8

Future Directions

Making Paradata Matter

Isto Huvila

8.1 Introduction

The starting point for this volume has been to demonstrate that paradata matters. At the same time, its aim has been to engage in a proper discussion on how and when it does so. However, as the previous chapters have demonstrated, the significance of paradata is not given and it can matter in different ways to different communities. In the final chapter of this volume, we are revisiting some of our original assumptions, considering conceptual and practical implications of paradata, and discussing directions for practical work with paradata, as well as future research on paradata, data-related practices and processes.

After a book-length examination, a fundamental question to ask is to what extent the concept of paradata really is meaningful and productive. Whilst not without its limitations, we see multiple theoretical and practical benefits in embracing it. In the earlier literature, paradata has often been portrayed as a new complementary data type and an artefact that solves the problems of understanding data-related decisions, practices, processes and their underpinnings. By contrast, our perspective on paradata has been less definite. Paradata is and remains a wicked problem in itself rather than an easy solution. It is not enough to acknowledge in passing that it is necessary to 'add paradata' to make data intelligible. Paradata deserves to be taken seriously.

In both theory and practice, we find that the concept of paradata and engaging with its applications is helpful in how it directs attention to practices and processes rather than attributes of data. As outlined in the first two chapters of this volume, engaging with paradata can benefit working with data in many ways, including:

211

- understandability
- accessibility
- interoperability
- trustworthiness
- reusability and reuse of data
- keeping track of its ownership
- improving the reproducibility of research

It can also help to open the black box of domain-specific practices and processes to transdisciplinary (Huvila, 2022) and non-specialist audiences (cf. Eichner et al., 2024) that are often excluded from the intricate working knowledge embedded in human practices.

At the same time, however, it also makes it increasingly evident that none of the anticipated outcomes of paradata are simple and straightforward to achieve, or uncontroversial in practice. Nothing is universally FAIR (Wilkinson et al., 2016) or MEAN (Huvila, 2017) and paradata cannot make them that. It can, however, function as a key ingredient in contributing to these goals while at the same time reminding and exposing in detail why, how, and to what extent they have been accomplished or remain out of reach.

8.2 Instrument of Knowledge Organisation

After stressing the limits and complexity of paradata and the need to take it seriously, we must ask how exactly it can take us closer to reaching the many expectations bestowed upon it. Even if the simplistic idea of paradata as a new form of auxiliary data is best abandoned, paradata is inherently a concept that belongs to the domain of knowledge organisation.

In a theoretical sense, rather than being a quick remedy, paradata can be helpful in directing attention to the diversity of means of how to inform and be informed about how data is created, managed and used, along with the underpinnings of these endeavours. As a concept, paradata has enough leeway to be complementary rather than an overlapping or a redundant notion. Compared to metadata, as conceptualised here, the object and objective of documentation are different with paradata. In a broad sense, both metadata and paradata are 'potentially informative' (Pomerantz, 2015, p. 26) and express 'ideas, feelings, emotions, and values' (Carbajal, 2021, p. 102). However, whereas metadata is typically, if not exclusively, conceptualised as being informative of resources, data or objects, the kernel of paradata is elsewhere. Its focal point of paradata is on shedding light on practices, processes and their underpinnings.

Paradata deserves to be saved from what Huggett criticises as a typical focus on technical aspect and 'digital background to digital data' (Huggett, 2022) rather than embracing it as a much broader and inclusive concept covering the whole entirety of practice and process information. Paradata engages primarily with other types of resources that are searched for and used in different ways. While metadata is conventionally, if not necessarily, expressed in words (cf. Carbajal, 2021), the brief overview of the diversity of expressions of paradata in Chapter 3 and the approaches to engage with paradata in Chapters 4–6 demonstrate how paradata is expressed and accessed in different terms than metadata. Even if a part of paradata might be structured and available for searching and retrieving like formal metadata and inscribed as metadata, much of what we conceive as paradata roams far beyond the scope of what is conventionally understood as metadata.

