IJDC | *Research Paper*

Data Makers' and Users' Views on Useful Paradata: Priorities in Documenting Data Creation, Curation, Manipulation, and Use in Archaeology

Isto Huvila Lisa Andersson

Olle Sköld Ying-Hsang Liu

Uppsala University

Abstract

Understanding and making data (re)usable requires adequate documentation of the data and information on how it has been created, curated, manipulated, and used, termed in data documentation literature as *paradata.* This paper reports the results of a survey study ($N = 79$) of data creating and (re)using archaeologists' views of what data creation, curation, manipulation, and use related information (termed here as *paradata*) they consider important when working with data. Data makers' and users' perceptions align to a considerable degree. It is important to explain the original context of data creation and to know the purpose, procedures and methods of data creation, analysis, and documentation. The findings underline that there is a need to continue developing and testing ideas on how to capture and document paradata and to find ways to help data makers adopt proven practices to facilitate this. Simultaneously, it is crucial that the paradata aimed at facilitating data use is relevant for data users rather than, for instance, technical or administrative details considered useful primarily by data makers.

Submitted 5 June 2023 ~ *Revision accepted* 18 April 2024

Correspondence should be addressed to Isto Huvila. Email: isto.huvila@abm.uu.se

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Understanding and making data (re)usable requires adequate documentation of the data and, as Leonelli (2020) suggests, of its 'data journey'. To date, a growing number of studies have surveyed researchers' and other stakeholders' general preferences and needs for data and its documentation (Gregory & Koesten, 2022). In contrast, only a handful of studies have focused on documentation and information specific to data creation, curation, manipulation and use, termed in data documentation literature as *paradata* (Huvila, 2022) and understood here as information on the practices and processes of data creation, processing, and use (cf. Sköld et al., 2022). Such process information is often a part of what is described as context (e.g., Thomer, 2022; Chapman et al., 2019; Gregory et al., 2019), but the specifics of what types of documentation are especially important and to whom are not well known.

This paper aims to address the research gap concerning the documentation of data processes and report findings from a survey study of data creating and (re)using archaeologists' views of what data creation, curation, manipulation, and use related information (termed here as *paradata*) they consider important when they are working with data. This paper addresses the following three research questions:

1. RQ1: What information on data-related processes and practices (paradata) do those who have used archaeological data consider important?
2. RQ2: What information on data-related processes and practices (paradata) do those who have made and/or deposited archaeological data consider important?
3. RQ3: What similarities and differences in the perceived usefulness of various types of paradata, if any, can be identified among data makers and users?

In this text, *data making* refers to practices of creating, collecting, producing, and literally 'making' data happen and exist. *Data use* refers to the secondary use of data after data creation. Data for (re)use can be created by the users or others.

Literature Review

Data Needs and Practices

A growing number of studies have inquired into data needs and practices in science and scholarship (Gregory & Koesten, 2022) as a complement to the long line of research on scholars and scientists (Case & Given, 2016) and to a certain extent, archaeologists (e.g., Jones et al., 2003; Huvila, 2014) general information needs.

Data needs vary by data user, their profession, and data use (Huvila, 2021; 2014), discipline and research theme (Gregory & Koesten, 2022), perceived concerns and norms of data reuse (Joo & Kim, 2017), and change over time (Gregory et al., 2020b). In an international multidisciplinary survey by Gregory et al. (2020a), half of the respondents indicated that they used multiple data types, with the arts and humanities associated with the largest number of data types. The most popular data type was

observational/empirical, although not in the arts and humanities. Half of the users also indicated that they were interested in data outside their discipline. Cross-disciplinary comparisons tend to focus on characterising data needs at the level of data types (e.g., empirical or experimental); domain-specific studies have generated more specific lists of what data users want to know, for example, location, time, data collection method, data collector, or, for example, the shape of physical entities represented in the data (Löffler et al., 2021; Suhr et al., 2020). From the perspectives of data users, research has been concerned with the factors of individual differences, such as intentions and goals of information seeking, user perceptions, and experience. In their study on engineers, Joo and Kim (2017) identified a correlation between attitudes towards data reuse and the perceived concerns and norms surrounding it. In addition, the intention to reuse data was found to be influenced by the availability of repositories and attitudes towards data reuse. A literature review focussing on data reuse from information-seeking perspectives suggests that stages of "... similar … tasks and activities, such as recognising the need, retrieval, selection, acquisition, and use, but different in goals and specific actions" were part of the researchers' routines (Wang et al., 2021, p. 1178). Faniel et al. (2016) found that in social sciences, the data quality of completeness, accessibility, ease of operation, and credibility had a positive association with user satisfaction. The evidence for the impact of experience on data and data-related information needs is inconclusive. Partly, it has been suggested that the complex data needs of experienced data users imply the need for more and more complex meta documentation (Friedrich, 2020); however, others (Faniel et al., 2012; Yoon, 2017) have suggested that novices need more thorough descriptions than seasoned experts. Overall, these studies indicate that disciplinary practices, users' perceptions of specific data attributes, goals or intentions, and experience play a significant role in influencing data reuse.

In archaeology, the earlier literature has voiced concerns that too little is known about digital archaeological practices (Huvila & Huggett, 2018). Similarly, it has been pointed out that a poor understanding of data needs is a risk that undermines the enterprise of preserving data (Huggett, 2015). A small number of studies have inquired specifically into archaeologists' data needs and practices (e.g., Condron et al., 1999; Selhofer & Geser, 2015; Kansa & Kansa, 2009) and as part of broader cross-disciplinary comparisons (e.g., Gregory et al., 2019). In certain comparative studies, due to the granularity of the categorisation of specific disciplines, it is unclear whether archaeologists have participated or not (e.g., Gregory et al., 2020a; Lee et al., 2020).

Studies show that archaeologists need information on archaeological primary material, including dates, finds spots, measurements, appearance, material (Huvila, 2014), artefact types (Brandsen et al., 2021), and identifiers (Faniel et al., 2018). Geographical search is important, as well as the possibility of exhaustive or high recall over high precision (Hessing et al., 2013; Brandsen et al., 2019; cf. similar tendencies with humanities researchers in general Bates et al., 1995) and combining different search options (Brandsen et al., 2021). A common complaint is the heterogeneity of small-scale databases and the lack of data integration and interoperability (Huvila, 2014). For visual material, high-quality descriptive information is important (Beaudoin & Brady, 2011).

A key finding from earlier research is the gap between what data creators make available and the data users need (Faniel et al., 2018). Some earlier studies suggest that data creators and users sometimes form two distinct groups (Borgman et al., 2019), and sometimes data makers are active users of secondary data, depending on whether data makers and users form two distinct groups (Gregory et al., 2023). Rather than forming a seamless continuum from data creation to reuse, the archaeological information process is characterised by discontinuities, handovers, and considerable efforts needed when

"taking information" in use (Buchanan, 2019; Huvila, 2018); for instance, work necessary to make information and data originating from their creators become useful for their diverse users.

Paradata Needs in Archaeology

The documentation of the results of archaeological work and "the processes by which the data were produced" (Berggren & Hodder, 2003, p. 430) as components of the data have been a self-evident part of archaeological fieldwork since the birth of archaeology as a branch of scholarship (Lucas, 2012). In contrast, until recently, the discussion of what it entails in practice has been rather limited. The professionalisation of fieldwork, increasing standardisation, and more recently, the digitalisation of field documentation, and the separation of data production and interpretation have raised concerns about the adequacy of how archaeological work is documented (Huggett, 2014; Reilly et al., 2021). Reflexive archaeology and its emphasis on archaeological fieldwork as interpretative rather than a data-producing activity explicitly emphasises "documenting the documentation" (Berggren & Hodder, 2003, p. 429). In addition, rather than withholding the paradata creation as a task for field directors and trench supervisors, reflexive archaeology has underlined the importance of including the research team when producing the documentation (Berggren & Hodder, 2003).

A variety of different methods have been used and proposed for producing paradata. Traditionally, the most prominent approach to producing paradata in archaeology has been to use personal notebooks and diaries. They typically incorporate textual reflections and notes, sketches, and diagrams (Sandoval, 2021; Berggren et al., 2015); however, diaries tend to vary considerably in detail and, for instance, their level of formality (Boddington, 2014). Earlier, the diaries were usually kept by field directors and trench masters. Additional techniques include tagging databases and collections using timestamps, filming digital videos (Berggren & Hodder, 2003), workflows (Lozić & Štular, 2021), models and metamodels (Gonzales et al., 2016), three-dimensional (3D) data-capturing (Jensen, 2018; Dell'Unto et al., 2017), and drawing sketches (Berggren et al., 2015). The different approaches serve different purposes. Timestamps help to follow the progress of the work; however, diary writing helps to capture archaeologists' thought processes and nuance the context of the research process (Berggren et al., 2015). Videos provide a record of the processes through audiovisual and spoken information (Berggren & Hodder, 2003).

