



# Patterns in paradata preferences among the makers and reusers of archaeological data

Isto Huvila<sup>\*</sup>, Lisa Andersson, Olle Sköld

Department of ALM, Uppsala University, Box 625, 75126, Uppsala, Sweden

## ABSTRACT

Knowledge of data reusers' and makers' preferences of data that describe processes and practices (paradata) remains limited, especially concerning broader patterns of such priorities. The aim of this study is to address this gap. Drawing on an exploratory factor analysis of a survey of makers and users of archaeological data, the study investigates 1) what patterns related to types of informational content can be identified in data makers' and users' views of the usefulness of specific types of paradata, 2) how the patterns differ between data makers and users, and 3) how the patterns can be explained in terms of information needs and preferences. The findings show that paradata preferences are patterned and there are differences between data-makers and data-users ideas of what is useful. However, the differences limit to details that make data related processes and practices understandable rather than to the broader patterns of what types of information is needed. We identified five broad categories of uses for paradata (Data collection procedures and tools, Data in context, Standards and guidelines, Credentials, Data processing), and corresponding, applicable types of paradata. The findings point also to indicative possibilities of linking paradata preferences to orientational, contextualising and content-oriented data practices. From a practical perspective, this study underlines the importance of approaching paradata not as a monolith but rather as an arrangement that is structured by different understandings of (para)data and how it is acted upon. Instead of caring for paradata in general, it is crucial to engage with specific types of paradata for different data practices. Keywords: paradata, archaeology, data management, data reuse, research data management.

## 1. Introduction

Data makers and users have different perceptions of the usefulness of specific types of paradata i.e. (meta) information relating to the process of making, collecting and processing of (research) data (Huvila, 2022). Earlier studies have started to identify different types of paradata needs, including reasons for why paradata is needed (Börjesson, Huvila, & Sköld, 2022) and preferences of particular sources of paradata (e.g., Durrant et al., 2017; Richards-Rissetto & von Schwerin, 2017; Huvila et al., 2021) i.e. that particular types of paradata are considered useful and some others not. However, so far the understanding of the broader patterns of how the preferences of specific types of paradata are linked to each other from data makers and users perspectives remains limited (Huvila, 2022). A better grasp of how the needs of particular types of information on data making, processing and use are related to each other would enhance the understanding of paradata, how it works, how it is used and what purposes it serves, data-related information behaviour, and could inform data documentation practices and development of data infrastructures.

The aim of this study is to create new knowledge on the perceptions of paradata among data makers and users with a focus on identifying patterns in what types of paradata are considered to be useful in

conjunction with each other. The three specific research questions addressed are: 1) what patterns related to types of informational content can be identified in data makers' and users' views of the usefulness of specific types of paradata, 2) how the patterns differ between data makers and users, and 2) how the patterns can be explained in terms of information needs and preferences.

The study draws on an exploratory factor analysis of international survey data collected from a convenience sample (n = 79) representing makers and users of archaeological data. A comprehensive overview of the survey findings that underline the general preference for comprehensive contextual information has been published elsewhere (Huvila et al., 2024). The present study expands the earlier study by focusing on broader patterns in paradata preferences rather than the perceived significance of individual information types. The diversity, temporal and disciplinary breadth of archaeological data ranging from material artefacts and features to historical and contemporary textual evidence, measurements, geographical, geological and for example biological and geophysical data (Kansa, 2005; Lucas, 2012) makes it a fruitful case for exploring diverse types of paradata related preferences.

In this text, we refer to *data making* as 'practices of creating, producing and literally making data to happen and exist' (Huvila et al., 2024) and data makers as individuals who have deposited data in an

<sup>\*</sup> Corresponding author.

E-mail address: [isto.huvila@abm.uu.se](mailto:isto.huvila@abm.uu.se) (I. Huvila).

archive or published it in a repository. *Data reuse* refers correspondingly in this study to ‘secondary use of data after data creation’ (Huvila et al., 2024). When discussing data makers and users, it is important to note that the roles are not mutually exclusive; data makers sometimes are data (re)users to others’ and their own data.

## 2. Literature review

In contrast to how data is frequently imagined and approached in everyday contexts as a crude uncomplicated resource, a long line of social and critical studies of data and data practices have unveiled its complexity and socio-technical and -material nature (Monteiro, 2022). Data is never raw and every time it is used, it is ‘cooked’ anew (Bowker, 2013) for the situation in hand. Moreover, besides being made, data also participates in the making of other things (Latour & Woolgar, 1986) meaning that data practices shape data, data shapes practices, and the practices, data and the methods assemblages they are a part of have long-reaching implications. Ruppert and Scheel exemplify this by showing how data practices in European population statistics make up what is understood as ‘European population’ (Ruppert & Scheel, 2021). Striving towards unpacking data, the feminist scholarship of science has been in the forefront by emphasising the significance of identifying makers and how the makers position themselves, i.e. account for their position in the society (Haraway, 1991), to the making of data and its implications (D’Ignazio & Klein, 2020; Shepard, 2022). The growing corpus of literature investigates data creation and use practices across scholarship, and beyond in professional work and leisure (Huvila, Douglas, et al., 2022; Gregory & Koesten, 2022b) links to the related endeavour of scrutinising the sociomaterialities of data in different states of making and becoming (e.g., Bates et al., 2016; Gitelman, 2013; Pink et al., 2018). The literature points to a diversity of data use(s). Wynholds et al. (2012) distinguish background use for comparison, instrument calibration and other research support purposes of data from foreground use to ask and answer new research questions. Among users of the Dutch data archive DANS, the background use of data was dominating (Borgman et al., 2019). Gregory et al. (2020) have developed another typology of data uses based on different phases of academic work consisting of data (re)use (as a basis for a new study), project creation and preparation, conducting research (i.e. using data as model inputs or calibration of instruments), data analysis and sense-making, and teaching. Koesten et al. propose a third categorisation, of three clusters of activity patterns in data sense-making: *inspecting* that provides data-users a broad overview of the data, *engaging with content* in detail, and *placing* ‘data in relation to the world and different contexts’ (Koesten et al., 2021). Earlier studies have also surveyed reasons of not using secondary data, concluding that the major reasons across disciplines tends to be its perceived irrelevance to the applied research approach (Gregory et al., 2023).

