



Concluding Discussion: Paradata for Information and Knowledge Management

Isto Huvila, Lisa Andersson, and Olle Sköld

Abstract

This concluding chapter draws together insights from the discipline-specific chapters to contrast and synthesise the diverse approaches to how the concept of paradata is conceptualised and used in the different cases covered in the volume “Perspectives to paradata”. Paradata as a concept that refers to process information resides firmly at the fringe of codified knowledge and organisational learning. Many different forms of information can function as paradata. There is a comparable variety in how the concept of paradata is understood. Due to the variety, having a unified definition can be debated. Major opportunities with paradata range from achieving reproducibility in data analysis and use and delivering the desired outcomes of the Open movement to increasing algorithmic and administrative accountability and transparency of artificial intelligence. However, transparency brought by paradata is not automatically virtuous. Paradata ethics and its relation to general information and knowledge management ethics is central to responsible use of paradata. Besides further inquiry into paradata concept, actual instances of paradata in the wild and how they are linked to social action, it is a key area that requires further research.

I. Huvila (✉) · L. Andersson · O. Sköld
Uppsala University, Uppsala, Sweden
e-mail: isto.huvila@abm.uu.se; lisa.andersson@ki.se; olle.skold@abm.uu.se

© The Author(s) 2024
I. Huvila et al. (eds.), *Perspectives on Paradata*,
Knowledge Management and Organizational Learning 13,
https://doi.org/10.1007/978-3-031-53946-6_14

1 Introduction

Process knowledge has many faces. Its diversity parallels with and simultaneously goes far beyond the multiplicity of processes and practices themselves. In this volume, our aim has been to delve into this diversity and its implications by introducing and exploring the notion of paradata as a concept and practical tool in and for managing knowledge—whether it is a matter of transferring knowledge between two archaeologists standing next to each other in the field, promoting societal accountability, or recording a computational research code to enable reproducibility of scientific discoveries. Doing so, we expand the insight into how the paradata concept is understood in different disciplines and cases, and why paradata have proven useful in practice to answer questions or achieve goals with data, information, and knowledge. By this exploration of process knowledge, how process information can support the management of information and knowledge in different forms, and of how process information and knowledge can be managed in a variety of contexts we hope to contribute to theoretical and practical advancement in the field of information and knowledge management.

The work with this volume commenced with a working definition of paradata that all the chapter authors were asked to reflect but not necessarily agree upon. It has served as a common ground and a point of departure to discipline- and context-specific explorations of what paradata can be in different settings, what the character of the processes is meant to describe, what methods are used to find or generate paradata, what paradata can do or enable, and what needs to be considered when creating and using paradata in different contexts.

This concluding chapter draws together insights from the discipline-specific chapters to contrast and synthesise the diverse approaches to how the concept of paradata is conceptualised and used in the different cases covered in this volume. Further, we proceed to three topics of discussion emerging from the synthesising analysis. First, we discuss how paradata are done in practice by various actors using different paradata creation methods for their specific purposes. Second, we delve into the implications of paradata for the theory and practice of information and knowledge management. Before concluding with brief remarks on future directions of paradata research and practice, this chapter spends a few words to discuss a third and crucial question that remains somewhat implicit in the chapters. It is the one of ethics of paradata and potential ethical hurdles that need to be considered when paradata—descriptions of activities—are put into practice.

2 Paradata: In Plural

While the punchline of this volume is that paradata, a potentially useful concept, has been under-researched so far, we discussed already in the introduction how the notion is by no means new. Chapter “Paradata in Surveys” (Schenk and Reuß) and chapter “A Leap of Faith: Revisiting Paradata in 3D Scholarship” (Papadopoulos)

elucidate the history of the paradata term in two disciplines where it has the longest history of use: survey research and 3D heritage visualisations. While it is impossible to pinpoint the exact reason why paradata emerged first in survey research and sometime later in heritage contexts, it is reasonable to state that it has been useful in relation to particular kinds of documentation needs evident in these two contexts. Survey scholars have needed to account for the processes going into a certain survey dataset, and heritage scholars making 3D reconstructions have needed to communicate the technical details and decisions going into creating a 3D model. The chapters of this volume show that related needs can be identified also in other contexts, although similarly to how both the understanding of what paradata are and how they are supposed to be acted upon differs between survey research and heritage studies, the desideratum relating to paradata and its practical role differs between particular contexts.

Some of the chapters note that the term paradata has been established in the disciplinary discourse of their respective domains. This applies to the above-mentioned pioneering fields of survey research and heritage visualisation but increasingly also to research data management and information research. In others, paradata as a concept is not a part of the conceptual apparatus of the discipline or has only recently been introduced. For example, Enqvist remarks forthright in chapter “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes” that in the legal domain, the paradata concept does not exist. When paradata is not a part of the formal requirements and vocabulary of particular fields, it is best described as a complementary lens to understanding specific types of information or data and their function (e.g., chapters “Making Research Code Useful Paradata”, “The Role of Paradata in Algorithmic Accountability”, “Adding Paradata About Records Processes via Information Control Plans”, and “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes”). In those cases, introducing paradata as a conceptual tool can be a way to delineate particular forms of knowledge that has not yet been systematically recorded (as for process information about computational code, AI algorithms and automated decision-making).