Earlier studies have also surveyed users' needs for archaeological and para data. The findings of a survey from the European Open Science Cloud (EOSC) initiative of a convenience sample of diverse stakeholders in research data management ($N = 155$) on user needs and expectations on data quality highlighted the importance of several key pieces of information. The respondents underlined the importance of information on how to cite datasets, terms of using them, and how to contact data providers. They also appreciated the availability of a "user guide" with information on how the data were collected, information on data provenance, versions of datasets, and the primary purpose of why the data was created. Further, the respondents called for examples of how to use data and appropriate tools for working with it (Lacagnina et al., 2023). Börjesson et al. (2022) identified four major categories of paradata needs in archaeology, including *scope paradata* on where to find data, details affecting data coverage and data quality; *provenance paradata* describing the disciplinary and time-bound origin of the data, its related epistemological culture, and the rationale of data generation; *methods paradata* on data generation methods, technical information, for example, on equipment, decision-making information, and *knowledge organisation and representation paradata* on the

rationale for the representation of information versus non-information, subsets within data, standards structuring data, semantics for representation, the rationale of location and dating data, and the relations between data entities. Huvila (2020) underlines the usefulness of paradata when assessing the credibility of information. The paradata correspond to those described in other disciplines. In survey research, the scope or coverage of datasets is crucial for reusers (David, 1991). Such general aspects as research questions, instrumentation (Gregory et al., 2019), scope and framing of research (Faniel et al., 2019), methods (Murillo, 2016; Koesten et al., 2019) and general contextual information have been reported elsewhere (Chapman et al., 2019; Gregory et al., 2019).

Another corpus of studies has focused on the metadata and paradata data that creators have documented. In a study of published data papers connected to the Global Biodiversity Information Facility, Li et al. (2020) identified 17 types of data events referred to in data papers, including data analysis, classification, collection, formatting, identification, modification, registration, removal, sharing, validation, and visualisation, databasing, georeferencing, metadata creation, taxonomy identification, and taxonomy validation. Kim et al. (2019) analysed data descriptions in 20 repositories, including six preserving archaeological data, and found an emphasis on data-level descriptions; however, usage rights and studies where the data originates are described to a much lesser extent. In addition, metadata schemes developed for documenting scientific and scholarly data carry evidence of data makers' intentions and preferences. Willis et al. (2012) identify 11 goals that the schemes are expected to achieve, including facilitation of data archiving and publication, simplicity, and sufficiency of data documentation. The goals vary between standards, similar to how individual standards have been adopted by individual data repositories (Austin et al., 2016).

Methods and Materials

An online survey targeted to individuals who, according to themselves, had previously used and/or created, collected, or deposited archaeological data (understood broadly as ranging from digital data to finds collections) was used to collect the views of what information users think they need to know about data to use it effectively, and what data producers consider to be important to know about their data. The survey focused on broadening the understanding of the variety of different paradata practices, opportunities, and problems and to gather basic information on the variety of paradata needs. The questionnaire included sections surveying the importance of different aspects of data considered important from data use and making perspectives, as well as questions on basic demographics, including professional background, the country where the respondent was professionally based, and the length of their research career. The survey was targeted to respondents who indicated having experience in data use and making, respectively. Survey participants were recruited using relevant archaeology-related mailing lists, social media, and personal contacts. The data is openly available in Swedish National Data Service data catalogue at https://doi.org/10.57804/fynk-v641 (Huvila et al., 2024).

In total, 79 respondents participated in the survey between February and October 2021. The survey data represents a purposive (selected based on experience of data making and/or use) convenience (availability) sample with an unknown bias. The eventual representativeness of the sample is difficult to assess. Earlier studies of data practices and needs have surveyed varying populations consisting (primarily) of researchers (e.g., Tenopir et al., 2020; Gregory, 2020) or researchers and other

stakeholders, including administrators (e.g., Lacagnina et al., 2023). In total, 44% (35/79) of the respondents were female, 49% (39/79) male, 1% (1/79) other, and the rest preferred not to say. Geographically, the survey covered all continents except for Australia and Antarctica, with 81% (64/79) being European respondents, with 33 countries represented. Overall, 52% (41/79) worked as researchers, and the rest in diverse, primarily administrative and data management-related duties. The latter group is labelled in the reporting as Administrators. Then, 56% (44/79) worked at a university or public research organisation, 20% (16/79) at a museum, 10% (8/79) at a government institution, and 8% (6/79) at a private company or institute. In addition, 4% (3/79) had no organisational affiliation, 57% (45/79) had permanent employment or were self-employed, 28% (22/79) had a temporary contract, and 10% (8/79) were students or trainees. Finally, 9% (7/79) worked as research centre, institute, or laboratory directors. The median career length among the respondents was 14 years (e.g., 1–45 years). The respondents were engaged in various types of archaeology-related research, from field research to spatial analysis, osteology, and artefacts studies. They used and produced various research materials, including spatial, photographic, 3D, finds, and geological data.

Descriptive statistics were calculated using cross-tabulations and the *describe* function from the package *psych* (R Core Team, 2023; Revelle, 2023). Associations between variables were studied using the Wilcoxon rank-sum (Mann–Whitney U) test with the function *wilcox.test*, Pearson's correlations (*cor.test*), and linear regression analysis (*lm*). A 0.001 significance level was used in all analyses.

Open-ended answers were analysed using applied thematic analysis (Guest et al., 2012), which is based on the iterative reading of the material, categorisation, and identification of themes with constant comparisons.

Results

Descriptive Statistics

The majority of the respondents (95% (75/79)) indicated that they had used data collected or created by others. In addition, 58% (46/79) found that they had been able to use data as they planned, and 34% (27/79) indicated the opposite. Several respondents commented that they had positive and negative experiences. One of them (#109) remarked that it was possible to write a paper but "that some ambitions had to be curtailed: either the amount of data was too limited (for modelling/statistical significance), or there were difficulties interpreting the data altogether".

Of the respondents, 58% (46/79) had experience depositing data, in contrast to 34% (27/79) who explicitly indicated a lack of such experience. In addition, 80% (63/79) indicated that their daily work duties encompassed data acquisition, 69% (55/79) data description, 86% (68/79) data use, 52% (41/79) data delivery, 73% (58/79) data reuse, 66% (52/79) data planning for their organisation, and 35% (28/79) data planning for others. Only 10% (8/79) of the respondents indicated that they had no influence on data management policies and practices that were directly relevant to their work. The rest participated to different degrees in decision-making.

Table 1 summarises the means for responses to "How useful it is to have the following information about the data you are using?" for all respondents, those who indicated they work as researchers and administrators with different levels of junior and senior data administrative duties, respectively. The highest-ranking piece of information

(Figure 1 for responses in decreasing order) with the lowest standard deviation (SD) pertains to the general context of data creation/collection (VAR05_1). Other key information (median 5) includes the names of methods and standard operating procedures (VAR05_3), references to documentation protocols and/or information systems (VAR05_5), explanation of the original purpose of data creation/collection (VAR06_1), narrative of how the data was collected/created (VAR06_2), data creators' annotations (VAR06_3), information on what data was not collected or included in the dataset (VAR06_5), references to measurement devices (VAR07_2), references to analysis procedures (VAR07_3), and information about versions of the dataset (VAR09_2). The lowest ranking information by mean, however, with a relatively high mean (3.16) and SD (1.39), refers to camera settings and where the data has been used previously (mean 3.32, SD 1.18), and by the median, to names (3) and credentials (3.5) of everyone involved in creating/collecting the data.

Table 1. Perceived importance of data-related information from a data use perspective.

