

Data Steward - training: Developing expertise for research support professionals

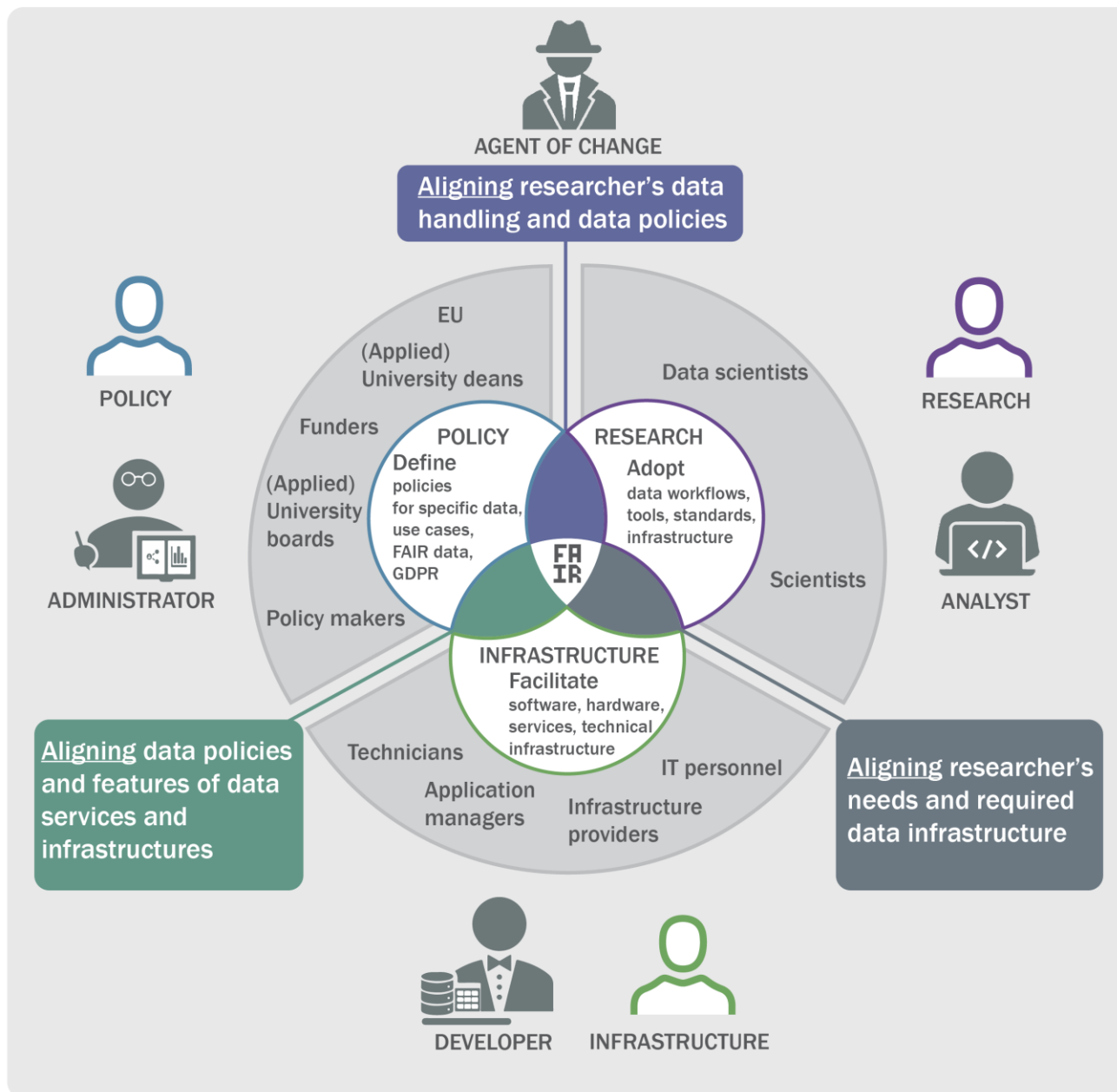
Päivi Malinen, Susanna Nykyri,
Teemu Rauhala, TAU



Mitä ovat datanhallinnan asiantuntijat?

Data steward – who are they?

