

FAIR data stewardship

The need for capacity building & the role of communities

Mijke Jetten, Community Manager Data Stewardship
DTL & Health-RI



Today's presentation

Link to presentation

<https://doi.org/10.5281/zenodo.7188375>

Content

- Challenges in FAIR data stewardship
- The importance of FAIR data stewardship capacity building
- The role of communities

After this presentation

- You are able to position yourself in the **data steward landscape**
- You are aware of **resources** on the required knowledge, skills and abilities
- You understand how **communities** may help becoming a better data steward
- You might even be inspired to start/participate in similar communities yourself

Our focus today: data stewardship



- **Data stewardship:** Responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data
- Data stewardship and data management skills are **essential in research**
 - The lack of consensus on the responsibilities, knowledge, and skills of data stewards hampers building adequate data steward capacity
- **Professionalisation of data stewardship is needed!**
 - (Inter)national alignment and coordination are needed to achieve coherent training/education, accompanied by a consistent human resource (HR)

Role	Task	FTEs needed per 1000 researchers
Data Steward	Assisting researchers with effective management of research data	26
Trainer on data stewardship	Training researchers on data management skills	4

WORLD VIEW · 25 FEBRUARY 2020

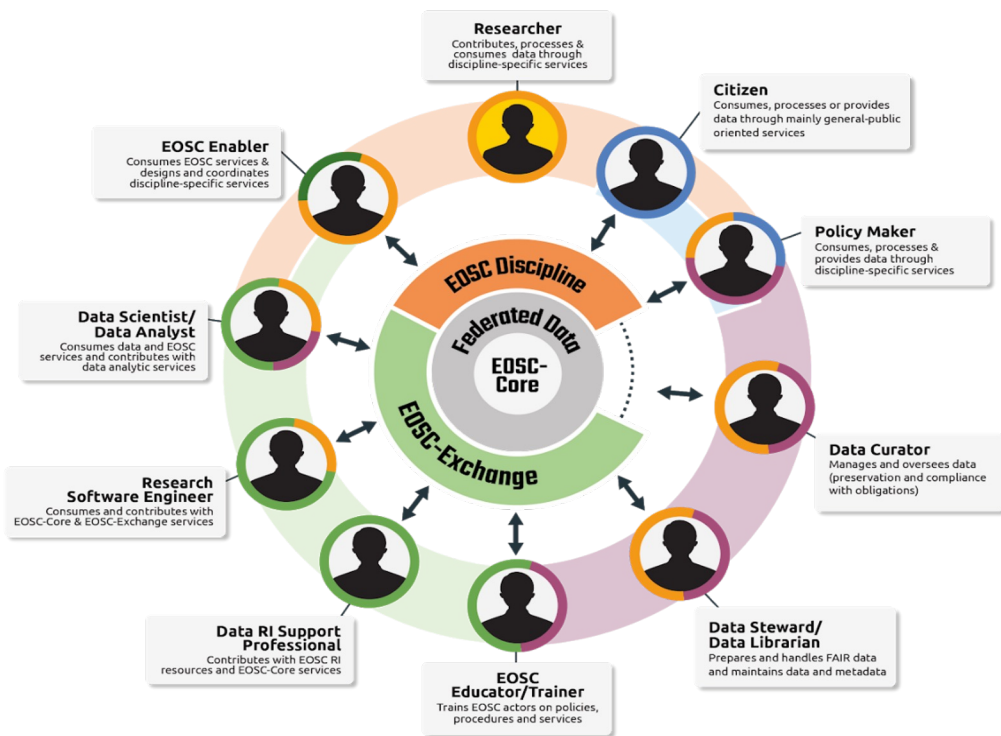
Invest 5% of research funds in ensuring data are reusable



It is irresponsible to support research but not data stewardship, says Barend Mons.