Rather than necessarily being created for the purpose, paradata is often discovered in existing material even if it is subsequently modified to fit its new purpose. While even informal metadata often retains a degree of formality, paradata is formal only occasionally and partially. Leveraging and combining the plethora of methods people use to find, access, use, make and manage paradata also presents a very different type of knowledge organisation challenge than those associated with using metadata. Their outcomes have differences, too. While both paradata and metadata are (potentially) informative, paradata is also very much performative, not only in theory but also in practice, in how it is enacted.

This multiplicity of paradata is critical to consider when it is put to work as an instrument of organising knowledge. Rather than assuming paradata to be a simple technique to organise knowledge, it warrants being approached as a form of *critical practice*. From this perspective (cf. Agre, 1997; Van Geenen et al., 2024), critique is not about disapproval or pessimism but about thinking and moving forward to make the most out of paradata both in theory and in practice.

While it can be tempting to suggest documenting everything, it is not feasible and most of the time it is impossible, as the examples in previous chapters have underlined. Trying to inscribe and collect too much can be highly detrimental in how it easily constructs a new facade very different from how the practices and processes unfolded in the first place. It both conceals and increases complexity, and takes time from other tasks. However, there are also sometimes small pieces of information that can truly make a difference and need to be inscribed to avoid losing them. Identifying them can be as difficult as determining where the fine line between enough and too much should be drawn. They both require a thorough understanding of the complex meshwork

of paradata as a whole and how this meshwork links to the complex ecology of practices and processes it documents.

The flip-side of seeking to avoid increasing the complexity of documentation is that its opposite can be equally detrimental. Conscious and unconscious data cleaning through harmonisation, standardisation and selective preservation can be controversial. While it might make data more manageable, help to save resources (cf. Pasquetto et al., 2017), and cast an aura of reliability, professionalism and quality, it runs a major risk of depriving it of much of its richness. A better approach is to make the complexity of data, practices and processes visible and to provide strategies and tools for dealing with it.

With paradata, the most critical knowledge organisation task is not to document everything or to provide meticulous instructions on how to reconstruct practices in the future. Rather, it is to provide an overview of all available traces and ingredients to help to keep track of their intersections. The goal should not be to attempt to provide direct shortcuts and minimise paradata users' need to think but to provide them with maps and navigational aids. A crucial decision is to choose methods and approaches that are likely to be most helpful and possible to implement.

8.3 Mindset

The limits to which paradata can be operationalised in knowledge organisation point to another parallel perspective to paradata. We are inclined to see paradata as a lens or mindset that makes it possible to expand and refine our understanding of what can and needs to be known about practices and processes. When framed as a mindset, rather than being a question of finding an exact definition of what paradata is, paradata turns to a question of what data, or in a broader sense, things, can be appropriated to function as paradata.

Instead of claiming that there is too little paradata, thinking of it in terms of mindset directs attention to the question of whether there is indeed enough relevant information even if we might fail to make full use of it. The crisis of a lack of paradata turns to a 'crisis of definition' (cf. Escobar, 1999) of what is capable of functioning as paradata.

As a mindset, it reminds us of how many practices, processes and their underpinnings are never documented explicitly. There is always more to know about data and data-related practices than is ever inscribed in the formal record or embodied in traces of data practices. This evident fact does not mean,

however, that having paradata would not matter, or that collecting it would be worthless. Rather it suggests that it is critical to reflect what such incompleteness means in practice and how it should be taken into account when pursuing a better understanding of practices and processes.

In this respect paradata is akin to archaeology, which also unfolds as an illustrative domain and has been remarkably successful in eliciting knowledge about past human practices on the basis of fragmented and incomplete (para) data. Archaeologists have developed methods to stitch together evidence, identifying marks of use in physical artefacts, using analogies, and experimentation to understand past practices and processes.

