DATA STEWARD training-info

Teemu Rauhala,
Päivi Malinen (TAU)

Avoin tiede

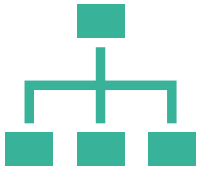
CSC

Tietoarkisto 25

universität wien

Today

Data Stewards –who are they?



Data Steward – training

Basics about the training
Content of the training
Training schedule
How to apply?

Time for questions

Data steward – who are they?

- Data Steward – wide range of supporting roles whose duties are related to the creation, maintenance and use of research data. Core responsibilities and tasks range from advising and consulting on different areas of data management, implementing open science and data management policies to technical ICT tasks. The tasks also vary between different universities and research institutes.
- Data steward is a person who is responsible for maintaining the quality, integrity and access arrangements of (research) data and metadata in a manner that is consistent with applicable law and organizational policies. His/her task is to ensure that the data is handled appropriately at all stages of the research life cycle.
  (Jetten et al. 2021)
- Data steward acts as a bridge builder, for example, supporting researchers in data management and complying with the FAIR principles, developing data management practices and policies in cooperation with research administration, promoting data sharing between different research infrastructures and integrating data management services into research infrastructures

National Forum for Research Data Management- Denmark

Four roles of a data steward

ZonMw/ELIXIR project- The Netherlands

Three roles of a data steward

[The roles of data stewards in the data stewardship landscape identified in Denmark and the Netherlands (zenodo.org)]

Why do we need data stewards?

- When research data is handled appropriately, researchers have significantly more time to conduct the research themselves.*"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."*- Mons, B (2020)
- Many research organizations are beginning to see that there are significant costs associated with not sharing knowledge. Good data management provides an excellent return on the investment in it. *"Invest 5% of research funds in ensuring data are reusable. It is irresponsible to do research but not data stewardship"* - Mons, B (2020)
- When researchers did data management themselves without utilizing the support provided by the organization, the costs of data management are higher than for researchers who use support services. Hofelich Mohr, A et. Al. (2024)

Mons, B. (2020) Invest 5% of research funds in ensuring data are reusable. Nature, Vol. 578, p. 491 https://doi.org/10.1038/d41586-020-00505-7
Hofelich Mohr, Alicia, Jake Carslon, Lizhao Ge, Joel Herndon, Wendy Kozlowski, Jennifer Moore, Jonathan Petters, Shawna Taylor, and Cynthia Hudson Vitale. Making Research Data Publicly Accessible: Estimates of Institutional & Researcher Expense. Washington, DC: Association of Research Libraries, February 2024.https://doi.org/10.29242/report.radsexpense2024

What competencies are needed?

- Data steward tasks are versatile and require extensive special expertise.
- Working as a data management expert requires, among other things, the ability to understand the data management needs of different researchers and research or development projects and develop sustainable and responsible operating methods for their implementation.
- Data protection and data security, in particular, require an in-depth understanding not only of current regulations but also of best practices in the field.

NPOS project “Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship”

- Final report (2021): https://doi.org/10.5281/zenodo.4320504

Data steward -course

Facts about the training 1/2

- Name of the training: Data Steward -course
- Training begins 3/2025 and ends 2/2026, implementation number 2 begins 3/2026
- Scope of the training 15 ETC
- Possible to combine work and studies. Estimated workload 10 h/week
- Online teaching approx. 4 days/month (break in the summer) + assignments. At the beginning of the training, a 2 days contact teaching period in Tampere
- The language of the training is English (possibility in flexilingual arrangements)
- The training costs 1500 euros/participant + travel and accommodation costs in Tampere at the beginning of the course
- At the end of the training the participants do a project work
- Completion of the training is subject to a certificate issued by the Tampere University

Facts about the training 2/2

- The participants form a peer learning and peer support community for each other during and after the training, → peer learning is implemented through various assignments and group work
- After the training, the person will be able to work in various research organizations (and possibly companies) in various expert tasks in data management
- The training expands the participants' employment opportunities beyond the research institutes as well
- The training has been designed in broad cooperation with The National Open Science and Research Coordination AVOTT, CSC, Finnish Social Sciences Data Archive and University of Vienna
- The training is partly funded by the Ministry of Education and Culture

Content of the training

Content of the training

- 5 modules, competence develops during the training from fundamentals to applied, practical skills.
- Online teaching throughout the course in Teams or Zoom
- In addition, before and after online teaching, self-study + assignments (reports/presentations/project work) in Moodle
- In each module (1-4) several different sessions, the progress of the participants is monitored throughout the training with separately assessed intermediate tasks
- A separate project work at the end

Module 1: Introduction to Research Data Management and Open Science (2 ECTS)

Content

Introduction to:

- Responsible conduct of research, good research practices and RDM
- Role and tasks of data stewards
- Legal and ethical aspects

Learning outcomes

- Define research data management and understand its importance in the research process and in implementing good scientific practice
- Identify common challenges and best practices in research data management, FAIR principles, and open science practices
- Understand the role of data stewards and other data (+methods) support personnel in research organizations
- Recognize and understand key ethical principles guiding research
- Understand the basic legal aspects in research data management

Module 2: Introduction to IT and Data Science (3 ECTS)

Content

Introduction to

- Data Science and data-driven research
- Hardware and software for Data Science
- Basic programming concepts
- Advanced programming concepts: FAIR Software, Reproducibility, and other advanced tools- Database systems

Learning outcomes

- Explain the fundamental concepts of data science, including data preprocessing, analysis, and visualization
- Identify and utilize appropriate hardware and software for data science tasks, including storage solutions, secure access, and computational resources.
- Learn how to read and run Python and R scripts and understand the importance of reproducibility and the FAIR principles in research software development.
- Familiarize with advanced tools for computational research such as version control, containers, and high-performance computing.