Barend Mons ^{ORCID}

<https://doi.org/10.1038/d41586-020-00505-7>



Description of 10 roles within the EOSC ecosystem, one situational example for each role and a list of required skills

	Researcher
Overview	The researcher is the main target of the EOSC ecosystem and interacts with it to obtain, process, produce, deposit and share research data, using mainly high-level services provided by the ecosystem.
Examples	A researcher would browse and identify data related to different genetic variants of the flu available in a FAIR data repository and perform a phylogenetic study on samples through a service from the EOSC marketplace, creating a graphical representation as a phylogenetic tree to include in an article, referencing the data sources and processing pipelines.
Required skills	<ul style="list-style-type: none"> • General knowledge on the EOSC ecosystem, covering the EOSC-Core and EOSC-Exchange interdisciplinary services for data access, sharing, reuse and processing, and relevant discipline-specific services. • Knowledge of the added value that EOSC services provide to research and publication workflows. • Understanding of how to assess the EAI of services when searching for and producing research data. • Skills to apply EOSC services wherever they support the research and innovation lifecycle and contribute to their development. • Training and communication skills to teach and educate other researchers and students on how to conduct research in the frame provided by EOSC.

	Data Steward/Data Librarian
Overview	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.
Examples	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.
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 fulfillment 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.

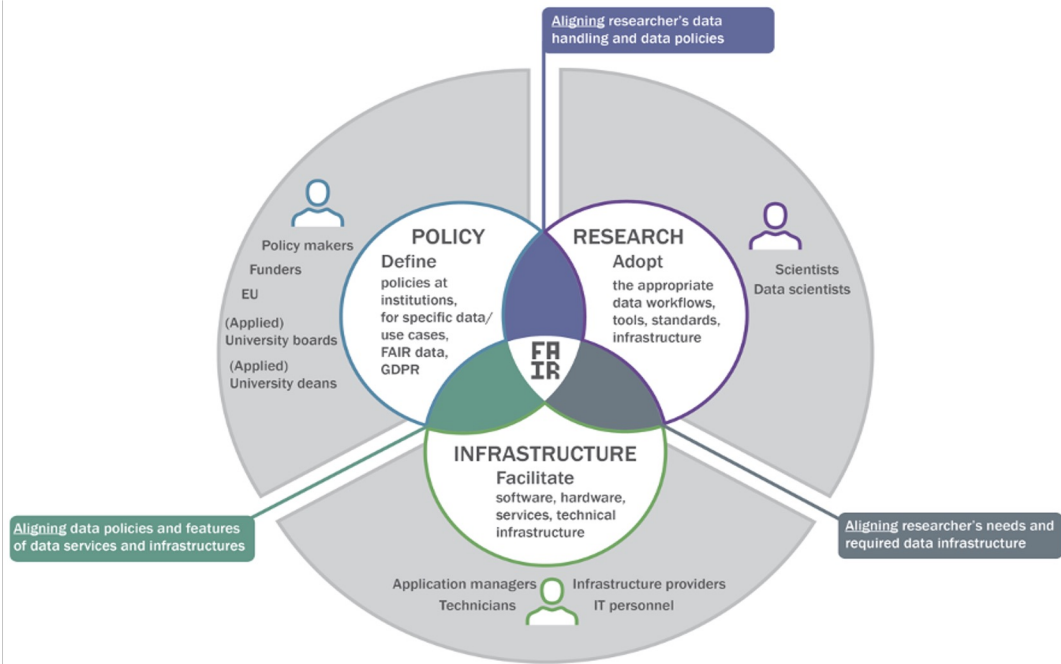
<https://www.eoscsecretariat.eu/news-opinion/digital-skills-fair-open-science-report-eosc-skills-training-working-group>



LCRDM data stewardship task areas

<https://doi.org/10.5281/zenodo.2669150>

<https://doi.org/10.5281/zenodo.3066366>



ZonMw/ELIXIR data stewardship roles in the data stewardship landscape

<http://doi.org/10.5281/zenodo.3474789>

<https://doi.org/10.5281/zenodo.3243909>

Competency frameworks

- NPOS/ELIXIR Data stewardship Competency Framework Responsibilities, activities, KSAs and learning objectives for three data steward roles

<https://competency.ebi.ac.uk>

- FAIR4S framework EOSCPilot D7.5

<https://eoscpilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities>

- Frameworks discussed in FAIRsFAIR D7.2 Briefing on FAIR Competences and Synergies

<https://doi.org/10.5281/zenodo.4009006>



Who is the Competency Hub for?

