



**EUROPEAN OPEN
SCIENCE CLOUD**

Data stewardship

Data Quality breakaway session

South-African Open Science Cloud workshop

Thursday 3 February 2022

Suzanne Dumouchel, EOSC Board of Directors

What national approaches could be implemented to support the development of data stewardship skills for researchers and support professionals?

What is the current state of documenting and reporting the data and information quality in your own discipline/organization, and how could this be improved?

What support structures could SAOSC put in place to ensure these practices are implemented in a reasonably standardized and consistent manner?.

Outline

1. Introduction
2. Defining data stewardship and data profiles
3. Data stewards and data quality
4. Data stewards as “issues solvers”
5. Knowledge infrastructures
6. Managing data stewardship in the EU ecosystem
7. Some hints for data stewards training
8. Conclusion
9. References

Support for Your Research

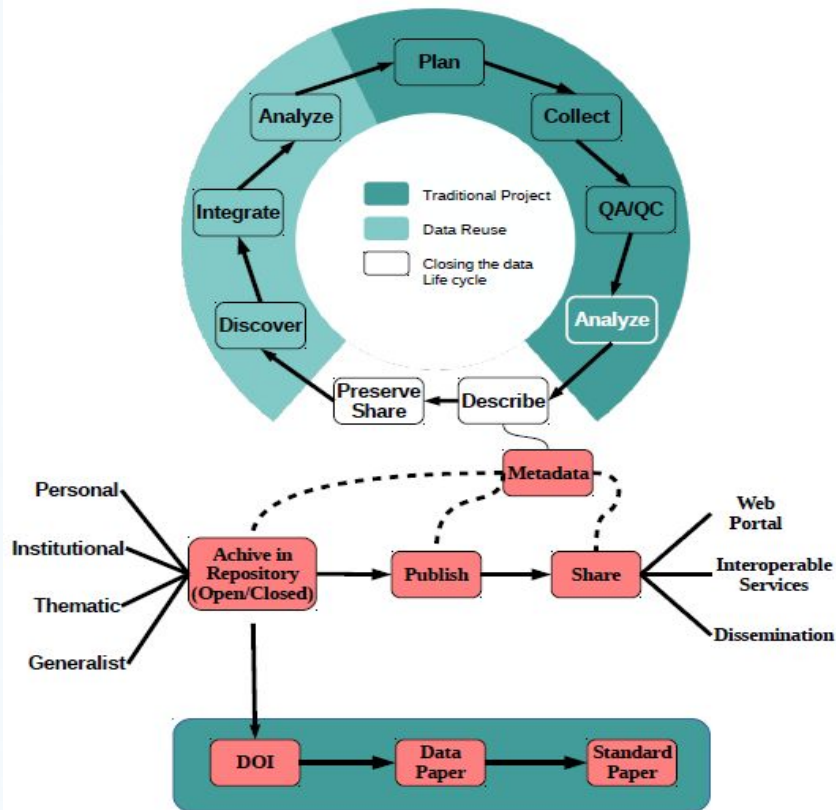
Researcher and Library Partnership

Research lifecycle



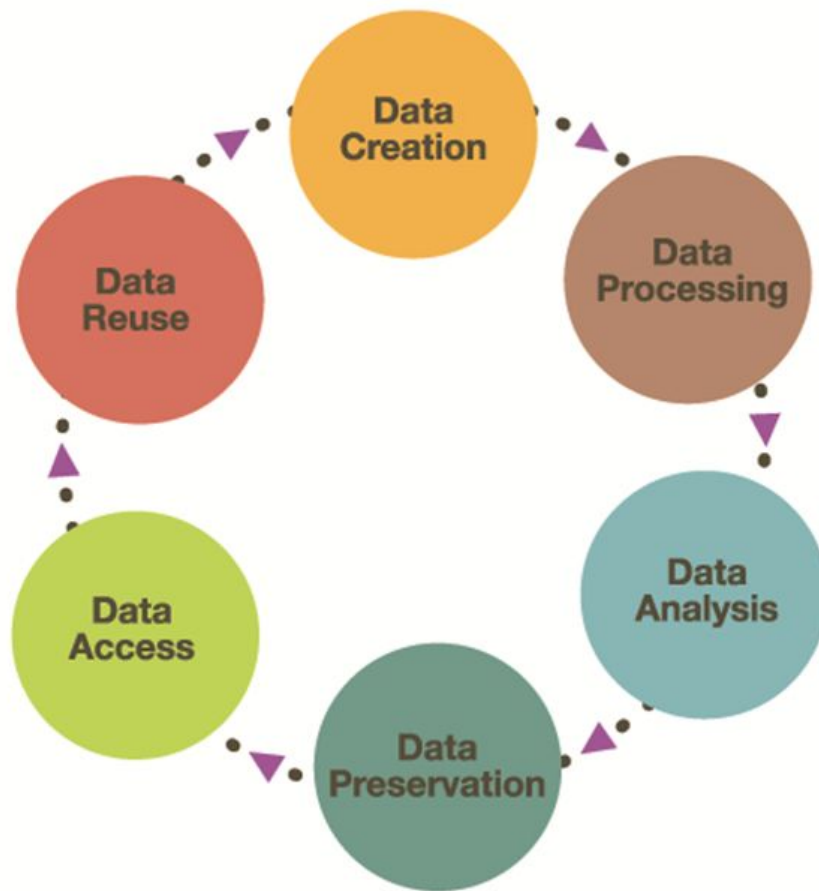
Research Data lifecycle

Research Data Life Cycle





Data lifecycle



Critical situation for OS implementation

A deluge of data (and repositories) but with almost no reproducibility, no reuse because of a lack of quality!

Importance of the selection process: maybe this is time to start considering what should be or not saved for scientific purposes...

Qualitative FAIR data and/or open data won't be possible without putting very important efforts on trainings for new data jobs such as data stewards and data scientists.

2. Data Scientists and Data Stewards

In 2018, the [Turning FAIR into Reality Final Report](#) and Action Plan identified the need for Data Scientists and Data Stewards.

Data Science has been defined as *“the ability to handle, process and analyze data to draw insights from it”*

Data stewardship defined as *“a set of skills to ensure data are properly managed, shared and preserved, both throughout the research lifecycle and for long-term preservation”*


Multiplication of actors in the EOSC ecosystem

*Digital skills for FAIR and Open Science :
report from the EOSC Executive Board
