Research Data Management

R

Tibor Koltay
Eszterházy Károly Catholic University, Hungary

INTRODUCTION

Desk research, followed by critical analysis was conducted. The interdisciplinary literature, identified Research Data Management (RDM) as a widely published subject, portrayed by a considerable number of papers that reflect on a wide variety of issues. RDM is related to varied subjects, described in this encyclopedia, as there is scholarly interest for RDM in the natural sciences, social sciences, as well as in the business sphere. The supposedly most data-intensive branch of the human sciences is digital humanities.

This chapter stresses the centrality of data sharing, whereby the ideas of the FAIR Principles are characterized as main drivers of data reuse. The importance of data management plans (DMPs) is emphasized. Other RDM processes, such as adding metadata, citing, retrieving, and curating datasets are also presented. The need for cooperating between disciplinary researchers and different data professionals is highlighted. Furthermore, it is underlined that a number of educational programs to data science also give attention to data management.

BACKGROUND

The consistent management of research data is especially crucial for the success of any long-term and large-scale collaborative research. RDM is also "the basis for efficiency, continuity, and quality of the research, as well as for maximum impact and outreach, including the long-term publication of data and their accessibility" (Finkel et al., 2020, p. 1).

As emphasized by several researchers, as well as by funding agencies and publishers, organizing and sharing research data is a fundamental part of the research process, thus comprehensive RDM is indispensable not only for ensuring reproducible and open scientific research, but increases citation rates for publications, and fosters research reproducibility (Borycz, 2021).

The main drivers of RDM are Open Science and Open Data. These initiatives gained momentum with the adoption of the Fair Access to Science and Technology Research Act in 2013 (US Congress, 2013), but this culture change is strongly supported also by the European Commission through the European Open Science Cloud (EOSC, n.d.).

RDM is often viewed as a set of mechanical, managerial, and technical handling processes (Ojanen et al., 2020). However, by encouraging the collaboration between researchers, and fostering better science, it can lead to better decision-making (USGS, n.d.).

RDM falls mainly into the domain and responsibility of researchers and data professionals. The latter group includes (data) librarians, data curators, and data stewards. It is unimaginable without services, offered by research offices, academic libraries, and computing (information technology) service units. Beside of these professionals, data scientists must be knowledgeable of RDM because it is one of the main services that provide them with the data that they are working with. Contrarily, RDM is also a

data science issue, which should not be restricted to machine learning or statistics, thus data scientists need to face the challenges of organizing and storing data. As declared by Davenport and Patil (2012, p. 73), data scientist's job is "bringing structure to large quantities of formless data and making analysis possible." In the light of this, we need to acknowledge that their job goes beyond formal data analysis, among others by including RDM.

Strongly related to the broad and varied aspects of RDM and data science are the activities of the Research Data Alliance, which is a "community-driven organization dedicated to the development and use of technical, social, and community infrastructure promoting data sharing and data-driven exploration. This organization is particularly important for "the global academic community where research infrastructure is often ad hoc, may have a short shelf-life, and hard to fund (Berman & Crosas, 2020).

Data stewardship also deserves attention because these professionals take care of data assets that do not belong to the stewards themselves. Data stewards' aim is ensuring that data-related work is performed in accordance with policies and practices. Together with preservation of data, their activities also fall into the domain of data science.

Data stewards should be involved in all phases of the research lifecycle, e.g. when designing and planning data-intensive experiments, co-supervising data capture, then running the data processing and curation. To be able to do all this, data stewards should have good knowledge of available methods and tools for data modelling and interpretation (Mons, 2018).

Among the key objectives of Open Science is the provision of Open Access to research results and Open Data. Both imply free use, reuse, and redistribution of data by anyone, who complies with the requirements, set by data licenses. These objectives can be achieved only if RDM and data stewardship meet criteria with respect to the collection, annotation, long-term preservation and archiving of data.

The efficient and consistent management of research data is crucial for the success of individual researchers and research teams (Finkel et al., 2020). As affirmed by Kanza and Knight (2022, p. 4) in their wittily titled and thought-provoking paper "Behind every great research project is great data management." They also underline that pursuing good research data management includes the following steps:

- Starting data management strategy early and covering it from data collection to publishing and sharing the results,
- Using sensible folder/file structures,
- Deciding early on the version control systems,
- Considering short- and long-term data storage by implementing the rule of keeping three copies of the data, within two types of media, with one stored at a separate site.

FOUNDATIONAL SKILLS, PRACTICES, AND PROCESSES

The most important facts about data sharing and reuse, the requirement for responsible conduct and research integrity will be presented here. We will also discuss general RDM skills and the importance of data management plans. The role of metadata and data citation will be touched on, as well.

Related Content

Comparative Analysis and Detection of Brain Tumor Using Fusion Technique of T1 and T2 Weighted MR Images

Padmanjali A. Hagargi (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 54-61).*

www.irma-international.org/article/comparative-analysis-and-detection-of-brain-tumor-using-fusion-technique-of-t1-and-t2-weighted-mr-images/266496

Recommendation System: A New Approach to Recommend Potential Profile Using AHP Method

Safia Baali (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-14).*

www.irma-international.org/article/recommendation-system/279278

Unsupervised Learning Techniques for Vibration-Based Structural Health Monitoring Systems Driven by Data: A General Overview

Francesco Colace, Brij B. Gupta, Angelo Lorusso, Alfredo Troiano, Domenico Santanielloand Carmine Valentino (2024). *Handbook of Research on AI and ML for Intelligent Machines and Systems (pp. 305-347).*

www.irma-international.org/chapter/unsupervised-learning-techniques-for-vibration-based-structural-health-monitoring-systems-driven-by-data/334478

Advanced Data-Driven Approaches for Intelligent Olfaction

Shiv Nath Chaudhri, Ashutosh Mishraand Navin Singh Rajput (2023). *Advanced Interdisciplinary Applications of Machine Learning Python Libraries for Data Science (pp. 124-154).*

www.irma-international.org/chapter/advanced-data-driven-approaches-for-intelligent-olfaction/330573

A New Approach of Deep Learning-Based Tamil Vowels Prediction Using Segmentation and U-Net Architecture

Julius Fusic S., Karthikeyan S.and Sheik Masthan S. A. R. (2021). *Handbook of Research on Deep Learning-Based Image Analysis Under Constrained and Unconstrained Environments (pp. 186-206).*

www.irma-international.org/chapter/a-new-approach-of-deep-learning-based-tamil-vowels-prediction-using-segmentation-and-u-net-architecture/268320