A parallel line of studies to the research on data and information creation practices discuss data makers’ and users’ perceptions of what makes data and various forms of data-on-data, including metadata and paradata, useful for them (e.g., Börjesson, Huvila, & Sköld, 2022; Löffler et al., 2021). Depending on whether data makers and users form two distinct groups (as e.g. in Borgman et al., 2019) or if data makers are also using secondary data (Gregory et al., 2023) is likely to have an impact on their views. Besides (meta)data on data, earlier studies have underlined the importance of adequate contextual information (Benardou et al., 2018; Wilkinson et al., 2016), including that conveying a sufficient understanding of how data was generated and how it has been reworked (Faniel & Yakel, 2017; Huggett, 2016) and used (Ninkov et al., 2023) ever since. Bishop and Kuula-Luomi (2017) report that the reusers of social science data tend to find information on data, data collection and analysis important whereas details of the larger project within which it was generated are less interesting. A part of such information is often available in metadata but paradata is a broader category of information available across data documentation and research outputs, secondary

materials and the data itself (Huvila, 2022). Börjesson and colleagues (2022a) have identified paradata needs relating to the scope and provenance of data, the methods of data generation, and the organisation and representation of data. Earlier research has also pointed to a gap between what information data makers usually provide and what data users need (Faniel et al., 2018). Unsurprisingly, the findings on the needs and perceived usefulness of particular types of data have close affinities with earlier findings from information needs and relevance research. ‘Visceral needs’ differ from the explicitly articulated ones (Taylor, 1968), needs may be directed towards sense-making (Dervin, 1998), orientation and reorientation (Chew, 1994), or construction of new information, knowledge or data to solve a particular problem or complete a task (Cole, 2012, 2018; e.g., Belkin et al., 1982; Byström & Järvelin, 1995).

Besides discussing tools and standards and developing conceptual frameworks, data reuse literature delved into the experiences, enablers and barriers of data reuse (Wang et al., 2021). The findings from the studies of data reuse points to a high-level of heterogeneity in practices with a notable disciplinary, usage and methods-based (Shen, 2017; Tenopir et al., 2015; Yan et al., 2020) and sometimes geographical variation (Murillo et al., 2020). Data-users need to be able to make sense of datasets on an overarching level, to be able to probe into a necessary level of detail and grasp both conventions and peculiarities in how data is made and documented (Koesten et al., 2021). Data users also sometimes appreciate narratives, and especially in STEM disciplines, metrics of how particular datasets have been used in previous work (Ninkov et al., 2023). Rather than being general, data needs tend also to be specific for particular tasks (Koesten et al., 2017) and dynamic and overlapping with each other (Gregory & Koesten, 2022b). Data makers also report difficulties in describing their decisions and their underpinning rationales in minute detail (Murillo et al., 2020). Data making and use in sciences is typically characterised by large-scale projects while in many humanities disciplines, the focus is often on small-scale datasets. However, large-scale data making and use does also take place in typically small-scale disciplines like history and classical studies (e.g., massive historical editorial projects and more recent work with digital corpus analysis cf. Cooper, 2020) similarly to how scientists conduct small-scale investigations in fields characterised by macro-scale data use, for instance, in astronomy (Wynholds et al., 2011).

Archaeologists’ data practices and data-related information needs have been studied to a certain extent both separately (Börjesson, Huvila, & Sköld, 2022; Huvila, 2020; Selhofer & Geser, 2015; Condron et al., 1999) and as a part of cross-disciplinary comparisons (e.g., Gregory et al., 2019). Archaeologists are often very specific about what type and form of data satisfies their needs (Borgman et al., 2019). While their data needs often concern information on archaeological primary material such as datings, finds spots, measures, appearance, material (Huvila, 2014), artefact types (Brandson et al., 2021), and identifiers (Faniel et al., 2018), archaeologists’ paradata needs tend to relate to making sense of the data and their scope, their provenance, the methods used in data making and manipulation (Börjesson, Huvila, & Sköld, 2022), and determining their credibility (Huvila, 2020). In a broader sense, the needs concern the possibility to follow the ‘data journey’ (Leonelli, 2020) of the data and the different events (incl. e.g. formatting, analysis, and sharing) on the way (Li et al., 2020). This helps not only to use data now or reuse it afterwards but also to ‘interuse’ (Curty in Murillo et al., 2020) it throughout the continuum of data. However, while the need for transparency is broadly acknowledged, archaeologists have criticised that the relative consensus has not led to generally accepted best practices. It is still not well understood what paradata should be collected (Oggenhaffen et al., 2021) and how to proceed beyond individual preferences to determine what exactly is the ‘something akin to paradata/provenance’ (Huggett, 2014) different data makers and (re)users find relevant and useful to know (Huggett, 2020).

### 3. Methods and material

An online survey was administered to individuals who, according to themselves, had used and/or created and deposited archaeological data—understood broadly and ranging from digital data to finds collections. The survey was used to collect views of what information data (re)users consider they need to know about data to use it effectively and what data makers deem important for others to know about their data. The focus of the survey was 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 rather than to determine their prevalence in the population. The questionnaire included sections surveying the importance of different aspects of data considered important from data reuse and making perspectives targeted to respondents who indicated that they had experience respectively on data reuse and making. Only those respondents who indicated having engaged in data making and/or use were asked their views from the respective perspectives. None of the questions were obligatory to answer. The survey instrument was carefully designed and tested to minimise response bias due to differences in how the respondents interpreted the survey questions and to make the questions as generally applicable as possible. Questions on basic demographics included professional background, the country where the respondent was professionally based, and the length of their research career. Survey participants were recruited using relevant archaeology-related mailing lists, social media and personal contacts. The more than 200 members of the COST Action ARKWORK (<https://www.cost.eu/actions/CA15201/>) representing 30 European countries and non-European partners were asked to distribute the survey in their respective countries. The survey data is openly available in DORIS repository at <https://doris.snd.se> (CAPTURE survey on making and using archaeological data) id '2024-429'.

In total 91 responses of which 79 were complete for analysis were collected in the survey between February and October 2021. The survey data represents a convenience sample with an unknown unmeasurable. It is conceivable that individuals with a specific interest in data making, sharing and reuse are overrepresented in the sample. However, as those with little interest and experience are less likely to have strong opinions on the matter, we consider that the sample is adequate for the exploratory purposes of this study. 44% (35/79) of the respondents were female, 49% (39/79) male, 1% (1/79) other while the rest preferred not to say. Geographically the survey covered all continents except Australia and Antarctica with 81% (64/79) European respondents and in total 33 countries represented. 52% (41/79) worked as researchers and the rest in diverse, primarily administrative and data management related duties. 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. 4% (3/79) had no organisational affiliation. 57% (45/79) had a permanent employment or were self-employed, 28% (22/79) had a temporary contract, and 10% (8/79) were students or trainees. 9% (7/79) worked as directors of a research centre, institute or laboratory. The median length of career among the respondents (75/79) was 14 years (range 1–45). The respondents were engaged in a broad variety of types of archaeology-related research from field research to spatial analysis, osteology and artefacts studies drawing on various types of research materials including spatial, photographic, 3D, finds and geological data.

