Data Stewards and Data Management Principles
(LTAT.02.014)

Priit Adler
Research Fellow, University of Tartu
Elixir EE, BIIT

2021.04.27
Recap

DMP
• DSW
• Knowledge model
Data Steward role

Based on “Professionalising data stewardship in the Netherlands: competences, training and education” https://zenodo.org/record/4623713#.Y1ebC30zaL4
“Invest 5% of research funds in ensuring data are reusable. It is irresponsible to support research but not data stewardship” - Barend Mons, Nature, Vol. 578, p. 491
https://doi.org/10.1038/d41586-020-00505-7

“Students in PhD programmes spend up to 80% of their time on ‘data munging’, fixing formatting and minor mistakes to make data suitable for analysis — wasting time and talent. With 400 such students, that would amount to a monetary waste equivalent to the salaries of 200 full-time employees, at minimum. So, hiring 20 professional data stewards to cut time lost to data wrangling would boost effective research capacity. Many top universities are starting to see that the costs of not sharing data are significant and greater than the associated risks. Data stewardship offers excellent returns on investment.”
**Lucy**

**Education and qualification**: Ma or PhD in a relevant scientific domain

**Work experience and background**: research experience in university or business; speaks the language of the researcher; aware of researchers’ needs; embedded in specific projects

**Responsibilities and tasks**: efficient and quick in handling data; analytical focus; responsible for data processing and analysing; manages big data; takes a consultancy role

**User story (example) for tool usage**:
- As a research oriented data steward, I can effectively support researchers, so that data management is an intrinsic part of their project
- As a research oriented data steward, I want to become more skilled in computing data, so I can help my researchers with managing data in a project
- As a research oriented data steward, I want to be able to identify my current knowledge of FAIR data management, so I know what training to take to improve my FAIR data competences
Domain areas, responsibilities and tasks of a data steward

For overarching data steward (combined over data steward policy, data steward research and data steward infrastructure)
Areas

1. policy/strategy
2. compliance
3. alignment with FAIR data principles
4. services
5. infrastructure
6. knowledge management
7. network
8. data archiving

Basic job profile components
Area 1: policy/strategy

Development, implementation and monitoring of research data management policy and strategy for the research institute

- Gives sound advice on RDM policy and strategy
- Is responsible for the availability of an up-to-date institutional data management plan (DMP)
- Is responsible for advice and awareness of RDM policy, FAIR principles and Open Science
- Is responsible for advice on an adequate research data infrastructure and proper tools
- Advises policy officers (on a strategic and tactical level)
Main tasks and activities

• Monitors the institute's RDM policy
• Advises the institute's management to advance
• Explores new needs, opportunities and trends in RDM
• Develops (together with others) DMP templates
• Implements RDM as a regular aspect of doing research
• Translates the RDM policy to data infrastructure and tools requirements
• Creates an overview of available data infrastructure and tools of the institute
• Development and operationalisation of products and services in the RDM domain
• Translates international developments into policies and practices at the university
• Leads the development and implementation of the faculty’s data management policy
• Liaison function at policy level
• Connects with stakeholders in the research domain and acts as sparring partner for leading scientists and board of directors
• Coordination of the innovatory information agenda as part of the complete research life cycle
• Safeguards the data architecture
• Optimising RDM processes
• Securing provenance and audit trail
Area 2: compliance

Compliance to relevant scientific, legal, and ethical standards

• Is familiar with, and works according to, the RDM policy to
  • the <national> Code of Conduct for Academic Practice,
  • the <national> Code of Conduct for Research Integrity and
  • the General Data Protection Regulation (GDPR),
  • Good Clinical Practice (GCP) and other laws and regulation.

Furthermore, a data steward continuously aligns with legal and ethical standards
• Is responsible for advice on compliance before, during and after a research project
• Is familiar with relevant legislation and field specific standards
• Is responsible for advice on the data infrastructure and tooling within legislation
Main tasks and activities

• Ensures compatibility of the RDM policy and monitors compliance
• Contacts the institute’s privacy officer, legal advisors or ethical board in case of questions regarding compliance
• Consults with privacy and security officers
• Translates policies from legal/privacy officer to the institutes practice
• Develops and/or guides standard solutions for recurring data issues and for data classification, including input for the Data Protection Impact Assessment (DPIA)
• Monitors and supervises the execution of a project or data collection in line with the DMP and relevant codes of conduct and legislation, including ethical, legal and social issues (ELSI)
• Identifies gaps in policy and takes action if needed
• Train researchers and research support employees on compliance requirements,  
• Monitors and supervises the use of the data infrastructure and tool landscape in line with the institute’s RDM policy, relevant codes of conduct and legislation  
• Advises researchers and research support staff about how to deal with privacy sensitive data in accordance with the (GDPR) guidelines  
• Monitoring up-time and security breaches of servers and services
Area 3: alignment with FAIR data principles