Variable		How useful it is to have the following information about the data you are using?	All				Researchers				Administrators			
			N	median	mean	SD	*N*	median	mean	SD	*N*	median	mean	SD
VAR04_1	Funding	References to project or funding body of data collection/creation	76	4.0	3.49	1.19	46	3.0	3.33	1.21	24	4.0	3.67	1.13
VAR04_2	Permissions	References to permissions granted for creating/collecting the data	76	4.0	3.47	1.33	46	3.0	3.33	1.49	24	4.0	3.83	1.01
VAR04_3	Legislation	References to legislation and official requirements directing the archaeological activity (e.g., national/federal legislation, requirements stated by a national heritage board)	76	4.0	3.36	1.35	46	3.0	3.09	1.49	24	4.0	3.79	1.06
VAR04_4	Guidelines	References to guidelines followed (e.g., The London Charter, EAA[1] Code of Practice for Fieldwork Training)	76	4.5	3.62	1.51	46	4.0	3.48	1.56	24	4.5	4.04	1.43
VAR05_1	Context	Explanation of the general context of data creation/collection (e.g., field survey, investigation, excavation, laboratory analysis, archival research)	75	5.0	4.87	0.38	45	5.0	4.84	0.42	24	5.0	4.88	0.34
VAR05_2	Methods lit	References to methods literature (e.g., field or lab manuals, handbooks) that informed data creation/collection when the data was created/collected (when applicable)	76	4.5	4.17	1.09	46	4.0	4.15	1.09	24	4.5	4.33	0.82
VAR05_3	Standards	References to names of methods and/or standard operating procedures that informed data creation/collection (e.g., single context method, probabilistic sampling)	75	5.0	4.53	0.74	45	5.0	4.53	0.81	24	5.0	4.58	0.58
VAR05_4	Formal lang	Reference to formal language used to structure the data (e.g., ontologies or classification systems like CIDOC-CRM)	75	4.0	3.93	1.41	46	4.0	3.76	1.58	24	4.0	4.25	0.68
VAR05_5	Protocols	References to documentation protocols and/or information systems (e.g., context sheets, field diaries, ArcGIS, Intrasis) used to structure the data collection process and/or the data	76	5.0	4.50	0.90	46	5.0	4.39	1.04	24	5.0	4.58	0.65
VAR06_1	Purpose	Explanation of original purpose of data creation/collection (e.g., research questions, aims of study or data collection)	76	5.0	4.47	0.66	46	5.0	4.46	0.66	24	5.0	4.42	0.72
VAR06_2	Narrative	Narrative of how the data was collected/created (e.g., field/excavation/lab diary or notebook, text in report)	76	5.0	4.42	0.68	46	4.0	4.39	0.65	24	5.0	4.50	0.72
VAR06_3	Annotations	Data creators' annotations (e.g., particularly noteworthy aspects of dataset)	76	5.0	4.21	0.90	46	4.0	4.07	1.00	24	5.0	4.46	0.72
VAR06_4	Env conditions	References to environmental conditions (e.g., weather, time of the day or year, conditions in laboratory/archive) when the data was created/collected	76	4.0	3.80	1.24	46	4.0	3.70	1.28	24	4.0	3.96	1.27

[1] European Association of Archaeologists.

Variable		How useful it is to have the following information about the data you are using?	All				Researchers				Administrators			
			N	median	mean	SD	*N*	median	mean	SD	*N*	median	mean	SD
VAR06_5	Not collected	Explanation of what data was not collected, created, or included in the dataset (e.g., incomplete or unreadable data files, auxiliary working notes, data of peripheral significance)	76	5.0	4.53	0.68	46	5.0	4.46	0.75	24	5.0	4.62	0.58
VAR07_1	Tools	References to tools (e.g., trowel, shovel) used in fieldwork	76	4.0	3.55	1.35	46	4.0	3.46	1.50	24	4.0	3.83	1.13
VAR07_2	Measure dev	References to measurement devices (e.g., total station, tape measure) used in fieldwork	76	5.0	4.08	1.37	46	4.5	3.85	1.65	24	5.0	4.54	0.59
VAR07_3	Analys proc	References to analysis procedures (e.g., chemical analysis, C14 dating) used in conjunction with the fieldwork	76	5.0	4.51	1.05	46	5.0	4.46	1.28	24	5.0	4.58	0.58
VAR07_4	Camera	References to type, model and settings of camera used in taking photographs	75	4.0	3.16	1.39	46	3.0	2.98	1.51	23	4.0	3.39	1.23
VAR07_5	Software	References to the names and versions of software packages, scripts and code used when processing digital data	76	4.0	3.93	1.19	46	4.0	3.83	1.37	24	4.0	4.17	0.82
VAR07_6	Data process	Details on data processing procedures (e.g., how a software package was used, including details on settings and extensions)	75	4.0	3.81	1.33	46	4.0	3.63	1.55	23	4.0	4.22	0.74
VAR08_1	Cred everyone	Credentials of everyone involved in creating/collecting the data	76	3.5	3.34	1.10	46	3.0	3.22	1.09	24	3.5	3.62	1.01
VAR08_2	Cred responsib	Credentials of the person who is responsible for the project and data creation/collection	75	4.0	3.93	1.11	46	4.0	3.83	1.23	23	4.0	4.09	0.90
VAR08_3	Names everyone	Names of everyone involved in creating/collecting the data	75	3.0	3.39	1.18	45	4.0	3.42	1.18	24	3.0	3.33	1.13
VAR08_4	Name responsib	Name of the person who is responsible for the project and data creation	76	4.0	4.28	0.81	46	4.0	4.30	0.84	24	4.0	4.21	0.78
VAR09_1	Management	How the data has been managed and curated since it was collected (incl. information on how it was eventually digitised)	76	4.0	4.24	0.95	46	5.0	4.24	1.04	24	4.0	4.29	0.75
VAR09_2	Versions	Information about versions of the dataset	76	5.0	4.09	1.10	46	4.0	4.00	1.19	24	5.0	4.38	0.92
VAR09_3	Use	How, where and by whom the data has been used so far	76	4.0	3.32	1.18	46	3.0	3.20	1.22	24	4.0	3.58	1.02
VAR09_4	Review	That the data has been reviewed through peer review/by journal reviewer/by data reviewer or repository curator	76	4.0	3.68	1.06	46	4.0	3.63	1.10	24	4.0	3.67	1.05
VAR09_5	Citing instr	Instructions for how the dataset should be cited	76	4.0	4.00	0.91	46	4.0	4.02	0.86	24	4.0	4.04	1.04
VAR09_6	Rel datasets	Information about related datasets	76	4.0	3.91	1.00	46	4.0	3.80	1.07	24	4.0	4.08	0.88
VAR09_7	Rel pubs	Information about related publications	76	4.0	4.16	0.75	46	4.0	4.26	0.68	24	4.0	3.96	0.86

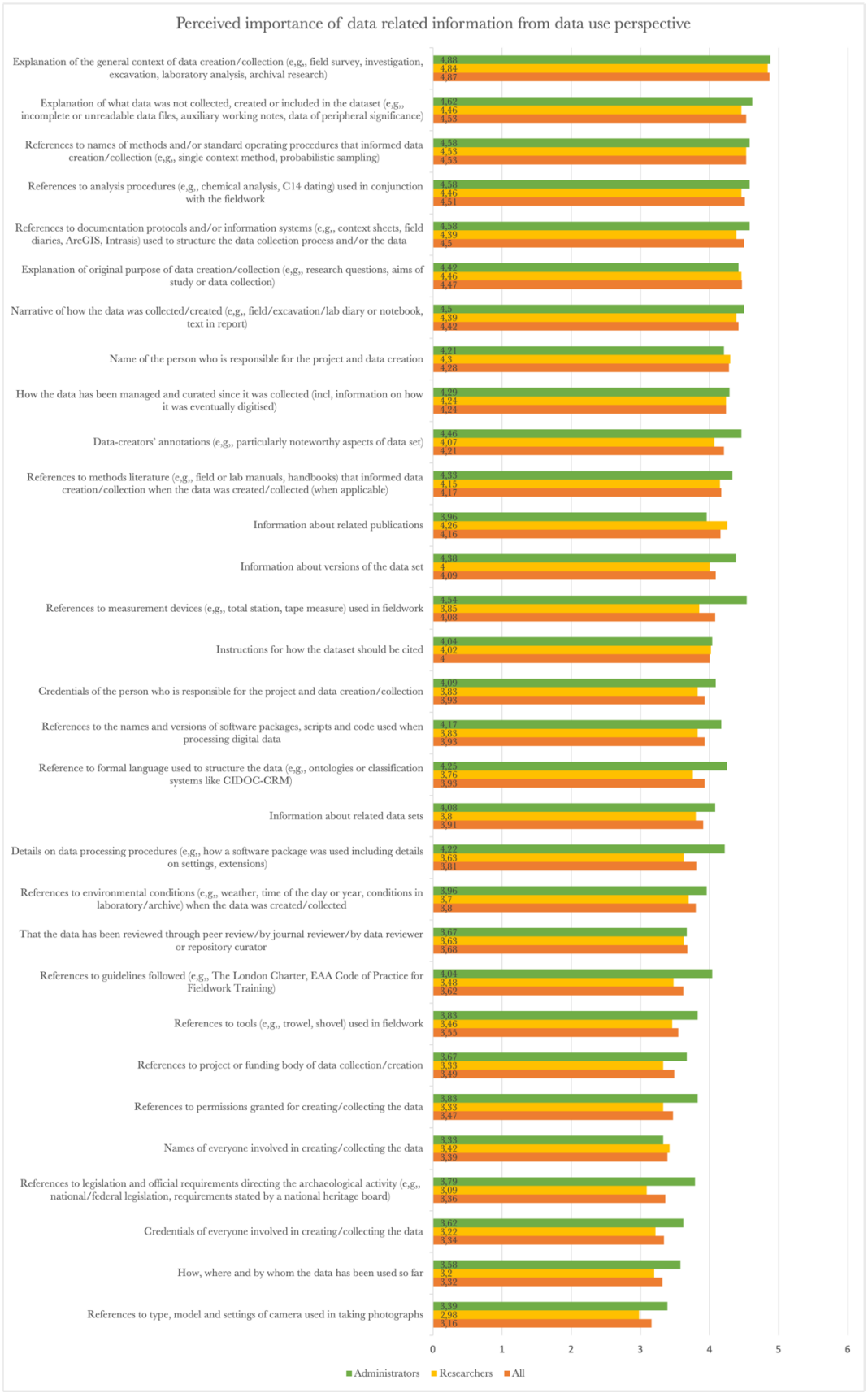

Figure 1. Perceived importance of data-related information from a data use perspective (means in decreasing order on a five-point Likert-like scale from one (Not at all useful) to five (Absolutely essential)).