In total 31 statements (Table 1) relating to paradata needed in data reuse and to what needs to be documented at creation time measured on a 5-point Likert-like scale (from disagree completely to agree completely) were analysed in R 4.2.2 statistical software using the package *factanal* (R Core Team, 2023) to carry out an exploratory maximum-likelihood factor analysis (EFA) of the data with Principal Component Analysis (PCA) as a method of extracting factors. The purpose of factor analysis was to identify patterns relating to preferred types of paradata. The data fulfil the commonly accepted criteria of EFA for

**Table 1**

Perceived importance of data related information from data reuse perspective.

Variables		How useful is it to have the following information about the data you are using?	n	median	mean	SD
VAR04_1	Funding	References to project or funding body of data collection/creation	76	4.0	3.49	1.19
VAR04_2	Permissions	References to permissions granted for creating/collecting the data	76	4.0	3.47	1.33
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
VAR04_4	Guidelines	References to guidelines followed (e.g. The London Charter, EAA Code of Practice for Fieldwork Training)	76	4.5	3.62	1.51
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
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
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
VAR05_5	Protocols	References to documentation protocols and/or information systems (e.g. context sheets, field diaries, ArcGIS, Intrasis) used to structure	76	5.0	4.50	0.90

(continued on next page)

Table 1 (continued)

Variables		How useful is it to have the following information about the data you are using?	n	median	mean	SD
VAR06_1	Purpose	the data collection process and/or the data 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
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
VAR06_3	Annotations	Data-creators' annotations (e.g. particularly noteworthy aspects of data set)	76	5.0	4.21	0.90
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
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
VAR07_1	Tools	References to tools (e.g. trowel, shovel) used in fieldwork	76	4.0	3.55	1.35
VAR07_2	Measure dev	References to measurement devices (e.g. total station, tape measure) used in fieldwork	76	5.0	4.08	1.37
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
VAR07_4	Camera	References to type, model and settings of camera used in taking photographs	75	4.0	3.16	1.39
VAR07_5	Software	References to the names and versions of software packages, scripts	76	4.0	3.93	1.19

Table 1 (continued)

Variables		How useful is it to have the following information about the data you are using?	n	median	mean	SD
VAR07_6	Data process	and code used when processing digital data Details on data processing procedures (e.g. how a software package was used including details on settings, extensions)	75	4.0	3.81	1.33
VAR08_1	Cred everyone	Credentials of everyone involved in creating/ collecting the data	76	3.5	3.34	1.10
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
VAR08_3	Names everyone	Names of everyone involved in creating/ collecting the data	75	3.0	3.39	1.18
VAR08_4	Name responsib	Name of the person who is responsible for the project and data creation	76	4.0	4.28	0.81
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
VAR09_2	Versions	Information about versions of the data set	76	5.0	4.09	1.10
VAR09_3	Use	How, where and by whom the data has been used so far	76	4.0	3.32	1.18
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
VAR09_5	Citing instr	Instructions for how the dataset should be cited	76	4.0	4.00	0.91
VAR09_6	Rel datasets	Information about related data sets	76	4.0	3.91	1.00
VAR09_7	Rel pubs	Information about related publications	76	4.0	4.16	0.75

the sample size (n), the Kaiser–Meyer–Olkin test of sampling accuracy (0.77 for data reuse and 0.63 for the data making), and significance (Sig.) by the Bartlett test ( $p < 0.0001$  for both). For data reuse, a four factor and data making a five-factor model was selected on the basis of the analysis of Eigen-values and the non-triviality of the factors (cf. Costello & Osborne, 2005). The solutions fulfil Hatcher's (Hatcher, 1994) recommendation of a 5:1 subject variable ratio.

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 Wilcoxon Rank Sum (Mann Whitney U) Test with the function *wilcox.test*, Pearson correlations (*cor.test*), and linear regression analysis (*lm*). A 0.001 significance level was used in all analyses.

#### 4. Results

##### 4.1. Descriptive statistics

The majority of the respondents 95% (75/79) indicated that they had used data collected or created by others while 58% (46/79) had experience of depositing data in contrast to 34% (27/79) who indicated explicitly a lack of such experience. As none of the questions were obligatory to answer, the n-value for each question varies.

Table 1 summarises means for responses to ‘How useful is it to have the following information about the data you are using?’ for all respondents, those who indicated they work as researchers, and as administrators or data managers respectively. The by far highest ranking piece of information with lowest standard deviation pertains to the general context of data creation/collection (VAR05\_1) while the lowest ranking information by mean, yet with 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 median, to names (3) and credentials (3.5) of everyone involved in creating/collecting the data.

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 were ‘[e]xplanation of general context of data creation/collection’ (VAR18\_1). Also here, 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) whereas by median, the lowest ranking information was the review status of documentation (VAR22\_4).

##### 4.2. Exploratory factor analysis

Exploratory factor analysis was used to identify categories of information on data the respondents considered important respectively from data making and use perspectives. Two factor models were produced using responses to questions relevant to respective categories i.e. VAR17\_\*-VAR22\_\* for data making (Table 2) and VAR04\_\*-VAR09\_\* for data reuse (Table 1).

The EFA of statements relating to the importance of information to include for data reuse resulted in a four-factor model (Table 4) with factors (in a decreasing order of importance) named *Data collection procedures and tools*, *Data in context*, *Data processing* and *Credentials*.

Factors and career length, employer and work-role did not show any significant associations.

The EFA of statements relating to the importance of information to include from data makers’ perspective resulted in a five-factor model (Table 3) with factors named *Data collection procedures and tools*, *Data in context*, *Standards and guidelines*, *Credentials*, and *Data processing* in decreasing order of their importance.

A regression analysis of factor scores and respondents career length ( $R^2 = 0.167$ ,  $F(1,41) = 9.419$ ,  $p < 0.01$ ) shows that respondents with longer careers are likely to associate with the factor *Standards and guidelines*. Other factors and career length, or employer and work-role did not show any associations.