Another indication of the varying status and genealogy of paradata are the different origin stories of paradata outlined in the different chapters, most explicitly in chapters “Paradata in Surveys”, “A Leap of Faith: Revisiting Paradata in 3D Scholarship”, and “Mapping Accessions to Repositories Data: A Case Study in Paradata”, and how they relate to earlier conceptual overviews in the literature (e.g., Denard, 2012; Edwards et al., 2017; Huvila, 2022; Cameron et al., 2023). While these accounts acknowledge cross-disciplinary influences to a varying degree, a common trait in the chapters in this volume and the earlier literature alike is that the common ground is still fairly weak and the concept is made meaningful through domain-specific appropriations rather than interdisciplinary consensus on the nature of the concept. Independent of the ultimate necessity or desirability of such a consensus, one of the reasons for the heterogeneity is undoubtedly the fact pointed out by Dawson and Reilly in chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach” that paradata have been theorised fairly

little so far. This has led to a contrast between distinctly practical and theoretical takes on paradata—for instance in how paradata are discussed largely as a hands-on matter of how to document development of datasets in survey research versus how it is approached by Dawson and Reilly as a manifestly theoretical concept. A parallel distinction can be sensed between a strive for general and definitive resource descriptions, for instance, in research data, archives and records management, and paradata created to situated data reuse needs in heritage visualisation and computational research.

Considering the aims of this volume to illustrate and elicit diversity, it is perhaps unsurprising that paradata appears as a rather amoeba-like concept both in theoretical and practical sense. Paradata, as data, is clearly a plural rather than a singular entity both literally and in what it can be and do in different settings. This does not mean, however, that the different threads do not entail related elements and starting points to a common exploration of what paradata as a whole can imply for information and knowledge management.

3 Doing Paradata

In the chapters much of the described documenting of processes is really about document*ing* rather than documents and repurposing information for new insights rather than generating entirely new data. Therefore, to untangle the knot of how it makes sense as a management concept it is appropriate to start by looking into how paradata are generated and how they come into being in different contexts. Similarly to how paradata are conceptualised in the texts and what is referred to as documentation, the chapters unfold a diverse array of methods on how paradata are or could be achieved in different contexts. Many of the methods overlap. Data that eventually can be useful as paradata may be a by-product of an administrative or a scholarly practice but before the data can function as paradata, they sometimes need to be extracted or (re)purposed as such.

In some of the chapters, paradata are *intentionally generated* through using specific methods for process description. Dillen (chapter “Paradata for Digitization Processes and Digital Scholarly Editions”) and Schenk and Reuß (chapter “Paradata in Surveys”) discuss several explicit approaches of paradata generation from taking notes and collecting ratings to providing a description of editorial principles in a digital scholarly edition. A common observation in chapters and contexts throughout this volume is that while purposeful generation of meaningful and useful paradata is often necessary, it is both difficult and resource-intensive to generate. As a partial remedy, multiple chapters, including those of Dillen (chapter “Paradata for Digitization Processes and Digital Scholarly Editions”), Rayburn and Thomer (chapter “Reconstructing Provenance in Long-Lived Data Systems: The Challenge of Paradata Capture in Memory Institution Collection Databases”), Jones and Bunn (chapter “Mapping Accessions to Repositories Data: A Case Study in Paradata”), and Cohen and colleagues (chapter “Paradata in Emergency Services

Communications Systems”) describe how paradata can be *extracted*—or perhaps, more correctly constructed—*post hoc* of the available data and documentation.

In other cases, paradata can *result* from data being collected using particular tools and methods. As for example, the chapters of Schenk and Reuß (chapter “Paradata in Surveys”), Dawson and Reilly (chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach”), Papadopoulos (chapter “A Leap of Faith: Revisiting Paradata in 3D Scholarship”), and Dillen (chapter “Paradata for Digitization Processes and Digital Scholarly Editions”) evince, datasets, images, and visualisations all tend to contain plentiful traces of how they were made. Paradata can also be a *by-product* of a process or practice. This is apparent not least in the work of Bilderbeek (chapter “Making Research Code Useful Paradata”) where computer code is to a certain extent both practice and documentation. Jones and Bunn (chapter “Mapping Accessions to Repositories Data: A Case Study in Paradata”) show further how paradata can be a by-product also in the sense that the original purpose of generating such data was completely different. Packalén and Henttonen’s (chapter “Adding Paradata About Records Processes via Information Control Plans”) information control plans are another example of how paradata can be a by-product of process planning rather than an independent task of its own. Finally, with a conceptually different approach to paradata, also Buchanan and Huntsman (chapter “Dustings of Paradata as Pedagogical Support at Four Archaeological Field-School Sites”) describe by referring to “dustings” something that can be linked to extracting paradata.