Table 2 provides an overview of the responses to how important it is to include specific types of information when archiving or publishing an archaeological dataset for all respondents, those who indicated they work as researchers and as administrators or data managers, respectively. The highest ranking by mean (see Figure 2 for responses in decreasing order) were explanations of the general context of data creation/collection (VAR18_1), explanations of the original purpose of data creation/collection (VAR19_1), references to names of methods and/or standard operating procedures (VAR18_3), references to analysis procedures (VAR20_3), name of the person responsible for the project and data creation (VAR21_4), references to documentation protocols and/or information systems (VAR18_5), and references to methods literature (VAR18_2). A further six variables had a median of five.

Table 2. Perceived importance of data-related information from a data making perspective.

Variables		How important is it to include the following information when you are archiving or publishing an archaeological dataset?	All				Researchers				Administrators			
			N	median	mean	SD	*N*	median	mean	SD	*N*	median	mean	SD
VAR17_1	Funding	Reference to project or funding body of data collection/creation	52	5.0	4.15	1.13	30	4.0	3.97	1.25	16	5.0	4.38	0.81
VAR17_2	Permissions	References to permissions granted for creating/collecting the data (when applicable) for creating/collecting the data	52	5.0	4.06	1.32	30	4.0	4.00	1.36	16	5.0	4.12	1.36
VAR17_3	Legislation	References to legislation and official requirements directing the archaeological activity (e.g., national/federal legislation, requirements stated by a national heritage board)	51	4.5	3.67	1.52	29	4.0	3.34	1.67	16	4.5	4.25	0.86
VAR17_4	Guidelines	References to guidelines followed (e.g., The London Charter, EAA Code of Practice for Fieldwork Training)	51	5.0	3.63	1.61	29	4.0	3.38	1.63	16	5.0	4.00	1.46
VAR18_1	Context	Explanation of general context of data creation/collection (e.g., field survey, investigation, excavation, laboratory analysis, archival research)	51	5.0	4.92	0.34	30	5.0	4.93	0.25	15	5.0	5.00	0.00
VAR18_2	Methods lit	References to methods literature (e.g., field or lab manuals, handbooks) that informed data creation/collection when the data was created/collected	51	5.0	4.51	0.76	30	5.0	4.40	0.77	15	5.0	4.67	0.82
VAR18_3	Standards	References to names of methods and/or standard operating procedures that informed data creation/collection (e.g., single context method, probabilistic sampling)	51	5.0	4.67	0.59	30	5.0	4.60	0.67	15	5.0	4.80	0.41
VAR18_4	Formal lang	Reference to formal language used to structure the data (e.g., ontologies or classification systems like CIDOC-CRM)	51	5.0	3.92	1.61	30	4.0	3.73	1.68	15	5.0	4.33	1.35
VAR18_5	Protocols	References to documentation protocols and/or information systems (e.g., context sheets, field diaries, ArcGIS, Intrasis) used to structure the data collection process and/or the data	51	5.0	4.53	1.03	30	5.0	4.30	1.24	15	5.0	4.87	0.52
VAR19_1	Purpose	Explanation of original purpose of data creation/collection (e.g., research questions, aims of study or data collection)	49	5.0	4.69	0.55	29	5.0	4.62	0.62	14	5.0	4.79	0.43
VAR19_2	Narrative	Narrative of how the data was collected/created (e.g., field/excavation/lab diary or notebook, text in report)	50	5.0	4.38	0.95	30	4.5	4.30	0.79	14	5.0	4.86	0.36

Variables		How important is it to include the following information when you are archiving or publishing an archaeological dataset?	All				Researchers				Administrators			
			N	median	mean	SD	*N*	median	mean	SD	*N*	median	mean	SD
VAR19_3	Annotations	Data creators' annotations (e.g., particularly noteworthy aspects of dataset)	50	5.0	4.38	0.83	30	5.0	4.30	0.92	14	5.0	4.50	0.76
VAR19_4	Env condits	References to environmental conditions (e.g., weather, time of the day or year, conditions in laboratory/archive) when the data was created/collected	49	4.0	3.86	1.00	30	4.0	3.70	1.06	14	4.0	4.00	0.96
VAR19_5	Not collected	Explanation of what data was not collected, created or included in the dataset (e.g., incomplete or unreadable data files, auxiliary working notes, data of peripheral significance)	50	5.0	4.44	0.70	30	5.0	4.47	0.68	14	5.0	4.43	0.76
VAR20_1	Tools	References to tools (e.g., trowel, shovel) used in fieldwork	50	4.0	3.84	1.22	30	4.0	3.87	1.17	14	4.0	3.79	1.42
VAR20_2	Measure dev	References to measurement devices (e.g., total station, tape measure) used in fieldwork	50	5.0	4.24	1.04	30	5.0	4.17	1.18	14	5.0	4.43	0.76
VAR20_3	Analys proc	References to analysis procedures (e.g., chemical analysis, C14 dating) used in conjunction with the fieldwork	50	5.0	4.64	0.63	30	5.0	4.60	0.72	14	5.0	4.79	0.43
VAR20_4	Camera	References to type, model and settings of camera used in taking photographs	50	4.0	3.36	1.31	30	3.0	2.97	1.27	14	4.0	4.00	1.04
VAR20_5	Software	References to the names and versions of software packages, scripts and code used when processing digital data	49	4.5	4.14	0.98	30	4.5	4.10	1.03	14	4.5	4.29	0.91
VAR20_6	Data process	Details on data processing procedures (e.g., how a software package was used, including details on settings, extensions)	50	5.0	4.10	0.99	30	4.0	3.93	1.05	14	5.0	4.43	0.94
VAR21_1	Cred everyone	Credentials of everyone involved in creating/collecting the data	51	4.0	3.76	1.29	30	4.0	3.80	1.21	15	4.0	3.93	1.16
VAR21_2	Cred responsib	Credentials of the person who is responsible for the project and data creation/collection	52	5.0	4.19	1.10	30	4.0	4.00	1.29	16	5.0	4.62	0.62
VAR21_3	Names everyone	Names of everyone involved in creating/collecting the data	52	4.0	4.08	1.13	30	4.0	4.20	0.96	16	4.0	4.12	0.96
VAR21_4	Name responsib	Name of the person who is responsible for the project and data creation	52	5.0	4.56	0.70	30	5.0	4.60	0.67	16	5.0	4.50	0.73
VAR22_1	Management	How the data has been managed and curated since it was collected (incl. information on how it was eventually digitised)	52	5.0	4.44	0.73	30	5.0	4.40	0.77	16	5.0	4.50	0.73
VAR22_2	Versions	Information about versions of the dataset	52	5.0	3.96	1.34	30	4.0	4.13	0.86	16	5.0	4.00	1.67
VAR22_3	Use	How, where and by whom the data has been used so far	51	4.0	3.49	1.29	29	4.0	3.79	0.98	16	4.0	3.25	1.39
VAR22_4	Review	That the data has been reviewed through peer review/by journal reviewer/by data reviewer or repository curator	52	3.0	3.67	1.15	30	4.0	3.87	0.90	16	3.0	3.25	1.39
VAR22_5	Citing instr	Instructions for how the dataset should be cited	52	4.0	3.90	1.16	30	4.0	4.07	0.87	16	4.0	3.81	1.38
VAR22_6	Rel datasets	Information about related datasets	52	4.0	3.90	1.12	30	4.0	4.03	0.89	16	4.0	3.94	1.24
VAR22_7	Rel pubs	Information about related publications	52	4.0	4.12	0.78	30	4.0	4.20	0.71	16	4.0	4.00	0.89