Students and professionals in computational biology

Explore the ISCB competencies and:

- Career profiles

Research infrastructure scientists

Explore the ARISE competencies

Data steward: policy, research and infrastructure

Explore the NPOS/ELIXIR competencies

Data steward: policy, research and infrastructure

Data Steward

1.0 **LIVE**

The NPOS/ELIXIR Data Stewardship Competency Framework distinguishes three data stewardship roles (policy/strategy, compliance, alignment with FAIR data principle archiving and transferable skills). For each of these three data stewardship role outcomes (including Bloom's level) are given.

Competencies Export

Filter competencies

Policy and Strategy: Development, implementation management policy and strategy of the institute

▼ [Policy Oriented Data Steward] Responsible for advice on and development, im policy and strategy for the institute, which includes the complete research data lif with the relevant stakeholders and within financial and legal constraints, within th basis for (project) data management plans

Activities and tasks

- Develops, implements and monitors the institute's RDM policy
- Advises the institute's management on short- and long-term actions to advanc
- Assesses and monitors the institute's time and financial investments in relator
- Explores new needs, opportunities and trends in RDM

Knowledge, skills and abilities (KSAs)

- Knowledge about the most important elements of a successful RDM policy
- Knowledge about internal policies and financial and legal constraints within the
- Knowledge about relevant internal and external (funder, publishers, governmer

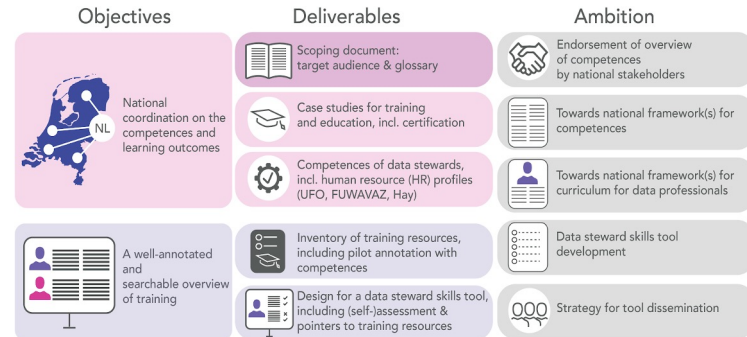
Professionalising data stewardship (NPOS-F)

- A collaborative effort of over 30 representatives of universities, university medical centres (UMCs), universities of applied sciences (UASs), service providers, and representatives of the major Dutch umbrella organisations
- Recommendations will be taken further by the Dutch NPOS 2021-2030 FAIR Data Programme
- For a quick overview, read the preamble, executive summary and Chapter 7 of the end report:

Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship. Mijke Jetten, Marjan Grootveld, Annemie Mordant, Mascha Jansen, Margreet Bloemers, Margriet Miedema, & Celia W.G. van Gelder. (2021). <https://doi.org/10.5281/zenodo.4320504>



Professionalising data stewardship: competences, training and education



Job profiles

NPOS-F recommendations:

- Define data stewardship competences and formalise the job profiles via job classification systems
- Recognise and reward data stewards, secure their position and include development and remuneration

2022: The basic job components have been transformed into a **formal UFO (university) data steward profile**, which was formalized in August 2021. In the meantime, the UMC & UAS data steward profiles presented in the report, are already informally adopted, with the expectation to be included in one of the next updates of the system.



Figure 4.2 Basic job profile components of a data steward

<https://doi.org/10.5281/zenodo.4320504>

What does a data steward do?

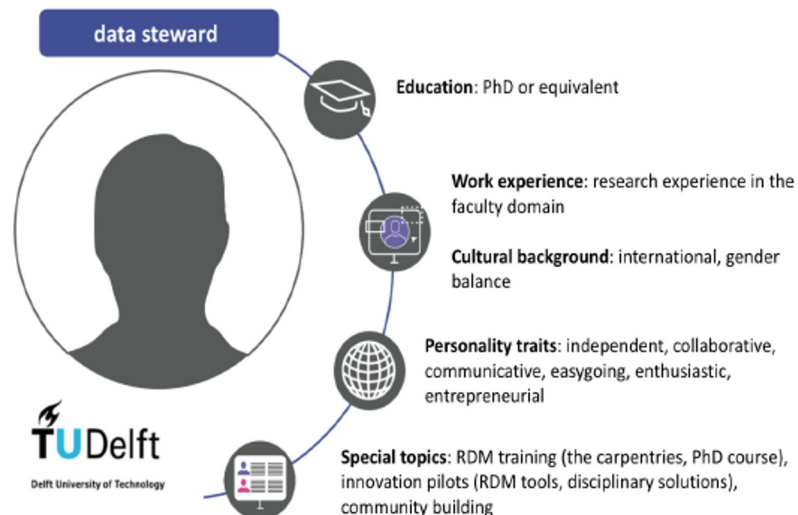


Figure 3.1 Delft University of Technology data steward

Data steward in the organisation



1. Appointment explicit part of RDM policy
2. Positioned at the research institute
3. Additional RDM support for centralised tasks at library
4. No formal central coordination, library is informal linking pin