The different mechanisms explored throughout the chapters of how paradata either potentially or actually can come into being have obvious affinities to previous categorisations. The earlier literature distinguishes between *ex ante* and *post hoc* (forensic) means of acquiring paradata as well as automatic and manual methods (Huvila, 2022). In fields where paradata—or process documentation in general—are recognised concepts and practices, there are established arrays of specific methods to work with them. In this volume, Schenk and Reuß (chapter “Paradata in Surveys”) provide a long list of typical techniques for capturing paradata in survey research including capturing time stamps, keeping call records, location and device paradata, tracking inputs and collecting ratings and observations. Also, here it is possible to see a distinction between purposive paradata generation and collecting by-products of especially digitally administered surveys. In heritage visualisation, where paradata has been discussed since the turn of the millennium, the array of preferred methods is still unfolding as Papadopoulos notes in chapter “A Leap of Faith: Revisiting Paradata in 3D Scholarship” and Dawson and Reilly in Chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach”. Disciplines where paradata and process documentation do not have a self-evident role, it is unsurprising that also the methodological apparatus remains less systematic and developed. The flipside of lack of established standards for paradata documentation is, as for example illustrated Rayburn and Thomer in Chapter “Reconstructing Provenance in Long-Lived Data Systems: The Challenge of Paradata Capture in Memory Institution Collection Databases”, the leeway to experiment with several different ways of documenting and visualising data processes.

The present volume contains several examples of the split between *ex ante* and *post hoc* methods. Trace and Hodges (chapter “The Role of Paradata in Algorithmic Accountability”) distinguish, on the one hand, different records created before, during, and after an AI system is deployed, and forms of paradata such as Explainability Fact Sheets and Data Statements to inform, among others, *ex ante* their designers and *post hoc* their users. Rayburn and Thomer (chapter “Reconstructing Provenance in Long-Lived Data Systems: The Challenge of Paradata Capture in Memory Institution Collection Databases”) show how careful *post hoc* analysis can provide a lot of insights in how a database was created and how it has been used. On the other hand, Packalén and Henttonen (chapter “Adding Paradata About Records Processes via Information Control Plans”) show how an *ex ante* planning and documentation of processes results in actionable paradata. Besides technical differences, they differ in a fundamental manner in what type of paradata they generate. Narrative descriptions of past activities result in different kind of paradata than a plan of action, or real-time timestamps and coordinates collected in the heat of action.

The chapters contain similarly a rich array of examples of the automatic and manual methods and approaches that are probably best described as hybrids. The examples of documents in the list of Trace and Hodges (chapter “The Role of Paradata in Algorithmic Accountability”) on documents that provide paradata for understandability and explainability demonstrate how a specific form of document can function as paradata both to describe previous undertakings and inform future actions. Similarly, even if plans—discussed both by Trace and Hodges (chapter “The Role of Paradata in Algorithmic Accountability”) and Packalén and Henttonen (chapter “Adding Paradata About Records Processes via Information Control Plans”)—constitute at face value a source of an *ex ante* form of paradata, depending on when they are finalised, they unfold as descriptions of how processes were planned to be. Buchanan and Huntsman’s (chapter “Dustings of Paradata as Pedagogical Support at Four Archaeological Field-School Sites”) concept of dustings as narrations of paradata provides a parallel perspective to the temporality and hybridity of paradata by underlining the different timelines of when practices and paradata take place independently of whether it is documented before, during, or after—or as a hybrid—before, during, *and* after the data happen.

Again similarly to how specific approaches operate on different levels of temporality, the different types of methods generate different types of paradata. In parallel, it is also evident how the means of making paradata happen link to diverse forms of information work discussed in the chapters. Where automatic paradata generation is often a straightforward process of collecting and ingesting paradata, it is evident in manual and hybrid processes how they are not only about stockpiling information but involve, incorporate, and prompt reflection and co-shape the data-making. Both Dawson and Reilly’s (chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach”) and Buchanan and Huntsman’s (chapter “Dustings of Paradata as Pedagogical Support at Four Archaeological Field-School Sites”) chapters show how this co-shaping can benefit from multidisciplinary perspectives and involvement in the process. At the same time, it is evident from the

chapters how involving an artist or a data archivist leads to very different types of observations and paradata. It differs by the techniques of how paradata are created, what types of artefacts are considered paradata, and on a fundamental epistemic level, how paradata come into being.

