

ACADEMIC INTEGRITY PRINCIPLES IN RESEARCH DATA MANAGEMENT CURRICULUM

Svitlana Chukanova

The National University of Kyiv-Mohyla Academy, Ukraine, chukanovaso@ukma.edu.ua

Annotation. Research Data Management and sharing (RDM) is an essential part of scholarly work in the world. This practice is aimed at working with the research data and usually performed by data librarian (data curator) or scholar him/herself. Processing data with scholarly purpose impossible without complying with principles of research ethics and academic integrity namely: honesty, trust, fairness, respect, responsibility, courage.

Methods. This abstract is based on the experience of the author in teaching Academic Integrity and Research Data Management in Doctoral School for PhD. Students at the National University of Kyiv-Mohyla Academy and overview with short theoretical analysis of the international experience in RDM studies.

Keywords: Research Data Management, FAIR, Academic Integrity, citation, PhD. Studies, Information Literacy, Scholarly Communication.

Academic integrity is defined in Ukrainian Law on Education as “...a set of ethical principles and rules defined by law, which should guide the participants of the educational process during training, teaching and conducting scientific (creative) activities to ensure confidence in learning outcomes and/or scientific (creative) achievements.” (“Zakon Ukrayiny "Pro osvitu", 2020).

Thus, processing the data should be transparent and legal. Ukraine is on the way of implementing the best practices of RDM. The National University of Kyiv-Mohyla Academy in terms of NaUKMA Doctoral School provides a course for PhD. students “Digital science and information literacy”. The course involves modules on scientometrics, scholarly communication, academic integrity, research data management, scholarly image, information search and information literacy.

In the long-term perspective, the RDM module might be transformed into a separate course and the following modules may construct it:

1. The notion of research data as a basis for scholarly activity.
2. The notion of research data management and research data lifecycle.
3. Data management planning and tools for it.
4. The organisation of the data.
5. Formats of the files.
6. Documenting and assigning metadata.
7. Preservation, storage, and securing data.
8. Data protection, licensing, and data ethics for sharing.

It is obvious from the list that academic integrity principles are connected to data sharing: citing data, using other authors' data, sensitive data protection etc. For the better understanding on which stages of RDM activity academic integrity becomes the main topic, let us explain and visualise research data lifecycle. As there are numerous interpretations and models of research data lifecycle for an example, I suggest looking closely at Data Lifecycle used for RDM training at the University of Edinburgh on figure 1.

This Data Lifecycle model is a 6-stage process of ongoing research and includes the stages of creating, documenting, using (including reusing licenses), storage, sharing, preservation.

Each step is very important and must be transparent and clear for the target audience or beneficiary of research:

- researchers themselves,
- other colleagues interested in the topic,
- public,
- granting organisations, etc.

The transparency and clarity of research conduct are provided by adhering to academic integrity principles.

Data lifecycle. University of Edinburgh RDM training

The image is developed and used with the permission of R. Rice

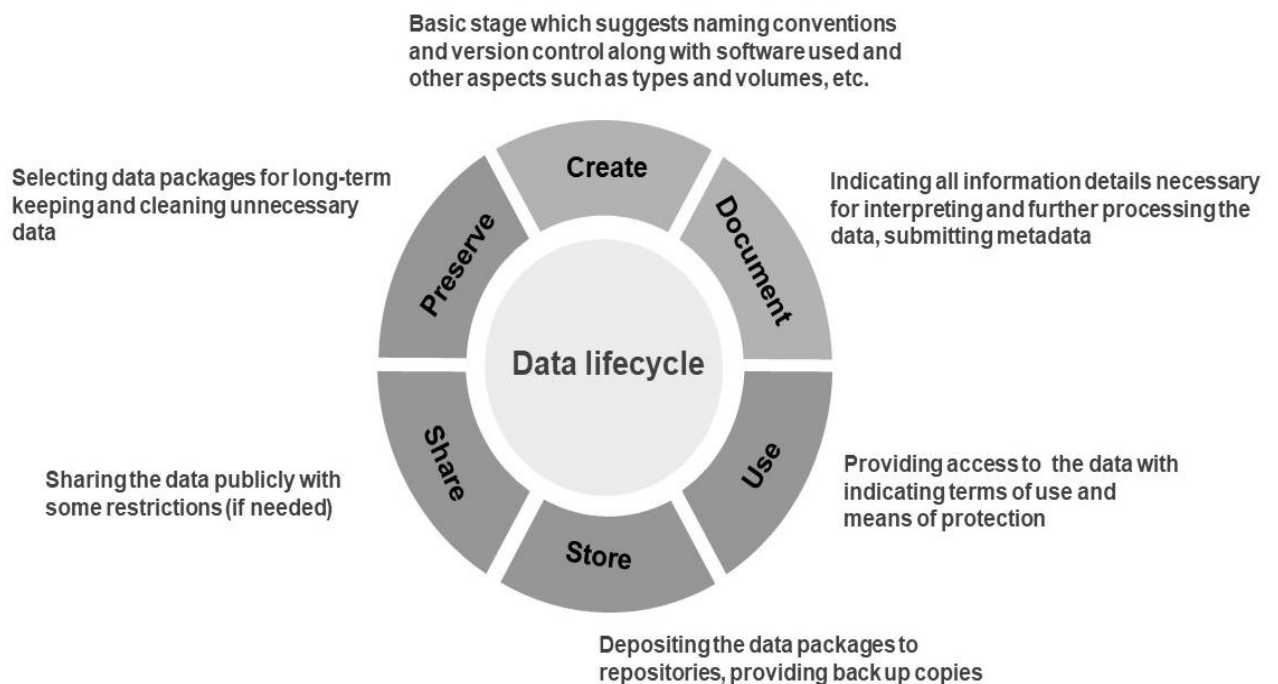


Figure 1. RDM Lifecycle. University of Edinburgh (Rice & Southall, 2016).