Skills and Training Working Group*



Figure 3. Framework of Actors in the EOSC Ecosystem

Data Steward

	Data Steward/Data Librarian
Overview	<p>A Data Steward is an expert on the preparation and treatment of data including data selection, storage, preservation, annotation provenance and other metadata maintenance, and dissemination. Data librarians are professional library staff who are experts on RDM, using research data as a resource or supporting researchers dealing with data (description, archiving and dissemination). Other closely related roles will also be considered under this category.</p>
Examples	<p>A Data Steward could be an expert who validates, recodes, trims or applies any other action on each source dataset of genomic samples related to influenza to guarantee that they can be properly used and integrated according to domain-specific standard formats.</p>
Required skills	<ul style="list-style-type: none">• Deep understanding of FAIR principles to ensure that research data from various domains is aligned with FAIR and CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles.• Ability to use EOSC-Core and EOSC-Exchange services for data publication and preservation and to facilitate the continued development of an infrastructure and library services to support data discovery, curation, preservation and sharing according to those principles.• Ability to validate the fulfilment of open science principles in EOSC-Core and EOSC-Exchange services related to data.• Ability to advise faculty and students on RDM according to the FAIR and CARE principles, including the discovery and reuse of existing datasets, through the EOSC services and ecosystem.

3. Data stewardship: the basis for data quality

Stewardship [...] embodies the responsible planning and management of [resources](#).

A **data steward** is an oversight or [data governance](#) role within an organization, and is responsible for ensuring the quality and fitness for purpose of the organization's data assets, including the [metadata](#) for those data assets. [...] A data steward may seek to improve the quality and fitness for purpose of other data assets their organization depends upon but is not responsible for.