Finally, as with definitions of paradata, specificity undeniably makes both paradata and the practices of making paradata easier recognisable. Having a named and known technique for collecting paradata like with Data Statements (Trace and Hodges, chapter “The Role of Paradata in Algorithmic Accountability”) or EER diagrams (Rayburn and Thomer, chapter “Reconstructing Provenance in Long-Lived Data Systems: The Challenge of Paradata Capture in Memory Institution Collection Databases”) contributes to the clarity of what is supposed to be collected and achieved. With some caution, also the digitality of generated paradata can help to contribute to the transparency of paradata generation and the findability of resulting paradata. A possible downside is how they might narrow down the understanding of what might count as paradata much similarly to how paradata have so far often been neglected in formal documentation. Less specific approaches such as encouraging individuals to narrate their doings or exploring artistic methods can sometimes be better in capturing the fluidity of processes and practices but at the same time, increase the variegation of how paradata are done and what the resulting paradata are and are capable of achieving.

An attempt to summarise the variety of approaches to doing paradata in this volume is obviously difficult. A glimpse to the means of how paradata are often literally done in practice—even if it is also sometimes explicitly, for example, collected, generated, or made—shows how understanding what paradata entail in diverse situations and contexts is not only dependent on how it is conceptualised. It is also as much dependent on how it is made and acted upon in practice. We posit that these two viewpoints, or perhaps rather constellations of vistas, are also a key in opening up perspectives to not only actual paradata but also where paradata as a concept and an arrangement of practices places itself in the context of information and knowledge management.

4 Paradata for Information and Knowledge Management

Considering the theorising and practical use cases for paradata explored in this volume, we propose that there is place for paradata in the conceptual apparatus of information and knowledge management. Depending on how it is understood conceptually and operationalised in practice, it can be fitted in the major discourses of the field as a tacit understanding of processes and practices turned to a data-thing that is manageable in a knowledge management system, or used as a concept that stands for the tacit understanding that needs to be managed through a social and socio-technical mesh of people and technologies (cf. Handzic, 2004). Before rushing into conclusions of how and where paradata might be placed in relation to the canon of information and knowledge management terminology, it is useful to step back and consider what the apparent openness and pliability of the

concept might imply for paradata in relation to managing of both information and knowledge.

We readily acknowledge that information termed in this volume as paradata can often be described using neighbouring concepts, especially if the scope of inquiry is limited to a single discipline, like archival science in chapters “Mapping accessions to repositories data: A case study in paradata” and “Adding paradata about records processes via Information Control Plans”, or archaeology in chapters “Reconstructing provenance in long-lived data systems: the challenge of paradata capture in memory institution collection databases” and “Towards Embodied Paradata. A diffractive art/archaeology approach”. Similarly, again in specific disciplinary contexts it is possible to see apparent overlap between what can be conceptualised and treated as paradata and what has been traditionally termed as something else. However, as Dawson and Reilly suggest in chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach” by discussing the notion of peridata, it is possible that even paradata might not be enough to address the complexities of documenting processes and practices in detail.

Even if there should be no rush to abandon earlier conceptualisations or perspectives to process knowledge, information, and data, the chapters in this volume point to a plethora of advantages of working *with* the notion of paradata. Most importantly, it can help to bring forth and make explicit aspects of processes that can be difficult to recognise, frame, and discuss when they are treated as a part of something else whether it would be a general description, context, or the historical origins of the object of interest. As Enqvist (chapter “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes”) remarks, while paradata is not a legal term and thus incapable of providing any formal guidance in the legal domain, it can still serve a pedagogical function. Another example is how Trace and Hodges (chapter “The Role of Paradata in Algorithmic Accountability”) use paradata as a lens to understand how accountability is conceptualised in relation to algorithmic systems, and a third one, how Jones and Bunn (chapter “Mapping Accessions to Repositories Data: A Case Study in Paradata”) have used the notion to inquire into a dataset and repurpose to a new use.

In this edited volume we have intentionally widened the range of disciplines in which the paradata concept is applied and approached, not only as a matter of information and knowledge management for the information and knowledge management field, but for the management of information and knowledge in and across disciplinary contexts far beyond that. Some of the disciplinary contexts feel perhaps more “natural” in this respect as the reflections on the novelty and currently more or less established status of the notion in different disciplines evince. Others required more explicit nudging by us and the chapter authors to open up for testing the term and its usefulness for the purpose of conceptual exploration of the discipline-specific practice. Similarly, in some of the chapters the links to information and knowledge management—in a broad sense including management of data, records, and other informational assets, processes, practices, and doings—are more explicit whereas in others, they might remain rather implicit. However, it is also equally evident that paradata are always to a certain extent

related to management of something independent of the disciplinary setting, like the management of a dataset (chapter “Making Research Code Useful Paradata”), a database (chapter “Reconstructing Provenance in Long-Lived Data Systems: The Challenge of Paradata Capture in Memory Institution Collection Databases”), or digitised texts (chapter “Paradata for Digitization Processes and Digital Scholarly Editions”).

