.org/keeptrismegistosalive_whatchanges.php)
- Tuan, Y.-F. (1975). Place: An Experiential Perspective. *Geographical Review*, 65(2), 151. <https://doi.org/10.2307/213970>
- Tupman, C. (2021). Where Can Our Inscriptions Take Us? Harnessing the Potential of Linked Open Data for Epigraphy. In I. Velázquez Soriano & D. Espinosa Espinosa (Eds.), *Epigraphy in the Digital Age: Opportunities and Challenges in the Recording, Analysis and Dissemination of Inscriptions* (pp. 115–128). Archaeopress.
- Tupman, C., & Jordanous, A. (2014). Sharing Ancient Wisdoms across the Semantic Web Using TEI and Ontologies. In T. L. Andrews & C. Macé (Eds.), *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches* (Vol. 1, pp. 213–228). Brepols Publishers. <https://doi.org/10.1484/M.LECTIO-EB.5.102572>
- UCL Centre for Digital Humanities. (2019, January 14). *LAIRAH checklist for a successful project*. UCL Centre for Digital Humanities. [https://www.ucl.ac.uk/digital-humanities/research\\_project\\_pages/lairah/lairah\\_dh\\_checklist](https://www.ucl.ac.uk/digital-humanities/research_project_pages/lairah/lairah_dh_checklist)
- UCL Centre for Digital Humanities. (2020, October 9). *New Digital Humanities Undergraduate Module*. UCL. <https://blogs.ucl.ac.uk/dh/2020/10/09/new-digital-humanities-undergraduate-module/>
- UK Data Service. (n.d.). *Transcription*. UK Data Service. Retrieved 11 August 2018, from <https://www.ukdataservice.ac.uk/manage-data/format/transcription>
- UK Data Service. (2015). *UK Data Service—Data management costing tool and checklist*. UK Data Service. <https://ukdataservice.ac.uk/media/622368/costingtool.pdf>

- University of Edinburgh. (2019, January 21). *Digital Humanities for Literary Studies* (ENLI10378). University of Edinburgh. <http://www.drps.ed.ac.uk/18-19/dpt/cxenli10378.htm>
- Unsworth, J. (2000, May 13). *Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?* Symposium on 'Humanities Computing: formal methods, experimental practice', King's College London. <http://people.virginia.edu/~jmu2m/Kings.5-00/primitives.html>
- van Hooland, S., & Verborgh, R. (2014). *Linked data for libraries, archives and museums: How to clean, link and publish your metadata*. Facet Publishing.
- Vannini, L. (2018). Review of Papyri.info. *RIDE*, 9. <https://ride.i-d-e.de/issues/issue-9/papyri-info/>
- Varga, R., Pázsint, A.-I., Boda, I., & Deac, D. A. (2018). Romans 1 by 1: Overview of a Research Project. *Digital Classics Online*, 37–63. <https://doi.org/10.11588/dco.2018.2.45878>
- Verborgh, R., Vander Sande, M., Colpaert, P., Coppens, S., Mannens, E., & Van de Walle, R. (2014). Web-Scale Querying through Linked Data Fragments. In C. Bizer, T. Heath, S. Auer, & T. Berners-Lee (Eds.), *Proceedings of the 7th Workshop on Linked Data on the Web* (Vol. 1184). <https://linkeddatafragments.org/publications/ldow2014.pdf>
- Verborgh, R., Vander Sande, M., Hartig, O., Van Herwegen, J., De Vocht, L., De Meester, B., Haesendonck, G., & Colpaert, P. (2016). Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. *Journal of Web Semantics*, 37–38, 184–206. <https://doi.org/10.1016/j.websem.2016.03.003>
- Verreth, H. (2017). Trismegistos Places, a geographical index for all Latin inscriptions. In S. Orlandi, R. Santucci, F. Mambrini, & P. M. Liuzzo (Eds.), *Digital and Traditional Epigraphy in Context: Proceedings of the EAGLE 2016 International Conference* (pp. 201–208). Sapienza Università Editrice. <http://www.editricesapienza.it/node/7610>
- Vierros, M. (2018). Linguistic Annotation of the Digital Papyrological Corpus: Sematia. In N. Reggiani (Ed.), *Digital Papyrology II: Case Studies on the Digital Edition of Ancient Greek Papyri* (pp. 105–118). De Gruyter. <https://doi.org/10.1515/9783110547450-006>

- Vitale, V. (2016). Transparent, Multivocal, Cross-disciplinary: The Use of Linked Open Data and a Community-developed RDF Ontology to Document and Enrich 3D Visualisation for Cultural Heritage. In G. Bodard & M. Romanello (Eds.), *Digital Classics Outside the Echo-Chamber* (pp. 147–168). Ubiquity Press.  
<https://doi.org/10.5334/bat>
- Vlachidis, A., Bikakis, A., Terras, M., Kontiza, K., Naudet, Y., Kyriaki, D., Antoniou, A., & Bassani, M. (2018). *D2.4 Refined digital cultural resource data & data structure*. CrossCult. <https://discovery.ucl.ac.uk/id/eprint/10069121/1/CrossCult-D2.4.pdf>
- Vogt, W. P., Gardner, D. C., & Haeffele, L. M. (2017). *When to Use What Research Design*. Guilford Publications.  
<http://ebookcentral.proquest.com/lib/open/detail.action?docID=873354>
- W3C Network Working Group. (2004, September 1). *Hypertext Transfer Protocol—HTTP/1.1*. W3C - Protocols.  
<https://www.w3.org/Protocols/rfc2616/rfc2616.html>
- W3C Web Accessibility Initiative. (2019, June 5). *Introduction to Web Accessibility*. Web Accessibility Initiative (WAI).  
<https://www.w3.org/WAI/fundamentals/accessibility-intro/>
- Warwick, C., Galina, I., Rimmer, J., Terras, M., Blandford, A., Gow, J., & Buchanan, G. (2009). Documentation and the users of digital resources in the humanities. *Journal of Documentation*, 65(1), 33–57.  
<http://dx.doi.org.libezproxy.open.ac.uk/10.1108/00220410910926112>
- Warwick, C., Terras, M., Huntington, P., Pappa, N., & Galina, I. (2012). Case Study—The LAIRAH Project: Log analysis of internet resources in the arts and humanities. In C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 10–12). Facet Publishing. <https://ebookcentral-proquest-com.libezproxy.open.ac.uk/lib/open/detail.action?docID=1167435>
- Wettlaufer, J., Johnson, C., Scholz, M., Fichtner, M., & Thotempudi, S. G. (2015). Semantic Blumenbach: Exploration of Text–Object Relationships with Semantic Web Technology in the History of Science. *Digital Scholarship in the Humanities*, 30(suppl\_1), i187–i198. <https://doi.org/10.1093/llc/fqv047>