Training for data stewards



1. Training organised by the coordinator
2. Formal RDM training
3. Training on TUD research support

Learning on the job



1. Structured mentoring
2. Internal peer support (data stewards team)
3. External peer networks
4. Soft skills development: training and on the job
5. Gaining relevant research expertise

Strengths and challenges



1. Strength: well-organised (coordination)
2. Strength: institutional buy-in
3. Strength: sustainable
4. Strength: team-feeling among the data stewards
5. Challenge: expensive (9 FTE)

Eight case studies (universities, UASs & UMCs)

Read these data stewards' full stories at <https://doi.org/10.5281/zenodo.4320504>

The role of communities

Four levels of community building

1. Local (universities, universities of applied sciences, university medical centres)
2. Disciplinary (for example Health-RI, Odissei, Clariah)
3. National (for example LCRDM, DSIG, GO FAIR)
4. International (for instance RDA, ELIXIR, EOSC)

Local communities

Research programmes > Implementation Plan Investments Digital Research Infrastructure

Implementation Plan Investments Digital Research Infrastructure

Digitalisation is an important development in all branches of science. On behalf of the Ministry of Education, Culture and Science (OCW), NWO is realising several activities in the area of digital research infrastructure. These investments are described in the 'The Implementation Plan Investments Digital Research Infrastructure'. Broadly speaking, NWO will use the ICT funding in two main areas: computer facilities and digitalisation.

Local DCCs

This call was a one-off stimulus for the setting up or further development of local Digital Competence Centers. Research institutions could use this funding to appoint data stewards and data managers for an existing DCC or the central setting up of a new DCC within the institution. With this call, NWO also wanted to ensure that the institution would safeguard the DCC concerned from both a policy and financial perspective.

<https://www.nwo.nl/en/researchprogrammes/implementation-plan-investments-digital-research-infrastructure>



<https://www.health-ri.nl/about-health-ri/organisation/nodes>

Community profile: TDCC-LSH [slide 1]



Harmonise good data access & stewardship

Develop the TDCC as data stewardship hub for the LSH domain, supporting data access across LSH stakeholders (and beyond)

Enhance Interoperability of digital solutions & resources

Develop the TDCC as "LSH interoperability-network", supporting the need to combine data, software and models across teams and organisations

Strengthen capacity & expertise base in digital research

Develop the TDCC as LSH community platform to strengthen the training & support network in digital LSH research, aligned with the international field



Community overview

TDCC-LSH is one of the three thematic DCCs that have been established by the scientific field and NWO (on behalf of the Ministry of Education, Culture and Science).

Aim: Strengthening digital competencies across Dutch science and harmonising digital practices across organisations and initiatives within the LSH domain.

URL: [Roadmap Thematic Competence Center, Thematic DCCs will strengthen digital competences in three science domains | NWO](#)

Keywords:

Connect knowledge & expertise, Harmonise collaboration, Facilitate expert communication, FAIR Data Stewardship, FAIR Implementation

Community profile: TDCC-LSH [slide 2]



How to get involved?
kimberley.zwiers@dtls.nl &
petra.aarnoutse@dtls.nl

Foreseen activities:

- Digital Life Sciences and Health Research Round tables
- Set up collaboration structure with DCC teams, with the other TDCCs and other stakeholders
- Continuous refreshment of the landscape analysis, e.g. every 2 years
- Facilitate the (co-) organisation of workshops, hackathons, focus meetings and conferences
- Communication and Outreach

The TDCC-LSH team:

- Programme Board: Ruben Kok, Gerrit Meijer
- Network Manager & Lead Training & Capacity Building: Celia van Gelder
- Lead FAIR Data Stewardship: Mijke Jetten
- Community Management:
 - For Health: Fieke Schoots
 - For Life Sciences: Kimberley Zwiers, Meike Bünger, Petra Aarnoutse,



Disciplinary → National → International communities



Making it easy for humans to make metadata for machines



Machine-actionable metadata are core to the FAIR Principles. GO FAIR and RDA members have launched the "Metadata for Machines" workshop series (M4M) to assess the state of metadata practices in data-related communities and stimulate the creation and re-use of FAIR metadata standards and machine-ready metadata templates (definitions of metadata

categories).

GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the **FAIR data principles**, making data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**). It offers an open and inclusive ecosystem for individuals, institutions and organisations working together through **Implementation Networks** (INs). The INs are active in three activity pillars: **GO CHANGE**, **GO TRAIN** and **GO BUILD**.



Here we provide information about open science, open access publications and FAIR data stewardship aiming at the research community that works on the current COVID-19 outbreak. The information is also a guidance for ZonMw's requirements and recommendations in its research programmes for COVID-19. It addresses the following research activities:

1. Transparency on research projects
2. Creating FAIR data
3. Access to research data
4. Sharing of research findings through open access pu



For more information visit: [Webinars for support & community building](#)

Support and community building

ZonMw's requirements and recommendations aim at converging the open science activities toward choices that the COVID-19 research community agrees on. The open science and FAIR data resources (standards, technologies and infrastructure) that are explained on this site are therefore a starting point.

To benefit from the latest developments, and to find out how these can be implemented in a research project, researchers can get support from their data steward and/or data support office that is available at most research institutes.

ZonMw collaborates with the [DSCC-implementation network](#), [Health-RI](#), and [LCRDM](#) to reach out to data stewards and data support offices to agree on the approach and resources to apply in COVID-19 research projects. We thereby build a research community with local data stewards and infectious disease experts at the research institutes to continue developing community specific open science and FAIR data practices.

<https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/webinars-for-support-community-building/>

National communities

- Data Stewards Interest Group (DSIG)

Community for data stewards to share experiences and foster the (Dutch) national implementation of data stewardship

You are kindly invited for the next DSIG meeting (see [webpage](#), [mailing list](#) and [slack community](#))

- Within Health-RI, a dedicated [Data Stewardship Community \(DSC\)](#) unites data stewards in the health domain

Data Stewardship Interest Group

A professional community for Data Stewards and alike in "Life Sciences"



Over 650 slack members



** UMC, University Utrecht, Lygature, SURF, The Hyve, Prinses Maxima Centrum
not depicted:
Ontoforce (Ghent, Belgium)



Data Stewardship Community

The Health-RI Data Stewardship Community (DSC) will establish a community hub for health data stewards to facilitate collaboration.



National Coordination Point Research Data Management

The National Coordination Point Research Data Management (LCRDM) is a national **network** of experts in the field of research data management (RDM).

The LCRDM forms the link between **policy and solution**. Close consultation between educational and research institutions is crucial for this. Within LCRDM, experts work together to put RDM subjects on the agenda that are too big for one institute to tackle and need a national plan of action.

LCRDM brings together research support services, policy makers, ICT specialists, managers of diverse research institutes and research funding organizations. The LCRDM coordinates and facilitates the collaboration between the various RDM stakeholders.



Pool of Experts



The LCRDM also organises **meetings** on RDM themes, such as November 3, 2020 in Maastricht on adoption and implementation, training and competences of data professionals, and RDM in times of Corona.

<https://www.lcrdm.nl/en>

RDM maillijst

Are you organizing a symposium, do you have a vacancy, do you want to discuss something with colleagues: use this list. Everyone involved in RDM is welcome. To



Do you have RDM questions that could benefit from a national approach? You will find our working method on the **Task groups** page.

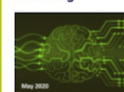


RDM in the Netherlands



Glossary

TG Software Archiving



Recommendations on encouraging of software archiving.

Pitch and team

TG FAIR enabling



Principles on the basis of which an organisation could be assessed on the degree of 'FAIR' enabling.

Pitch and team

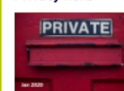
Digital consent



A legal and ethical framework for obtaining digitized informed consent.

Pitch and team

Privacy risks



Implement and maintain a DPIA per research scenario, so that a scenario can be selected for new research proposals.

