

Academic Integrity Principles in Research Data Management Curriculum

**SVITLANA CHUKANOVA, National University
of “Kyiv-Mohyla Academy” (Ukraine)**

RDM as a subject of PhD Course

- Research Data Management is an essential part of scholarly work in the world.
- This practice is aimed at working with the research data through the data lifecycle
- Processing data impossible without complying with principles of research ethics and academic integrity namely:
 - **honesty,**
 - **trust,**
 - **fairness,**
 - **respect,**
 - **responsibility,**
 - **courage.**

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).



RDM course should include the following topics:

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.

Resources for RDM training

- RDM training can be performed with the use of different approaches and with the involvement of open resources and courses.
- The most known platforms for RDM training are:
 - **Research Data MANTRA** – free online course developed by the University of Edinburgh;
 - **DataONE Education Modules**;
 - **FOSTER**.
 - **Data Tree**



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

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.

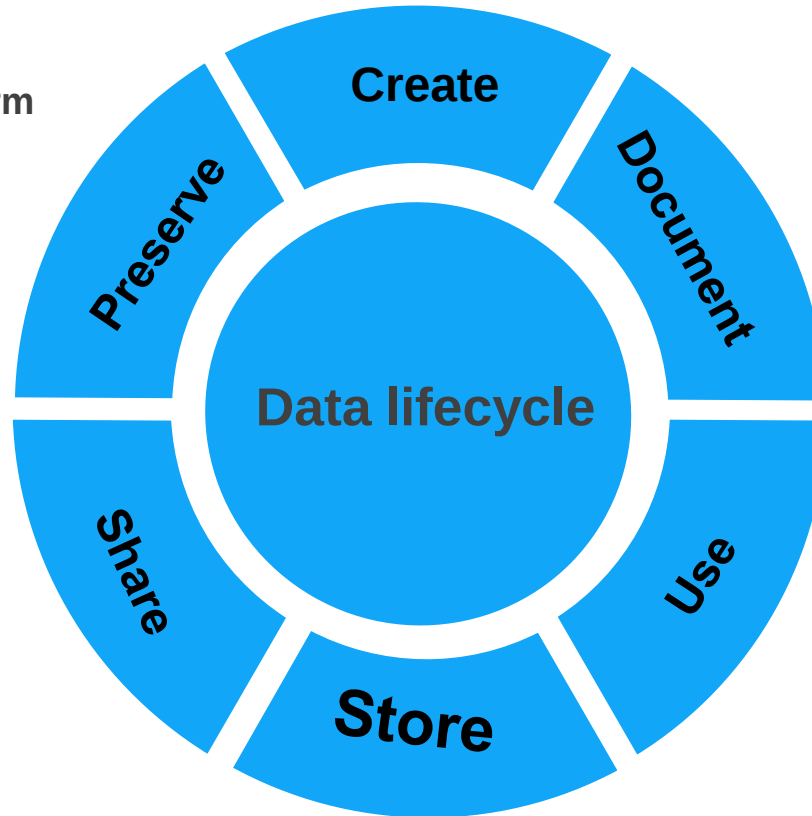
Data lifecycle. University of Edinburgh RDM training

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

Basic stage which suggests naming conventions and version control along with software used and other aspects such as types and volumes, etc.

Selecting data packages for long-term keeping and cleaning unnecessary data

Indicating all information details necessary for interpreting and further processing the data, submitting metadata



Sharing the data publicly with some restrictions (if needed)

Providing access to the data with indicating terms of use and means of protection

Depositing the data packages to repositories, providing back up copies

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.**



Academic and research ethics

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

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).



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

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.



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 one 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.

Data Citation (Martone (ed.), 2014).

Research 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 8 principles.

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.

Credit and Attribution

Each author of the data package should be acknowledged.

Evidence

Data serve as evidence of research and thus it must be cited in related papers which describe this research.

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.

Access

Citing the data provides access to data packages, metadata, and supplementary materials either for machines or for humans.

Persistence.

Identifiers and metadata must be persistent, and event outlives the data packages themselves.

Specificity and Verifiability.

Data descriptions must provide information for data packages verification (versions, numbers etc.) if needed.

Interoperability and Flexibility.

Data citation styles must be commonly accepted throughout the scholarly community



Conclusion

- 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 behavior.
- 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.

Thank you!

chukanovaso@ukma.edu.ua