International Journal of Digital Humanities (2023) 5:353–371
https://doi.org/10.1007/s42803-023-00072-z

ORIGINAL PAPER

A documentation checklist for (Linked) humanities data

Sarah Middle[1,2]

Received: 1 March 2023 / Accepted: 4 October 2023 / Published online: 6 November 2023

Abstract
Documentation, including information, instructions, and use cases, is key to reproducibility in Digital Humanities research and usability of resulting tools and resources. However, despite multiple studies that support this assertion, clear and comprehensive documentation is often lacking due to fundamental incompatibilities with existing funding models and the resulting prioritisation of project tasks. Through a user study of researchers involved in using and producing Linked Ancient World Data, supplemented by existing literature, I identified components for inclusion in documentation to facilitate use of these tools and resources, as well as the reproducibility of methods used in their production. At the same time, it became clear that producers would benefit from a solution to simplify the process of documentation creation. As a result, I assembled a documentation checklist whose scope for application reaches beyond Linked Ancient World Data to other Humanities disciplines and digital methods. This paper starts by discussing previous work, before providing an overview of my survey and interview methods. I then present my findings and discuss their implications for future research and development, including an introduction to the checklist and its implementation. To conclude, I draw together threads from the preceding sections and suggest wider structural changes to further facilitate and promote transparency and reproducibility in Digital Humanities.

Keywords Documentation · Digital humanities · Linked data · Reproducibility · Usability · User research

1 Introduction

Documentation is often key to the usability and reproducibility of digital tools and resources, and exists in different forms. For the purposes of this article, documentation

✉ Sarah Middle
sarah.middle@open.ac.uk

1 Classical Studies, Open University, Walton Hall, Kents Hill, Milton Keynes MK7 6AA, UK

2 Science and Technology, National Museums Scotland, Chambers Street, Edinburgh EH1 1JF, UK

can firstly include information about the tool or resource, such as an explanation of its underlying data structures[1], 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. 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. Documentation of this type can reduce the need for specialist training to use the tool or resource. Thirdly, use cases can serve to demonstrate the potential of a tool or resource, providing examples of how it has been used by others.

Various studies have made recommendations for Digital Humanities tool and resource documentation. Topics include components that should be incorporated into the documentation, as well as how it should be presented and shared. However, these studies also acknowledge that such documentation is often lacking due to fundamental incompatibilities with existing funding models and the resulting prioritisation of project tasks.

In this article, I present relevant findings and recommendations from a study into Linked Ancient World Data usability (see also Middle, 2022). Linked Data (Berners-Lee, 2010) is a set of technologies to facilitate the connection of digital objects, via machine-readable description and identification. In recent years, this approach has increasingly been applied in a Humanities research context, with uptake being relatively high in the Ancient World domain. Examples include the identification of places in Pleiades (Elliott and Gillies, 2009) and chronological periods in PeriodO (Rabinowitz et al., 2018). Other initiatives have aggregated data about sources and artefacts, such as EAGLE (Panciera and Orlandi, 2017), which focuses on inscriptions, and Papyri.info (Reggiani, 2017) on papyri. Furthermore, tools such as Recogito (Simon et al., 2019) have been developed for semantic annotation of texts and images. My study sought to investigate the usability of tools and resources such as these by conducting a survey and a series of interviews, which revealed (among other findings) the crucial role of effective documentation. To consolidate these findings into practical recommendations, I used participant responses, as well as previous studies discussed in the next section, to produce a checklist that will facilitate documentation creation for future projects. In addition to the Linked Ancient World Data domain, this checklist has been designed for broader applicability to other Humanities disciplines and data structures.

2 Background: digital humanities documentation

Clear, transparent, and open documentation is widely recognised as a key component of effective digital tools and resources, facilitating (re)use and reproducibility. For example, Edmond and Morselli (2020) recommend that projects are transparent about their technical architecture by providing open documentation. They go on to state that this documentation should incorporate and share the knowledge gained through

[1] While the underlying data itself might also be considered as a form of documentation, this is not within the scope of the current article.

the project, to assist future users as well as producers undertaking similar work in future. Similarly, in describing the *Digital Humanities ToolKit (DHTK)*, Picca and Egloff (2017) emphasise the attention they paid to producing clear and informative documentation, with the aim of encouraging users to contribute to potential future developments. Indeed, elsewhere in this special issue, Schöch highlights the importance of clear documentation to facilitate code reuse. More broadly, the provision of appropriate documentation promotes compliance with the FAIR (Findable, Accessible, Interoperable, Reusable) Guiding Principles for data management (Wilkinson et al., 2016), which recommend applying appropriate metadata standards and including sufficient descriptive information to enable reuse.

Vitale (2016) makes the additional observation that effective documentation promotes critical thinking about a digital tool or resource, thereby enabling scholarly engagement. Similarly, Fafinski (2022) discusses how every digital resource has its own narrative (e.g., decisions relating to production, digitisation, modelling and description) that must be communicated through documentation to ensure that reliable scholarly analysis and interpretation can take place. Indeed, Warwick et al. (2009) found that researchers tended to dismiss tools and resources with poor documentation, considering this a marker of low quality content.