- Datanhallinnan asiantuntija (Data steward) yleistermi lukuisille eri tukitoiminnoissa ja rooleissa toimiville henkilöille, joiden työtehtävät liittyvät tutkimusdatan luomiseen, ylläpitoon ja käyttöön. Ydinvastuut ja -tehtävät vaihtelevat datanhallinnan eri osa-alueiden neuvonnasta, avoimen tieteen ja datanhallinnan linjausten jalkauttamisesta teknisiin tehtäviin. Tehtävät vaihtelevat myös eri korkeakoulujen ja tutkimuslaitosten välillä.
- Datanhallinnan asiantuntija on henkilö, joka on vastuussa (tutkimus)datan ja metadatan laadun, eheyden ja pääsyjärjestelyjen ylläpitämisestä tavalla, joka on soveltuvan lain ja organisaation linjausten mukainen. Data stewardin tehtävänä on taata, että dataa käsitellään asianmukaisesti kaikissa tutkimuksen vaiheissa. (Jetten et al. 2021)
- 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)



National Forum for Research Data Management- Denmark



Four roles of a data steward

ZonMw/ELIXIR project- The Netherlands



Three roles of a data steward

Mitä osaamista tarvitaan? What competencies are needed?



- Datanhallinnan asiantuntijatehtävät ovat monipuolisia ja vaativat laajaa erityisosaamista.
- Datanhallinnan asiantuntijana toimiminen edellyttää muun muassa kykyä hahmottaa eri tutkimus- tai kehittämisprojektin datanhallinnan tarpeet ja kehittää kestävät ja vastuulliset toimintatavat niiden toteuttamiseksi.
- Erityisesti tietosuoja ja tietoturva edellyttävät syvällistä ymmärrystä paitsi voimassa olevista säännöksistä myös alan parhaista käytännöistä.
- 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 projects and develop sustainable and responsible operating methods for their implementation.

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 protection and data security, in particular, require an in-depth understanding of current regulations, and of best practices in the field.

Mihin tarvitaan datanhallinnan asiantuntijoita?

Why do we need data stewards?

- Kun tutkimusdataa käsitellään asianmukaisesti, tutkijoilla on huomattavasti enemmän aikaa tehdä itse tutkimusta/ 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.”*- Barend Mons*
- Monet tutkimusorganisaatiot ovat alkaneet nähdä, että tiedon jakamatta jättämisestä aiheutuu merkittäviä kustannuksia. Hyvä datanhallinta tarjoaa erinomaisen tuoton siihen sijoitettuun panokseen nähden. / 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 research but not data stewardship”* - Barend Mons*

*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>

Mihin tarvitaan datanhallinnan asiantuntijoita?

Why do we need data stewards?

- Jotta vastuullisesta avoimesta tieteestä tulisi normi, tutkimusta tekevien organisaatioiden, rahoittajien ja poliittisten päättäjien tulee olla varmoja, että he voivat luottaa siihen, että tutkimusprojekteilla on riittävää datanhallinnan osaamista.
- Tutkimushankkeet, joissa on panostettu datanhallintaan →
 - Tutkimus suoritetaan kustannustehokkaasti
 - Mahdollistetaan tutkimustulosten jatkokäyttö ja hyödynnettävyys
- Datanhallinnan ja asiantuntijan kustannuksia voidaan sisällyttää projektin kustannuksiin → hyvä ottaa asia huomioon jo rahoitushakemusta tehtäessä
- For responsible open science to become the norm, research organizations, funders, and policymakers must be confident that they can trust that research projects have sufficient data management expertise.
- Research projects that focus on data management →
 - Research is carried out cost-effectively
 - Further use and usability of research results are enabled
- Data management and expert costs can be included in the project costs → it is a good idea to take this into account when applying for funding

Data steward –koulutus / training

- Tampereen yliopisto sai 12/23 OKM:n kolmivuotisen tieteen edistämisen erityisrahoituksen hankkeelle ”Datanhallinnan asiantuntijoiden ammattiin pätevöittävän koulutuksen suunnittelu ja toteutus”.
- Järjestämällä koulutusta datanhallinnan asiantuntijoille nopeutetaan alan kehitystä ja vahvistetaan TKI-toiminnan laatua ja vaikuttavuutta.
- Koulutuskokonaisuuden kehittäminen vastaa kasvavaan tarpeeseen varmistaa, että datanhallinnan alalla toimivilla on vahva teoreettinen pohja sekä käytännön osaaminen vastaamaan työelämän tarpeisiin.
- Tampere University received a three-year special funding for the promotion of science from the Ministry of Education and Culture for the project “Design and implementation of training for data management experts to qualify for the profession”.
- By organizing training for data management experts, the development of the field will be accelerated and the quality and effectiveness of RDI activities will be strengthened.
- The development of the training responds to the growing need to ensure that those working in the field of data management have a strong theoretical foundation and practical skills to meet the needs of working life.