Fundamentally, thinking of paradata in terms of a mindset is a matter of developing a methods discourse for data reuse and practice in order to process knowledge in different research contexts. This is crucial everywhere but especially in domains where such discussion has so far remained unarticulated. Knowing how to discuss such matters entails a particular set of literacies and competencies, in a much broader sense than how data literacy is portrayed in a part of the literature (Koltay, 2015) as a straightforward skillset of being able to work with and understand a thing called data.

Rather it requires a comprehensive insight into the interrelation of data creation and reuse (Kansa and Kansa, 2021): a helix Mathieu and Pruulmann-Vengerfeldt (2020) describe in communication research as a data loop. It goes beyond the ability to collect and analyse data to master its effects, opportunities and constraints across domains and time. It incorporates both encoding and decoding of data and acknowledging that data users are also encoders (cf. Livingstone, 2019; Mathieu, 2023).

As a mindset paradata does also remind everyone engaged in working with data of the importance, to be *'involved* in the capturing, processing, and linking of any data they plan to use for their work' (Christen and Schnell, 2024, p. 7) (added emphasis). As discussed in Chapter 6, data literacies that incorporate a paradata mindset turn, like information literacies (Hicks et al., 2023; Lloyd, 2010; Tuominen et al., 2005), into meshworks of interlinked and overlapping practices of being with data.

Further, as suggested in Chapter 3, there are thresholds both in relation to how and when paradata is technically and epistemically useful and for whom. Such thresholds are built into data literacies but are also something data literacies can help to overcome. We have touched upon multiple aspects of such literacies in this volume from what types of things it applies to and where they can be found in Chapter 3, to methods, approaches and competences to master, to generate, identify and curate in Chapters 4–6.

8.4 Limits and Ethics of Paradata

Acknowledging that paradata materialises as a distinct form of mindset, literacy and practice obliges us to turn attention to where paradata ends and what might be its associated risks. We have already recognised that rather than assuming that paradata is a universal remedy, straightforward to achieve, and uncontroversial in practice, paradata should be approached as a form of critical practice.

Implementing paradata through standards and blanket expectation to produce comprehensive documentation leads easily to what Stengers (2018) describes, aptly considering the acronym of our project, in terms of a *capture*. Capture is for Stengers a situation in which the dominant institutions of research, or more broadly in case of paradata data work, lead to mediocrity and a sincere belief that a particular representation is genuinely consistent with a particular practice or process. A capture of practices and processes in the spirit of an 'ideology of information management' (Kamin, 2023, p. 191), substituting them with descriptions comes with a risk of an increasing sense of estrangement from the practices and processes themselves.

However, working with paradata does not need to follow dataism in its 'belief in the objective quantification and potential tracking of all kinds of human behaviour and sociality' (Van Dijck, 2014, p. 198). Embracing paradata as a form of critical practice and critique of itself, renders it capable of facilitating the opposite: to question and problematise itself and its premises, to learn and change. Doing so is hardly effortless and not a problem solvable with technology even at the time when the lightning fast development of artificial intelligence techniques has again put many old assumptions of the capabilities of technical systems to a question. At best, paradata is a partial solution. It is vital to be realistic with one's expectations of what and how it can be a useful tool. Otherwise it easily raises hopes of being capable of doing more than it can deliver.

A parallel question relating to the limits of paradata pertains to its social consequences. Just as transparency and openness are difficult to argue against in general terms, paradata also becomes easily shrouded in a veil of hard-to-criticise consensus. While much of the documenting and preservation of information on data-related practices and processes is beneficial, unveiling individuals and their practices can also have adverse effects. They can be used to harm people even beyond those individuals and groups who have participated in data practices and consented to be involved, documented and preserved. This applies both to living beings and non-living things and for example, in case of cultural and natural resources like archaeological sites or nature reserves, their combinations, all deserving protection and care.

In research with human subjects, paradata can expose study participants, individuals related to them and their activities, data collectors, and researchers in ways that can be difficult to anticipate. It does not require a sudden takeover of an oppressive regime to make paradata dangerous. Even something as seemingly uncontroversial as a new paradigm of measuring the effectiveness of data work can turn paradata into an instrument of exploitation and unjust treatment. Moreover, even if paradata would be harmless alone, when brought together and combined with other available information, the meshwork of paradata can have unintended consequences both at the present and in the future.