However, many tools and resources are still lacking clear or adequate documentation, an issue compounded by its absence from many funder requirements for Digital Humanities projects. Back in 2008, the Log Analysis of Internet Resources in the Arts and Humanities (LAIRAH) project found that documentation of Digital Humanities resources was often inadequate and/or difficult to access (Warwick et al., 2008, 2009), findings echoed in later research by the Roy Rosenzweig Center for History and New Media (Gibbs and Owens, 2012). Although some progress has been made since then, a recent survey of digital editions found that only 20% provided "complete editorial and technical documentation" and infer that this suggests little has changed since the LAIRAH project results were reported over ten years previously (Franzini et al., 2019).

Reasons for the continued lack of effective documentation largely focus on the financial and time costs of its production, among the competing demands of managing a fixed-term digital project. In particular, Warwick et al. (2009) found during the LAIRAH project that participants tended to place a low priority on documentation because their main focus was the production of peer-reviewed publications. To address this issue, Flanders and Jannidis (2018) recommend that projects are planned in a way that recognises and incorporates the work involved. Other suggestions involve securing support from information professionals. For example, Poole and Garwood (2018) suggest that involving librarians or archivists from the early stages of a project will help to ensure reusability of data; one of the ways they can facilitate this process is by advising on the production of effective documentation. Galina Russell (2011) also recommends that the lack of (access to) documentation for Digital Humanities projects might be mitigated through library collaborations.

Further recommendations, such as those from the LAIRAH project (Warwick et al., 2008, 2009), involve making documentation production a mandatory condition of receiving funding for a digital project. While this is indeed the case for some funders, such as the US National Endowment for the Humanities (NEH) (Office of Digital Humanities, 2021), others do not consider documentation to be a sufficiently

high priority. For example, documentation was previously required in the UK's Arts and Humanities Research Council's (AHRC) 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). However, as indicated above, mandating documentation production alone is insufficient without the requisite support. Recognising this issue, both Franzini et al. (2019) and Warwick et al. (2008) recommend that any such mandate should be accompanied by funder provision of a "documentation template". Franzini et al. (2019) also suggest some elements that might be included, such as a project description, provenance and technical information, and reuse conditions.

In the domain of ontologies, several tools facilitate the production and publication of documentation, most notably *Live OWL Demonstration Environment* (*LODE*)[2] and *OWLDoc*[3], both of which were included in an evaluation of such tools by Peroni etal. (2013) and found to improve ontology usability. Additionally, the *Minimum Information for Reporting an Ontology* (*MIRO*) guidelines (Matentzoglu et al., 2018) provide a comprehensive list of components that should be included in ontology documentation, some of which might be applied in a Digital Humanities context (e.g., licence, scope, and methodology). However, a separate list of more general requirements for Digital Humanities might be a more helpful starting point in this area.

The closest initiative to a documentation template for Digital Humanities projects is the *Digital Documentation Process (DDP) Archiving Wizard* (Fostano and Morreale, 2019a, b), designed to facilitate preservation of data produced by Digital Humanities projects. The *Archiving Wizard* prompts the user to provide key information for inclusion in a catalogue record to accompany the data, as well as an "archiving dossier narrative" (a more in-depth explanation of the rationale behind the project, its achievements and publications). These two components are combined with a persistent identifier to ensure discoverability of the preserved project materials (Morreale, 2019). Although the resulting specification does not explicitly provide a template for documentation that supports the use of fully functional digital tools and resources, its catalogue record component does include various metadata fields that could equally be applied in this context.

Various initiatives (including those cited above) have also made recommendations for components of effective documentation; these include:

- General information about the tool/resource (e.g. topic, authors/developers, dates when it was produced and most recently updated) (Fostano and Morreale, 2019a; Morreale, 2019; Warwick et al., 2009; Sahle et al., 2014);
- Disciplinary and linguistic scope of the tool/resource (Fostano and Morreale, 2019a; Sahle et al., 2014);
- Key tasks for which the tool/resource is designed, with a site map to show where they can be accessed (Fostano and Morreale, 2019a; Gibbs and Owens, 2012; Morreale, 2019; Presner, 2012; Warwick et al., 2009; Sahle et al., 2014);
- Provenance of physical and digital sources used to create the data, explaining how they were selected and processed, and providing transparency about any limitations

[2] https://essepuntato.it/lode/

[3] https://code.google.com/archive/p/co-ode-owl-plugins/wikis/OWLDoc.wiki

(Arts and Humanities Research Council, 2018; Barats et al., 2020; Birnbaum, 2020; Fostano and Morreale, 2019a; Hering, 2014; Tóth-Czifra, 2020; UK Data Service, 2015; Warwick et al., 2008, 2009; Wilkinson et al., 2016; Sahle et al., 2014);

- Explanation of the data model(s), standards and formats, and the use of any pre-existing software, such as content management systems (Arts and Humanities Research Council, 2018; Barats et al., 2020; Fostano and Morreale, 2019a; Hering, 2014; Presner, 2012; Rockwell, 2012; Warwick et al., 2009; Wilkinson et al., 2016; Sahle et al., 2014);
- Licence(s) for (re)use, e.g., Creative Commons[4] (Birnbaum, 2020; Fostano and Morreale, 2019a; Wilkinson et al., 2016);
- Information about how users might contribute to the tool/resource (Rockwell, 2012);
- Use cases, to demonstrate how the tool/resource has been applied in practice (Gibbs and Owens, 2012);
- Links to publications about the tool/resource (Fostano and Morreale, 2019a).