Data Collaboration



Inventory of trans-institutional research projects (use cases), identification of bottlenecks and incompatible policy and description of implications and solutions.

Pitch and team

"23 Things" adoption



A wider adoption of the existing RDA (Research Data Alliance Europe) guide "23 Things - Libraries for Research Data".

Pitch and team

Anonymisation



√2019

Pseudonymisation



√2019

Data Curation



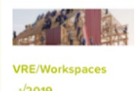
√2019

Task groups



The task groups are composed by the **advisory group** for each submitted issue. The team members are drawn from the **pool of experts** by means of a pitch, and possibly supplemented with experts. The task groups are ideally supervised by a process supervisor or consultant. The motivation of the task group members is: 'I also have to do something with my subject in my own working environment.'

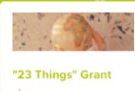
Read more about **pitching** new task groups.



VRE/Workspaces
√2019



Data Stewardship
√2019



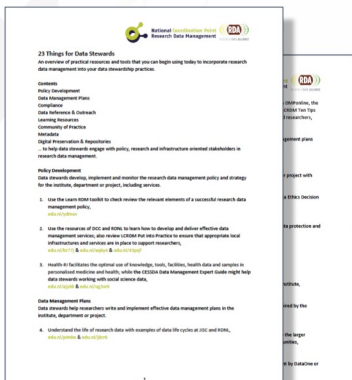
"23 Things" Grant
√2019



Working groups 2015
- 2017

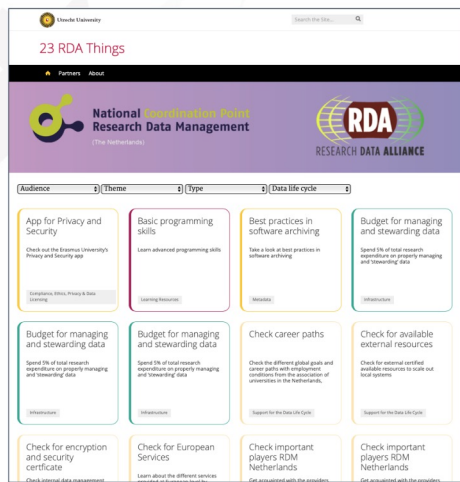
National → International communities

23 Things (a.o. for data stewards)



Digital sheets for training purposes

- Things for researchers & PhD candidates
- Things for Bachelor & Master students
- Things for data & subject librarians
- Things for data stewards
- Things for IT support staff & IT specialists
- Things for research software engineers
- Things for policy makers



Filter by:

- Audience (7 options)
- Theme (16 options)
- Type (12 options)
- Data life cycle (6 options)

<https://23things.sites.uu.nl> (beta)



<https://www.lcrdm.nl/en/23things>

<https://doi.org/10.5281/zenodo.3773662>

<https://23things.sites.uu.nl/>



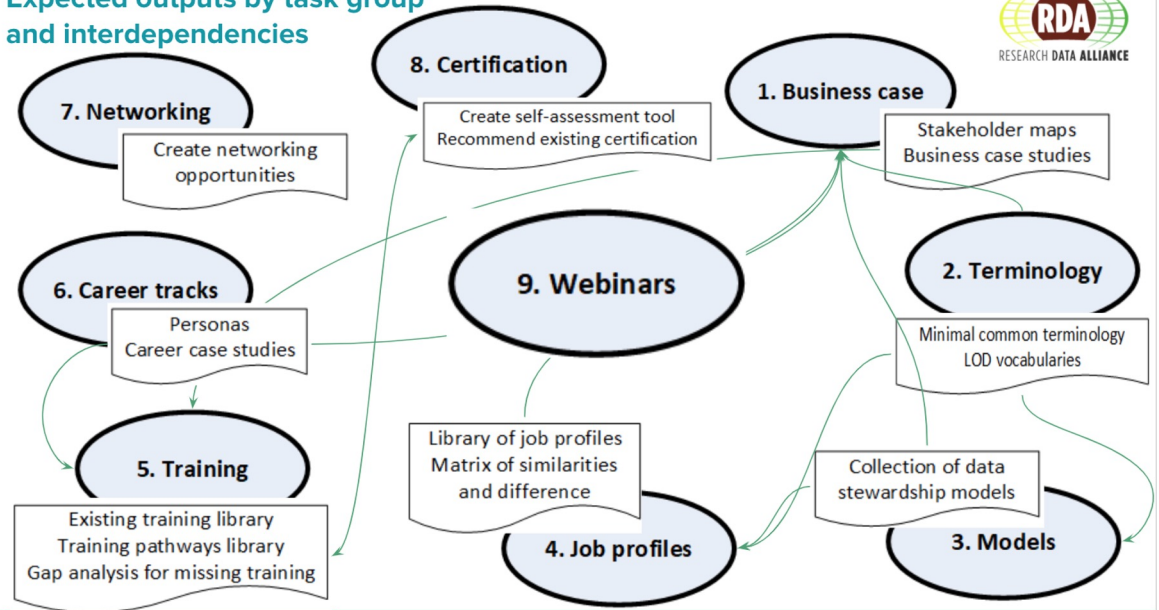
Landelijk Coördinatiepunt
Research Data Management