The fact that paradata can also cause damage should be taken as seriously as its possible benefits. Leonelli and Williamson (2023) advocate for data linkage that covers technical and legal infrastructures, guidelines and mechanisms for follow-up, and an open mindset for different perspectives to what constitutes good, bad and acceptable. There is critical need for a comparable 'responsible practice' in paradata.

Just as paradata as a whole is a complicated matter, its consequences are too complicated to anticipate and consequently, to be dismissed at the outset. While releasing paradata should not be restricted beyond reason, it is important to acknowledge that making and keeping paradata is a matter of trust and responsibility. Finding parallels to paradata, what it can achieve and where its limitations are is demanding. It can perhaps be compared to a certain extent to patents that ideally provide a mechanism to open information while retaining a necessary level of control of its particularly pertinent aspects and implications. As with the requirement of documenting practices and processes, the mechanisms to disclose them should be weighted against both imaginable and unimaginable risks and provide necessary protections to plausible social, political, epistemic and economic concerns in the particular contexts and situations. Only in carefully considered cases should they be made a non-negotiable requirement.

8.5 Conclusions

In the closing of this volume, it is evident that there are many loose threads to follow in the future. The question of (para)data literacy is only one of them. This book and our work in the CAPTURE project and the parallel work of colleagues elsewhere has obviously only started to provide insights into paradata and the broader questions of understanding and documenting, preserving and utilising information on data-related practices and processes.

Even if paradata is situated and deeply contextual, our aim with this volume has been to focus on issues that have resonance in multiple domains. However, as much of the empirical work conducted on paradata in the context of the CAPTURE project focused on archaeology, it is undoubtedly overrepresented here and in our conclusions. Therefore, even if we still believe that archaeology in its diversity and cross-disciplinarity provided us a useful starting point to inquire into paradata, we anticipate that future studies of paradata in specific domains will produce new knowledge that can further nuance the understanding of paradata and provide actionable insights in those contexts. Rather than helping to develop a blanket solution, archaeology has probably worked best as a healthy reminder of how such panaceas do not exist. Not only every field of research and practice but also diverse varieties of data making, management and use as particular types of undertakings have their own often contradicting priorities. They all deserve to be taken seriously from their very own premises.

In this volume we have shown how it can function both as a referent to a particular category of things that can be appropriated as informative about practices and processes but also as a mindset to aid thinking about practices, processes, their underpinnings and implications. For the time being, paradata is unquestionably a factish (Latour, 2011; Stengers, 2018), a preliminary term used to refer to a phenomenon but also perhaps a figuration (Braidotti, 2011; Haraway, 1992) in how it materialises an arrangement of ideas about practices and processes.

We are not sure if it is going to stick, if it will or should be replaced by something else, and (or) how it will evolve in the future. What we do think, however, is that the question of how to understand and document processes and practices of making, processing and using cultural artefacts is crucial to understanding how they are knitted and are knitting themselves into the social fabric. For doing so, we need appropriate concepts but also a lot of empirical work.

After investigating paradata with data creators and users, we have also observed that data management practices and data repositories are sites with a crucial impact to paradata. An earlier comprehensive body of work has developed means to document and preserve information on processes and transformations in the curatorial context, including curatorial provenance and management metadata. However, what remains less explored is how curatorial work and data governance and its underpinning political and normative ideals affects paradata in practice. The same applies to wider contexts of paradata in relation to how exactly it works in the intricate meshwork of social reality.

References

Agre Philip (1997). Toward a critical technical practice: Lessons learned in trying to reform AI. In Bowker Geoffrey et al. (eds.), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Mahwah, NJ: Lawrence Erlbaum, 131–158.

Braidotti Rosi (2011). *Nomadic Subjects: Embodiment and Sexual Difference in Contemporary Feminist Theory,* 2nd ed. Columbia University Press.