Alignment to the FAIR data principles and the principles of Open Science

- Is responsible for alignment with the FAIR data principles and the principles of Open Science
- Facilitates and supports FAIR data and FAIR software
- Is responsible for advice on an adequate research data infrastructure and proper tools
- Is responsible for alignment of the data infrastructure and tool landscape with the FAIR principles
Main tasks and activities

• Advises, supports and provides guidelines to researchers on the Findability (F) of data, including adequate data infrastructure and tools, persistent identifiers and rich (institute-specific) metadata standards
• Advises, supports and provides guidelines to researchers on the Accessibility (A) of (meta)data to potential reusers
• Advises, supports and provides guidelines to researchers on the Interoperability (I) of data, including broadly applicable languages, vocabularies and other standards
• Advises, supports and provides guidelines to researchers on the Reusability (R) of data, including documentation and licenses with the conditions for reuse and IP rights
• Engages researchers in developing metadata schemes and documentation standards to improve FAIR data and software
• Monitors and supervises the use of the data-infrastructure and tool landscape on alignment with the FAIR data principles
• Identifies gaps and takes action if needed
• Advise researchers on careful management of research data with the FAIR principles,
• Ensuring that research data from various domains is adapted in accordance with FAIR principles
Area 4: services

Availability of adequate support on research data management, in staff or services

- Is responsible for advice for the availability of adequate support on RDM for the researchers, PhDs and research support staff of the institute
- Is responsible for identifying the requirements of adequate data infrastructure for RDM for researchers
- Provides training in the field of RDM
- Analyses data management needs of a researcher/research group
- Proposes, implements and monitors workflows to improve RDM practices
- Oversees and assists in executing data collection, description, cleaning, merging, licensing, sharing and use of metadata standards
**Main tasks and activities**

- Advises board, management, researchers and support staff of the institute on RDM support
- Initiates or supervises the set-up and update of suitable support facilities or services in the institute
- Provides training in the field of RDM
- Analyse data management needs of a researcher and/or research group
- Propose, implement and monitor workflows to improve RDM practice
- Oversee and assist in executing data collection, description, cleaning, merging, licensing, sharing and use of metadata standards
- Provides guidance and instruction on discovery, acquisition and use/reuse of data
- Solves (together with others) practical RDM problems
- Advises and supports researchers on data infrastructure and tools
- Advises, develops and takes care (standard) ICT solutions for recurring issues and questions are implemented
- Provides guidance and instruction on the use of data infrastructure and tools
- Support researchers, PhDs, and students to store, (re)use and analyse research data and information
- ... will continue on the next slide 😊
Main tasks and activities – continued

• Advise researchers on careful management of research data and about procedures and technical aspects that are important for the quality of (meta)data
• Takes the lead in developing further service provisions at research group, discipline or faculty level
• Contributes to scientific articles and helps to draft subsidy applications
• Performing statistical analyses
• Processing data
• Building scripts to assist data control and data cleaning
• Correct data extraction
• Data capturing/data harvesting
• Data enrichment/linking datasets
• Reusing and developing (machine learning) algorithms
• Visualising data
• Working on complex data(processing) issues, providing input to improve processes or systems
• Developing new workflows for ingestion, aggregation and export of data
Area 5: infrastructure

Availability of adequate data infrastructure for research data management

- Is responsible for identifying the requirements of adequate data infrastructure for RDM to comply with the institute's RDM policy and alignment to (inter)national data infrastructure and tools
- Monitors the department's or project group's needs including supporting adequate access, in accordance with the DMP
- Identifies the requirements of adequate data infrastructure and tool landscape that fits with the needs of the researchers, with the institute's RDM policy and supports FAIR data and Open Science
Main tasks and activities

• Initiates and supervises requests and acquires data infrastructures and tools for RDM within the institute
• Supervises monitoring of the need, use and availability of data infrastructures and tools
• Advises the management of the institute on data infrastructures and tools
• Is aware of and advises the institute on relevant (inter)national data infrastructures
• Monitors the needs regarding data infrastructure and tools for RDM within the department, project or data collection
• Supports access to data infrastructure and tools for RDM
• Sets requirements for data infrastructure and tools for RDM
• Requests and acquires data infrastructure and tools for RDM within the institute
• Monitors the need, use and availability of data infrastructure and tools
• Is aware of and advises the institute, department and researchers on relevant (inter)national data infrastructure and tools
• Takes part in initiatives to further develop data infrastructure and tools
• ... will continue on the next slide 😊
Main tasks and activities – continued