- Whitelaw, M. (2015). Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly*, 9(1).  
<http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. 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., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1).  
<https://doi.org/10.1038/sdata.2016.18>
- Williamson, G. (2014, January 28). Transcribing Conversation. *SLTinfo*.  
<https://www.sltinfo.com/transcribing-conversation/>
- WissKI. (n.d.). FAQ. WissKI. Retrieved 12 March 2021, from <http://wiss-ki.eu/faq-page>
- World-Historical Gazetteer. (2017, December 21). *About*. World-Historical Gazetteer.  
<http://whgazetteer.org/home/about/>
- World-Historical Gazetteer. (2018, January 2). Progress and Next Steps, Jan. 2018. *World-Historical Gazetteer*. <http://whgazetteer.org/2018/01/02/progress-and-next-steps-jan-2017/>
- Wusteman, J. (2017). Usability testing of the Letters of 1916 Digital Edition. *Library Hi Tech*, 35(1), 120–143. <https://doi.org/10.1108/LHT-10-2016-0111>

## Appendix 1. Survey: Using Digital Tools and Resources for Ancient World Research

### About This Survey

Digital tools and resources are increasingly being used as part of the research process in the Humanities in general and the Ancient World in particular. As part of my CHASE<sup>192</sup>-funded PhD in Classical Studies at the Open University, I am investigating the integration of digital tools and resources with methods used for researching the Ancient World. This survey is aimed at anyone involved in Ancient World research, with any level of digital expertise. Questions relate to your experiences of using digital tools/resources and their effectiveness for different research activities.

One particular approach in which I am interested is Linked Data. If you are familiar with using or producing Linked Data resources, you will be asked some additional questions.

For the purposes of this survey:

- Digital resources are defined as any material that can be consumed in an electronic format, including digitised or born-digital texts, images or artefacts, as well as websites, databases, catalogues, and interactive visualisations
- Digital tools are defined as software that enables the user to carry out a specific function relating to a digital resource (such tools may be online or installed on the user's computer)
- Ancient World research refers to the study of any civilisations existing prior to the end of the Western Roman Empire in 476 AD

All questions are optional; you are welcome to leave blank those which you do not wish to answer, and you can withdraw from the survey at any time before submitting your responses.

---

<sup>192</sup> <http://www.chase.ac.uk/>

Information you provide in this survey will be used for research purposes only. It will be anonymised and stored securely; subsequent analysis will form part of my PhD thesis, in which no individuals will be identifiable. An anonymised version of the results dataset will be made available to other researchers via the Humanities Commons CORE repository<sup>193</sup> after completion of this study.

The survey should take approximately 15 minutes to complete for those unfamiliar with Linked Data, and 20-30 minutes for those with experience of Linked Data use/production.

You are free to withdraw from the study without explanation or prejudice and to request the destruction of any data that have been gathered from you until 30 May 2018. After this point data will have been anonymised for the purpose of analysis and it will not be possible to withdraw any further data.

This research has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee<sup>194</sup> - HREC reference number: HREC/2018/2807/Middle. If you have any questions about this survey or my PhD research, please contact me at [sarah.middle@open.ac.uk](mailto:sarah.middle@open.ac.uk), or my primary supervisor Elton Barker at [elton.barker@open.ac.uk](mailto:elton.barker@open.ac.uk).

### About You

1. Region/state: [free text – short answer]

2. Country: [dropdown menu populated with the full list of countries at <http://www.textfixer.com/tutorials/dropdowns/country-list-iso-codes.txt>]

---

<sup>193</sup> <https://hcommons.org/core/>

<sup>194</sup> <http://www.open.ac.uk/research/ethics/>

3. Age group: [radio buttons]

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-54
- ☐ 55-64
- ☐ 65-74
- ☐ 75-84
- ☐ 85+

4. Gender: [radio buttons]

- ☐ Female
- ☐ Male
- ☐ Prefer not to say
- ☐ Prefer to self-describe: [free text – short answer]

5. Current role: [radio buttons]

- ☐ Student
- ☐ Early-career researcher
- ☐ Academic, e.g. Lecturer, (Assistant/Associate) Professor
- ☐ Museum/archive/library professional
- ☐ Developer
- ☐ Other – please state: [free text – short answer]

## Using Digital Resources for Ancient World Research

6. For which of the below research activities do you use digital tools/resources (tick all that apply)?

- ☐ Adding or creating metadata
- ☐ Annotating or marking up documents
- ☐ Collaborating
- ☐ Communicating
- ☐ Content analysis
- ☐ Contextualising, i.e. creating associations between resources
- ☐ Converting from one format to another
- ☐ Crowdsourcing
- ☐ Extracting pieces of information from a text (e.g. names, places)
- ☐ Gathering together related resources
- ☐ Geographic/spatial analysis
- ☐ Identifying, i.e. creating identifiers for digital objects
- ☐ Modelling data/information
- ☐ Network analysis
- ☐ Relational analysis, i.e. discovering relationships between resources
- ☐ Programming/coding
- ☐ Project management
- ☐ Publishing (making content available online, not necessarily in an official publication)
- ☐ Search/discovery
- ☐ Storage
- ☐ Structural Analysis
- ☐ Stylistic Analysis
- ☐ Teaching
- ☐ Transcribing
- ☐ Translating
- ☐ Visualising data (e.g. maps, graphs, timelines)
- ☐ Other activities (please state) [free text – short answer]



7. To what extent do you agree with the following statements about your use of digital tools/resources for research purposes?