RDA Professionalising Data Stewardship IG

The screenshot shows the RDA Professionalising Data Stewardship IG webpage. At the top, there is a navigation bar with the RDA logo and 'RESEARCH DATA ALLIANCE'. Below this, there are sections for 'O&A Members' (63) and 'MEMBERSHIP' (Members: 10956). A 'Register now' button is visible. The main content area features a 'Professionalising Data Stewardship IG' header with a 'Taxonomy' dropdown. Below the header is a grid of icons for 'Posts', 'Create Wiki index', 'Events', 'Repository', 'Outputs', 'Charter', 'Plenaries', and 'Members'. A 'Group Status' section indicates 'Not yet endorsed'. At the bottom, there is a 'Status' section with 'In Group Revisions' and 'Chair (s): Mijke Jetten, Marta Teperek, Peter Neish'. A paragraph of text describes the IG's purpose and current activities.

Expected outputs by task group and interdependencies



The Research Data Management toolkit for Life Sciences

Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)

<https://rdmkit.elixir-europe.org/>



What can we help you find?

Browse all topics by



Data life cycle

Start here to get an overview of research data management based on stages in the data life cycle.



Your role

Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.



Your domain

Learn about data management tasks that affect your domain or research community, and the solutions adopted to address them.



Your tasks

Find guidelines and solutions for tackling common data management tasks.



Tool assembly

Find concrete combinations of tools and resources assembled into an ecosystem for research data management.



National resources

Find pointers to country specific information resources and national research data management practices.



All tools and resources

Browse the RDMkit's catalogue of tools and resources for research data management.



All training resources

Browse all training resources mentioned in RDMkit pages.



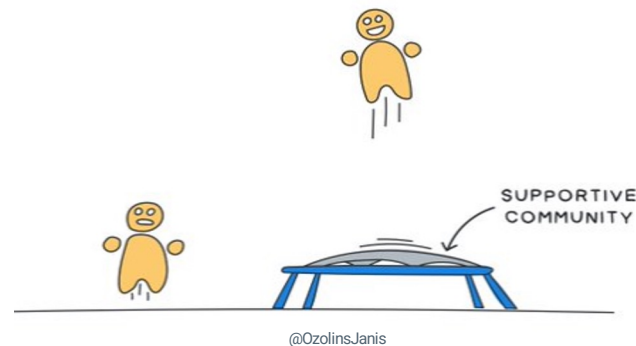
We welcome contributors!

This project would not be possible without the many **amazing community contributors**. RDMkit is an open community project, and you are welcome to join us!

[Start contributing](#)

Takeaways

- Creating FAIR data implies making well informed choices about
 - ... the number of data stewards
 - ... where in the organisation
 - ... and with what competences
 - ... including training
- Data stewardship is a new profession and things are not set in stone yet
 - There is a lot going on in Europe (and beyond) where you can build on, both for yourself as well as for your institute
 - Discuss your current/future roles, responsibilities and tasks in your local teams and organisations
 - Become part of the data stewards community!
- Recommended next steps
 - Competencies and skills: what roles do you have in the team, what additional roles do you need?
 - Capacity: identify capacity (fte) needed and work towards installing that capacity
 - Build your local data stewardship community
 - Look at the reports, tools and training resources in this presentation



Thank you for listening!

Interested to learn more about DTL,
Health-RI, ELIXIR-NL and NPOS
activities? Contact me via
mijke.jetten@dtls.nl