Our interpretation of the factors is to construe them as patterns of conceptions of what makes data making understandable. *Data collection procedures and tools* represents a perspective that portrays data as an outcome of the use of specific instruments and procedures, and data-making as systematic use of such tools and following procedures and protocols. Consequently, it emphasises the importance of knowing in detail what tools, devices, procedures and protocols were used in data-

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

Variables		How important is it to include the following information when you are archiving or publishing an archaeological dataset?	n	median	mean	SD
VAR17_1	Funding	Reference to project or funding body of data collection/creation	52	5.0	4.15	1.13
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
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
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
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
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
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

(continued on next page)

Table 2 (continued)

Variables		How important is it to include the following information when you are archiving or publishing an archaeological dataset?	n	median	mean	SD
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
VAR18_5	Protocols	References to documentation protocols and/or information systems (e.g. context sheets, field diaries, ArcGIS, Intras) used to structure the data collection process and/or the data	51	5.0	4.53	1.03
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
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
VAR19_3	Annotations	Data-creators' annotations (e.g. particularly noteworthy aspects of data set)	50	5.0	4.38	0.83
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
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
VAR20_1	Tools	References to tools (e.g. trowel, shovel) used in fieldwork	50	4.0	3.84	1.22
VAR20_2	Measure dev	References to measurement devices (e.g. total	50	5.0	4.24	1.04

Table 2 (continued)

Variables		How important is it to include the following information when you are archiving or publishing an archaeological dataset?	n	median	mean	SD
VAR20_3	Analys proc	station, tape measure) used in fieldwork References to analysis procedures (e.g. chemical analysis, C14 dating) used in conjunction with the fieldwork	50	5.0	4.64	0.63
VAR20_4	Camera	References to type, model and settings of camera used in taking photographs	50	4.0	3.36	1.31
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
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
VAR21_1	Cred everyone	Credentials of everyone involved in creating/ collecting the data	51	4.0	3.76	1.29
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
VAR21_3	Names everyone	Names of everyone involved in creating/ collecting the data	52	4.0	4.08	1.13
VAR21_4	Name responsib	Name of the person who is responsible for the project and data creation	52	5.0	4.56	0.70
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
VAR22_2	Versions	Information about versions of the data set	52	5.0	3.96	1.34
VAR22_3	Use	How, where and by whom the data has been used so far	51	4.0	3.49	1.29
VAR22_4	Review	That the data has been reviewed through peer review/by journal	52	3.0	3.67	1.15

(continued on next page)

Table 2 (continued)

Variables		How important is it to include the following information when you are archiving or publishing an archaeological dataset?	n	median	mean	SD
VAR22_5	Citing instr	reviewer/by data reviewer or repository curator Instructions for how the dataset should be cited	52	4.0	3.90	1.16
VAR22_6	Rel datasets	Information about related data sets	52	4.0	3.90	1.12
VAR22_7	Rel pubs	Information about related publications	52	4.0	4.12	0.78

Table 3

Results of the exploratory factor analysis with four factors interpreted as major categories of information types considered important to include for facilitating data reuse.

Variable	Data collection procedures and tools	Data in context	Data processing	Credentials
VAR05_5	Protocols			
VAR07_1	Tools			
VAR07_2	Measure dev			
VAR07_3	Analys proc			
VAR07_4	Camera			
VAR05_3	Standards			
VAR09_2	Versions	0.68		
VAR09_3	Use	0.85		
VAR09_6	Rel datasets	0.78		
VAR05_2	Methods lit		0.55	
VAR07_5	Software		0.90	
VAR07_6	Data process	0.53	0.65	
VAR08_1	Cred everyone			0.96
VAR08_2	Cred responsib			0.73
VAR08_3	Names everyone			0.61

making. From the data-making perspective, additional information associated with the factor are environmental conditions of data-making, what data was not collected and how the data has been curated since it was made.

*Data in context* represents a perspective that frames data in terms of datasets, their versions, their prior (re)use, and related datasets. From the data-making perspective, additional factors deemed useful are information on how the dataset should be cited and whether it has been reviewed or not.

*Standards and guidelines* factor emerged only in the data-making context. It frames data as an outcome of following legislative and administrative guidelines, and formal language and ontology standards.

According to the factor *Credentials*, data is framed by the actors who made it happen. From the data reuse perspective, the key element is knowing who was responsible for data-making and their credentials. In contrast, from the data-making perspective, the focus is on the body that funded data-making and credentials rather than the identity of individual data-makers.

*Data processing* reminds of *Data collection procedures and tools* by its focus on processes. However, while *Data collection procedures and tools* focuses on specific hands-on tools and procedures, *Data processing* operates on the level of methods and modus operandi. In the context of data-use, it orients toward references to methods literature, software, and data-processing procedures, and in data-making towards methods literature and names of methods rather than specific instruments or protocols.

## 5. Discussion

### 5.1. Perspective to data influences usefulness of paradata

With regards to the first research question, *what information type related patterns can be identified in data makers' and users' views of the usefulness of specific types of paradata*, the analysis produced two factor models interpreted as preferential patterns i.e. what broader categories of types of paradata are considered useful respectively from data making and use perspectives. In general, the findings underline the perceived importance of understanding and conveying understanding of the broad context of data-making. The significance of information both on data collection and processing and its contexts in the results is not specific to archaeology but has been observed across disciplinary contexts (e.g., Dixit et al., 2018; Faniel et al., 2018; Papenmeier et al., 2021). In contrast to some earlier research, which has linked data and data-related meta-information preferences to disciplinary domains and cultures (Gregory & Koesten, 2022b; Tenopir et al., 2015), the presence of multiple factors representing paradata related preferences identified in the present study both support the general observations of archaeology as a data-wise diverse field (cf. Kansa, 2005; Lucas, 2012) and in broader terms that there are internal differences within domains in how data and data reuse are conceptualised. The findings do also align with earlier discipline-wise comparisons where archaeology as a whole remains close to humanities and social sciences but shows a particular interest in materiality, technical measurements, typologies (Huvila, 2014), artefact types (Brandson et al., 2021), and identifiers (Faniel et al., 2018). Conceiving useful paradata in data-level terms of tools and procedures versus credentials or dataset level information points not only to a variety of data practices but also to diverging ideas of the data itself as data points generated by instrumentation, outcomes of human interventions, or larger aggregations of individual data points.