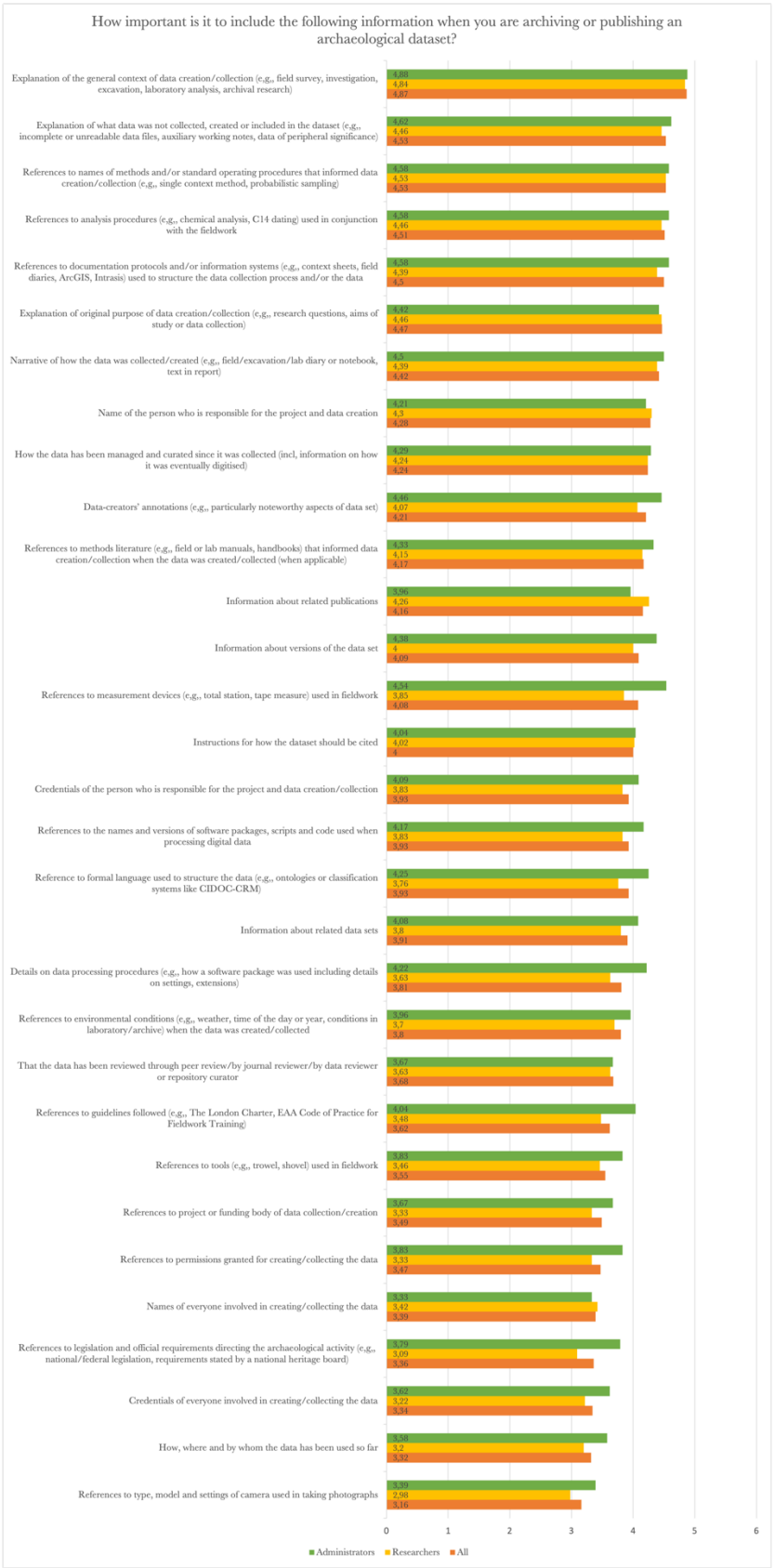

Figure 2. Perceived importance of data-related information from a data making perspective (means in decreasing order on a five-point Likert-like scale from one (Not at all useful) to five (Absolutely essential).

In addition, by mean, the camera settings ranked lowest (mean 3.36, SD 1.31) with a comparably high SD, how the data has been used previously (mean 3.49, SD 1.29); however, by median, the lowest ranking information was the review status of documentation (VAR22_4).

In the open-ended answers to the question of what information is important to document from data users' perspective, the respondents emphasised the importance of documenting methods of data creation (34 respondents), factors influencing the reliability, uncertainties, bias, errors, mistakes, limitations and exclusions related to data (18), the extents, scope, sampling, systematicity and, in general, what was observed, documented, and preserved (13), and description of data point, which included terms, codes, formats, and controlled vocabularies used with data (12). In addition, multiple respondents emphasised the importance of knowing the location of data creation and referred in broader terms to the importance of documenting and preserving the "context" of data creation (e.g., #21, #60, #87, and #91). From a data creation perspective, the respondents referred to a variety of information that is important to document. The main categories were the context of the origin of the data (40), methods (35), and description of the data (31).

Finally, Table 3 lists tools the respondents use for documenting data collection and creation procedures. In total, 83% (43 versus 9 non-users, or 43/52) use spreadsheets and take photographs, and 81% (42 versus 10 non-users, or 42/52) use text editors or word processing software. Relatively few indicated that they use dedicated diary or notebook software (10%, 5/52) or audio recording (8%, 4/52). Figure 3 provides an overview of the tools in decreasing order of popularity.

Table 3. Tools used by the respondents to document data collection and creation procedures ($N = 52$).

Which of the following tools do you use to document your data collection and creation procedures	Yes	No	% Yes
Handwritten notebooks	32	20	62%
Word processor or text editor (e.g., Microsoft Word, text editor)	42	10	81%
Online word processor or editor (e.g., Google Docs)	12	40	23%
Spreadsheet online or offline (e.g., Microsoft Excel, LibreOffice Calc)	43	9	83%
Dedicated notebook or diary software (e.g., Jupyter Notebook, DayOne)	5	47	10%
Project management software	22	30	42%
Taking photographs	43	9	83%
Audio recorder or recording software	4	48	8%
Video camera or recording software	12	40	23%
I don't document how I collect or create data	0	52	0%

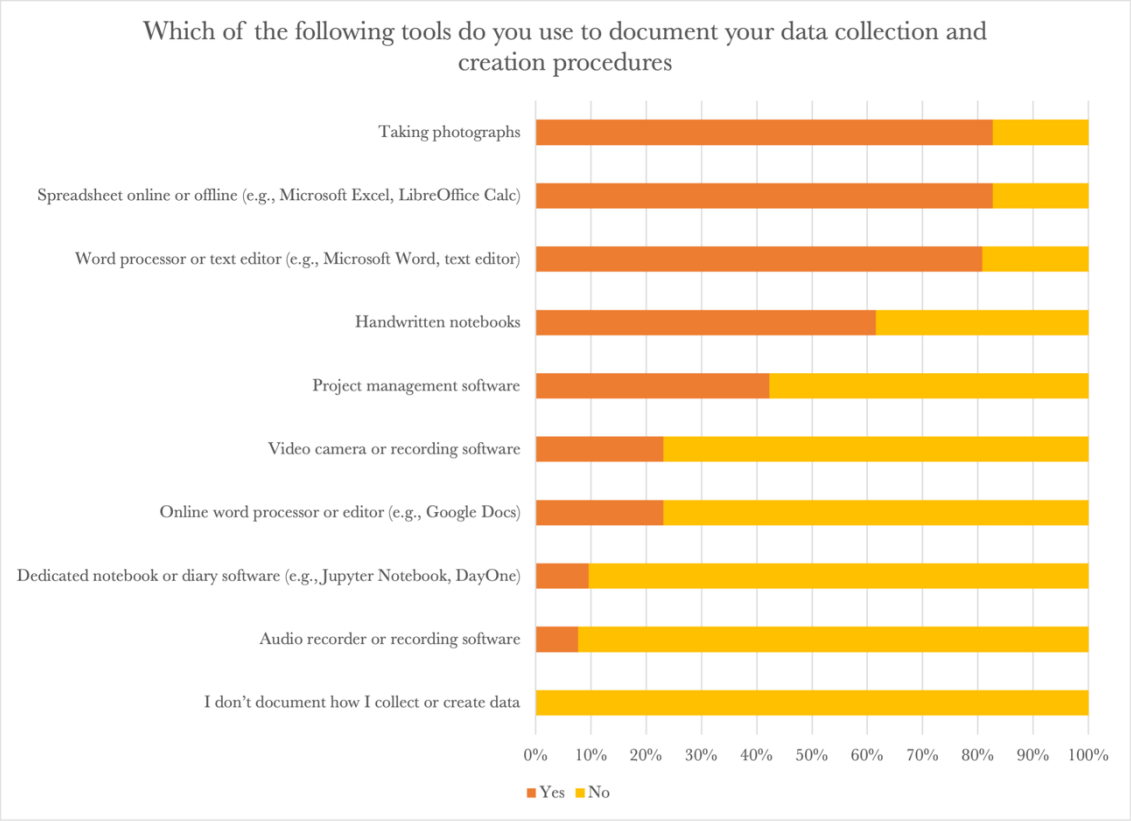

Figure 3. Tools used by the respondents to document data collection and creation procedures ($N = 52$).

Group and Response-wise Differences

Differences in the respondents' views were tested using the Wilcoxon rank-sum test in relation to the length of their career, their employer (university/research institution, or other) and work tasks (researchers versus administrators). In addition, some differences could be identified between the importance of providing particular information from data making and use perspectives. No gender-wise differences in views could be detected.