4. Issues that can be solved by data stewards

- “Mobility problem” -> Importance of documenting provenance
- GDPR compliance and personal issues, sensitive data
- Releasing data
- Multilingualism issue but also multiculturalism => concepts translation depending on the scientific context
- Moving from research collections to larger repositories (in libraries?) and then archives

=> Bringing values (economic, politic, scientific, technical, etc.) to scientific data

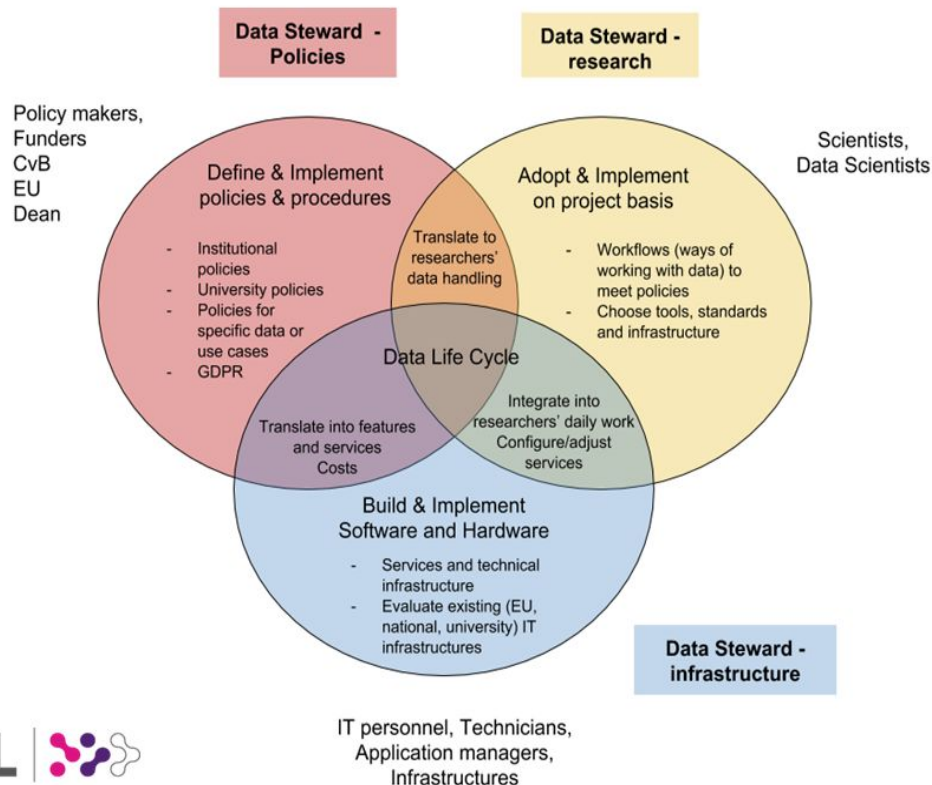
Data stewardship and trust

- Data in their context -> misleading interpretation when reused
- transparency of the algorithms for searching new data
- bias in the algorithm coding: gender, regional, career, etc.
- selection of metadata and metadata naming
- URLs naming

5. Complex environment of (FAIR) Data Stewards

Christine Staiger. (2019, May 21).
FAIR Data Stewardship - Data
sharing, archiving and publishing.
<https://doi.org/10.5281/zenodo.3066870>

Defining FAIR Data Stewards



5. Knowledge infrastructures

Data stewardship is part of a more complex ecosystem

Data (quality) cannot be thought without the associated softwares, hardwares, instruments, tools, protocols, documentations, standards in a collective and complementarity approach.

=> This creates a complex environment that can be called a “knowledge infrastructure” which can be defined as “Ecologies” “Complex adaptive systems” or “networks of people, artifacts and institutions that generate, share, and maintain specific knowledge”

6. Managing data stewardship in the EU

- Reports from funded projects and upcoming projects
- Trainings sessions (see the zenodo repository) that are made everywhere in Europe

And collaborations on an international level...

- Implementation Networks on different topics under the Go FAIR umbrella
- RDA working groups:

[Professionalising Data Stewardship IG | RDA](#)

[Education and Training on handling of research data IG | RDA](#)

6. Managing data stewardship in the EU : the EOSC-A AG

Research Careers and Curricula

The most important stakeholders for EOSC are the researchers. The Advisory Group Research Careers and Curricula is mainly focused on the researchers.