As the regression analysis shows, in the present data only the data making related factor *Standards and guidelines* showed a statistically significant association with longer career length of the respondents. While the present data does not lend itself to elaborate on the observed association, the finding supports the idea that preferential patterns cannot directly be traced back to disciplinary or organisational factors only but require attending to different understandings, conceptualisations and practices relating to data and data reuse. Simultaneously, it also points to the earlier identified complexity and situatedness of data and data-related information needs (Gregory & Koesten, 2022a) that refute un-complex explanations and require further qualitative in-depth inquiries. Unlike in Friedrich (2020) where the differences between junior and senior data users was traced back to different uses of data respectively for training and more complex research tasks, the present results could be speculated to depend on respondents' longer experience of working with archaeological data, insights in the requirements and consequences of standardisation from senior professional perspective, and in general, of documentation standards and guidelines.

The statistical analysis shows that the respondents vary in their views of what constitutes the central component of data-making. Our interpretation is that the two models form two concurrent sets of views—from data-making and data-reuse perspectives—of what aspects of data-making are crucial to be documented and communicated for data reusers to understand it. In parallel, the individual factors within the two factor models can be interpreted as different views from data-making

**Table 4**

Results of the exploratory factor analysis with four factors interpreted as major categories of information types considered important to include from data makers' perspective.

Variable		Data collection procedures and tools	Data in context	Standards and guidelines	Credentials	Data processing
VAR19_4	Env condits	0.72				
VAR19_5	Not collected	0.60				
VAR20_2	Measure dev	0.50				
VAR20_4	Camera	0.72				
VAR20_5	Software	0.61				
VAR20_6	Data process	0.77				
VAR22_1	Management	0.52				
VAR22_2	Versions		0.85			
VAR22_3	Use		0.84			
VAR22_4	Review		0.57			
VAR22_5	Citing instr		0.77			
VAR22_6	Rel datasets		0.71			
VAR17_3	Rel pubs			0.66		
VAR17_4	Guidelines			0.79		
VAR18_4	Formal lang			0.88		
VAR17_1	Funding				0.60	
VAR21_1	Cred everyone				0.88	
VAR21_2	Cred responsib				0.77	
VAR18_2	Methods lit					0.63
VAR18_3	Standards					0.94

and data reuse perspectives of what is most important to know about the data-making process for data reuse. The diversity of such elements can be linked to the positionality of data-making and use (D'Ignazio & Klein, 2020; Shepard, 2022) and to what Gregory et al. (2020) underline in data reuses: how they are enacted in different manners and have different meanings depending on how data and data reuse are conceived. The perspectives have similarities with earlier identified types of data use. They include the categories of data sense-making in Koesten et al. (2021) of 'engaging with content' that can linked to *Data processing* and *Data collection procedures and tools* i.e. a pursuit of close understanding of data and how it was made and processed. Similarly, *Data in context* aligns with 'placing' data in relation to its context. According to the present results, the constituent can vary from being the use of particular tools and procedures, knowledge of earlier dataset-level operations, compliance to established standards, guidelines and regulations, appropriate credentials of data-makers, or data-making and -processing related methods.

Regarding the second research question, *how the patterns differ between data makers and users*, the analysis shows that while there are certain differences between the two factor models (data-making and data reuse), they are conspicuously close to each other. This appears to be at odds with earlier observations of a discrepancy between what data-makers do and data reusers indicate they need (e.g., Faniel et al., 2018; also Borgman et al., 2019). A closer look at the factor models reveals, however, differences within the factors that can explain the seeming contradiction by the simultaneous presence of broader preferential patterns (factors) and divergence in details (factor loadings). The most apparent difference between the models is that in the EFA of the preferences from a data-making perspective, *Standards and guidelines* emerged as a separate factor whereas it did not stand out for data-use. There are also differences within factors. In comparison to the data-use specific factor model, *Data collection procedures and tools* and *Data in context* factors are associated with additional types of paradata in the data-making specific model.

Some of the variation between factor models can be explained by differences in what is considered a useful narrative, what needs to be named and on what level of detail when documenting and using data. However, the differences in factor loadings might also suggest a difference in how formal aspects of data documentation such as standards, curatorial provenance, quality assurance, and credentials are more crucial from a data-making perspective while data-use relies more, on partly technical and partly social, information. For example, *Standards and guidelines* that did not emerge as a separate factor in relation to data-

use, probably are more central in guiding how to generate paradata than in understanding it. The variety points to diverging roles of paradata in what Koesten et al. (2021) term 'clusters of activity patterns' in data sense-making. Of the identified factors, *Data collection procedures and tools* and *Standards and guidelines* have affinities with Koesten and colleagues' (2021) cluster engaging with content, factor *Data processing* is reminiscent of inspecting, and *Data in context* and *Credentials* have kinship with placing (Table 5). Depending on the characteristics of data and what data-users do when making sense of data, the relevant paradata occasionally needs to be detailed and technically well-standardised and sometimes overarching and orientational. The absence of a specific type of paradata needed in a particular type of data sense-making

**Table 5**

Uses of paradata and corresponding useful paradata types.

	Key question	Tentative links to associated data (sense-making) practices	Potentially useful paradata
Data collection procedures and tools	Technical means and steps of data-making	Engaging with content (Koesten et al., 2021); Foreground use (Wynholds et al., 2012)	Names of devices, tools, procedures and protocols used in data-making
Data in context	Context of making and prior use of a dataset	Placing (Koesten et al., 2021); Background use (Wynholds et al., 2012)	Information on the versions and prior use of datasets and related datasets
Standards and guidelines	Producing formally correct datasets from standards and regulations perspectives	Engaging with content (Koesten et al., 2021); Foreground use (Wynholds et al., 2012)	Standards, guidelines and regulations followed in data-making
Credentials	Dataset has been produced by an actor with necessary credentials	Placing (Koesten et al., 2021); Background use (Wynholds et al., 2012)	Names and credentials of data-makers
Data processing	What known, documented methods and processes have been followed	Inspecting (Koesten et al., 2021); Background use (Wynholds et al., 2012)	References to methods literature, software and data-processing procedures



activity is obviously a problem but diverging understandings among data-makers and data-users of what level of detail in paradata or kind of explanation or information makes various aspects of data-making intelligible can be troublesome for effective data reuse as well.