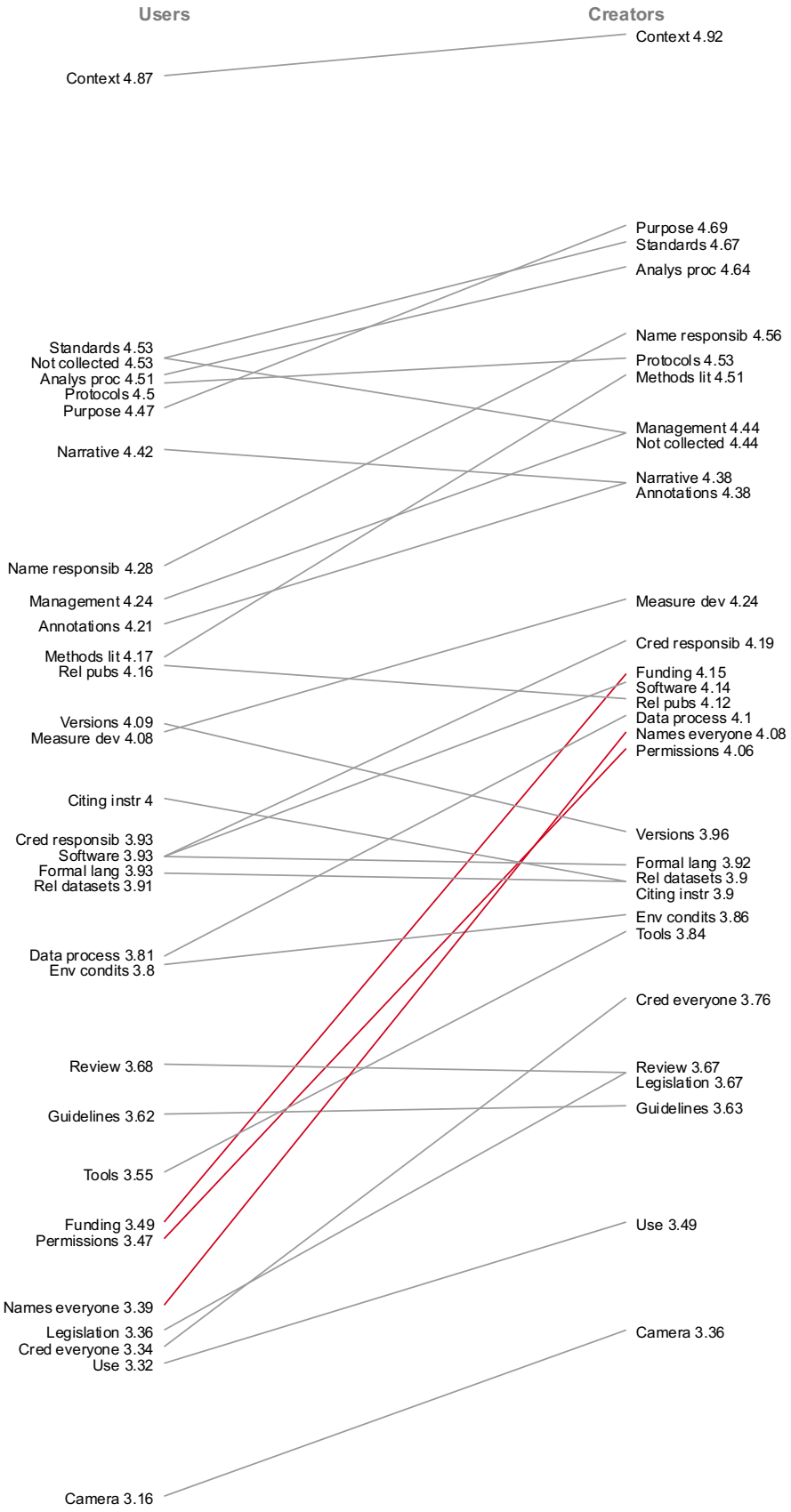

Figure 4: A slope graph showing differences between use and creation perspectives. Differences marked with red are statistically significant.

From the data user perspective, the career length of the respondents correlated negatively with the perceived importance of references to legislation and official requirements ($t = -3.0386$, df = 71, $p<.01$). From data making perspective, the career length of the respondents correlated negatively with the perceived importance of references to legislation and official requirements ($t = -5.0382$, df = 46, $p<.0001$), reference to the formal language used to structure the data ($t = -3.721$, df = 46, $p<.001$), references to documentation protocols and/or information systems ($t = -3.7105$, df = 46, $p<0.001$), data creators annotations ($t = -3.0811$, df = 46, $p<.01$), references to analysis procedures ($t = -2.8643$, df = 46, $p<.01$), references to the names and versions of software packages, scripts and code used when processing digital data ($t = -4.1858$, df = 45, $p<.001$), and details on data processing procedures ($t = -3.3412$, df = 46, $p<.01$).

Some differences exist from the data making perspective between university and research institute employees, and others that included museums, government, and private organisations (for descriptives see Table 2). University employees consider that it is less important to provide an explanation of the original purpose of data creation/collection (4.50 versus 4.88 W = 191, $p<0.01$) and the name of the person who is responsible for the project and data creation (4.27 versus 4.85 W = 209, $p<.01$) but more important to give references to the names and versions of software packages, scripts, and code used when processing digital data (4.50 versus 3.80 W = 426.5, $p<.01$) and instructions for how the dataset should be cited (4.27 versus 3.54 W = 489.5, $p<.01$).

The differences between use and creation perspectives are visualised in Figure 4. References to project or funding body of data collection/creation (4.15 versus 3.19, W = 1321.5, $p<.001$), references to permissions granted for creating/collecting the data (4.06 versus 3.47, W = 1396, $p<.01$), names of everyone involved in creating/collecting the data (4.08 versus 3.39, W = 1261.5, $p<.001$) were considered to be more important from data- making than use perspective.

Several differences relating to information are considered to be important from data making and data use perspectives between researchers and administrators could be observed; however, none of them were statistically significant (for descriptives, see Tables 1 and 2).

Discussion

The findings confirm several earlier observations of the preferences of data makers and reusers. Even if the survey dataset was of moderate size, the perceptions diverged a lot, suggesting a high variety of data and paradata needs and preferences (cf. Gregory & Koesten, 2022; Börjesson et al., 2022; Borgman et al., 2019). Many respondents indicated, similarly to earlier cross-disciplinary studies (Gregory et al., 2020a), that they used multiple types of data in their work. In contrast to Borgman et al. (2019), earlier observations that archaeology-heavy Dutch data archive DANS, data creators and users did not as clearly form two distinct groups even if the data creation and use perspectives differed from each other to some degree. Data reusers and creators find explanations of the general context of data creation (e.g., field survey, investigation, excavation, laboratory analysis, archival research) as the most useful type of information. This corresponds with earlier observations on the importance of diverse contextual information (e.g., data collection methods, environmental conditions, project goals, and hypotheses in Thomer, 2022), despite the intuitiveness of paradata (i.e., information on

processes and practices), as an information category it remains vague and literally context-dependent. Rather than describing an information need, it can be argued to refer to a knowledge need.

In addition, data users value references to methods, tools, protocols, and the information systems used. These are all aspects of data creation emphasised as important in the literature but likewise criticised for not necessarily being documented in enough detail (Huggett, 2014; Reilly et al., 2021). This might be changing as the online publishing of instruments for data collection and research tools becomes more popular as supplementary materials to journal articles, datasets, and data papers. Nonetheless, to date, the submission of these materials is often optional.

The purpose and narrative description of how and why data was created or collected and what was not collected was also considered useful (cf. e.g., diary-keeping Berggren et al., 2015). In contrast, previous reuse and the names and credentials of everyone involved in data creation/collection were considered the least useful, in addition to camera settings that are of interest in particular instances of data reuse. The low scoring of the names of those involved in data collection aligns with earlier observations of the disciplinary norms of how data is perceived and attributed, even if it goes against the ideals of reflexive archaeology (Berggren & Hodder, 2003) and the often emphasised practical importance of first-hand information when interpreting archaeological fieldwork data. However, it conforms with the practices of attributing data work in field reports and datasets to those who are responsible rather than to everyone involved in the work (Huvila, 2017).

The respondents' views largely aligned; however, there were some differences. Overall, the usefulness of the different categories of paradata was generally rated higher from the data creation than the use perspective (Figure 4) even if most of the differences remained statistically insignificant. Explanations of what data was not collected, created, or included (e.g., similar to Börjesson et al., 2022; David, 1991) and a narrative description of how the data was created were valued higher from the use than from the maker's perspective. This applies to version information and instructions on how to cite the dataset. Both are understandable and can be linked to the earlier documented tendency to focus on documenting the data in datasets rather than studies that made data happen (Kim et al., 2019); however, there are recent exceptions (e.g., Thomer, 2022; Liu et al., 2022). Data users' interest in narrative descriptions has also been observed in earlier studies (e.g., Ninkov et al., 2023). For data makers, considering what was not done or made is far less intuitive than focussing on the actual data in hand. The importance of version information and citation instructions are also, to some extent, aspects of datasets that have apparent relevance first when the dataset is (re)used rather than at the moment of its making.