Alongside the above recommendations, Gibbs and Owens (2012) emphasise the use of clear and simple language throughout, avoiding making assumptions about the technical backgrounds of end users.

Beyond Digital Humanities, in the domain of software development, there is general agreement that factors such as accessibility, accuracy, completeness, consistency, currency, maintainability, readability, trustworthiness, usability, and usefulness are key to effective documentation (Aghajani et al., 2019; Chen and Huang, 2009; Forward and Lethbridge, 2002; Plösch et al., 2014; Zhi et al., 2015). However, Zhi et al. (2015) caution that more work is needed on cost-benefit analysis of documentation production. While it is recognised that substantial time and effort is involved, this has not been quantified in a manner that would allow comparison with the extent of benefits that documentation can provide. Nevertheless, Aghajani et al. (2019) emphasise that documentation is intrinsic to the development of effective software. In particular, they suggest that quality issues might be addressed by introducing automated processes, such as the On-Demand Developer Documentation (OD3) system proposed by Robillard et al. (2017). However, in doing so, they additionally recommend ensuring that documentation is organised, presented and updated in a way that is accessible to humans. Similarly, Chen and Huang (2009) stress the importance of a long-term maintenance plan that prioritises keeping documentation (and code) up to date.

The issues discussed in this section are therefore not unique to Digital Humanities, nor have they been effectively resolved elsewhere. However, it should be noted that the audience for the documentation discussed in the software engineering literature tended to be practitioners or developers. In contrast, the Digital Humanities literature (and, indeed, my study) makes recommendations based on a broader range of end user needs, some of whom do not have substantial technical experience. The current article builds on these previous recommendations by presenting a documentation checklist for future Linked Humanities Data projects, based on evidence collected during a study of the user needs of Ancient World researchers.

[4] https://creativecommons.org/licenses/

3 Methods

Several recent projects have conducted user research into how digital tools and resources are being used as part of the wider Digital Humanities ecosystem, incorporating quantitative and qualitative approaches, to achieve both breadth and depth. While various methods have been effectively implemented during these projects, some of the most prominent have involved surveys, interviews, or a combination of the two. These include the Europeana Cloud initiative (Angelis et al., 2015); the research by Kemman et al. (2014) on searching behaviours of scholars, and user studies on the DanteSources (Bartalesi et al., 2018) and Newstracker (Kleppe and Otte, 2017) platforms. In learning from their experiences, I opted to conduct a survey followed by a series of interviews, to build up a broad picture of the research area in question (Linked Data use and production in Ancient World research), supplemented by more nuanced information on participants' experiences.

This combined approach facilitated the gathering of quantitative data from 212 Ancient World researchers via an online survey[5] (during Spring 2018), which was then enriched by qualitative information from detailed semi-structured interviews[6] (between Autumn 2018 and Spring 2019) with 16 survey participants. To reach potential, as well as actual, Linked Data users and producers, I targeted the survey at Ancient World researchers who use digital tools and resources in their work, with additional questions for those participants who had used and/or produced Linked Data. The study was approved by the Open University's Human Research Ethics Committee (ref. HREC/2018/2807/Middle), with informed consent obtained from all participants.

The survey started with standard demographic questions, after which survey participants were asked to express their level of agreement with a series of statements to assess their digital competence/confidence, in a similar way to the "information retrieval self-efficacy" questions asked in the Kemman et al. (2014) survey. Participants' responses were scored between 1 (Strongly Disagree) and 5 (Strongly Agree), with the majority scoring between 4 and 5 for statements such as "I regularly use digital tools/resources" and "I am confident in using digital tools/resources". For more complex statements, such as "I have sufficient skills/experience to teach others about digital tools/resources" and "I have the ability to create my own digital tools/resources", the majority of participants scored between 2 and 4. On calculating the mean score for each participant across all statements, I found that the majority scored between 3.14 and 4.57, indicating that participants were generally confident about their digital competence.

In the main body of the survey, participants were asked about their use of digital tools and resources, with additional questions for those who had knowingly used Linked Data (49 participants, 27%) and those who had been involved in digital tool or resource production (66 participants, 31%). Of these broad, open-ended questions, the following yielded responses relating to the topic of documentation:

- What features should be included in a good digital tool/resource?

[5] Survey data is available via the Humanities Commons CORE repository at https://doi.org/10.17613/1291-4r33.

[6] Interview transcripts are available from the author on request.

- What barriers have you encountered when attempting to use digital tools/resources in your research?
- How did the experience of using [your preferred Linked Data tool/resource] compare with other digital tools/resources with which you are familiar?
- How could [your preferred Linked Data tool/resource] be improved for future users?