As per the third research question, *how the patterns can be explained in terms of information needs and preferences* that focuses on the question of how to interpret the differences between the factor models, the findings suggest that the perceived usefulness of paradata is not (only) related to the utility of individual types of information. Rather, the usefulness of specific types of paradata is linked to particular understandings of the key constituents of the data-making process. This can be interpreted to evince of epistemic differences not only between the practices of data making and use but also between their underlying perspectives to what is data or how it is enacted, how data works and how it is practised in a broader sense in specific situations (cf. e.g., Monteiro, 2022; Ruppert & Scheel, 2021). The perspective to paradata that focuses on *Data collection procedures and tools* appears to conceptualise data as a result of using specific sets of tools and following particular technical procedures. In contrast, *Standards and guidelines* appears to frame data as an outcome of following particular standards and legislative regulations, whereas *Credentials* attributes it to particular ‘authors’. Further, *Data in context* focuses on datasets, their provenance and relations, whereas *Data processing* emphasises methods and data-processing rather than individual tools and procedures as their key constituents.

It is possible to sense comparable paradigmatic differences identified in the earlier literature (Gregory et al., 2020) between the outspoken needs and emphasis of structural versus contextual and methodological meta-information, and how particular perspectives put weight on varying dataset-related details. Even if not necessarily directly corresponding with the categorisation of foreground and background use in Wynholds and colleagues’ (2012) study, the focus on data and datasets could suggest exploratory foreground use of data while the interest in formal details could signal instrumental background use (Table 5). Another, parallel division that reminds of corresponding categorisations of information needs (Chew, 1994; Cole, 2018) can be observed between inclinations towards technical, task and construction oriented (e.g. in *Data collection procedures and tools*, *Standards and guidelines*) and orientational information (e.g., in *Data in context*, *Credentials*). A further division the current study was unable to document due to the research design, can be undoubtedly traced back to differences in paradata needs in relation to different types of primary data (cf. Gregory & Koesten, 2022b), including qualitative or quantitative data, finds, or for example, spatial coordinate information. With the most obvious case, *Standards and guidelines* are relevant only with data types for which standards and guidelines exist. Similarly, for example, physical archaeological finds and samples in natural sciences (e.g., Ramdeen, 2015) require more descriptive both contextual and procedural information metadata whereas information on tools use is most relevant when specialised instruments are used.

## 5.2. Theoretical implications

The major theoretical implication of the findings is that data and data-related meta-information preferences cannot be explained solely on the basis of disciplinary domains and cultures (Gregory & Koesten, 2022b; Tenopir et al., 2015). Instead, the findings show that there are internal differences within domains in how data and data reuse are conceptualised. Archaeology exemplifies such internal diversity regarding data and data documentation related preferences and practices. A key difference was identified between those who conceptualise data as data points generated by instrumentation, outcomes of human interventions, or larger aggregations of individual data points. The findings similarly show how data-makers and users think differently regarding what is useful paradata. This suggests of fundamental differences in how makers and users think about data and the practice(s) through which it is generated. The findings suggest that for data-makers,

documenting data-making is about formal framing of the practice, whereas for data-users its crucial characteristic is to what degree it is capable to convey an adequate understanding of how the data came into being. Considering that many of the respondents reported having both created and (re)used data, the perspectives appear to relate to their work roles (cf. Huvila, 2008) as data makers and users rather than their individual preferences. Moreover, we are inclined to suggest that the perspectives are linked to particular epistemic understandings of data and how it is makeable, how it is acted upon in different contexts and situations, and in a fundamental sense, how it works as an ingredient of information and knowledge. Because data practices shape data and data shapes data practices, the perspectives and their related approaches to paradata have consequences far beyond research documentation. Similarly to how Ruppert and Scheel (2021) describe how data practices in population statistics enact what is ‘European population’, the identified perspectives to data and paradata and how they are acted upon in archaeology have direct influence on archaeological knowledge—whether it is an outcome of tool use, authors positioning themselves according to a particular set of credentials (cf. Haraway, 1991), or datasets being positioned against other datasets.

## 5.3. Practical implications

The major practical implication of the identified preferential clusters is that paradata is patterned. First, the presence of multiple contrasting views (i.e. factors) from both data-making and reuse perspectives on what paradata is considered to be useful suggests that succeeding in identifying and developing appropriate best practices of process documentation for all archaeological data-making and (re)use is unlikely or has to encompass multiple types of paradata, for example, as optional components of a paradata standard. Second, the findings support earlier suggestions that pursuing data management solutions that are particular to specific scientific or scholarly disciplines have limitations and should be questioned. Instead, it can be more important to consider similarities and dissimilarities in data-makers’ and (re)users’ data practices and their associated epistemological and ontological perspectives, explicit and implicit needs and preferences.

Instead of trying to develop lists covering discrete pieces of information that together form a necessary body of paradata for a particular dataset, we suggest on the basis of the present findings that it is more relevant to focus on broader functional constellations of paradata types addressing specific types of paradigmatic paradata needs and wants. Table 5 outlines proposed uses and useful paradata types based on the results of the factor analysis together with associated data (sense-making) practices identified in the literature.

If the identified factors are taken as a starting point, a comprehensive set of paradata would include an adequate documentation of 1) technical means and steps of data-making, 2) the context of making and prior use of the documented dataset, 3) standards and regulations followed during data-making, 4) names and credentials of key data-makers, and 5) references to methods used in data-making and processing. A comparison to earlier observations of what types of paradata is available, for example, in archaeological reports (e.g. (Huvila, Börjesson, & Sköld, 2022, Huvila, Börjesson, & Sköld, 2022),) and datasets (Börjesson, Sköld, et al., 2022) show that information relevant to all factors is available to a certain extent, albeit unevenly and spread out in different documents. In this respect, future data documentation and documentation guidelines would benefit from a more systematic coverage of the specific facets of paradata to be capable of comprehensively addressing their associated information (or, paradata) needs rather than attempts at exhaustive coverage of individual pieces of paradata.

The differences in perspectives suggest also that data-makers should put more focus on the narrative description of not only earlier data use (cf. Ninkov et al., 2023) but also data-making and on what data were not collected, created or included in the final dataset. While administrative information is important for management and evidential record-keeping

purposes—that are not always addressed in enough detail in archaeological documentation—it does not necessarily suffice for data reusers as an account of the data making process. Similarly, knowing funders, permissions and data-makers identity can be crucial for being able to manage data, replicate a study as a whole and to determine whether it can be used but it is not necessarily the information that makes data useable for their users (cf. Bishop & Kuula-Luumi, 2017).