Another clear difference arising from the paper was that administrative information on funding, permissions, and the names of the individuals involved in the work was perceived as more important from the data creators' than users' perspectives. While such information is important from archiving, record-keeping, and data publishing (cf. Willis et al., 2012) perspectives for data makers, their sponsors and for justifying the making and publishing of data (e.g., Huvila, 2016b), administrative details are not necessarily as crucial for using a particular dataset work practice as long as it is available and technically usable. Much of archaeological data is available from public authorities, which reduces the practical need for dataset-level information on, for example, permissions and funding.

From a data making perspective, the respondents also put weight on documenting references to methods literature. This contrasts with earlier observations of the scarcity of references to methods papers and manuals in the archaeological grey literature and

publications (Huvila et al., 2021; Huvila, 2022; Börjesson, 2015). At least a part of this seeming discrepancy can possibly be explained by different views of what types of methods literature should be cited and where. Earlier studies suggest that in contrast to standard handbook literature, citing non-archaeological and non-standard methods is more common and could be assumed to be considered important even if it would not necessarily be performed very often.

Among the demographic factors, the career length of the respondents was associated with differences in the views of what information was considered important. Experienced respondents rated the importance of information lower than junior ones, similar to, for instance, Faniel et al., (2012) and Yoon (2017) but in contrast to Friedrich (2020). A plausible explanation for this tendency is that experience and tacit knowledge make explicit documentation redundant; however, less experienced individuals need more explicit cues to make sense of data, supposedly even when experts would be working on more complex research questions. The differences between researchers and administrators could suggest the differences in work duties and priorities. Administrators are more interested in the data on a dataset level, including its purpose and creators (e.g., Huvila, 2021); however, researchers focus on how they can use the data directly in their research work (including how to use and cite it).

Typical tools used for data documentation include spreadsheets, photography, and text or word processing software. Audio recording and dedicated diary software were rare; however, over half (62%) of the respondents took handwritten notes. While there is little previous quantitative evidence of archaeologists' use of specific documentation methods, the results are reasonable in light of earlier remarks on the continuing popularity of handwritten documentation (Huvila, 2016a; 2019) and the digitalisation of everyday information processing in archaeology (Huggett, 2017). Similarly, it suggests that many of the digital methods to capture paradata including videos, 3D data-capturing, and digital drawing (e.g., Berggren & Hodder, 2003; Jensen, 2018; Dell'Unto, et al., 2017; Berggren et al., 2015) have not (yet) become mainstream.

When interpreting the findings and their implications, some obvious limitations need to be taken into account. The paper was based on a convenience sample with an unknown bias, meaning that the views are not representative of the general population of data makers or users. However, considering the exploratory rather than the confirmatory aims of the paper, it provides a workable basis for studying the variety of perspectives to process documentation.

Conclusions

The survey results show that while data makers' and users' perceptions of which information on data-related processes and practices (paradata) is important to document and data had differences, the preferences aligned to a considerable degree. The respondents agreed on the importance of having an explanation of the original general context of data creation (e.g., field survey, investigation, excavation, laboratory analysis, archival research) and knowing the purpose, procedures and methods of data making, analysis and documentation. Data users rated narrative descriptions of data creation higher than data makers; however, the latter were more inclined than the users to value administrative information and references to methods literature. A plausible explanation is that the perceived importance of particular types of information is contextual to the practices of data making and use. Data users find it important to make data usable in the context of use; however, data makers focus on information that is available to be

documented and is relevant from the perspective of making data. The respondents with a longer professional career rated the importance of several types of information lower than their junior colleagues, suggesting that experience and tacit knowledge can compensate for an eventual lack of explicit paradata. Overall, the findings underline that there is a need to continue developing and testing ideas on how to capture and document paradata and to find ways to help data makers adopt proven practices to facilitate this. Simultaneously, it is crucial that the paradata aimed at facilitating data use is relevant for data users rather than, for instance, technical or administrative details considered useful primarily by data makers.

Acknowledgements

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 818210 as a part of the project CApturing Paradata for documenTing data creation and Use for the REsearch of the future (CAPTURE).

References

Austin, C. C., Brown, S., Fong, N., Humphrey, C., Leahey, A., & Webster, P. (2016). Research Data Repositories: Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements. *IASSIST Quarterly*, *39*(4), 24–24.

Bates, M., Wilde, D. N., & Siegfried, S. (1995). Research practices of humanities scholars in an online environment: The Getty Online Searching Project Report no. 3. *Library and Information Science Research*, *17*, 5–40.

Beaudoin, J. E., & Brady, J. E. (2011). Finding visual information: A study of image resources used by archaeologists, architects, art historians, and artists. *Art Documentation*, *30*(2), 24–36.

Berggren, Å., Dell'Unto, N., Forte, M., Haddow, S., Hodder, I., Issavi, J., Lercari, N., Mazzuccato, C., Mickel, A., & Taylor, J. (2015). Revisiting reflexive archaeology at Çatalhöyük: Integrating digital and 3D technologies at the trowel's edge. *Antiquity*, *89*(344), 433–448.

Berggren, Å., & Hodder, I. (2003). Social Practice, Method, and Some Problems of Field Archaeology. *American Antiquity*, *68*(3), 421–434. Retrieved from http://www.jstor.org/stable/3557102

Boddington, M. (2014). *Truth and Archaeology: Justification in Archaeology*. [Doctoral dissertation, University of Cambridge] University of Cambridge Research Repository. Retrieved from https://doi.org/10.17863/CAM.35319

Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, *70*(8), 888–904.

Börjesson, L. (2015). Grey literature – grey sources? Nuancing the view on professional documentation: The case of Swedish archaeology. *Journal of Documentation*, *71*(6), 1158–1182.

Börjesson, L., Huvila, I., & Sköld, O. (2022). Information needs on research data creation. *Information Research*, *27*(Special Issue), Paper isic2208.

Brandsen, A., Lambers, K., Verberne, S., & Wansleeben, M. (2019). User requirement solicitation for an information retrieval system applied to Dutch grey literature in the archaeology domain. *Journal of Computer Applications in Archaeology*, *2*(1), 21–30.

Brandsen, A., Verberne, S., Lambers, K., & Wansleeben, M. (2021). Usability evaluation for online professional search in the Dutch archaeology domain. *arXiv preprint.* Retrieved from http://arxiv.org/abs/2103.04437

Buchanan, S. (2019). The assemblage of repository and museum work in archaeological curation. *Information Research*, *24*(2), paper 816. Retrieved from http://www.informationr.net/ir/24-2/paper816.html

Case, D. O., & Given, L. M. (2016). *Looking for Information*. Bingley: Emerald.

Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L.-D., Kacprzak, E., & Groth, P. (2019). Dataset search: A survey. *The VLDB Journal*, *29*(1), 251–272.

Condron, F., Richards, J., Robinson, D., & Wise, A. (1999). *Strategies for digital data: Findings and recommendations from digital data in archaeology: A survey of user needs*. York: ADS. Retrieved from http://ads.ahds.ac.uk/project/strategies/toc.html

David, M. (1991). The science of data sharing. Documentation. In J. E. Sieber. (Ed.) *Sharing Social Science Data. Advantages and Challenges*, (pp. 91–115). Newbury Park: SAGE.

Dell'Unto, N., Landeschi, G., Apel, J., & Poggi, G. (2017). 4D recording at the trowel's edge: Using three-dimensional simulation platforms to support field interpretation. *Journal of Archaeological Science: Reports*, *12*, 632–645.

Faniel, I. M., Austin, A., Kansa, E., Kansa, S. W., France, P., Jacobs, J., Boytner, R., & Yakel, E. (2018). Beyond the archive: Bridging data creation and reuse in archaeology. *Advances in Archaeological Practice*, *6*(2), 105–116.

Faniel, I. M., Frank, R. D., & Yakel, E. (2019). Context from the data reuser's point of view. *Journal of Documentation*, *75*(6), 1274–1297.

Faniel, I. M., Kriesberg, A., & Yakel, E. (2012). Data reuse and sensemaking among novice social scientists. In A. Grove (Ed.) *Proceedings of the American Society for Information Science and Technology*, *49*, (pp. 1–10).

Faniel, I. M., Kriesberg, A., & Yakel, E. (2016). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, *67*(6), 1404–1416.

Friedrich, T. (2020). *Looking for Data: Information Seeking Behaviour of Survey Data Users*. [Doctoral dissertation, Humboldt-Universität zu Berlin, Berlin]. Retrieved from https://doi.org/ 10.18452/22173

Gonzalez, C., Martín-Rodilla, P., & Viorica Epure, E. (2016). Formalization and reuse of methodological knowledge on archaeology across European organizations. In *Presentation at the 44th Computer Applications and Quantitative Methods in Archaeology Conference*. Oslo. Retrieved from https://hal-paris1.archives-ouvertes.fr/hal-01425579

Gregory, K., Groth, P., Cousijn, H., Scharnhorst, A., & Wyatt, S. (2019). Searching data: A review of observational data retrieval practices in selected disciplines. *Journal of the Association for Information Science and Technology*, *70*(5), 419–432.