Module 3: Research Data Life Cycle (6 ECTS)

Content

- Data management planning
- Data documentation
- Data sharing and preserving
- Domain-specific characteristics of research data

Learning outcomes

- Recognize the phases of the research data lifecycle.
- Understand the importance of a data management plan (DMP), evaluate a DMP and provide feedback.
- Understand how ethical and legal issues impact data collection and handling.
- Understand how good documentation supports data reuse.
- Identify appropriate storage and backup solutions.
- Become familiar with open science principles, understand the benefits and risks of data sharing and different levels of openness.
- Recognize the requirements for long-term preservation of research data.

Module 4: Research Data Management Support (2 ECTS)

Content

- Building, maintaining and developing RDM support services
- Skills and expertise in RDM
- Teaching RDM
- Assessing needs, service design

Learning outcomes

- Understand the role of RDM Support Services in different kinds of organizations and the role of data steward in this context
- Identify ways of developing RDM support services in an organization
- Identify different target groups for support services & implement a variety of services suitable for those groups
- Understand the role of the wider RDM network

Module 5: Project Work (2 ECTS)

- The participants can complete the project either individually or in a team
- The project work starts immediately during the first module, and time is reserved for its completion at the end of the training. Students receive guidance on doing project work.
- The project work topics should allow the practical application of the skills and knowledge gained from the training. Project work can be done in connection with participants own work tasks or, for example, development projects of your own work organization, so that the topic is applicable to the participant's own needs and thus serves the learning process in the best possible way.
- Topics can be e.g. Demo use of national data services together with a detailed description about the work, concept for training or training materials addressing different aspects of RDM, RDM needs assessment of an institution …

Training schedule

and the leading teachers

Module 1: Introduction to Research Data Management and Open Science

Leading teacher: Katja Fält, Tampere University

Dates:

- Monday, 3 March 2025, 9am to 3pm (in Tampere)
- Tuesday, 4 March 2025, 9am to 3pm (in Tampere)
- Friday, 31 March 2025, 12 noon to 3pm
- Saturday, 22 March 2025, 9am to 3pm

Module 2: Introduction to IT and Data Science

Leading teacher: Enrico Glerean, Aalto University

Dates:

- Friday, 4 April 2025, 12 noon to 3pm
- Saturday, 5 April 2025, 9am to 3pm
- Friday, 25 April 2025, 12 noon to 3pm
- Saturday, 26 April 2025, 9am to 3pm

Module 3: Research Data Life Cycle

Leading teachers: Mari Elisa Kuusniemi, University of Helsinki, Lucie Hradecká, Aalto University

Dates:

- Friday, 9 May 2025, 12 noon to 3pm
- Saturday, 10 May 2025, 9am to 3pm
- Friday, 23 May 2025, 12 noon to 3pm
- Saturday, 24 May 2025, 9am to 3pm
- Friday, 22 August 2025, 12 noon to 3pm
- Saturday, 23 August 2025, 9am to 3pm
- Friday, 5 September 2025, 12 noon to 3pm
- Saturday, 6 September 2025, 9am to 3pm

Module 4: Research Data Management Support

Leading teacher: Manna Satama, University of Eastern Finland

Dates:

- Friday, 26 September 2025, 12 noon to 3pm
- Saturday, 27 September 2025, 9am to 3pm

Module 5: Project work

- Introduction to project work: Friday, 10 October 2025, 12 noon to 3pm

How to apply?

- Fill in the electronic application and include a motivation letter. https://www.lyyti.in/DataSteward2025
- Application period: 1.10.2024-15.1.2025
- In the motivation letter (maximum one A4) you should explain in more detail why you are applying for the training. In the motivation letter, it is recommended to describe
  - how you will benefit from the training
  - how you can use the benefits of the course in your own work
  - your plans to bring your domain expertise to the role of Data Steward and
  - your plans to develop your data management skills
  - what kind of future plans and employment plans do you have.
- Admission: 20 participants are selected for the course.
- If there are many applicants, doctoral researchers will be prioritized. In selecting participants, emphasis is placed on the need for training highlighted in the motivation letter.

Training site: https://www.tuni.fi/en/study-with-us/tautree-data-steward

Teemu Rauhala

Key Account Manager

Continuous Learning Services

teemu.rauhala@tuni.fi

phone number+358503232083

Päivi Malinen

Information Specialist, Research Data Services

Tampere University Library

paivi.malinen@tuni.fi

phone number+358 50 318 3408