Whether explicitly referring to information and knowledge management or not, the chapters make multiple references to diverse examples of what paradata can make manageable, what can be managed with paradata, and how paradata themselves can be managed. Paradata can be described as a management concept comparable to others that function as a prism in surfacing and aid thinking about process descriptions, highlighting many otherwise invisible or forgotten facets of cognition, interactions, negotiations, intangible, embodied, unconscious, unregarded, or blinded processes across contexts from archaeology (e.g., chapters “Dustings of Paradata as Pedagogical Support at Four Archaeological Field-School Sites” and “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach”) to artificial intelligence (chapters “Mapping Accessions to Repositories Data: A Case Study in Paradata” and “The Role of Paradata in Algorithmic Accountability”). It can similarly be conceptualised as and compared to a boundary object (chapter “Dustings of Paradata as Pedagogical Support at Four Archaeological Field-School Sites”), friction point (chapter “Making Research Code Useful Paradata”) and as throughout the volume, an interface between practices, processes, and their related entities. Perhaps in the least formal sense, paradata can serve as a tool for self-reflection, reflecting upon reliability, validity, reproducibility of processes and practices across contexts.

The chapters that take an empathetically technical rather than theoretical take on paradata in highlighting how it can matter in practice. Chapter “Paradata for Digitization Processes and Digital Scholarly Editions” shows how paradata can be helpful managing library collections when they are being digitised. Like paradata can in conceptual sense make processes thinkable, it can help to make putting processes into words, accessible, analysable, and changeable. Chapters “Paradata in Emergency Services Communications Systems” and “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes” underline how paradata are intimately linked to information and control in the context of legal processes and emergency services, and chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach” how paradata can systematise the understanding of the temporality and ephemerality of processes beyond stating the fact. As, for example, the chapters of Bilderbeek (chapter “Making Research Code Useful Paradata”) and Jones and Bunn (chapter “Mapping Accessions to Repositories Data: A Case Study in Paradata”) evince how paradata can function as a measure against the obsolescence of information and knowledge by providing means to understand it and sometimes, in practice, (re)create it, if necessary, from new premises.

What then makes paradata pertinent for information and knowledge management right now? The ubiquity of large-scale data processing in broad areas of social life has increased the calls for the importance of reproducibility of analyses that underpin decisions and knowledge-making. Both Trace and Hodges (chapter “The Role of Paradata in Algorithmic Accountability”) and Bilderbeek (chapter “Making Research Code Useful Paradata”) highlight the opportunities with paradata in achieving reproducibility.

Reproducibility ties also to the broader issue of the new advent of Artificial Intelligence that has made paradata perhaps more pertinent than ever. Blackboxing has become increasingly apparent in society and linked to an unprecedented array of modes of social action. An advantage of the paradata concept has grown and is growing especially in contexts where blackboxing has been experienced as a problem already before. While much more than mere paradata are needed to address the conundrums of the transparency of algorithms, thinking with paradata can help to work towards a greater transparency of artificial intelligence techniques and algorithms (Trace and Hodges, chapter “The Role of Paradata in Algorithmic Accountability”) and Bilderbeek, chapter “Making Research Code Useful Paradata”; cf. Cameron et al., 2023). In a wider sense, from the perspective of using paradata as a lens to documentation and management of process information and knowledge, it can help to take necessary steps to shed light at least to parts of black boxes without losing the sight of the complexity of what it takes to make a process transparent enough to be intelligible.

The opposite face of the pertinence of paradata in the contemporary information and knowledge landscape is how it can also help to distinguish between the knowledge about processes someone wants to share and withhold. Enqvist’s chapter “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes” draws attention to how legislation affects and reflects the strive to open and close. Outside of the scope of the themes touched upon in this volume, for example patents and lab notebooks provide further examples of settings and techniques of how to regulate transparency with something that could be termed paradata.

While we acknowledge that we have but touched the possible settings and situations where paradata might matter, as we suggested already in the introduction, we posit that in the context of information and knowledge management the paradata as a concept resides firmly at the fringe of codified knowledge and organisational learning. Paradata makes sense both as a lens to make visible the complexity of processes and practices, and the limits to what extent they can be codified, and as a form of documentation that catches the complexity, systematises it, and makes it intelligible across time, as is illustrated, for example, by the paradata for database maintenance in chapter “Reconstructing Provenance in Long-Lived Data Systems: The Challenge of Paradata Capture in Memory Institution Collection Databases”. When successfully implemented, paradata-as-codified-knowledge (or paradata-as-data) unfolds as an asset for pushing forward the ideals of Open Science, friction-free data publishing and sharing, and systematic data governance.

At the same time, a closer look at paradata-as-lens underlines the limits of such attempts and directs attention to the situated nature of processes, practices, and how they can be knowable to anyone not part of them.