Of the above questions, the first two were aimed at all participants and sought to identify features and barriers of digital tools and resources in general, to contextualise my findings on Linked Data in particular. The third question was aimed more specifically at Linked Data users, to allow them to compare their experiences of digital tools and resources based on different technologies. Similarly, the fourth question aligns directly with one of the key aims of the study, to identify recommendations to improve the usability of Linked Ancient World Data tools and resources.

After initial analysis of survey responses, I selected a sample of 16 interview participants ranging from occasional users of digital tools and resources to experienced Linked Data users and producers, and who were broadly representative of the survey population demographics. I developed interview scripts to investigate issues raised in the survey in more detail and tailored them for each participant based on their responses. In addition to further discussion of participants' responses to the survey questions mentioned above, I included some questions that asked specifically about experiences of using and producing documentation:

- (For all participants) 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?
- (For participants who had produced a digital tool/resource) What factors did you consider when putting together the documentation?

– 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?

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 into their use and production of documentation (among other topics), which I will present in the following section.

4 Findings

In terms of quantitative survey findings, clear documentation was the second most popular feature of a good digital tool or resource mentioned by participants, after ease of use/installation. The inverse, unclear or non-existent documentation, appeared in sixth place among barriers to digital tool or resource use, behind cost, training, technical issues, incompatibility, and usability issues. Documentation was mentioned by more participants who had knowingly used Linked Data than those who had not, with

Linked Data users generally having a higher self-assessed level of digital competence. However, overall results suggest that unclear or non-existent documentation can pose a significant barrier for both groups, indicating that its absence is noticed, particularly where a tool or resource is unreliable or difficult to use.

Although most interview participants stated that they would attempt to use an unfamiliar tool or resource immediately, without prior consultation of the documentation, many additionally expected that they would be able to access and use sufficient documentation to resolve any issues they encountered (e.g., "I expect to be able to go straight and use it and then I go to the documentation when I fail." (PART061)). One participant, however, was an exception to this rule, by first checking the documentation to assess its comprehensiveness, trustworthiness, and currency. If the documentation is lacking in these areas and there does not appear to be a way to obtain 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:

> "Before investing a lot of time in the tool, I… do then the tool evaluation, which is to say, 'Is there good documentation?'… When… there's something I want to do that I don't know how to do, is there going to be… some resource… that can… get me over those hurdles?" (PART008)

Several responses implied an assumption that consulting the documentation first was the 'correct' thing to do, with some participants seeming almost guilty when admitting that they operate differently (e.g., "I'm a bad user - I dive straight in and then just start tapping things into it, which is probably not what you're supposed to do." (PART078)). 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 often assumed that documentation would be too long and difficult to understand, suggesting previous confusion resulting from overly technical language (e.g., I "just ignore the documentation… it's written in a language - in a jargon that I don't have the time or patience to deal with" (PART089)). 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 sometimes found that tool and resource producers had made incorrect assumptions about users' understanding and computing environments:

> "The research tools I use, more often than not it's a frustrating experience, because the developers haven't had somebody to write clear, extensive documentation, and they, in the worst cases, will just write in this cryptic code that only… programmers understand, like shell commands… I'll enter those things in the shell, and nothing happens, so… they've made some assumptions about what my environment is." (PART037)

Time was another factor that affected participants' digital tool or resource use, where the existence or quality of documentation could have a significant impact. For example, one participant 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 ("I will sometimes keep myself from going further into a tool or technology that lacks documentation, just because I've got other things I need to prioritise." (PART037)). Even where documentation is available, participants commented that they did not have sufficient time to consult it in detail, and that if it took too long to find the answer to their question, they would stop using the tool or resource (e.g., "If the documentation is… very long, or very difficult to approach, then I just quit." (PART109)).

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 usually intended for their documentation to assist less technically experienced users (e.g., "The documentation needs to be… straightforward for researchers who do not know or care about the… technology" (PART061)). To ensure that documentation makes sense to others, participant recommendations included a focus on communicating to the user and writing collaboratively where possible, or to at least ask other people for feedback before publication (e.g., "I do approach it with a… focus on communicating to another user, but… I might be doing that in an echo chamber, and I look forward to having… some collaborators who might get me out of that." (PART037)). One participant 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:

> "The researcher… can describe [the tool] according to her perspective and… what she needs it to achieve for her research, rather than being the technologist describing the tool from the point of view of how it's designed and… the functionalities it provides." (PART061)

Several participants recommended that creating documentation should be an integral part of the production process, as it can prompt changes that result in a more usable tool or resource. However, some admitted that they did not enjoy the process of producing documentation and found that, in practice, it was often left until the end of a project or was not produced at all (e.g., "It's frankly something I find quite boring to produce and that's sometimes maybe led me to leave it till last… whereas in fact it should probably be something that you create early on and then update as the project goes on" (PART054)). Other obstacles to documentation production mentioned by participants included time constraints of needing to publish a tool or resource for presentation at a conference, or competing priorities among the project team around producing documentation versus expanding the dataset.