#### 5.4. Limitations

The key finding of the present study relates to uncovering patterns of what types of paradata are considered useful in conjunction with each other among the makers and users of archaeological data. When interpreting the findings, it is necessary to take into account certain limitations. First, the analysis is based on a rather small convenience sample. As a result, the responses are not representative and generalisable to all archaeologists, data makers and users. The lack of statistics on individuals working with archaeological data and country-wise differences make it difficult to assess how well the study sample corresponds with the target population. What can be noted is that a large number of countries, different levels of education, types of employers, genders, both fresh and experienced individuals are represented. However, while the data has limited value for studying prevalence of views in the population, it is useable for identifying patterns in perspectival differences. Also the likely overrepresentation of individuals with a specific interest in data making, sharing and reuse is unlikely to have a major implications to the existence of specific patterns even if it undoubtedly affects their prevalence.

Second, survey studies have generally limitations in eliciting nuances and reasoning underpinning the responses. For example, respondents' actual visceral needs and preferences that can be sometimes unknown for the respondents themselves can be only partially (re)constructed from explicitly expressed formalised needs (Taylor, 1968). Individuals can also be engaged in a broad variety of data practices associated with different perspectives that are difficult to capture in a survey study. Therefore, future surveys and qualitative research with other groups and qualitative research are recommended to nuance the understanding of paradata preferences.

#### 6. Conclusions

The findings show that paradata preferences are patterned and there are differences between data-makers and data-users ideas of what is useful. However, it appears that the differences are what details make data related processes and practices understandable rather than in the broader patterns of what types of information is needed. We identified five broad categories of uses for paradata (Data collection procedures and tools, Data in context, Standards and guidelines, Credentials, Data processing), and corresponding, applicable types of paradata. We argue that focusing on broader categories rather than specific types of paradata uses is necessary for deciphering data practices in general, and more specifically unpacking of how paradata is meaningful in facilitating understanding of data, data making and processing, and supporting data use. The findings point also to indicative possibilities to link paradata preferences with specific types of data-use practices: orientational, contextualising and direct engagement with content. From a practical perspective, this study underlines the importance of approaching paradata not as a monolith but rather as an arrangement that is structured by different understandings of (para)data and how it is acted upon. Instead of caring for paradata in general, it is crucial to engage with specific types of paradata for distinct data practices.

#### Funding

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).

#### CRedit authorship contribution statement

**Isto Huvila:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Lisa Andersson:** Methodology, Investigation, Data curation, Conceptualization. **Olle Sköld:** Writing – review & editing, Methodology, Investigation, Conceptualization.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Isto Huvila reports financial support was provided by European Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

We would like to thank two anonymous reviewers for their invaluable detailed and constructive comments on earlier versions of this text. We are also grateful for our colleagues Dr Ying-Hsang Liu and Dr Jessica Kaiser for comments on the manuscript.

#### References

- Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, 3(2), Article 2053951716654502.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I: Background and theory. *Journal of Documentation*, 38(2), 61–71.
- Benardou, A., Champion, E., Dallas, C., & Hughes, L. M. (2018). Introduction: A critique of digital practices and research infrastructures. In A. Benardou, E. Champion, C. Dallas, & L. M. Hughes (Eds.), *Cultural heritage infrastructures in digital humanities*. London: Routledge.
- Bishop, L., & Kuula-Luumi, A. (2017). Revisiting qualitative data reuse: A decade on. *Sage Open*, 7(1).
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *JASIST*, 70(8), 888–904.
- Börjesson, L., Huvila, I., & Sköld, O. (2022a). Information needs on research data creation. *Information Research*, 27(Special Issue), Article isic2208.
- Börjesson, L., Sköld, O., Friberg, Z., Löwenborg, D., Pålsson, G., & Huvila, I. (2022b). Repurposing excavation database content as paradata: An explorative analysis of paradata identification challenges and opportunities. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, 6(3), 1–18.
- Bowker, G. C. (2013). Data flakes: An afterword to “raw data” is an oxymoron. In L. Gitelman (Ed.), *“Raw data” is an oxymoron* (pp. 167–172). The MIT Press.
- Branden, A., Verberne, S., Lambers, K., & Wansleben, M. (2021). *Usability evaluation for online professional search in the Dutch archaeology domain*. <http://arxiv.org/abs/2103.04437>.
- Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191–213.
- Chew, F. (1994). The relationship of information needs to issue relevance and media use. *Journalism Quarterly*, 71(3), 676–688.
- Cole, C. (2012). *Information need: A theory connecting information search to knowledge formation*. Medford, NJ: Information Today.
- Cole, C. (2018). *The consciousness' drive: Information need and the search for meaning*. Cham: Springer.
- 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*.
- Cooper, C. (2020). Introduction: Old and new approaches to ancient material culture. In C. Cooper (Ed.), *New approaches to ancient material culture in the Greek & roman world* (pp. 1–44). BRILL.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10(7).
- Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2), 36–46.
- D'Ignazio, C., & Klein, L. F. (Eds.). (2020). *Data feminism*. Cambridge, MA: MIT Press.