- The [Data stewardship, curricula and career paths Task Force](#) will focus on the Data Stewards role and their core activities. To help researchers to make FAIR data it is necessary to have professional staff. To keep professional staff, we need to have a common curriculum for their skills and possibilities for career paths.
- The [Research careers, recognition and credit Task Force](#) will address incentives and rewards for researchers to manage and share their data, code and other research outputs, activities, and processes. To make Open Science happen it is necessary that criteria of Open Science and FAIR principles are an integral part of academic career progression and grant assessment processes.
- The [Upskilling countries to engage in EOSC Task Force](#) is mainly focused on developments in Open Science education being addressed at Member State level within research performing organizations. The task force will assist in aligning skills initiatives and supporting the onboarding of these into EOSC. Without education in Open Science skills, it would be harder for researchers to live up to the necessary transformation of their working life.

Focus on the Data stewardship, curricula and career paths Task Force

The **Data stewardship, curricula and career paths Task Force** will focus on the Data Stewards role and their core activities. To help researchers to make FAIR data it is necessary to have professional staff. To keep professional staff, we need to have a common curriculum for their skills and possibilities for career paths.

Chairs



Ilire Hasani-Mavriqi
TU Graz



Vera Matser
EMBL



Wilhelm Widmark
Stockholm University

Board Liaison

22
members

Main tasks of the TF

- Task 1: Stakeholder management and dissemination.

Clarify terminology around Data Stewards/Data Stewardship and how these terms might be used in different international, disciplinary and institutional contexts. Clarify overlapping terms (e.g. data management activities).

- Task 2: Definition of a minimal data stewardship curricula

Define Data Stewards roles, their core activities, possible specialisations or extension activities, and the context in which these roles operate

Develop a competency profile for data stewardship core activities

Define levels of training needed by the different Data Steward roles

Main tasks of the TF

Produce guidelines for a Data Stewardship curricula, which could include university curricula as well as short training.

Refined version of the the competency profile and curricula with input from the implementation use cases

- Task 3: Career paths for Data Stewards and associated roles:

Develop Data Steward career paths;

Recommendations for research performing institutions and organisations on how to establish career paths for Data Stewards

Recommendations for the recognition and rewards for data stewardship activities (e.g. credits, incentives);

- Task 4: Implementation examples using use cases

7. Some hints for data stewards training

1. Minimum Viable Skillsets (MVS)
2. Training-of-Trainers (ToT)
3. FAIR-by-design methodology for learning materials -
4. Harmonised curricula and learning paths
5. Lifelong learning through professional networks
6. Creation of a wide network across Competence Centres

8. Conclusive words

“Scholars can collect, discover, retrieve, analyze, and disseminate data at scales never before possible. Some of those data are worth keeping forever; others have transient value. Some are more readily recreated if needed than kept. Throughout human history, keeping everything was never an option. Future uses of information can never be fully anticipated. [...] Sustaining data is a much higher bar than simply storage and backup. The challenge is to make data discoverable, usable, assessable, intelligible, and interpretable, and to do so for extendable periods of time. [...] The hardest part is to determine who may be willing to make those investments on behalf of the interested parties. **The value proposition for access to data is the value proposition for knowledge infrastructures**”. (p.287, Big data, Small data, No data, Christine L. Borgman)

9. References

Christine L. Borgman. (2015). *Big data, Small data, No data*, MIT Press.

European Commission, Directorate-General for Research and Innovation. (2021). *Digital skills for FAIR and Open Science : report from the EOSC Executive Board Skills and Training Working Group*, Lazzeri, E.(editor), Barker, M.(editor), Manola, N.(editor), Gaillard, V.(editor), Kuchma, I.(editor), Stoy, L.(editor), Publications Office, <https://data.europa.eu/doi/10.2777/59065>

Peng, G., Privette, J.L., Kearns, E.J., Ritchey, N.A. and Ansari, S.,(2015). A Unified Framework for Measuring Stewardship Practices Applied to Digital Environmental Datasets. *Data Science Journal*, 13, pp.231–253. <https://datascience.codata.org/articles/abstract/16/>

Sarretta, Alessandro. (2018). Research Data Life Cycle. Zenodo. <https://doi.org/10.5281/zenodo.1149040>

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



**EUROPEAN OPEN
SCIENCE CLOUD**

Thank you! Ready for questions

Suzanne Dumouchel

EOSC Board of Directors

suzanne.dumouchel@huma-num.fr