As an alternative to external, standalone documentation, one participant's development approach incorporated instructions within the interface and focused primarily on ease of use ("The documentation could almost be thought of as part of the tool… the initial search box has the instructions inside it… throughout the tool there are little

pop-ups and things like that, that might suggest ways to use it" (PART119)). However, they did acknowledge the importance of providing some separate documentation, such as provenance information or instructions for more complex operations.

Participants also gave their views on potential components of documentation, many of which involved informing users about data provenance and structure. Such documentation is crucial for building user trust in a tool or resource, by providing sufficient detail to allow its critical evaluation. One participant 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:

> "A lot of the… processed data that you get from things that other people may have done… they're quite often black boxes, in that you get presented with a nice… image, or a nice dataset, or something like that, but all the stages… the data's gone through to… end up in that nice, processed state aren't usually made very clear… I guess that's a documentation problem, but it also then becomes an interpretive problem in that if you were to want to use that data, you would have to bear in mind all the things that have been done to it, and they might not have been the kind of things you might have done to it yourself." (PART001)

Accompanying documentation or metadata can assist the user in making an informed decision whether to use a dataset and provide support with its analysis and interpretation. The process of creating a dataset should therefore be accompanied as far as possible by the production of documentation containing information about its structure, as well as key decisions made, theoretical approaches taken, and any known limitations, such as gaps or inaccuracies. The same participant considered inaccurate and incomplete data to be "a fact of… digital life" that all 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. Transparency about data provenance and processing, through documentation, can mitigate concerns about inaccurate or incomplete data, thereby facilitating reproducibility and reuse.

Tutorials and examples were also mentioned frequently, particularly in relation to complex search activities. For example, one participant expected tutorials to demonstrate how to search effectively if specific terminology or dialects are required, or if input involves a non-Latin character set ("Some dictionaries, for example, online, in Ancient Greek, they… have different dialects... If you don't know what Doric is, or Ionian… or if you don't know how to search for those languages or dialects - if you don't know how to use the Greek keyboard, that's a problem... So… there should be an easy tutorial." (PART254)). Documentation can also assist the discovery process by providing example queries for customisation by users. One participant found such examples to be particularly helpful when using the numismatic resource Nomisma[7], while also mentioning that their inclusion is relatively rare across Linked Ancient World Data tools and resources:

[7] http://nomisma.org

> "Very few endpoints provide queries to look at and… copy… for your own purposes. Nomisma does this very, very well, they have a long list of possible queries that I have used many times, converting them in my own benefit." (PART109)

Example queries ensure use of the correct ontologies and terms, while providing efficiency with a valid query for users to adapt. 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.

Various tools and resources invite contributions from users, e.g., by providing mechanisms to suggest corrections or updates, or add new data. To facilitate user contributions, participants recommended that accompanying documentation should provide clear and current information and instructions (e.g., "It is possible to get a number of people contributing if… your documentation… is good, if your explanations are good." (PART008)). One participant also mentioned that it is helpful when such documentation is complemented by feedback from data producers or moderators, to ensure that contributions meet quality and format requirements ("I've had feedback from [Pleiades]… on the places I've contributed, which has been great." (PART054)).

Participants also spoke about the importance of ascertaining conditions of a dataset's reuse and their occasional difficulties in locating information about copyright and licencing on the digital tools and resources they use. One participant stated that, in this situation, they tended to operate on the principle that it is "better to ask forgiveness than permission… you can just take it down if they complain" (PART041). 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 material unless they are explicitly told not to. Making data available under an open licence, such as those provided by Creative Commons, provides an explicit indication that reuse is permitted (and any conditions that should be applied), thereby increasing reusability.

Finally, several participants commented on the potential unreliability of tools and resources that have seemingly been abandoned by their creators (e.g., "You find that nobody's updated the website in five years… nobody responds to your emails." (PART008)), suggesting that information about the most recent updates would be helpful to users, as well as providing a reliable point of contact.

Clear documentation can therefore mitigate usability and reliability issues, while managing user expectations. Furthermore, it can reduce 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.

5 Discussion

My findings demonstrate the importance of documentation production to ensure usability and reusability of Humanities data, particularly where the underlying model is relatively complex, as is often the case with Linked Data. However (in agreement with findings from the existing literature), I also found that documentation production is often not given the attention it needs during the development stage, leading to actual

and potential usability issues with the resulting resources. My study therefore confirms the usefulness of a documentation template or checklist to facilitate this process (as recommended above by Franzini et al. (2019) and Warwick et al. (2008)), while also providing indications of the types of components that such a checklist should contain.

Transparency is key to the process of documentation development, i.e., providing effective, open communication about what the tool or resource enables the user to do and how this is achieved. In a Linked Data context in particular, such transparency ensures that users can understand how this data has been modelled and the rationale behind these particular modelling decisions. 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.

Depending on the scope or format of the resource, documentation might incorporate instructions, examples and use cases, ideally negating the need for resource-specific training. 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 (re)usability, particularly where substantial changes to the interface would be prohibitively expensive or time-consuming. When producing documentation, it is advisable to consider the domain/technical knowledge of anticipated users (or lack thereof) by explaining terminology and using clear, accessible language, to avoid limiting the potential audience.