	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	N/A
I regularly use digital tools/resources						
I am confident in using digital tools/resources						
I actively seek out new digital tools/resources						
My research would not be possible without digital tools/resources						
I am aware of the underlying data structures behind those digital tools/resources I regularly use						
I have sufficient skills/experience to teach others about digital tools/resources						
I have the ability to create my own digital tools/resources						

8. Which research tasks do you perform regularly that could not be replicated digitally, or that you prefer not to do on a computer? [free text – long answer]

9. Which tool(s)/resource(s) do you use most regularly, and why? [free text – long answer]

10. What features should be included in a good digital tool/resource? [free text – long answer]

11. What barriers have you encountered when attempting to use digital tools/resources in your research? [free text – long answer]

#### Linked Data

12. Are you familiar with the term 'Linked Data'? [radio buttons]

- ☐ Yes
- ☐ No
- ☐ Unsure

13. Have you ever knowingly used one or more tools or resources that are based on Linked Data? [radio buttons]

- ☐ Yes
- ☐ No
- ☐ Unsure

#### Using Linked Data Resources [displayed if answer to 13 is 'Yes']

14. Which Linked Data tools/resources have you used in your research (please include the names and URLs, if known)? [free text – long answer]

For the following questions in this section, please think about the Linked Data tool/resource with which you are most familiar.

15. Name of tool/resource: [free text – short answer]

16. URL of tool/resource: [free text – short answer]

17. Were the potential advantages of Linked Data made clear to you? [radio buttons]

- ☐ Yes
- ☐ No
- ☐ Unsure

17a. [displayed if answer to 17 is 'Yes'] What did you understand those advantages to be? [free text – long answer]

18. For which research activities did you intend to use this tool/resource (tick all that apply)?

- ☐ Adding or creating metadata
- ☐ Annotating or marking up documents
- ☐ Collaborating
- ☐ Communicating
- ☐ Content analysis
- ☐ Contextualising, i.e. creating associations between resources
- ☐ Converting from one format to another
- ☐ Crowdsourcing
- ☐ Extracting pieces of information from a text (e.g. names, places)
- ☐ Gathering together related resources
- ☐ Geographic/spatial analysis
- ☐ Identifying, i.e. creating identifiers for digital objects
- ☐ Modelling data/information
- ☐ Network analysis
- ☐ Relational analysis, i.e. discovering relationships between resources
- ☐ Programming/coding
- ☐ Project management
- ☐ Publishing (making content available online, not necessarily in an official publication)
- ☐ Search/discovery
- ☐ Storage
- ☐ Structural Analysis
- ☐ Stylistic Analysis
- ☐ Teaching
- ☐ Transcribing
- ☐ Translating
- ☐ Visualising data (e.g. maps, graphs, timelines)
- ☐ Other activities (please state) [free text – short answer]

19. To what extent do you agree with the following statements about this tool/resource?

	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
I think that I would like to use this tool/resource frequently					
I found the tool/resource unnecessarily complex					
I thought the tool/resource was easy to use					
I think that I would need the support of a technical person to be able to use this tool/resource					
I found the various functions in the tool/resource were well integrated					
I thought there was too much inconsistency in this tool/resource					
I would imagine that most people would learn to use this tool/resource very quickly					
I found the tool/resource awkward to use					
I felt very confident using the tool/resource					
I needed to learn a lot of things before I could get going with this tool/resource					

20. How did the experience of using this tool/resource compare with other digital tools/resources with which you are familiar?

21. To what extent did this tool/resource assist you in the research activities selected above?

22. How could the tool/resource be improved for future users?

### Producing Digital Resources

23. Have you been involved in the production of any digital research tools/resources?  
[radio buttons]

- ☐ Yes
- ☐ No

23a. [displayed if answer to 23 is 'Yes'] What is your preferred approach to structuring data (e.g. relational databases, text encoding, Linked Data) and why? [free text – long answer]

23b. [displayed if answer to 23 is 'Yes'] Have you have ever been involved in the production of a tool or resource that is based on Linked Data? [radio buttons]

- ☐ Yes
- ☐ No

24. [displayed if answer to 23b is 'No'] What are your reasons for not choosing a Linked Data approach? [free text – long answer]

For the following questions in this section, please think about the Linked Data tool/resource with which you were involved most recently.

25. [displayed if answer to 23b is 'Yes'] Name of tool/resource: [free text – short answer]

26. [displayed if answer to 23b is 'Yes'] URL of tool/resource: [free text – short answer]

27. [displayed if answer to 23b is 'Yes'] For which research activities was this tool/resource intended (tick all that apply)?

- ☐ Adding or creating metadata
- ☐ Annotating or marking up documents
- ☐ Collaborating
- ☐ Communicating
- ☐ Content analysis
- ☐ Contextualising, i.e. creating associations between resources
- ☐ Converting from one format to another
- ☐ Crowdsourcing
- ☐ Extracting pieces of information from a text (e.g. names, places)
- ☐ Gathering together related resources
- ☐ Geographic/spatial analysis
- ☐ Identifying, i.e. creating identifiers for digital objects
- ☐ Modelling data/information
- ☐ Network analysis
- ☐ Relational analysis, i.e. discovering relationships between resources
- ☐ Programming/coding
- ☐ Project management
- ☐ Publishing (making content available online, not necessarily in an official publication)
- ☐ Search/discovery
- ☐ Storage
- ☐ Structural Analysis
- ☐ Stylistic Analysis
- ☐ Teaching
- ☐ Transcribing
- ☐ Translating
- ☐ Visualising data (e.g. maps, graphs, timelines)
- ☐ Other activities (please state) [free text – short answer]

28. [displayed if answer to 23b is 'Yes'] Who was the target audience for the tool/resource (tick all that apply)?

- ☐ Ancient World researchers with a technical background
- ☐ Ancient World researchers with a non-technical background
- ☐ Developers
- ☐ Museum/archive/library professionals
- ☐ General public
- ☐ Other – please state: [free text – short answer]

29. [displayed if answer to 23b is 'Yes'] To what extent does your data comply with Tim Berners-Lee's 5 Star model<sup>195</sup> (tick all that apply)?

- ☐ ★ Available on the web with an open licence
- ☐ ★★ Available as machine-readable structured data
- ☐ ★★★ Available as a non-proprietary format
- ☐ ★★★★ Use open standards from W3C (RDF and SPARQL) to identify things
- ☐ ★★★★★ Link your data to other people's data to provide context

30. [displayed if answer to 23b is 'Yes'] Based on your experiences, would you choose Linked Data if producing a similar resource in future? [radio buttons]

- ☐ Yes
- ☐ No
- ☐ Unsure

31. [displayed if answer to 23b is 'Yes'] Do you have any other comments about the advantages/disadvantages of producing a Linked Data resource, and why you might (not) choose this approach in future? [free text – long answer]

#### Any Other Comments

32. Do you have any other comments about the use or production of digital resources for Ancient World research? [free text – long answer]

---

<sup>195</sup> <https://www.w3.org/DesignIssues/LinkedData.html#fivestar>

### Further Research

I will be following up this survey by interviewing a sample of respondents in more detail (either in person or via Skype, depending on location). If you would be willing to take part in a follow-up interview, please provide your name and email address.