While these two standpoints might appear irreconcilable, they do also represent two perspectives to how and to what extent paradata are difficult and possible to achieve. In this sense, paradata is clearly a disruptive concept as it both suggests and resists the idea that data processing information can be codified, captured, and passed along on a time-space continuum. Both perspectives make paradata transformative in their respective manners. Chapter “Mapping Accessions to Repositories Data: A Case Study in Paradata” highlights how paradata-as-data has a power to datafy, or to turn data to data in a particular sense whereas, perhaps especially, the work of Dawson and Reilly in chapter “Towards Embodied Paradata. A Diffractive Art/Archaeology Approach” points to how considering the conceptual premises and implications of paradata points to the opposite. However, as Stephanie Bunn et al. (2022) notes, while it is tempting to juxtapose “the kinds of knowledge we think can be extracted or condensed from a craft process and transferred into a diagram for a novice learner” and the ones impossible, it is necessarily not the most productive approach. The chapters show how in line with the technical perspective to managing knowledge, paradata unfold as a potential tool helping to make tacit knowledge explicit (cf. Polanyi, 2009; Nonaka & Takeuchi, 1995), and from a human perspective, paradata are something to make tacit knowledge, that is fundamentally non-codable, easier to understand, and following the classification of knowns and unknowns of Huggett (2020), to make unknown unknowns at least known unknowns. The chapters themselves feature multiple examples how the two perspectives can be combined in a reflexive dialogue, for example, in digital textual and visual scholarly editions. The bottom line is not paradata itself, but about doing things with data, critical reflection on such doings and their implications, and making them understandable to an extent that is deemed desirable.

5 The Idea of Transparency and Ethics of Paradata

When considering the implications and opportunities of paradata, it is necessary to direct attention to an aspect of paradata that has remained largely implicit throughout much of this volume. This is the extent to which processes and practices are indeed desirable to be made transparent and what ethical concerns paradata arises in par with trying to solve others. While paradata is admittedly a pro-transparency concept and intuitively about (positive) openness, increased trust (e.g., chapters “Dustings of Paradata as Pedagogical Support at Four Archaeological Field-School Sites” and “Mapping Accessions to Repositories Data: A Case Study in Paradata”), and accountability (chapter “Adding Paradata About Records Processes via Information Control Plans”), there is nothing in any form or notion of paradata that makes it automatically virtuous. Even if it would be tempting to rally for what Hess (2005) describes as a technology- or product-oriented movement to promote paradata as a definite means for positive change, there is reason for

caution. There can be both too much and in different terms “wrong” kind of paradata that is inappropriate, or for example, difficult to navigate for their users. Enqvist’s discussion in chapter “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes” raises a legitimate and highly pertinent question of whether transparency is always desirable. It is necessary to hold back and consider for a moment what eventually makes paradata ethical and responsible.

Similarly to how paradata is conceptually multi-faceted and knotty, also its ethical underpinnings and repercussions are diverse. An obvious dividing line is how and to what extent paradata is approached as a lens or a form of data. They can be roughly seen following what Mickel and colleagues (2023) distinguish as two cultures of ethics “one focused on the relationship between science and social justice, and the other focused on clean data and accuracy as an ethical issue”. However, as they continue, also the mundane interactions shape “at a decidedly local level” how those who engage in everyday work “understand and ultimately approach ethical decision-making”. They, for their part, shape everyday practices with paradata that shape paradata itself and what is considered as good practice.

In addition, such issues as the availability and unavailability of paradata have implications to how (para)data is understood, what kind of new (para)data is generated, what it makes manageable and how, and what consequences it has on different groups and individuals. The reasons to create and eventually (not) share data have also implications to paradata, what they imply and to whom. Legal professionals have theirs, archivists theirs, and researchers their multiple context and situation influenced reasons to embrace the concept and particular practices to operationalise it. For researchers, data sharing for the purposes of gaining credit or economic benefits has very particular implications to what paradata end up being and doing with an emphasis on the gain rather than documentation. Altruistic sharing and documentation motivated by helping others by sharing an existing resource has others—perhaps focused on very particular ideas of transparency—similarly to obligatory creating and sharing that might lead to prioritising conformity with guidelines rather than usefulness and responsibility.

There is a plethora of conceivable ethical risks with both poor and too specific paradata. While paradata can counter obsolescence also it can be and become obsolescent and indecipherable. Paradata is easy to use for measurement, assessment, and evaluation of practices and processes beyond what is fair and reasonable. Producing paradata can be used as a token for underlining general commitment to transparency even if the documentation itself would remain of limited quality. Paradata comes similarly with a risk of misleading people by emphasising aspects of processes and data that might not be the most pertinent ones. Paradata come with an imminent risk of surfacing people’s cognition beyond what might appear comfortable and necessary. Similarly to other forms of meta-information, it can indirectly reveal too much about processes and the datasets, information, and knowledge it is supposed to describe. Risks are apparent with, for example, various types of archival records, health information, and knowledge belonging or relating to vulnerable communities.