Koulutuksesta / About the training 1/3

- Koulutuksen nimi: Data Steward training
- Ensimmäinen toteutus 3/2025 – 2/2026 ja toinen 3/2026-2/2027 (pilotin aikana täsmennetään koulutuksen pysyvämpi jatkumo)
- Haku koulutuksen 1. toteutukseen alkoi syksyllä 2024 (2. toteutuksen haku alkaa syksyllä 2025). Haku oli kansainvälinen, eli kuka tahansa ja mistä tahansa maasta voi hakea koulutukseen. Itse koulutus alkoi 3/2025 Kick-off päivillä Tampereella (seuraava toteutus alkaa 3/2026). Muutoin opetus online (Teams/Zoom)
- Koulutuksen aloitti 16 osallistujaa eri taustoilla: datanhallinnan ja tutkimushallinnon asiantuntijoita ammattikorkeakouluista ja yliopistoista, tohtoriväitöskirjan tekijöitä sekä post-doc vaiheen tutkijoita
- Name of the training: Data Steward training
- First implementation 3/2025 – 2/2026 and second 3/2026-2/2027 (a more permanent continuation of the training will be specified during the pilot period)
- Application for the 1st implementation of the training began in Autumn 2024 (application for the 2nd implementation begins in Autumn 2025). The application was international, meaning anyone from any country can apply for the training. The training itself began in 3/2025 with the Kick-off days in Tampere (the next implementation begins in 3/2026). Otherwise, teaching is online (Teams/Zoom)
- The training began with 16 participants with different backgrounds: data management and research administration experts from universities of applied sciences and universities, doctoral dissertation authors and post-doc researchers

Koulutuksesta / About the training 2/3

- Koulutus laajuudeltaan 15 op
- Koulutuksen oppimisalustana Moodle
- Materiaalit tallennettu Zenodoon: [Search Data Steward -training \(Tampere University\)](#)
- Kustakin moduulista opiskelijat saavat osaamismerkkin, kokonaisuudesta kokoavan osaamismerkkin sekä sertifikaatin
- Koulutus sopii suoritettavaksi töiden ohella. Verkko-opetusta noin 4 pvä/kk + tehtävät (itseopiskelua + ryhmätöitä).
- Koulutuksen kieli on englanti
- Koulutus maksaa 1500 euroa/osallistuja + matka- ja majoituskulut Tampereella kurssin alussa
- Koulutuksen yhtenä tavoitteena on laajentaa tulevaisuuden työllistymismahdollisuuksia väitöskirjatutkijoille
- Scope of the training 15 ECTS
- The training platform is Moodle
- Materials stored in Zenodo: [Search Data Steward -training \(Tampere University\)](#)
- For each module, students receive a competency badge, a comprehensive competency badge and a certificate
- The training is suitable for completion alongside work. Online teaching approximately 2 days/month + assignments (self-study + group work).
- The language of the training is English
- The training costs 1500 euros/participant + travel and accommodation costs in Tampere at the beginning of the course
- One of the goals of the training is to expand future employment opportunities for doctoral researchers

Koulutuksesta / About the training 3/3

- Koulutus suunnitellaan ja toteutetaan laajassa yhteistyössä
- Koulutusta koordinoi Tampereen yliopisto, apuna neuvottelukunta, jossa jäseniä 4 yliopistosta (Aalto, HY, TY, UEF) 1 tutkimusorganisaatiosta (LuKe), 2 tutkimusta tukevasta organisaatiosta (CSC ja FSD) ja 1 ammattikorkeakoulusta (Metropolia)
- The training is planned and implemented in broad cooperation
- The training is coordinated by the University of Tampere, assisted by an advisory board with members from 4 universities (Aalto, HY, TY, UEF), 1 research organization (LuKe), 2 research support organizations (CSC and FSD) and 1 university of applied sciences (Metropolia)