33. Name: [free text – short answer]

34. Email address: [free text – short answer]

### End of Survey

Thank you for taking the time to complete this survey.

If you have any questions about this survey or my PhD research, please contact me at [sarah.middle@open.ac.uk](mailto:sarah.middle@open.ac.uk) or my primary supervisor Elton Barker at [elton.barker@open.ac.uk](mailto:elton.barker@open.ac.uk).



## Appendix 2. Information Leaflet for Interview Participants



*Sarah Middle*  
*Faculty of Arts and Social Sciences, Open University*  
*sarah.middle@open.ac.uk*

### Further information (Q&A) about:

#### **INTEGRATION OF LINKED DATA RESOURCES WITH EXISTING RESEARCH METHODOLOGIES IN CLASSICS AND RELATED DISCIPLINES - INTERVIEWS**

#### **What is the aim of this research?**

The purpose of this study is to understand the Ancient World researchers' experiences of using and/or producing digital tools/resources. The study will focus particularly on the use and production of Linked Data tools/resources, and the research methods with which such tools/resources might best be integrated.

#### **Who is conducting the research and who is it for?**

I (Sarah Middle) am carrying out this research on behalf of the Open University and the Consortium for the Humanities and the Arts South-East England, as part of a PhD in Classical Studies. I have received training in carrying out interview research.

#### **Why am I being invited to participate in this research?**

Last year, you participated in the survey 'Using Digital Tools and Resources for Ancient World Research' and indicated you would be willing to take part in a follow-up interview by providing your name and email address.

#### **If I take part in this research, what will be involved?**

I will be conducting interviews from November 2018 to February 2019. The interview will take approximately one hour and will be conducted at your workplace or via Skype, at a date and time that is convenient to you. To ensure your safety, I will carry photographic identification.

#### **What will we be talking about?**

The interview will consist of a series of questions and prompts, many of which will be based on your survey responses. I am particularly interested in finding out about how your use of digital tools/resources fits with your research

processes (as well as your experiences of producing digital tools/resources, if applicable). If you additionally answered questions in the survey about the use and/or production of Linked Data resources, I would like to find out more about your experiences by exploring some of your responses in greater detail.

### **Is it confidential?**

*[The following text was added for Linked Data producers only:* Part of the interview will involve discussing a Linked Data tool/resource that you have been involved in producing. Specific mention of this tool/resource in my thesis alongside your comments could therefore compromise your anonymity. As such, please let me know if you are willing to be identified in any comments relating to this tool/resource. If not, I will ensure that this tool/resource is mentioned only in general terms and not referred to by name. If you change your mind after the interview has taken place, please inform me before 28 February 2019.

For all other questions,] your participation will be treated in **strict confidence** in accordance with the Data Protection Act. No personal information will be passed to anyone outside my supervisory team. I will include a report of the findings from this study in my PhD thesis, but no individual will be identifiable in published results of the research.

### **What happens now?**

Please contact me at [sarah.middle@open.ac.uk](mailto:sarah.middle@open.ac.uk) to let me know when would be the most convenient time for you to take part in an interview for this study. If you would prefer not to take part in this research, please notify me via email at the above address. Your participation is entirely voluntary.

### **What if I have other questions?**

If you have any other questions about the study I would be very happy to answer them. Please email me at [sarah.middle@open.ac.uk](mailto:sarah.middle@open.ac.uk). If you have concerns that you would prefer to raise with an alternative contact, please email my primary supervisor Elton Barker at [elton.barker@open.ac.uk](mailto:elton.barker@open.ac.uk).

This research has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee - HREC reference number: HREC/2018/2807/Middle/2 (<http://www.open.ac.uk/research/ethics/>).

## Appendix 3. Consent Form for Interview Participants

### **CLASSICAL STUDIES, FACULTY OF ARTS AND SOCIAL SCIENCES**

#### **Consent form for persons participating in a research project**

#### **INTEGRATION OF LINKED DATA RESOURCES WITH EXISTING RESEARCH METHODOLOGIES IN CLASSICS AND RELATED DISCIPLINES - INTERVIEWS**

Name of participant:

Name of principal investigator(s): Sarah Middle

1. I consent to participate in this project, the details of which have been explained to me, and I have been provided with a written statement in plain language to keep.
2. I understand that my participation will involve an interview and I agree that the researcher may use the results as described in the plain language statement.
3. I acknowledge that:
  - a. the possible effects of participating in this research have been explained to my satisfaction
  - b. I have been informed that I am free to withdraw from the project without explanation or prejudice and to request the destruction of any data that have been gathered from me until it is anonymized at the point of transcription on 28 February 2019, After this point data will have been processed and it will not be possible to withdraw any unprocessed data I have provided
  - c. the project is for the purpose of research
  - d. I have been informed that the confidentiality of the information I provide will be safeguarded subject to any legal requirements
  - e. I have been informed that with my consent the data generated will be stored securely on the Open University's OneDrive system
  - f. I have been informed that other genuine researchers may request access to de-identified data in the future. Access will only be granted if they agree to preserve the confidentiality of the information as requested in this form.
  - g. If necessary any data from me will be referred to by a pseudonym in any publications arising from the research
  - h. I have been given contact details for a person whom I can contact if I have any concerns about the way in which this research project is being conducted
  - i. I have been informed that a summary copy of the research findings will be forwarded to me, should I request this.

I consent to this interview being audio-recorded

☐ **yes** ☐ **no**  
(please tick)

*[The following text was added for Linked Data producers only:*

I consent to being identified in the resulting thesis in relation to  
specific comments I have made regarding the production of  
named Linked Data tool/resource]

☐ **yes** ☐ **no**  
(please tick)

I wish to receive a copy of the summary project report on research findings ☐ **yes** ☐ **no**  
(please tick)

Email or postal address to which a summary should be sent:

Participant signature:

Date:

---

Contact details for the Principal Investigator (PI) and Research organisation and Faculty:

Sarah Middle, Classical Studies, Faculty of Arts and Social Sciences, Open University  
sarah.middle@open.ac.uk

Contact details for an alternative contact if you have any concerns about the way the research project is being conducted:

Elton Barker, Classical Studies, Faculty of Arts and Social Sciences, Open University  
elton.barker@open.ac.uk

This research has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee - HREC reference number: HREC/2018/2807/Middle/2 (<http://www.open.ac.uk/research/ethics/>).