Academic integrity principles must be applied throughout the whole research process (Fishman (ed.), 2014), but special emphasis must be put on the stages of sharing, using, reusing and, preservation. Great attention should be paid to the research which involves sensitive data about health, personal data, humans as research subjects etc. The main academic integrity principles on these stages are the following:

- Avoiding plagiarism of data packages when reusing other authors' data by proper citation as one would do with publication or research paper.
- Use personal data only with the permission of the participant (consent to processing the data) with further anonymisation of information which can disclose the person.
- All the data especially vulnerable must be protected and kept in the repositories with a seal of approval or on the institutional servers.
- Data sharing must correspond with FAIR principles.

FAIR is an acronym which deciphered as Findable, Accessible, Interoperable, Reusable. It is used to show that data are provided with detailed descriptive metadata, can be easily found and reused with a proper citation to be verified or used as a basis to another research ("FAIR Principles - GO FAIR", 2020).

GO FAIR initiative generalised and worked out a detailed explanation on each component of FAIR principle based on the work of M. D. Wilkinson et al. in Scientific Data in 2016. Let us visualise this generalisation in table 1 and look closely what characteristics of metadata pertain to each FAIR criteria.

Table 1. FAIRification of the data (Wilkinson et al, 2016).

Criteria	Description 1	Description 2	Description 3	Description 4
Findable	Metadata has persistent identifiers	Data accompanied by rich metadata (detailed descriptive information)	Datasets have persistent identifiers and correlate with metadata	Metadata are searchable
Accessible	Metadata can be retrieved via standardised communications protocol which provides authentication procedure if needed	Metadata still can be accessed after data is no longer available	-	-
Interoperable	Metadata can be integrated as it shares common language and principles with other metadata	Metadata vocabulary shares common terminology in terms of FAIR principles	Metadata have links and references to other metadata	-
Reusable	Metadata have rich descriptions	Data can be accessed easily	Metadata has clear information for citation in case of reuse	Metadata organising performed following a common template

As we can see from the description of FAIR the usage of data should be regulated according to the principles of openness and transparency, data should be cited and used or reused with legal purposes only either for verification or for building up new researches.

It was mentioned above that data should be cited, so, let us regard the basic principles of data citation. According to the Data Citation Synthesis Group: Joint Declaration of Data Citation Principles, there are eight principles of data citation:

1. Importance. The data must be considered legitimate research result or product of the research process which serves as a basis for hypothesis thus it should be cited as well as any other research product e.g. paper or article.
2. Credit and Attribution. Each author of the data package should be acknowledged.
3. Evidence. Data serve as evidence of research and thus it must be cited in related papers which describe this research.

4. Unique Identification. Data packages must be available and assigned an identifier which is easy to process by computers and which is broadly accepted in the scholarly community.
5. Access. Citing the data provides access to data packages, metadata, and supplementary materials either for machines or for humans.
6. Persistence. Identifiers and metadata must be persistent, and event outlives the data packages themselves.
7. Specificity and Verifiability. Data descriptions must provide information for data packages verification (versions, numbers etc.) if needed.
8. Interoperability and Flexibility. Data citation styles must be commonly accepted throughout the scholarly community (Martone (ed.), 2014).

The conclusion from all mentioned above is the following: PhD. programs should include in their curriculums Information Literacy and Scholarly Communication components. These components in their turn must highlight the topics of Academic Integrity and Research Data Management. AI and RDM topics have a common element of research ethics which is a proper citation and ethical behaviour. In the long-term perspective, it is suggested to launch a separate course on research data management with more attention to research ethics in terms of RDM while Academic Integrity module enriches by the information of scholarly publishing culture and Academic Integrity values.

REFERENCES

FAIR Principles - GO FAIR. GO FAIR. (2020). Retrieved 20 November 2020, from <https://www.go-fair.org/fair-principles/>.

Fishman (ed.), T. (2014). The Fundamental Values of Academic Integrity [Ebook] (2nd ed., pp. 1-36). Clemson University. Retrieved 20 November 2020, from <https://www.academicintegrity.org/wp-content/uploads/2017/12/Fundamental-Values-2014.pdf>.

Martone (ed.), M. (2014). Joint Declaration of Data Citation Principles - FINAL. FORCE11. Retrieved 20 November 2020, from <https://www.force11.org/datacitationprinciples>.

Rice, R., & Southall, J. (2016). The data librarian's handbook (p. 78). Facet Publishing.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Zakon Ukrayiny "Pro osvitu". Zakon.rada.gov.ua. (2020). Retrieved 20 November 2020, from <https://zakon.rada.gov.ua/laws/show/2145-19#Text>.

Bio of the author. Svitlana Chukanova, chukanovaso@ukma.edu.ua

Head of the Viktor Kytasty American Library at the National University of Kyiv-Mohyla Academy, Head of a Sector at the NaUKMA Library

Graduated from Kyiv National Linguistic University (Translators' / Interpreters' Department) in 2011. PhD in Education (Kyiv Borys Grinchenko University), defended a dissertation on the topic: "Professional training of the Library and Information Science specialists in the higher education system of the USA" in 2016. Has 14 years' experience in library and information science and practice. Open World 2013 Alumna (Education and Innovations). Teaches Academic Integrity Module in terms of the Introduction to Kyiv-Mohyla Studies for BA students of the first year, and Research Data Management and Academic Integrity for PhD students of the first year in terms of Digital Science and Information Literacy course.