Carbajal Itza A. (2021). Historical metadata debt: Confronting colonial and racist legacies through a post-custodial metadata praxis. *Across the Disciplines* 18(1-2), 91–107.

Christen Peter and Schnell Rainer (2024). When data science goes wrong: How misconceptions about data capture and processing causes wrong conclusions. Harvard Data Science Review 6(1).

Eichner Katrina C. L., Campbell Renae J. and Warner Mark S. (2024). Archaeological collections and the public: It isn't all about us. *Advances in Archaeological Practice* 12(1), 43–52.

Escobar Arturo (1999). After nature: Steps to an antiessentialist political ecology. *Current Anthropology* 40(1), 1–30.

Haraway Donna (1992). Ecce homo, ain't (ar'n't) I a woman, and inappropriate/d others: The human in a post-humanist landscape. In Butler J. and Wallach J (eds.), *Feminists Theorize the Political*. Routledge. 86–100.

Hicks Alison et al. (2023). Leveraging information literacy: Mapping the conceptual influence and appropriation of information literacy in other disciplinary landscapes. *Journal of Librarianship and Information Science* 55(3), 548–566.

Huggett Jeremy (2022). Digital Tools or Knowledge Devices? *Introspective Digital Archaeology.* https://introspectivedigitalarchaeology.com/2022/09/29/digital-tools-or-knowledge-devices/.

Huvila Isto (2017). *Being FAIR When Archaeological Information Is MEAN: Miscellaneous, Exceptional, Arbitrary, Nonconformist. In Presentation at the Centre for Digital Heritage Conference 2017,* Leiden June 14–16, 2017.

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

Kamin D. (2023). *Picture-Work: How Libraries, Museums, and Stock Agencies Launched a New Image Economy*. The MIT Press.

Kansa Eric and Kansa Sarah Whitcher (2021). Digital data and data literacy in archaeology now and in the new decade. *Advances in Archaeological Practice* 9(1), 81–85.

Koltay Tibor (2015). Data literacy: In search of a name and identity. *Journal of Documentation* 71(2), 401–415.

Latour Bruno (2011). *On the Modern Cult of the Factish Gods*. Durham, NC: Duke University Press.

Leonelli Sabina and Williamson Hugh F. (2023). Introduction: Towards responsible plant data linkage. In Williamson Hugh F. and Leonelli Sabina (eds.), *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*. Cham: Springer International Publishing, 1–24.

Livingstone Sonia (2019). Audiences in an age of datafication: Critical questions for media research. *Television & New Media* 20(2), 170–183.

Lloyd Annemaree (2010). Framing information literacy as information practice: site ontology and practice theory. *Journal of Documentation* 66(2), 245–258.

Mathieu David (2023). Deconstructing the notion of algorithmic control over datapublics. In Møller Hartley Jannie, Sørensen Jannick Kirk and Mathieu David (eds.), *DataPublics: The Construction of Publics in Datafied Democracies*. Bristol: Bristol University Press, 27–48.

Mathieu David and Pruulmann-Vengerfeldt Pille (2020). The data loop : How audiences and media actors make datafication work. *MedieKultur* 69, 116–138.

Pasquetto Irene V., Randles Bernadette M. and Borgman Christine L. (2017). On the reuse of scientific data. *Data Science Journal* 16(8), 1–9.

Pomerantz Jeffrey (2015). *Metadata*. Cambridge, MA: MIT Press.

Stengers Isabelle (2018). *Another Science Is Possible: A Manifesto for Slow Science*. Cambridge: Polity.

Tuominen K., Savolainen R. and Talja S. (2005). Information literacy as a sociotechnical practice. *The Library Quarterly* 75(3), 329–345.

Van Dijck Jose (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2), 197–208.

Van Geenen Daniela, Van Es Karin and Gray Jonathan W. Y. (2024). Pluralising critical technical practice. *Convergence* 30(1), 7–28.

Wilkinson, Mark D. et al. (Mar. 2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, 160018.