## Appendix 4. Interview Script

### Introduction (All Groups)

Thank you for completing my survey and for offering to take part in the interview stage of my research. This interview should take approximately one hour and will:

- *Group 1*: discuss your experiences of using digital tools and resources for Ancient World research.
- *Group 2*: start with some brief introductory questions before discussing your experiences of producing digital tools and resources for Ancient World research. If time allows, I'll ask some additional questions about your use of digital tools and resources.
- *Groups 3, 6*: start with some brief introductory questions before discussing your experiences of producing Linked Data for Ancient World research. If time allows, I'll ask some additional questions about your use of digital tools and resources.
- *Group 4*: start with some brief introductory questions before discussing your experiences of using digital tools and resources in general, then moving on to talk about Linked Data in particular.
- *Group 5*: start with some brief introductory questions before discussing your experiences of producing digital tools and resources for Ancient World research. If time allows, I'll ask some additional questions about your use of digital tools and resources in general and Linked Data in particular.

As I mentioned in the information I sent, I'll be audio-recording this interview.

*Linked Data producers who are happy to be identified by name (Groups 3, 6)*: Thank you for agreeing to be identified by name in relation to comments about specific tools or resources you have produced. Any more general comments, for example those about your use of digital resources, will remain anonymous. If you say anything during the interview that you would prefer to be anonymised, please let me know.

*All other participants (Groups 1, 2, 4, 5):* Please be assured that your interview transcript will be anonymised and any comments of yours that I include in my thesis will not be identifiable.

You are welcome to ask questions or pause or stop the interview at any point. Also, please let me know if you would prefer for your answer to a particular question not to be recorded. Do you have any questions that you'd like to ask before I start the recording?

<Start recording>

The recording has started. Please could you confirm that you are happy to proceed with the interview?

<Pause to test the sound level of the recording>

Do you have any initial questions?

1. To start off with, please could you tell me a bit about what your role entails?

*(Follow-up questions:* How long have you been in this role? What aspects of the Ancient World particularly interest you? To what extent does your role involve the use or production of digital tools or resources?)

2. How confident are you generally with digital activities?

*(Follow-up questions:* Do you feel you are equipped with sufficient skills to use digital tools or resources in your research? What training might be most useful?)

3. Have you experienced any pressure to produce, or not to produce, digital tools or resources, as opposed to more traditional research outputs?

## Linked Data Production (Groups 3, 6)

As you mentioned in the survey that you've been involved in the production of Linked Data tools or resources, my first set of questions will focus on your production of digital tools or resources in general and Linked Data in particular. Then, if we have time, I'll ask some questions about your use of digital tools and resources.

4. What is your understanding of the term 'Linked Data'?

5. Do you think it's important to understand what Linked Data is to be able to use a Linked Data tool or resource effectively?

*(Follow-up questions: Why [not]?)*

6. You mentioned that your preferred data structure(s) is/are [DATA STRUCTURES]. Why is this?

*(Follow-up questions: Have you performed any research into which data structures might be most suitable for your work? [If so, what were your findings?] How has using [DATA STRUCTURES] helped you achieve your aims in producing digital tools or resources?)*

7. In the survey, you said that the most recent Linked Data tool or resource you'd been involved in producing was [RESOURCE]. Please could you tell me a bit more about it?

*(Follow-up questions: In what way does [RESOURCE] incorporate Linked Data? Why was Linked Data chosen over other technologies? What were the main aims in producing [RESOURCE]?)*

8. Before producing [RESOURCE], what training or experience did you have in producing digital tools or resources?

*(Follow-up questions: Do you feel that this training was sufficient? [Why/why not?] What impact did these experiences have on your approach?)*

9. Did you work individually or as part of a team?

9a. *If the participant worked as part of a team:* What was the composition of that team in terms of Ancient World/Humanities researchers, Digital Humanists, Computer Scientists, developers or information professionals? What was your role?

*(Follow-up questions:* What impact did the composition of the team have on your experiences?)

10. Did you receive any additional advice or support during the production process?

10a. *If the participant received additional advice/support:* What form did this take and what impact did it have on the project?

10b. *If the participant did not receive additional advice/support:* Could this have benefited you? What form would you have liked this advice or support to take?

11. When creating a specification for [RESOURCE], was this based on a particular topic or research question, or was the intention to have a broader remit?

*(Follow-up questions:* How do you think this affected the production process, the end result and the user experience? Would you approach this differently next time?)

12. In the survey, you associated the following research activities with [RESOURCE]: [RESEARCH ACTIVITIES]. Was it the intention from the outset that [RESOURCE] would incorporate all these research activities or did this develop over the course of the production process?

*(Follow-up questions:* What was the impact of this on the production process and the end result?)

13. What challenges did you encounter during the production process for [RESOURCE] and how did you address them?

*(Follow-up questions:* To what extent does [RESOURCE] differ from the original specification? Were there any barriers to departing from the original specification?)



14. In the survey, you identified that [RESOURCE] complies with [NUMBER] stars in Tim Berners-Lee's five star model. To what extent was this model considered during the production process?

15. In the survey, you associated multiple audiences with [RESOURCE], which included [AUDIENCES].

15a. *If the resource was developed for multiple audiences:* Why did you choose to develop [RESOURCE] for multiple audiences?

*(Follow-up questions:* Were all audiences considered equally throughout the production process or was one of these your primary audience? To what extent were the other audiences considered? How did having multiple potential audiences affect the production process?)

16. Did the development process incorporate user testing?

16a. *If 'Yes':* What form did this take, did you feel this was an important part of the process, and what difference did it make?

16b. *If 'No':* What were the barriers to incorporating user testing?

17. How was [RESOURCE] disseminated to potential users [i.e. what steps were taken to ensure that the relevant audiences would find out about it]?

18. What factors did you consider when designing the user interface?

19. What factors did you consider when putting together the documentation?

*(Follow-up questions:* What did you feel that users need to know when using the tool or resource? Is the documentation primarily usage instructions or did you include information about the underlying data structures? Why did you choose to take this approach?)