• Explores new trends in data infrastructure and tools for RDM
• Monitors a process, system or the data architecture to optimise it
• Building databases
• Caring for structured and secure data storage
• Securing provenance and audit trail
• Advising on the improvement of the data architecture
• Working on complex data(processing) issues, providing input to improve processes or systems
• Developing new workflows for ingestion, aggregation and export of data
• Finding optimal solutions for use of existing data infrastructure
• Extending existing data infrastructure
Area 6: knowledge management

Adequate level of knowledge and skills on research data management within the institute, department or project

• Is responsible for determining the adequate level of knowledge and skills on RDM within the department or project group in order to comply with the institute's RDM policy
Main tasks and activities

- Monitors RDM skills of the researchers and research support staff within the institute
- Identifies knowledge and skill gaps and ensures appropriate training
- Ensures a sufficient level of awareness among researchers and research support staff of the institute
- Monitors RDM skills in the department or project
- Identifies knowledge and skill gaps of the DMP, and takes action if needed
- Initiates and provides training on RDM, tailored to the needs of the researchers and research support staff of the institute
- Creates awareness on RDM among researchers and research support staff
- Explains the added value of RDM
- Introduces to new employees the institute's RDM
- Monitors technical RDM related skills in the institute, department or project
- Identifies technical knowledge and skill gaps, and takes action if needed
- Initiates and provides training on data infrastructure and tools for RDM
- Creates awareness on data infrastructure and tools among the institute, department and researchers and explains the added value to RDM
- Advise researchers and research support staff on careful management of research data and about procedures and technical aspects that are important for the quality of (meta)data
- Develop or give training courses that relate to the RDM field
- Is proactive in knowledge dissemination, for example, by organising events
- As spokesman for the faculty has a mission to convince researchers of the added value of good research data management
Area 7: network

Obtaining and maintaining a network of aligned expertise areas and relevant departments and organisations inside and outside the institute, department or project

• Is responsible for obtaining and maintaining a network of aligned expertise areas and relevant departments and organisations inside and outside the institute with regard to RDM
Main tasks and activities

• Within the organisation, builds and maintains a network of researchers interested in RDM
• Refers researchers to other RDM related facilities and services (legal, financial and/or operational), inside and outside the institute, the department or project
• Liaises with (technical) experts inside and outside the institute, the department or project
• Maintains a network with RDM related colleagues and other relevant departments and organisations
• Connects data support people with each other
• Refers researchers and research support staff to RDM related data infrastructure and tools, inside and outside the institute
• Take the initiative to establish contacts with discussion partners from both research or specialist domain groups and colleagues at faculty or discipline level in order to consult on subjects from the data management field
• Deploys his or her knowledge of (inter)national developments in the field of data management and Open Science
• Works to build up an (inter)national network in the RDM field
Area 8: data archiving

Adequate support and data infrastructure for FAIR and long-term archiving of data of the institute, department or project

- Is responsible for identifying the requirements of adequate support and data infrastructure for FAIR and long-term archiving of data of the institute, department or project group by researchers, including selection of data, and sustainable and legitimate access to data sources of the department or project group, for the required period
Main tasks and activities

- Develops, implements and monitors the institute's internal and/or external data archiving and access policy
- Monitors the internal and/or external archiving of data by researchers of the institute, department's, project's
- Assesses whether internal and/or external data storage and archiving facilities meet the applicable requirements
- Advises and supports researchers in the selection of data to be archived
- Assesses whether internal and/or external data storage and archiving facilities meet the applicable requirements
- Advises on data infrastructure and tools for data archiving services
- Identifies the institute's, department's or researcher's needs into infrastructural requirements
- Assesses whether internal and/or external data storage and archiving facilities meet the applicable requirements
- Monitors and evaluates data infrastructure and tools that best fit the institute's RDM policy
- Advises on (meta)data formats for data archiving
- Support researchers and students to store, use/reuse and analyse research data and information
- Caring for structured and secure data storage after a research project
RDMkit role pages

https://rdmkit.elixir-europe.org/data_steward_policy.html
https://rdmkit.elixir-europe.org/data_steward_research.html
https://rdmkit.elixir-europe.org/data_steward_infrastructure.html
Discussion and questions