Gregory, K., Groth, P., Scharnhorst, A., & Wyatt, S. (2020a). Lost or found? Discovering data needed for research. *Harvard Data Science Review 2*(2).

Gregory, K., Cousijn, H., Groth, P., Scharnhorst, A., & Wyatt, S. (2020b). Understanding data search as a socio-technical practice. *Journal of Information Science*, *46*(4), 459–475.

Gregory, K., & Koesten, L. (2022). Data Needs. In K. Gregory, & L. Koesten (Eds.) *Human-Centered Data Discovery*, Synthesis Lectures on Information Concepts, Retrieval, and Services, (pp. 19–32). Cham: Springer International Publishing.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied Thematic Analysis*. Thousand Oaks, CA: Sage.

Hessing, W., Waugh, K., van Heeringen, R., & Visser, C. (2013). Evaluatie en optimalisatie waarderingssystematiek Kwaliteitsnorm Nederlandse Archeologie. Fase 1: Evaluatie. Amersfoort: Vestigia.

Huggett, J. (2014). Promise and Paradox: Accessing Open Data in Archaeology. In C. Mills, M. Pidd, & E. Ward (Eds.) *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities*. Sheffield: HRI Online Publications.

Huggett, J. (2015). Challenging Digital Archaeology. *Open Archaeology*, *1*(1), 79–85.

Huggett, J. (2017). The apparatus of digital archaeology. *Internet Archaeology*, *44*.

Huvila, I. (2014). Archaeologists and their information sources. In I. Huvila (Ed.) *Perspectives to Archaeological Information in the Digital Society*, (pp. 25–54). Uppsala: Department of ALM, Uppsala University.

Huvila, I. (2016a). Awkwardness of becoming a boundary object: Mangle and materialities of reports, documentation data and the archaeological work. *The Information Society*, *32*(4), 280–297.

Huvila, I. (2016b). If we just knew who should do it,' or the social organization of the archiving of archaeology in Sweden. *Information Research*, *21*(2), Paper 713. Retrieved from http://www.informationr.net/ir/21-2/paper713.html

Huvila, I. (2017). Archaeology of no names? The social productivity of anonymity in the archaeological information process. *ephemera*, *17*(2), 351–376.

Huvila, I. (2018). Ecology of archaeological information work. In I. Huvila (Ed.) *Archaeology and Archaeological Information in the Digital Society*, (pp. 121–141). London: Routledge.

Huvila, I. (2019). Learning to work between information infrastructures. *Information Research*, *24*(2), paper 819. Retrieved from http://www.informationr.net/ir/24-2/paper819.html

Huvila, I. (2020). Information-making-related information needs and the credibility of information. *Information Research*, *25*(4), paper isic2002.

Huvila, I. (2021). Ambiguity, standards and contextual distance: Archaeological heritage administrators and their information work. *Open Information Science*, *5*(1), 190–214.

Huvila, I. (2022). Improving the usefulness of research data with better paradata. *Open Information Science*, *6*(1), 28–48.

Huvila, I., Andersson, L., & Sköld, O. (2022). Citing methods literature: Citations to field manuals as paradata on archaeological fieldwork. *Information Research*, *27*(3), paper941.

Huvila, I., & Huggett, J. (2018). Archaeological practices, knowledge work and digitalisation. *Journal of Computer Applications in Archaeology*, *1*(1), 88–100.

Huvila, I., Sköld, O., & Börjesson, L. (2021). Documenting information making in archaeological field reports. *Journal of Documentation*, *77*(5), 1107–1127.

Huvila, I., Sköld, O., & Andersson, L. (2024). *CAPTURE enkät om skapande och användning av arkeologisk data (Version 1) [Dataset]*. Uppsala universitet. Retrieved from: https://doi.org/10.57804/fynk-v641

Jensen, P. (2018). Semantically enhanced 3D: A web-based platform for spatial integration of excavation documentation at Alken Enge, Denmark. *Journal of Field Archaeology*, *43*(sup1), S31–S44.

Jones, S., MacSween, A., Jeffrey, S., Morris, R., & Heyworth, M. (2003). From the Ground Up The Publication of Archaeological Projects: A user needs survey. A summary. *Internet Archaeology*, *4*, Article 14. Retrieved from https://intarch.ac.uk/journal/issue14/4/

Joo, Y. K., & Kim, Y. (2017). Engineering researchers' data reuse behaviours: A structural equation modelling approach. *The Electronic Library*, *35*(6), 1141–1161.

Kansa, S. W., & Kansa, E. C. (2009). Yes, it is all about you: User needs, archaeology and digital data. *CSA Newsletter*, *22*(1). Retrieved from http://csanet.org/newsletter/spring09/nls0902.html

Kim, J., Yakel, E., & Faniel, I. (2019). Exposing standardization and consistency issues in repository metadata requirements for data deposition. *College & Research Libraries*, *80*(6), 843–875.

Koesten, L., Kacprzak, E., Tennison, J., & Simperl, E. (2019). Collaborative practices with structured data: Do tools support what users need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, (pp. 1–14). New York: ACM.

Lacagnina, C., David, R., Nikiforova, A., Kuusniemi, M. E., Cappiello, C., Biehlmaier, O., Wright, L., Schubert, C., Bertino, A., Thiemann, H., & Dennis, R. (2023). *Towards a data quality framework for EOSC*. Luxembourg: EOSC Association.

Lee, L. D. J., Besiki, S., & Shuheng, W. (2020). Toward a metadata model for research information management systems. *Library Hi Tech*, *38*(3), 577–592.

Leonelli, S. (2020). Learning from Data Journeys. In S. Leonelli, & N. Tempini (Eds.) *Data Journeys in the Sciences*, (pp. 1–24). Cham: Springer International Publishing.

Li, K., Greenberg, J., & Dunic, J. (2020). Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. *JASIST*, *71*(2), 172–182.

Liu, Wu, M., Power, M., & Burton, A. (2022). *Elicitation of data discovery contexts: An interview study. ARDC user study of data discovery project report*. Caulfield East: NCRIS/ARDC.

Löffler, F., Wesp, V., König-Ries, B., & Klan, F. (2021). Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? *PLOS ONE*, *16*(3), e0246099.

Lozić, E., & Štular, B. (2021). Documentation of Archaeology-Specific Workflow for Airborne LiDAR Data Processing. *Geosciences*, *11*(1), 26.

Lucas, G. (2012). *Understanding the Archaeological Record*. Cambridge: Cambridge University Press.

Murillo, A. P. (2016). *Data Sharing and Data Reuse: An Investigation of Descriptive Information Facilitators and Inhibitors*. [Doctoral dissertation, UNC] Chapel Hill, Chapel Hill, NC.

Ninkov, A. B., Gregory, K., Ripp, C., Roblin, E., Peters, I., & Haustein, S. (2023). A survey of researchers on rewarding data citation and reuse. In *Proceedings of the International Conference on Scientometrics & Informetrics (ISSI 2023)*. Genève: Zenodo. Retrieved from https://zenodo.org/records/7823626

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing [Software]. Retrieved from https://www.R-project.org/

Reilly, P., Callery, S., Dawson, I., & Gant, S. (2021). Provenance illusions and elusive paradata: When archaeology and Art/Archaeological practice meets the phygital. *Open Archaeology*, *7*(1), 454–481.

Revelle, W. (2023). Psych: Procedures for Psychological, Psychometric, and Personality Research [Software]. Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Sandoval, G. (2021). In pursuit of a reflexive recording. An epistemic analysis of excavation diaries from the Çatalhöyük research project. *Norwegian Archaeological Review*, *53*(2), 135–153.

Selhofer, H., & Geser, G. (2015). *D2.2 Second Report on Users' Needs*. Salzburg: ARIADNE.

Sköld, O., Börjesson, L., & Huvila, I. (2022). Interrogating paradata. *Information Research, 27*(Special Issue), Paper colis2206.

Stahlman, G. R. (2022). From nostalgia to knowledge: Considering the personal dimensions of data lifecycles. *Journal of the Association for Information Science and Technology*, *73*(12), 1692–1705.

Suhr, B., Dungl, J., & Stocker, A. (2020). Search, reuse and sharing of research data in materials science and engineering–A qualitative interview study. *PLOS ONE*, *15*(9), Paper e0239216.

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, *15*(3), e0229003.

Thomer, A. K. (2022). Integrative data reuse at scientifically significant sites: Case studies at Yellowstone National Park and the La Brea Tar Pits. *Journal of the Association for Information Science and Technology*, *73*(8), 1155–1170.

Wang, X., Duan, Q., & Liang, M. (2021). Understanding the process of data reuse: An extensive review. *Journal of the Association for Information Science and Technology*, *72*(9), 1161–1182.

Willis, C., Greenberg, J., & White, H. (2012). Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology*, *63*(8), 1505–1520.

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, *68*(4), 946–956.