20. To what extent does [RESOURCE] cater for users with different levels of technical skill?

*(Follow-up questions: Is usage confined to the user interface or can users access the underlying data to perform more complex operations – such as via an API or SPARQL endpoint?)*

21. To what extent did you ensure that your data was complete before publishing [RESOURCE]?

*(Follow-up questions: How did you ensure this? Would you feel comfortable with publishing a resource online that you knew wasn't complete or needed further changes? Are you continuing to add to the data now the resource is published? [What is the process for this?])*

22. Did you consider incorporating user-generated content or crowdsourcing?

*(Follow-up questions: How have you managed this process? What impact has this had on [RESOURCE] and its usage?)*

23. To what extent are users able to reuse content or data from [RESOURCE]?

*(Follow-up questions: What measures are you taking to facilitate this?)*

24. What feedback have you received from your users?

25. To what extent does [RESOURCE] meet your original aims?

*(Follow-up questions: Do you consider it to be successful? [If so, how did you evaluate this success?] What factors do you think contributed to its success or otherwise?)*

26. Is there anything more that [RESOURCE] could do to exploit the full potential of Linked Data?

27. What measures have you taken to ensure the sustainability of [RESOURCE]?

*(Follow-up questions: To what extent are you able to plan for the long-term? How much autonomy do you have in making decisions about the future of [RESOURCE]? What is required for a tool or resource like [RESOURCE] to be self-sustaining? Are there any practical changes that could be made to funding models that would facilitate this?)*

28. How did your experiences of producing [RESOURCE] compare with your experiences of producing other digital tools or resources, either using Linked Data or other technologies?

*(Follow-up questions: Was your experience generally positive or negative? What lessons have you learned from producing [RESOURCE] that will improve the process for next time? Are there any barriers to implementing these improvements? What were the outcomes of choosing Linked Data over any other approach to structuring data? Did anything surprise you? Would using a different approach have made a significant difference?)*

29. In what contexts does Linked Data work well, and what are its limitations?

*(Follow-up questions: Where might it not be the right approach? In these cases, what other technologies or approaches might it be integrated with?)*

30. *If the participant has not used Linked Data (Group 3):* In the survey, you said you hadn't knowingly used a Linked Data tool or resource. How did this affect the process of producing a Linked Data tool or resource?

31. What do you feel are the main barriers to producing Linked Data resources for Ancient World research?

*(Follow-up questions: Do you have any ideas for potential solutions to these barriers?)*

Do you have anything more to add about producing Linked Data tools or resources before we move on?

### Non-Linked Data Tool/Resource Production (Groups 2, 5)

As you mentioned in the survey that you've been involved in the production of digital tools or resources, my first set of questions will focus on their production. Then, if we have time, I'll ask some questions about your use of digital tools and resources.

[Some of the following questions are duplicated from the above section and have been given the relevant numbers for ease of reference]

6. You mentioned that your preferred data structure(s) is/are [DATA STRUCTURES].

Why is this?

*(Follow-up questions: Have you performed any research into which data structures might be most suitable for your work? [If so, what were your findings?] How has using [DATA STRUCTURES] helped you achieve your aims in producing digital tools or resources?)*

32. Please could you tell me a bit about the tool or resource that you've produced most recently?

*(Follow-up questions: What data model was used in this resource? What were the main aims in producing this resource? What research activities was it intended for?)*

33. What types of research activity did you expect users to carry out with this tool or resource?

8. Before producing [RESOURCE], what training or experience did you have in producing digital tools or resources?

*(Follow-up questions: Do you feel that this training was sufficient? [Why/why not?] What impact did these experiences have on your approach?)*

9. Did you work individually or as part of a team?

9a. *If the participant worked as part of a team:* What was the composition of that team in terms of Ancient World/Humanities researchers, Digital Humanists, Computer Scientists, developers or information professionals? What was your role?

*(Follow-up questions:* What impact did the composition of the team have on your experiences?)

10. Did you receive any additional advice or support during the production process?

10a. *If the participant received additional advice/support:* What form did this take and what impact did it have on the project?

10b. *If the participant did not receive additional advice/support:* Could this have benefited you? What form would you have liked this advice or support to take?

11. When creating a specification for [RESOURCE], was this based on a particular topic or research question, or was the intention to have a broader remit?

*(Follow-up questions:* How do you think this affected the production process, the end result and the user experience? Would you approach this differently next time?)

13. What challenges did you encounter during the production process for [RESOURCE] and how did you address them?

*(Follow-up questions:* To what extent does [RESOURCE] differ from the original specification? Were there any barriers to departing from the original specification?)

34. Who were the intended audiences for [RESOURCE]?

34a. *If the resource was developed for multiple audiences:* Why did you choose to develop [RESOURCE] for multiple audiences?

*(Follow-up questions:* Were all audiences considered equally throughout the production process or was one of these your primary audience? To what extent were the other audiences considered? How did having multiple potential audiences affect the production process?)

16. Did the development process incorporate user testing?

16a. *If 'Yes'*: What form did this take, did you feel this was an important part of the process, and what difference did it make?

16b. *If 'No'*: What were the barriers to incorporating user testing?

17. How was [RESOURCE] disseminated to potential users [i.e. what steps were taken to ensure that the relevant audiences would find out about it]?

18. What factors did you consider when designing the user interface?

19. What factors did you consider when putting together the documentation?

*(Follow-up questions: What did you feel that users need to know when using the tool or resource? Is the documentation primarily usage instructions or did you include information about the underlying data structures? Why did you choose to take this approach?)*

20. To what extent does [RESOURCE] cater for users with different levels of technical skill?

*(Follow-up questions: Is usage confined to the user interface or can users access the underlying data to perform more complex operations – such as via an API or SPARQL endpoint?)*

21. To what extent did you ensure that your data was complete before publishing [RESOURCE]?

*(Follow-up questions: How did you ensure this? Would you feel comfortable with publishing a resource online that you knew wasn't complete or needed further changes? Are you continuing to add to the data now the resource is published? [What is the process for this?])*

22. Did you consider incorporating user-generated content or crowdsourcing?

*(Follow-up questions: How have you managed this process? What impact has this had on [RESOURCE] and its usage?)*

23. To what extent are users able to reuse content or data from [RESOURCE]?

*(Follow-up questions: What measures are you taking to facilitate this?)*

24. What feedback have you received from your users?

25. To what extent does [RESOURCE] meet your original aims?

*(Follow-up questions: Do you consider it to be successful? [If so, how did you evaluate this success?] What factors do you think contributed to its success or otherwise?)*

27. What measures have you taken to ensure the sustainability of [RESOURCE]?

*(Follow-up questions: To what extent are you able to plan for the long-term? How much autonomy do you have in making decisions about the future of [RESOURCE]? What is required for a tool or resource like [RESOURCE] to be self-sustaining? Are there any practical changes that could be made to funding models that would facilitate this?)*

35. How did your experiences of producing [RESOURCE] compare with your experiences of producing other digital tools or resources?