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 (see Huskey, in this special issue, for an example of how this might be achieved, using Git's commit log as a research journal). 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.

6 Documentation checklist for (Linked) humanities data

In terms of specific elements that documentation should contain, I propose the following Documentation Checklist for (Linked) Humanities Data. Its components were developed following analysis of the survey and interview findings detailed in Section 4, and were enhanced and verified through consultation of the Digital Humanities literature discussed in Section 2. As my study focused on Linked Ancient World Data, the checklist is likely to be most relevant to this domain; however, by additionally incorporating findings from the wider literature, it should also be more broadly applicable to other Humanities disciplines and data formats.

Although serious consideration should be given to each point in the checklist, not all points will be relevant to all projects. The full checklist might be used when developing an interactive digital tool or resource; however, selected components (in particular 6.6, 6.7 and 6.9) could be used as a template for documenting raw datasets.

6.1 General information

- 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 the project on funder website(s), where available;
- Project start and end dates.

6.2 Scope

- 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.

6.3 Technical requirements

- 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).

6.4 Functionality

- Key tasks for which the tool/resource is designed;
- Associated research methods to which these tasks might best apply (e.g. using TaDiRAH[8] 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.

6.5 Tutorials

- Tutorials for each aspect of the tool or resource's functionality, from basic to advanced;

[8] https://tadirah.info/

- Links to relevant (ideally free) training resources, or the producers' own tutorials available via external sites.

6.6 Data provenance

- 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).

6.7 Data model

- Standards and formats;
- URI structure(s) and namespace(s);
- Diagram of data model(s) to illustrate structure and relationships between terms;
- Diagram of any new ontology(ies) created for the tool/resource[9];
- 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.

6.8 Data access

- 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.

[9] See MIRO (Matentzoglu et al., 2018) for a more detailed set of components for ontology-specific documentation

6.9 Data reuse

- Copyright information;
- Licence(s), e.g., Creative Commons;
- Example citation;
- Contact form or instructions for how to (optionally) inform the data producer(s) of any reuse.

6.10 User contributions

- 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.

6.11 System development and reuse

- 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.

6.12 Publications

- 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).

6.13 Maintenance

- 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.

7 Conclusions

Through studying users and producers of digital tools and resources for Ancient World research, particularly those involving Linked Data, I found that one of the key criteria

for usability was documentation, on which this article has focused. While my findings largely cement those from previous studies, I have developed these further by identifying and consolidating specific components that users require and expect from such documentation. The resulting checklist should serve to facilitate future documentation production, thereby potentially mitigating the time constraints that can have a significant impact on the process.

To further develop and strengthen the documentation checklist, future work should involve its implementation on existing projects. Through such real-world applications, areas for improvement might be identified, alongside any omissions. The checklist might then gradually be refined and customised over time, as a result of collaborative discussion between its users. Implementation would be further assisted by funding bodies recognising the importance of documentation in a Digital Humanities context, requiring its production as part of its funding conditions, and ensuring that sufficient time has been allocated to it in project plans. While this might result in allocating more time (and, therefore, money) to a particular project, this should be considered an investment in future (re)usability and sustainability, thereby preventing the tool or resource from being subsequently reinvented at an even greater cost.

Acknowledgements Special thanks are owed to my PhD supervisors Elton Barker, Phil Perkins, Mathieu D'Aquin and Alessandro Adamou, for their advice and support in carrying out and writing up the research on which this article is based. Thank you also to the two anonymous reviewers, whose constructive feedback has helped to strengthen the article.

Author Contributions S.M. is the sole author of this article.

Funding This work was supported by the Arts and Humanities Research Council (AHRC) through the Consortium for Humanities and Arts South-east England (CHASE) Doctoral Training Partnership (ref. AH/L503861/1).

Data Availability Survey data is available via the Humanities Commons CORE repository at https://doi.org/10.17613/1291-4r33. Interview transcripts can be requested via the Open University's research data repository at https://doi.org/10.21954/ou.rd.22101905.v1.

Declarations

Competing interests The authors declare no competing interests.

Ethics Approval This study was approved by the Open University's Human Research Ethics Committee (ref. HREC/2018/2807/Middle).

Consent Informed consent was obtained from all individual participants included in the study.