Koulutuksen sisältö / Training content

- **Module 1: Introduction to research data management and open science (2 ECTS)**
 - Role and tasks of data stewards, Processing Personal Data in Scientific Research, Good scientific practice and ethical principles, Ethical review, ethical aspects in different data (social media) AI and ethics and Legislation impacting RDM, IPR issues connected to data, ownership issues, RDM agreements
- **Module 2: Introduction to IT and data science (3 ECTS)**
 - Introduction to Data Science and Data-Driven Research, Hardware and Software for Data Science and Basic programming concepts and advanced programming concepts: FAIR Software, Reproducibility, and other advanced tools
- **Module 3: Research Data Life Cycle (6 ECTS)**
 - Data management planning, Data documentation, Supporting data documentation, Data sharing & preserving and Domain-specific characteristics of research data
- **Module 4: Research data management support (2 ECTS)**
 - Skills and expertise in RDM, Teaching RDM, Building and developing RDM support services and Sharing knowledge/networks
- **Module 5: Project work (2 ECTS)**

Koulutuksen sisältö / Training content

Module 1: Introduction to Research Data Management and Open Science

- 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
- Recognise the role of data stewards and other data (+methods) support personnel in research organisations
- Identify the key ethical principles guiding research
- Recognise the basic legal aspects in research data management

Module 3: Research Data Life Cycle

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

Module 2: Introduction to IT and Data Science

- Recognise and explain fundamental data science concepts, including data preprocessing, analysis, and visualisation.
- Identify and utilise appropriate hardware and software for data science tasks, including storage solutions, secure access, and computational resources.
- Create and execute Python and R scripts while understanding the importance of reproducibility and the FAIR principles in research software development.
- Recognise advanced computational research tools, such as version control systems, containers, and high-performance computing.

Module 4: Research Data Management Support

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

Lisätietoja / Further information

- The training's own websites: [Data Steward | Tampere universities](#)
- The project's websites: <https://www.tuni.fi/fi/tutkimus/datanhallinnan-asiantuntijan-koulutushanke>
- [Advancing RDM Careers: A Framework for Expert Education in Finland \(zenodo.org\)](#)



Keskustelupöydät

Discussion tables

3 kpl 15 min keskustelupöytää

3 15-minute discussion tables

Pöytä 1 / Table 1

Perustutkintokoulutuksen kehittäminen

Aihe 1: Mitkä ovat ne tieteenaalat, joihin Data Steward -koulutus voitaisiin integroida?

Aihe 2: Opiskelijoiden valmiudet ja tarpeet

- Mitä valmiuksia ja taitoja perustutkinto-opiskelijoilla tulisi olla ennen Data Steward -koulutuksen aloittamista?
- Miten koulutus voi vastata opiskelijoiden ja työelämän tarpeisiin?

Developing Undergraduate Education

Topic 1: What are the disciplines into which Data Steward training could be integrated?

Topic 2: Student capabilities and needs

- What capabilities and skills should undergraduate students have before starting Data Steward training?
- How can education meet the needs of students and working life?

Pöytä 2 / Table 2

Tohtorikoulutuksen ja jatko-opintojen kehittäminen

Aihe 1: Data Steward -koulutuksen rooli tohtorikoulutuksessa

Aihe 2: Yhteistyö ja verkostoituminen: yhteistyöverkostot eri yliopistojen ja tutkimusorganisaatioiden välillä.

Developing doctoral education and postgraduate studies

Topic 1: The role of Data Steward education in doctoral education

Topic 2: Collaboration and networking: collaborative networks between different universities and research organizations.

Pöytä 3 / Table 3

Jatkuvan oppimisen kehittäminen

Aihe 1: Data Steward -koulutuksen integrointi jatkuvan oppimisen ohjelmiin

Aihe 2: Koulutuksen markkinointi ja tunnettuuden lisääminen

Developing Continuous Learning

Topic 1: Integrating Data Steward training into continuous learning programs

Topic 2: Marketing and raising awareness of the training

Kiitos 😊!