*(Follow-up questions: Was your experience generally positive or negative? What lessons have you learned from producing [RESOURCE] that will improve the process for next time? Are there any barriers to implementing these improvements? Did anything surprise you? Would using a different approach have made a significant difference?)*

36. *If participant is familiar with the term 'Linked Data':* In the survey, you also said that you were familiar with the term 'Linked Data' – what is your understanding of this term?

37. *If participant is familiar with the term 'Linked Data':* Have you ever considered producing a Linked Data tool or resource, or would you consider producing such a tool or resource in the future?

Do you have anything more to add about producing digital tools or resources before we move on?

#### Digital Tool/Resource Use (All Groups)

I'm now going to ask some questions about your research processes and your use of digital tools and resources generally.

38. Approximately, what proportion of your time is spent using digital tools or resources?

39. When using a digital tool or resource for the first time do you expect to be able to dive straight in or do you go first to the documentation about how to use it?

40. Can you tell me a bit about some of the digital tools or resources that you use most often, and in what ways they help or hinder your research process?

41. In the survey, you mentioned that you use digital tools or resources for [ACTIVITY – one of Data Recognition/Annotating/Visualization/Gathering] – please could you tell me a bit more about what tools or resources you use, how your use of these tools or resources fits in with your research process, and how effective they are for this research activity?

42. You also mentioned [ACTIVITY – one of Data Recognition/Annotating/Visualization/Gathering] – please could you tell me about your experiences using digital tools or resources for this research activity?  
[use above sub-questions as prompts if needed]



43. *If participant selected 'Discovering' in the survey:* You also mentioned that you use digital tools or resources for search and discovery. Which digital search tools have you found to be particularly effective and why?

*(Follow-up questions:* Do you tend to use basic or advanced search options provided by the tool or resource, or do you actively seek out ways of accessing the data or performing more complex queries? How would you ideally like the processes of discovering primary or secondary material to work?)

Before the interview, I asked you to put a series of barriers to digital tool or resource use into the order of how much they affect you [Cost, Training required, Bugs/technical issues, Incompatibility with device/operating system, Usability issues, Documentation unclear/non-existent, Functionality inconsistent with other tools/resources, Inaccurate/incomplete data, Scope too specialised, Difficult to relate to research goals, Unaware of what is possible].

44. Please could you tell me how you arrived at this ordering?

*(Follow-up questions:* Why are these barriers towards the top and these towards the bottom? Is there anything missing from the list? [If yes, where would this fit in your order?]) Are there any barriers here that you don't consider to be an issue for you?)

45. Which of these are more likely to prevent you from using a tool or resource completely, and which mean that you would persist in using the tool or resource while recognising that your experience would be more difficult as a result?

46. Can you give me any examples of digital tools and resources you've used that have any of these barriers?

*(Follow-up questions:* To what extent do these barriers affect how you use – or not use – these digital tools or resources?)

Before the interview, I asked you to put a series of features of good digital tools or resources into the order of how important they are to you [Ease of use/installation, Clear documentation, Search functionality, Export functionality, Accessibility, Open access, Citation instructions, Open standards, Reliability, Understands how Humanities researchers work].

47. Please could you tell me how you arrived at this ordering?

*(Follow-up questions: why are these features towards the top and these towards the bottom? I'm interested in where you've placed clear documentation and ease of use – please could you tell me a bit more about your reason for this ordering? Is there anything missing from the list? [If yes] where would this fit in your order?)*

48. Thinking about the digital tools and resources you use most often, which of them have the features you identified as being most important?

*(Follow-up questions: To what extent does the presence or absence of these features affect how you use a digital tool or resource?)*

49. *All producers (Groups 2, 3, 5, 6):* How does this list and your ordering compare with your priorities when producing digital tools and resources?

Before the interview, I also asked you to put a series of more specific features into the order of what would be of most interest to you [Provides access to multiple resources (e.g. collections) from a central point; Connects data about digital objects (e.g. texts, images, artefacts) based on their common features; Uses existing data about digital objects (e.g. texts, images, artefacts) and the relationships between them; Provides unique identifiers for digital objects (e.g. texts, images, artefacts); Provides contextual information about digital objects (e.g. texts, images, artefacts); Machine-readable data; Interoperable with other tools, resources, collections and datasets; Includes data visualisation tools; Disambiguates places, people or objects with similar names; Uses open standards].

50. Please could you tell me how you arrived at this ordering?

*(Follow-up questions: Why are these features towards the top and these towards the bottom?)*

51. Do you think you've used any tools or resources that have any of these features?

*(Follow-up questions: If so, which tools or resources? [If not already discussed] How was your experience of using these tools or resources?)*

52. *Non-producers familiar with the term 'Linked Data' (Groups 1, 4):* In the survey you mentioned that you were familiar with the term 'Linked Data' – what is your understanding of this term?

53. *Non-Linked Data users familiar with the term 'Linked Data' (Groups 1, 2, 3):* Based on what you already know, how helpful do you think Linked Data tools or resources would be for your research?

*(Follow-up questions: Is there anything that has prevented you from using a Linked Data tool or resource (if so, what?)*

54. How do you normally find out about digital tools or resources that might be relevant to your research?

*(Follow-up questions: What attracts you to a new digital tool or resource? How could the process of finding digital tools or resources for Ancient World research be improved?)*

55. Is there anything you'd like to be able to do with digital tools or resources that doesn't seem to be possible at the moment?

Do you have anything more to add about your research process and your use of digital tools or resources in general [before we move on to talk about Linked Data specifically]?

## Linked Data Use (Groups 4, 5, 6)

My final set of questions will be about your use of Linked Data resources.

56. You answered some questions about your experiences using [RESOURCE] – please could you tell me a bit about this tool/resource?

57. How was your experience of using [RESOURCE]?

*(Follow-up questions: Based on the available information, what were your expectations about [RESOURCE] before you used it? Did the fact that it was based on Linked Data affect your expectations? To what extent did it meet your expectations? What were its advantages, challenges or limitations? What could you achieve using the resource and what did you find difficult? Were any specific technical skills or training required to use [RESOURCE]? Do you think that basing [RESOURCE] on Linked Data made a difference to your experience as a user? [in what way?])*

58. Have you used any other tools or resources based on a Linked Data approach? [If so, which tools or resources have you used?

