EMMA JOHANSEN[1]

Research Data Management: A Disambiguation

Introduction

> "Data! Data! Data!" he cried impatiently. "I cannot make bricks without clay." (Doyle, 2003).

Traditionally speaking, the management of research data has not been an especially well-funded or supported endeavour. Nonetheless, it has often been expected to be comprehended and carried out by researchers without much instruction (Kennan, 2017). However, since an increased amount of data is produced across all disciplines in recent decades, interest in practices *surrounding* data has grown (Ray, 2013). After all, it only seems logical that large quantities of data would require a great deal of management in order to be, well, manageable. Palsdottir (2020), Ray (2013) and Kennan (2017) all touch on the importance of research data management (RDM) in order to improve access to data, the shareability and reusability of data, and, ultimately, the longevity of data. It is for this very reason that it is becoming increasingly common for funders to require data managements plans (DMPs) from researchers, which in turn has led to academic

---

1 Emma Johansen is a second-year student at the Master's Programme in ALM specializing in Library and Information Science at Uppsala University. She is also pursuing a master's degree in English Literature at Dalarna University. Email: emma.johansen.4138@student.uu.se

libraries offering more services related to data. The aim of this short paper is to examine some of the ways in which RDM has been defined in literature and move towards a disambiguation for the reader.

Disambiguating RDM

In order to break RDM down into the sum of its parts, it seems fruitful to work from the middle out, beginning with 'data', a term which has proven challenging to define in literature on the subject. The National Science Board (2005, as cited in Koltay, 2017) defines data as "any information that can be stored in digital form, including text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations etc." (p. 4). However, this definition of data may prove limiting as there is a focus on the digital aspect of data as well as a reliance on the term 'information'. Dalkir (2005, p. 7) makes a distinction between 'data' and 'information', claiming that information is content that "represents analyzed data" whereas data is content that is "directly observable or verifiable." With this definition in mind, it is easy to understand why data is so important for research, as this understanding places data in the position of verifiable evidence and as something necessary in order to create information.

Schöpfel, Prost and Rebouillat (2017) remark that there is similarly little consensus when it comes to defining 'research data'. Rice (2009, as cited in Kennan, 2017) defines research data as data "collected, observed or created for the purposes of analysing to produce original research results" (p. 506). This is a fairly straight-forward definition of the purposes of research data although Schöpfel, Prost and Rebouillat (2017) highlight another important aspect of research data, namely that it is necessary in order to validate research findings. However, Gregory et al. (2020) note that not all data needed for research is what is traditionally thought of as research data and that, for example, metadata, texts, and social media posts can also be used during the research process. Moreover, Koltay (2017) brings information back into play as he suggests that the increased interest in contextualized data rather than raw data removes some of the differences between research data and information as the data has, to some extent, been analyzed by the creator. This is to emphasize that data, and research data, can take many

different forms, and that there may be upsides to keeping some ambiguity and fluidity in the terms.

This brings us back to the full phrase—research data *management*—and why the management of research data has become a key issue. Kennan (2017) notes that it is only recently that data has been shared with people beyond the initial research team. According to Ray (2013), trust must originate from how data is managed and documented if those without affiliations with the original researcher or research team are going to be able to confidently utilize the data in their own research. If data is indeed content that is "directly observable or verifiable" as Dalkir (2005) claims, then proper RDM ensures that data can be verified by outsiders. Another aspect of data management has to do with discovery. Koltay writes that: "Discovery includes the knowledge of how to select data and needs to be supplemented by the ability to synthesize data and combine it with other information sources and prior knowledge. All these abilities need to be based on the skills of identifying the context where data is produced and reused" (2017, p. 11). RDM helps to contextualize the data, however, as Koltay notes, a degree of data literacy is also required. Another aspect of RDM that is not to be overlooked is the management of sensitive data, such as data containing personal information. It is of course vital that this type of data is managed with established security measures in place (Cox & Pinfield, 2014). Ray (2013) states that some of the challenges surrounding the implementation of good RDM practices regard lack of time, effort, and resources. Additionally, she notes that there is a general focus on reaching the requirements put forth by funders rather than trying to achieve the best data management practices. The result of this can be a subpar product and a lack of knowledge about the benefit of these practices.

Conclusion

The terms 'data' and 'research data' have both proven quite difficult to define in the literature. However, it should also be acknowledged that the terms may be malleable by design as data can be any number of things. Expanding and exemplifying what we mean when we discuss research data may in fact help more researchers manage their data in responsible ways. As has been touched on previously,

research data has become an increasingly relevant topic in academic libraries and as such it is worthwhile to delve further into topics such as RDM. This note shows that a solid definition is perhaps not feasible, but sheds light on different strands of thought about research data management, central to all scientific practice.

References

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316.

Dalkir, K. (2005). Introduction to Knowledge Management in Theory and Practice. In: *Knowledge Management in Theory and Practice*, Dalkir K. (Ed.). Oxford: Elsevier, pp. 1–22.

Doyle, A. C. (2003). The Adventure of the Copper Beeches. In *Sherlock Holmes: The Complete Novels and Stories Volume I*. New York: Bantam Classics, pp. 492–518.

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

Kennan, M. A. (2017). Managing Research Data. In: *Research Methods: Information, Systems, and Contexts*, Williamson, K & Johanson, G (Eds.). Cambridge: Chandos Publishing, pp. 505–515.

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science,* 49(1), 3–14.

Palsdottir, A. (2020). Data literacy and management of research data – a prerequisite for the sharing of research data. *Aslib Journal of Information Management,* 73(2), 322–341.

Ray, J. M. (2013). *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, IN: Purdue University Press.

Schöpfel, J., Prost, H., & Rebouillat, V. (2017). Research Data in Current Research Information Systems. *Procedia Computer Science*, 106, 305–320.