The studies of algorithmic accountability and legal ramifications of paradata point succinctly to difficulties to know whose interests paradata serve now and in the future. With automated paradata and using paradata for advancing algorithmic accountability, an obvious but admittedly, a convoluted question is what ethical questions are relevant when describing human involvement in processes and to what extent they are also relevant when describing automated or machine supported processes. Further, by increasing transparency of specific aspects of practices and processes, paradata might simultaneously hide and blackbox others either by accident or on purpose. Similarly to misinformation, there can also be non-truthful or even malign motivations to create paradata to promote particular narratives, distort the public image of a certain process or practice, or to “paradata-wash” a truly messy and poorly designed and executed endeavour. However, like many forms of misinformation, paradata-washing does not need to be malicious. Similarly to how Ruokolainen and Widén (2020) remind of misinformation, the dividing line between paradata and “mis-paradata” is not sharp. It is to an equal extent produced in a particular social situation.

However, in spite of all the thinkable caveats, if responsible, paradata have a capacity to do good. It has an apparent capability to realign some of the very fundamental “lines of accountability” (Guston, 1999) of how processes and practices become and remain dependable. By increasing transparency, both as form of codified knowledge and a lens, paradata can increase trust, shared understanding, equitability, and transparency. In attempts to decolonise data, paradata can tell about how they were collected, what wrongdoings were committed that could be repaired but also who was represented and to what degree a dataset actually is and is not biased. While this might entail an in-depth exploration of practices and processes, using paradata as a mere reminder to avoid overly simplistic conclusions can sometimes be enough. Similarly, even if the contemporary ideas of transparency and the aims of the Open movement are interpreted in terms of a close to unrestricted access to information, it is important to consider that there is nothing inherent in paradata suggesting that everything needs to be released or to be directly exploitable by others—whether they are multinationals, governments, individuals, or communities. Archival secrecy, privacy policies, patents, and the provisions in data sharing principles to make data as “open as possible and as closed as necessary” (Wilkinson et al., 2016) are all examples of how paradata can be generated but kept responsibly closed whenever needed.

Paradata ethics and its relation to general information and knowledge management ethics is without any doubt an important line of future research and practical attention. Incontestably, with ethics as well as with keeping up with paradata in general, the most important point of departure is to acknowledge the need of critical reflection. Like transparency is not given with or without paradata, the meritoriousness of paradata or any specific idea or form of transparency is equally little self-evident. Schenk and Reuß (chapter “Paradata in Surveys”) point to the on-going debate on the requirement of informed consent about paradata collection and the limits of their confidentiality. Enqvist (chapter “Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes”) and Trace and Hodges (chap-

ter “The Role of Paradata in Algorithmic Accountability”) stress how extensive transparency easily conflicts with privacy. This is not least the case when legacy data is repurposed for new uses like in Cohen and colleagues’ work of modelling of emergency service communications in chapter “Paradata for Digitization Processes and Digital Scholarly Editions”. Often it is possible reach a trade-off but sometimes it is difficult to draw a line between what paradata are reasonable to keep and what needs to be discarded. In the current volume, survey studies and human subject research provide example of by definition benevolent activity that epitomises much of the intricacy of balancing between process transparency and keeping involved individuals non-identifiable. Comparable and even greater challenges are apparent in many other fields from healthcare to public and private security.

6 Future Perspectives on Paradata

It is obvious that with this volume we have only scratched the surface of the theory, practice, and implications of paradata for information and knowledge management and for the diverse settings where it is applied. The conceptual field is still very open, so is the field of practice. With perhaps the exception of survey research, the notion of paradata is still very much in the making without consolidated theory and practices of producing and exploiting it neither as a theoretical lens nor descriptive resource. However, the chapters throughout this volume show that it is stabilising, especially in fields like heritage visualisation, archivistics and artificial intelligence, systematising implications of the concept and its concrete instantiations. The forms and formats of paradata and the conceptual understanding of what the concept and paradata entail is a challenge that is approached in different disciplines from varying angles. For a legal scholar, paradata can work as a pedagogical concept, for survey researcher it is an established part of the scholarly toolbox, and for many fields of practice and scholarship it comes with a promise of shedding light and emphasis on matters that might have previously been under the radar. There is much work to be done to frame what counts as paradata and where to draw a line between paradata and other concepts. While the usefulness of theoretical uniformity can be debated, there is room for contrasting and comparing different theoretical understandings of paradata, finding synergies and complementarities.

The chapters shed light to different ways of collecting, extracting, creating, and curating paradata. Similarly to the need for a continuing theoretical discussion, there is a methodological discussion to be had. Relevant questions include, for instance, how interviews can be used to generate paradata in comparison with the on-going discussion about ethnography of fieldwork in archaeology. Another key line of future inquiry pertains to expectations and implications. Paradata is a concept that holds a great promise in helping to deliver the desired outcomes envisioned in the Open movement (European Commission, 2016) and the many of the contemporary societal aspirations for transparency, accountability and effective, responsible and equitable sharing and use of information and knowledge.