References

Aghajani, E., Nagy, C., Vega-Márquez, O.L., Linares-Vásquez, M., Moreno, L., Bavota, G., & Lanza, M. (2019). Software Documentation Issues Unveiled. 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE) (pp 1199–1210). (ISSN: 1558-1225) https://doi.org/10.1109/ICSE.2019.00122

Angelis, S., Benardou, A., Chatzidiakou, N., Constantopoulos, P., Dallas, C., Dunning, A., . . . Waterman, K. (2015). Europeana cloud deliverable 1.3 user requirements analysis and case studies report. content strategy report (Tech. Rep.). Europeana. Retrieved 2017-07-03, from 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

Arts and Humanities Research Council (2018). Technical plan. Retrieved 2021-04-13, from 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. Retrieved 2021-03-10, from https://ahrc.ukri.org/documents/guides/research-funding-guide1/

Barats, C., Schafer, V., & Fickers, A. (2020). Fading Away... The challenge of sustainability in digital studies. *Digital Humanities Quarterly*, *14*(3). Retrieved from http://digitalhumanities.org/dhq/vol/14/3/000484/000484.html

Bartalesi, V., Meghini, C., Metilli, D., Tavoni, M., & Andriani, P. (2018). A web application for exploring primary sources: The DanteSources case study. *Digital Scholarship in the Humanities*, *33*(4). https://doi.org/10.1093/llc/fqy002

Berners-Lee, T. (2010). Linked Data - Design Issues. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Birnbaum, D.J. (2020). Project checklist. Retrieved 2021-03-30, from https://dh.obdurodon.org/project-checklist.xhtml

Chen, J.-C., & Huang, S.-J. (2009). An empirical analysis of the impact of software development problem factors on software maintainability. *Journal of Systems and Software, 82*(6), 981–992.

Edmond, J., & Morselli, F. (2020). Sustainability of digital humanities projects as a publication and documentation challenge. *Journal of Documentation*, *76*(5), 1019-1031. 10.1108/JD-12-2019-0232

Elliott, T., & Gillies, S. (2009). Digital Geography and Classics. *Digital Humanities Quarterly*, *3*(1). Retrieved 2023-06-15, from http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html

Fafinski, M. (2022). Facsimile narratives: Researching the past in the age of digital reproduction. *Digital Scholarship in the Humanities, 37*(1), 94–108. https://doi.org/10.1093/llc/fqab017

Flanders, J., & Jannidis, F. (2018). Data Modeling in a Digital Humanities Context: An Introduction. J. Flanders & F. Jannidis (Eds.), The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources (pp 3–25). London: Routledge. https://doi.org/10.4324/9781315552941

Forward, A., & Lethbridge, T.C. (2002). The relevance of software documentation, tools and technologies: a survey. Proceedings of the 2002 ACM symposium on Document engineering (pp 26–33). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/585058.585065

Fostano, K., & Morreale, L.K. (2019a). PART 1 Active Projects Digital Documentation Process (DDP) Archiving Wizard. Retrieved 2021-03-30, from https://digitalhumanitiesddp.com/part-1-active-projects-digital-documentation-process-ddp-archiving-wizard/

Fostano, K., & Morreale, L.K. (2019b). PART 2 Active Projects Digital Documentation Process (DDP) Archiving Wizard. Retrieved 2021-03-30, from https://digitalhumanitiesddp.com/part-2-active-projects-digital-documentation-process-ddp-archiving-wizard/

Franzini, G., Terras, M., & Mahony, S. (2019). Digital Editions of Text: Surveying User Requirements in the Digital Humanities. *Journal on Computing and Cultural Heritage*, *12*(1), 1:1-1:23. https://doi.org/10.1145/3230671

Galina Russell, I. (2011). The role of libraries in Digital Humanities (El papel de las bibliotecas en las humanidades digitales). San Juan, Puerto Rico. Retrieved 2022-02-22, from https://www.ifla.org/past-wlic/2011/104-russell-en.pdf

Gibbs, F., & Owens, T. (2012). Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly*, *6*(2). Retrieved from http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html

Hering, K. (2014). Provenance Meets Source Criticism. *Journal of Digital Humanities*, *3*(2). Retrieved from http://journalofdigitalhumanities.org/3-2/provenance-meets-source-criticism/

Kemman, M., Kleppe, M., & Scagliola, S. (2014). Just Google It - Digital Research Practices of Humanities Scholars. C. Mills, M. Pidd, & E. Ward (Eds.), Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities. Sheffield: HRI Online Publications. Retrieved from https://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman, arXiv:1309.2434

Kleppe, M., & Otte, M. (2017). Analysing and understanding news consumption patterns by tracking online user behaviour with a multimodal research design. *Digital Scholarship in the Humanities*, *32*(supplement 2), ii158–ii170. https://doi.org/10.1093/llc/fqx030

Matentzoglu, N., Malone, J., Mungall, C., & Stevens, R. (2018). MIRO: guidelines for minimum information for the reporting of an ontology. *Journal of Biomedical Semantics*, *9*(1). https://doi.org/10.1186/s13326-017-0172-7

Middle, S. (2022). A Five-Star Model for Linked Humanities Data Usability. Digital Humanities 2022: Conference Abstracts (pp. 313–314). Tokyo, Japan: University of Tokyo. Retrieved 2023-06-15, from https://dh2022.dhii.asia/dh2022bookofabsts.pdf

Morreale, L. (2019). Medieval Digital Humanities and The Rite of Spring: Thoughts on Performance and Preservation (Tech. Rep.). BodoArXiv. Retrieved 2021-03-30, from https://osf.io/preprints/bodoarxiv/7p2t6/, https://doi.org/10.34055/osf.io/7p2t6