- Dixit, R., Rogith, D., Narayana, V., Salimi, M., Gururaj, A., Ohno-Machado, L., Xu, H., & Johnson, T. R. (2018). User needs analysis and usability assessment of DataMed – a biomedical data discovery index. *Journal of the American Medical Informatics Association*, 25(3), 337–344.
- Durrant, G. B., Maslovskaya, O., Phoenix, A., & Edwards, R. (2017). Paradata for nonresponse investigations in social surveys. In R. Edwards, J. Goodwin, H. O'Connor, & A. Phoenix (Eds.), *Working with paradata, marginalia and fieldnotes* (pp. 20–39). Cheltenham: Edward Elgar.
- 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., & Yakel, E. (2017). Practices do not make perfect: Disciplinary data sharing and reuse practices and their implications for repository data curation. In L. R. Johnston (Ed.), *Curating research data, volume one: Practical strategies for your digital repository* (pp. 103–126). Chicago: ACRL.
- Friedrich, T. (2020). *Looking for data: Information seeking behaviour of survey data users*. Berlin: Humboldt-Universität zu Berlin. Ph.D. thesis.
- Gitelman, L. (Ed.). (2013). *Raw data<sup>2</sup> is an oxymoron*. Cambridge, MA: MIT Press.
- Gregory, K., Groth, P., Cousijn, H., Scharnhorst, A., & Wyatt, S. (2019). Searching data: A review of observational data retrieval practices in selected disciplines. *JASIST*, 70(5), 419–432.
- Gregory, K., Groth, P., Scharnhorst, A., & Wyatt, S. (2020). Lost or found? Discovering data needed for research. *Harvard Data Science Review*.
- Gregory, K., & Koesten, L. (2022a). 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.
- Gregory, K., & Koesten, L. (2022b). *Human-centered data discovery*. Synthesis lectures on information concepts. *Retrieval, and services*. Cham: Springer.
- Gregory, K., Ninkov, A., Ripp, C., Roblin, E., Peters, I., & Hausteine, S. (2023). Tracing data: A survey investigating disciplinary differences in data citation. *Quantitative Science Studies*, 1–51.
- Haraway, D. J. (1991). *Simians, cyborgs, and women: The reinvention of nature*. New York: Routledge.
- Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.
- 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. (2016). Digital haystacks: Open data and the transformation of archaeological knowledge. In A. T. Wilson, & B. Edwards (Eds.), *Open source archaeology, ethics and practice* (pp. 6–29). Berlin: de Gruyter Open.
- Huggett, J. (2020). Capturing the silences in digital archaeological knowledge. *Information-an International Interdisciplinary Journal*, 11(5), 278.
- Huvila, I. (2008). Work and work roles: A context of tasks. *Journal of Documentation*, 64(6), 797–815.
- Huvila, I. (2014). Archaeologists and their information sources. In I. Huvila (Ed.), *Perspectives to archaeological information in the digital society* (pp. 25–54). Uppsala: Uppsala University. Department of ALM.
- Huvila, I. (2020). Information-making-related information needs and the credibility of information. *Information Research*, 25(4). paper isic2002.
- 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. (2022a). Citing methods literature: Citations to field manuals as paradata on archaeological fieldwork. *Information Research: An International Electronic Journal*, 27(3).
- Huvila, I., Andersson, L., Sköld, O., & Liu, Y.-H. (2024). *Data makers' and users' views on useful paradata: Priorities in documenting data creation, curation, manipulation and use in archaeology*. *International Journal of Digital Curation* (in press).
- Huvila, I., Börjesson, L., & Sköld, O. (2022b). Archaeological information-making activities according to field reports. *Library & Information Science Research*, 44(3), Article 101171.
- Huvila, I., Douglas, J., Gorichanaz, T., Koh, K., & Suorsa, A. (2022c). Guest editorial: Advances in research on information creation. *Library & Information Science Research*, 44(3), Article 101178.
- Huvila, I., Sköld, O., & Börjesson, L. (2021). Documenting information making in archaeological field reports. *Journal of Documentation*, 77(5), 1107–1127.
- Kansa, E. (2005). A community approach to data integration: Authorship and building meaningful links across diverse archaeological data sets. *Geosphere*, 1(2), 97–109.
- Koesten, L., Gregory, K., Groth, P., & Simperl, E. (2021). Talking datasets - understanding data sensemaking behaviours. *International Journal of Human-Computer Studies*, 146, Article 102562.
- Koesten, L., Kacprzak, E., Tennison, J. F. A., & Simperl, E. (2017). The trials and tribulations of working with structured data. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM.
- Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
- 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.
- 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), Article e0246099.
- Lucas, G. (2012). *Understanding the archaeological record*. Cambridge: Cambridge University Press.
- Monteiro, E. (2022). *Digital oil: Machineries of knowing*. Cambridge, MA: MIT Press.
- Murillo, A. P., Curty, R. G., Jeng, W., & He, D. (2020). Confronting the challenges of computational and social perspectives of the data continuum. *Data and Information Management*, 4(2), 119–126.
- Ninkov, A. B., Gregory, K., Ripp, C., Roblin, E., Peters, I., & Hausteine, S. (2023). *A survey of researchers on rewarding data citation and reuse*.
- Oppenhaffen, L., Lami, M. R., & Mickleburgh, H. (2021). Art, creativity and automation. From charters to shared 3D visualization practices. *Open Archaeology*, 7(1), 1648–1659.
- Papenmeier, A., Krämer, T., Friedrich, T., Hienert, D., & Kern, D. (2021). Genuine information needs of social scientists looking for data. *Proceedings of the Association for Information Science and Technology*, 58(1), 292–302.
- Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, 5(1), Article 2053951717753228.
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Ramdeen, S. (2015). Preservation challenges for geological data at state geological surveys. *GeoResJ*, 6, 213–220.
- Revelle, W. (2023). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University.
- Richards-Rissetto, H., & von Schwerin, J. (2017). A catch 22 of 3D data sustainability: Lessons in 3D archaeological data management & accessibility. *Digital Applications in Archaeology and Cultural Heritage*, 6, 38–48.
- Ruppert, E., & Scheel, S. (Eds.). (2021). *Data practices: Making up a European people*. London: Goldsmiths Press.
- Selhofer, H., & Geser, G. (2015). *D2.2 second report on users' needs*. Salzburg: ARIADNE.
- Shen, Y. (2017). Burgeoning data repository systems, characteristics, and development strategies: Insights of natural resources and environmental scientists. *Data and Information Management*, 1(2), 115–123.
- Shepard, M. (2022). *There are No facts: Attentive algorithms, extractive data practices, and the quantification of everyday life*. Cambridge, Massachusetts: The MIT Press.
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29, 178–194.
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, 10(8), Article e0134826.
- Wang, X., Duan, Q., & Liang, M. (2021). Understanding the process of data reuse: An extensive review. *JASIST*, 72(9), 1161–1182.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018.
- Wynholds, L., Fearon, D., Borgman, C. L., & Traweek, S. (2011). When use cases are not useful: Data practices, astronomy, and digital libraries. In *Proceedings of the JCDL 2011, June* (pp. 13–17). Ottawa.
- Wynholds, L., Wallis, J. C., Borgman, C. L., Sands, A., & Traweek, S. (2012). Data, data use, and scientific inquiry: Two case studies of data practices. In *Proceedings of the 12th ACM/IEEE-CS joint conference on digital libraries* (pp. 19–22). New York, NY, USA: ACM.
- Yan, A., Huang, C., Lee, J.-S., & Palmer, C. L. (2020). Cross-disciplinary data practices in earth system science: Aligning services with reuse and reproducibility priorities. In *Proceedings of the association for information science and technology. Association for information science and technology* (Vol. 57, p. e218), 1.