Finally, before departing to pursue future research, we feel that it is relevant to highlight two aspects of paradata that form a red thread through the texts in this volume. Creating paradata is an integral part of scientific method and in a broader sense, systematic knowledge-making. From the perspective of information and knowledge management, it is similarly an integral part of understanding and knowing what is managed. Another similarly crucial aspect of paradata is that it quite apparently changes the epistemological base for information and knowledge management. Paradata turns attention to what Prusak described as “thoroughly adopted” and invisible knowledge of processes (Prusak, 2001, p. 1006). Rather than being content with managing acontextual dichotomously true or false knowledge-things, paradata direct awareness to managing knowledge and information with history and future. With paradata we need and want to know how knowledge came into being.

References

- Bunn, S., Bruun, M. H., Wahlberg, A., Douglas-Jones, R., Hasse, C., Hoeyer, K., Kristensen, D. B., & Winthereik, B. R. (Eds.) (2022). Technology as skill in handwork and craft: Basketwork and handweaving. In *Palgrave handbook of the anthropology of technology* (pp. 61–83). Palgrave Macmillan.
- Cameron, S., Franks, P., & Hamidzadeh, B. (2023). Positioning paradata: A conceptual frame for AI processual documentation in archives and recordkeeping contexts. *Journal of Computing and Cultural Heritage*, 3594728. <https://doi.org/10.1145/3594728>
- Denard, H. (2012). A new introduction to the London Charter. In H. Denard & D. Baker (Eds.), *Bentkowska-Kafel Anna* (pp. 57–71). Paradata and transparency in virtual heritage. Ashgate.
- Edwards, R., Goodwin, J., O'Connor, H., & Phoenix, A. (2017). *Working with paradata, marginalia and fieldnotes*. Edward Elgar.
- European Commission. (2016). *Open innovation, open science, open to the world | Shaping Europe's digital future*.
- Guston, D. H. (1999). Stabilizing the boundary between US politics and science: The rôle of the office of technology transfer as a boundary organization. *Social Studies of Science*, 29(1), 87–111. <https://doi.org/10.1177/030631299029001004>
- Handzic, M. (2004). *Knowledge management: through the technology glass*. World Scientific.
- Hess, D. J. (2005). Technology- and product-oriented movements: Approximating social movement studies and science and technology studies. *Science, Technology, & Human Values*, 30(4), 515–535. <https://doi.org/10.1177/0162243905276499>
- Huggett, J. (2020). Capturing the silences in digital archaeological knowledge. *Information-an International Interdisciplinary Journal*, 11(5), 278. <https://doi.org/10.3390/info11050278>
- Huvila, I. (2022). Improving the usefulness of research data with better paradata. *Open Information Science*, 6, 28–48. <https://doi.org/10.1515/opis-2022-0129>
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.
- Polanyi, M. (2009). *The tacit dimension (Revised)*. The University of Chicago Press.
- Prusak, L. (2001). Where did knowledge management come from? *IBM Systems Journal*, 40(4), 1002–1007. <https://doi.org/10.1147/sj.404.01002>
- Ruokolainen, H., & Widén, G. (2020). Conceptualising misinformation in the context of asylum seekers. *Information Processing and Management*, 57(3). <https://doi.org/10.1016/j.ipm.2019.102127>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Isto Huvila PhD, is professor in information studies at the Department of ALM at Uppsala University in Sweden. He received an MA degree in cultural history at the University of Turku in 2002 and a PhD degree in information studies at Åbo Akademi University (Turku, Finland) in 2006. Huvila was chairing the recently closed COST Action ARKWORK and is directing the ERC funded research project CAPTURE. His primary areas of research include information and knowledge management, information work, knowledge organisation, documentation, and social and participatory information practices.

Lisa Andersson Ph.D., works as a researcher at the Department of ALM at Uppsala University in Sweden. She received her MA degree in library and information science in 2011, and her doctoral degree in library and information science in 2017, both at the Department of ALM at Uppsala University in Sweden. Her research focuses on data and information management including research data and information management systems, knowledge organisation and data descriptions, data publishing and use. Andersson has published in library and information science journals but also in cross-disciplinary journals from the fields of archaeology and digital humanities.

Olle Sköld Ph.D., works as a Senior Lecturer at the Department of ALM at Uppsala University in Sweden. He received his MA degree in archival studies in 2010 at Lund University, and a doctoral degree in information studies in 2018 at Uppsala University. His research is characterised by a broad interest in the ALM field, knowledge organisation and production, research data creation and use, and digital humanities. Sköld has published in information studies journals including the *Journal of the Association for Information Science and Technology*, *Journal of Documentation* and *Information Research*.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