59a. Why did you decide to use these tools or resources in particular?

*(Follow-up questions: What did they offer you? Are there any alternatives (that do or do not use Linked Data) and how do they compare? How do they compare to [RESOURCE]?)*

59b. Are there any features that seem to characterise the Linked Data tools or resources you've used?

*(Follow-up questions: What do they have in common? Are they noticeably different from other digital tools or resources you're familiar with? Which ones have you found particularly effective or ineffective?)*

59c. What has your research gained from using Linked Data tools or resources?

*(Follow-up questions: Where, if at all, do you think a Linked Data tool or resource might best fit within your research process?)*

60. Thinking back to the features we talked about earlier, which of these were present in these Linked Data tools or resources?

61. What do you feel are the main barriers to using Linked Data resources?

*(Follow-up question: Do you have any ideas for potential solutions to these barriers?)*

62. What steps could developers take to ensure that future Linked Data resources are useful and usable for Ancient World research?

*(Follow-up questions: What would you like to do that these tools or resources can't?)*

Do you have anything more to add about your experiences using Linked Data tools or resources?

[Close](#)

Is there anything else you think I should be asking as part of this interview?

Is there anyone else you can think of who might be interesting to talk to about these topics?

That's the end of my questions. Thank you very much for taking part in this interview, it's been really helpful for my research. Do you have any more questions or comments before I stop the recording?

## Appendix 5. Documentation Checklist

The below checklist was created with Linked Ancient World Data tools and resources in mind but should be more broadly applicable across the Digital Humanities. Please note that, although serious consideration should be given to each point in the checklist, not all points will be relevant to all projects.

<b>General Information</b>	Brief overview of the tool/resource, including its main topic(s), research aims and key aspects of its technical architecture	
	Current and previous producers, with brief biographical details and institutional affiliation(s)	
	Up to date contact details, either for individuals or the production team as a whole	
	Current and previous funders, including link(s) to project on funder website(s), where available	
	Project start and end dates	
<b>Scope</b>	Research aims or questions that the tool/resource has been developed to address, with reference to existing publications and/or digital tools/resources	
	Disciplinary, geographic, temporal or thematic scope	
	Available languages, and how to access different language versions	
	Key areas of the above that users might expect to be included, but which are not	
	Links to smaller digital tools/resources that are included within this tool/resource	
	Links to larger digital tools/resources within which this tool/resource is also included	
	Future plans for expansion or further development, if any	
<b>Technical Requirements</b>	Compatible browser versions and/or operating systems	
	Software/plugins required to run the tool/resource (accompanied by a clear download link, with installation instructions, if required)	

<b>Functionality</b>	Key tasks for which the tool/resource is designed	
	Associated research methods to which these tasks might best apply (e.g. using TaDiRAH <sup>196</sup> terms)	
	Site map to demonstrate where key tasks can be performed and where documentation can be found	
	Explanation for why the functionality has been presented in this particular way	
	Mechanism for reporting bugs and technical issues	
<b>Tutorials</b>	Tutorials for each aspect of the tool or resource's functionality, from basic to advanced	
	Links to relevant (ideally free) training resources, or the producers' own tutorials available via external sites	
<b>Data Provenance</b>	Physical and digital sources used to create the data, with links if available (including the provenance of any physical objects digitised as part of tool/resource production)	
	Explanation for why these sources were selected (and why any possible alternatives were not selected)	
	Known limitations of these sources (e.g., if any are incomplete, exhibit historical/political bias, or are now considered to be out of date)	
	Tool(s), resource(s) and methodology(ies) used for data production and processing (accompanied by an explanation for why these were selected)	
	Alterations that have been made to the data as part of its inclusion in this tool/resource	
	How uncertainty in the data is represented (and how any associated probability values were calculated)	
	Known limitations of the resulting data, such as quality issues or gaps (e.g., if the resource is still a work in progress)	
<b>Data Model</b>	Standards and formats	
	URI structure(s)	

<sup>196</sup> <http://tadirah.dariah.eu/vocab/index.php>

	Diagram of data model(s) to illustrate structure and relationships between terms	
	Diagram of any new ontology(ies) created for the tool/resource	
	Links to external ontologies or authority files to which the data is connected	
	Explanation for data modelling decisions made by the producer(s)	
	Dependencies on, or integration with, external datasets	
<b>Data Access</b>	Instructions for interacting with the data via an API or SPARQL endpoint (e.g., variable names and data types with accompanying explanations/scope notes)	
	Example queries	
	Available export formats	
	Link(s) to download entire dataset (or parts thereof), either within the tool/resource itself or via a trusted repository	
	Details of (and explanations for) any access restrictions	
<b>Data Reuse</b>	Copyright information	
	Licence(s), e.g., Creative Commons <sup>197</sup>	
	Example citation	
	Contact form or instructions for how to (optionally) inform the data producer(s) of any reuse	
<b>User Contributions</b>	Instructions for how users might contribute their own data, e.g., adding or editing individual records, or batch upload of an entire dataset	
	Information about how to contribute code and/or join the development community, e.g., standards, programming language(s), commenting conventions (NB: in addition to documentation, code should include clear comments)	
	Details of the editorial/moderation process	

<sup>197</sup> <https://creativecommons.org/licenses/>



<b>System Development and Reuse</b>	Content management system (and version) on which the tool/resource is built	
	Reasons for selecting (or developing) this system	
	Instructions for installing and customising a new instance of the system	
<b>Publications</b>	Scholarly or informal publications about development and use of the tool/resource (e.g., journal articles, conference papers, blog posts)	
	Use cases for the tool/resource, ideally written by members of the user community to demonstrate its practical application, (at the outset these might need to be 'manufactured' by the producers)	
<b>Maintenance</b>	Details of who is responsible for maintaining the tool/resource, e.g., individual(s), institution(s) or infrastructure(s) (or a clear statement that the tool/resource is not currently being maintained)	
	Date when the tool/resource/dataset was last updated	
	Information about any significant changes to functionality or user experience	

Although this checklist was predominantly informed by participant responses from my study, I was assisted in its creation by existing publications from the Arts and Humanities Research Council (2015, 2020), Baker (2014), Barats et al. (2020), Birnbaum (2020), Fostano and Morreale (2019b, 2019a), Hering (2014), Morreale (2019), Presner (2012), Rockwell (2012), UCL Centre for Digital Humanities (2019), UK Data Service (2015), and Warwick et al. (2009).