Office of Digital Humanities (2021). Notice of Funding Opportunity, Funding Opportunity Title: Digital Humanities Advancement Grants. National Endowment for the Humanities. Retrieved 2021-04-08, from https://www.neh.gov/sites/default/files/inline-files/Digital%20Humanities%20Advancement%20Grants%20NOFO%20June%202021%2020210115-HAA.pdf

Panciera, S., & Orlandi, S. (2017). EAGLE: Past, Present and Future. S. Orlandi, R. Santucci, F. Mambrini, & P.M. Liuzzo (Eds.), Digital and Traditional Epigraphy in Context: Proceedings of the EAGLE 2016 International Conference (pp 1–10). Rome: Sapienza Universitá Editrice. Retrieved 2020-05-14, from http://www.editricesapienza.it/node/7610

Peroni, S., Shotton, D., & Vitali, F. (2013). Tools for the automatic generation of ontology documentation: a task-based evaluation. *International Journal on Semantic Web and Information Systems*, *9*(1), 21–45. https://doi.org/10.4018/jswis.2013010102

Picca, D., & Egloff, M. (2017). DHTK: The Digital Humanities ToolKit. A. Adamou, E. Daga, & L. Isaksen (Eds.), Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II) co-located with 16th International Semantic Web Conference (ISWC 2017) (pp 81–86). Vienna, Austria. Retrieved from http://ceur-ws.org/Vol-2014/paper-09.pdf

Plösch, R., Dautovic, A., & Saft, M. (2014). The Value of Software Documentation Quality. 2014 14th International Conference on Quality Software (pp 333–342). (ISSN: 2332-662X) https://doi.org/10.1109/QSIC.2014.22

Poole, A. H., & Garwood, D. A. (2018). “Natural allie Librarians": Larchivists, and big data in international digital humanities project work. *Journal of Documentation, 74*(4), 804–826. https://doi.org/10.1108/JD-10-2017-0137

Presner, T. (2012). How to Evaluate Digital Scholarship. *Journal of Digital Humanities*, *1*(4). Retrieved from http://journalofdigitalhumanities.org/1-4/how-to-evaluate-digital-scholarship-by-todd-presner/

Rabinowitz, A., Shaw, R., & Golden, P. (2018). Making up for Lost Time: Digital Epigraphy, Chronology, and the PeriodO Project. A. De Santis & I. Rossi (Eds.), Crossing Experiences in Digital Epigraphy (pp 202–215). De Gruyter. https://doi.org/10.1515/9783110607208-017

Reggiani, N. (2017). *Digital Papyrology I: Methods*. Tools and Trends: De Gruyter. https://doi.org/10.1515/9783110547474

Robillard, M.P., Marcus, A., Treude, C., Bavota, G., Chaparro, O., Ernst, N., . . . Wong, E. (2017). On-demand Developer Documentation. 2017 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp 479–483). https://doi.org/10.1109/ICSME.2017.17

Rockwell, G. (2012). Short Guide To Evaluation Of Digital Work. *Journal of Digital Humanities*, *1*(4). Retrieved from http://journalofdigitalhumanities.org/1-4/short-guide-to-evaluation-of-digital-work-by-geoffrey-rockwell/

Sahle, P., Vogeler, G., Broughton, M., Cummings, J., Fischer, F., Steinkrüger, P., & Scholger, W. (2014). Criteria for Reviewing Scholarly Digital Editions, version 1.1. Retrieved 2023-03-01, from urlhttps://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1/

Simon, R., Vitale, V., Kahn, R., Barker, E., & Isaksen, L. (2019). Revisiting Linking Early Geospatial Documents with Recogito. *e-Perimetron*, *14*(3), 150–163. Retrieved 2020-03-26, from http://www.e-perimetron.org/Vol143.htm

Tóth-Czifra, E. (2020). The Risk of Losing the Thick Description: Data Management Challenges Faced by the Arts and Humanities in the Evolving FAIR Data Ecosystem. J. Edmond (Ed.), Digital Technology and the Practices of Humanities Research (pp 235–266). Open Book Publishers. https://doi.org/10.11647/OBP.0192

UK Data Service (2015). UK Data Service - Data management costing tool and checklist. UK Data Service. Retrieved 2021-03-29, from https://ukdataservice.ac.uk/media/622368/costingtool.pdf

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. G. Bodard & M. Romanello (Eds.), Digital Classics Outside the Echo-Chamber (pp 147–168). London: Ubiquity Press. https://doi.org/10.5334/bat

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. https://doi.org/10.1108/00220410910926112

Warwick, C., Terras, M., Huntington, P., & Pappa, N. (2008). If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Literary and Linguistic Computing*, *23*(1), 85–102. https://doi.org/10.1093/llc/fqm045

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., . . . & 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

Zhi, J., Garousi-Yusifoğlu, V., Sun, B., Garousi, G., Shahnewaz, S., & Ruhe, G. (2015). Cost, benefits and quality of software development documentation: A systematic mapping. *Journal of Systems and Software*, *99*, 175–198. https://doi.org/10.1016/j.jss.2014.09.042