Towards FAIR Data Steward as profession for the Life Sciences

Programma bijeenkomst Personalised Medicine
20 juni 2019

Salome Scholtens
UMCG
What is a data steward?

..... and what has it to do with FAIR data and personalised health?
Background of the project

- Sufficient high-quality data steward expertise is essential for FAIR data management in Lifesciences

- All research data has to be collected, maintained, processed, analyses, stored, shared, linked, archived, etc

- Challenges:
  - These persons are currently insufficiently available
  - No consensus on what a data steward is or does
  - No job definitions/profiles available (e.g. in UFO)
  - No/too little tailored education & training
Project details

• 1 year project (August 2018 – July 2019)

• ZonMw/KWF/Zilveren Kruis funded (50k), complemented with in-kind contributions of the core team members

• Core team: DTL/ELIXIR-NL, UMCG/RUG, Radboudumc/RU, UMCU

• Consultation committee: representatives of main stakeholders: LCRDM, Data4Lifesciences, Health-RI, 4TU, SURF, NFU, DAS (hogescholen), ZonMw

• Active connection with data steward networks
Aim

- To professionalise the data steward function within the life-sciences domain, with a special focus on implementation of the FAIR principles
- In order to increase the quality and capacity of data stewards for life sciences
- By developing training and a standardized job description, and creating more visibility of the function

A collaborative approach built on existing expertise
Method

• The content was defined in multiple iterative cycles

• In general following these steps:
  – Analysis of about 40 job descriptions
  – Analysing and mapping existing data stewardship competency frameworks (EOSCpilot, Purdue, DAMA, EDISON) and to the FAIR principles
  – Consultation of experts (Consultation Committee)
  – Consultation of persons with data steward functions
Open science in practice

Intermediate results are already available, are shared and we received feedback

https://zenodo.org/communities/nl-ds-pd-ls/?page=1&size=20
Data Steward functions

• Types:
  • Data Steward A - institute and policy focused
  • Data steward B - project and research focused
  • Data Steward C - data and e-infrastructure focussed

• Knowledge areas:
  Policy/strategy
  Alignment with FAIR data principles
  Infrastructure
  Network
  Compliance
  Services
  Knowledge
  Data archiving
Data steward landscape

Policy makers, Funders, CvB/RvB, EU, Dean

Define & Implement policies & procedures
- Institutional policies
- Policies for specific data or use cases
- FAIR data & GDPR

Data Steward Policies

Data Steward Research

Adopt & Implement on project basis
- Workflows (ways of working with data) to meet policies
- Choose tools, standards and infrastructure

Translates researchers' data handling

FAIR Data

Facilitate & Implement Software and Hardware
- Services and technical infrastructure
- Evaluate existing (EU, national, university) IT infrastructures

IT personnel, Technicians, Application managers, Infrastructures

Data Steward Infrastructure

Source: https://zenodo.org/record/2585691#.XPYDucgzaUI
## Function description

### Definition of responsibilities and activities per type

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy/strategy</td>
<td>Responsible for advice on and development, implementation and monitoring of a RDM policy and strategy for the research institute, which includes the complete research data life cycle, and supports FAIR data and Open Science, in alignment with the relevant stakeholders and within financial and legal constraints, within the institute and in the context of the institute.</td>
</tr>
<tr>
<td>Compliance</td>
<td>Responsible for compliance of the RDM policy to the Netherlands Code of Conduct for Academic Practice, the Netherlands Code of Conduct for Research Integrity and the GDPR, as well as continuous alignment with legal and ethical standards.</td>
</tr>
<tr>
<td>Alignment with FAIR data</td>
<td>Responsible for alignment of the RDM policy to the FAIR data principles and the principles of Open Science.</td>
</tr>
<tr>
<td>Services</td>
<td>Responsible for the availability of sufficient support on RDM, in staff or services, for the researchers and research support staff of the institute.</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Responsible for the availability of adequate e-infrastructure for RDM to comply with the institute’s RDM policy and alignment to (inter)national data and e-infrastructures.</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Responsible for an adequate level of knowledge and skills on RDM within the institute in order to comply with the institute’s RDM policy.</td>
</tr>
<tr>
<td>Network</td>
<td>Responsible for obtaining and maintaining a network of aligned expertise areas and relevant departments and organizations inside and outside the institute with regard to RDM.</td>
</tr>
<tr>
<td>Data archiving</td>
<td>Responsible for policy and adequate support and e-infrastructure for FAIR and long-term archiving of data of the institute, stored internally as well as externally, and for sustainable and legitimate access to data sources of the institute, for the required period.</td>
</tr>
</tbody>
</table>


Function description

LCRDM report
Competences

- Definition of KSAs: Knowledge, Skills and Abilities of a data steward
- Definition of Learning Objectives for training for data stewards
- Based on Blooms levels

<table>
<thead>
<tr>
<th>RESPONSIBILITY:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KNOWLEDGE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge about legislation, ethics, code of conducts and research codes with regard to (research)data, (medical) research and privacy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SKILL/ABILITY:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate legislation and codes of conducts with regard to research data to practical implications and guidelines that researchers can understand</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LEARNING OBJECTIVES:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Explain the legislation and codes of conduct with regard to research data to researchers (applying)</td>
</tr>
<tr>
<td>- Examine and list the practical implications of legislation and codes of conduct with regard to research data (analyzing)</td>
</tr>
<tr>
<td>- Create guidelines based on legislations and codes of conduct with regard to research data (creating)</td>
</tr>
</tbody>
</table>
Workshops

- June 11 and June 12, 2019, Utrecht
- Evaluate and pilot with 50 data stewards
Training

• Current approach
  – Identify existing training
  – Gap analysis
  – Develop taxonomy of training aspects
  – Set-up for a script for an eLearning module

• Guided by ElevateHealth

• To be continued in the coming year
Example of taxonomy
Next steps

• Sustainability plan and recommendations to ensure national embedding

• Continuation in NPOS context
  - sept 2019-sept 2020
  - two focal points:
    - establishing nationally endorsed competencies/skills for training data professionals
    - overview training

• Further alignment with LCRDM approach
• Broadening towards other domains and towards Open Science
Thank you!

s.scholtens@umcg.nl