Professionalising the data steward roles in the Netherlands

FAIR data stewardship capacity building & the role of communities

Mijke Jetten, Community Manager Data Stewardship
DTL & Health-RI, mijke.jetten@dtls.nl

DCC RDMF21, July 12, 2021

Outline

- The need for skills and capacity building in data stewardship
  - How much capacity is needed?
  - Data stewardship roles and competences

- The Netherlands
  - Dutch roadmap towards professionalising data stewardship
  - Data stewardship training
  - Data stewardship community building

- International initiatives

Link to slide deck: https://doi.org/10.5281/zenodo.5091300

CC BY

Part 1

The need for skills & capacity building in data stewardship

CC BY

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

health RI enabling data driven health

elixir NETHERLANDS

nationaal programma open science

The need for skills and capacity building in data stewardship

Data stewardship and data management skills are essential in research.
Challenges:

- There is no consensus on the **responsibilities** and **tasks** of data stewards, and **formal profiles** (including knowledge, skills and abilities (KSAs)) are lacking
- No/too little and/or not findable tailored **education & training**
- This hampers the **adequate capacity building** and complicates efficient data management which are necessary for open science implementation
- (Inter)national **alignment and coordination** are needed to achieve coherent training and education, accompanied with a consistent human resource (HR) policy

In the Netherlands, by **bundling expertise and manpower**, supported by the **research funders**, and **embedded in (inter)national initiatives**, we are making steps to tackle FAIR data handling challenges

- Joint capacity building efforts will help build a skilled workforce that is ready for professional data handling
- Our approach: by the community, for the community

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

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

CC BY

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES | health RI enabling data driven health | elixir NETHERLANDS | nationaal programma open science

How much capacity is needed?

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

nationaal programma open science

That is 3 FTE per 100 researchers. Number taken from OECD report

The EC High Level Expert Group EOSC estimates 5 FTE per 100 researchers

Landelijk Coördinatiepunt Research Data Management

Results of LCRDM Do I PASS for FAIR 2020 survey

CC BY

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

health RI enabling data driven health

elixir NETHERLANDS

nationaal programma open science

Data stewardship roles and competences

NPOS/ELIXIR/ZonMw competency framework:
responsibilities, activities, KSAs and learning objectives of data steward roles

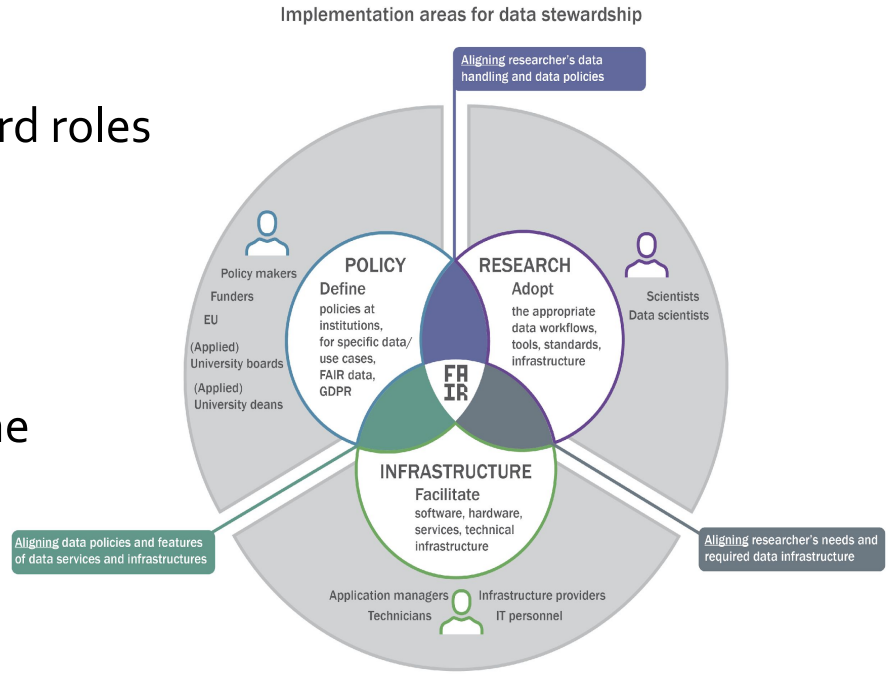
- **Data steward policy**: institute and policy focused
- **Data steward research**: project and research focused
- **Data steward infrastructure**: data & e-infrastructure focussed

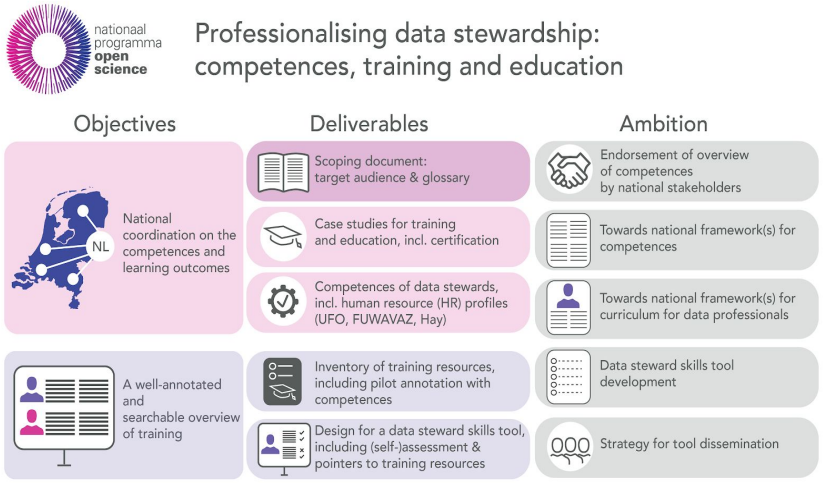
ZonMw report (2019) "Towards FAIR data steward as profession for the life sciences", with matrices and a Zenodo archive with output

- Implemented in the EBI Competency Hub

Work continued in the Dutch National Programme Open Science (NPOS) and the NPOS F professionalising data stewardship report

- A collaborative effort of over 30 representatives of universities, UMCs, UASs, service providers, and representatives of the major Dutch umbrella organisations





CC BY
DTL DUTCH TECHCENTRE FOR LIFE SCIENCES
health RI enabling data driven health
elixir NETHERLANDS
nationaal programma open science

Part 2

The Netherlands

CC BY

Dutch roadmap towards professionalising data stewardship

Creating FAIR data implies

Making well informed choices about

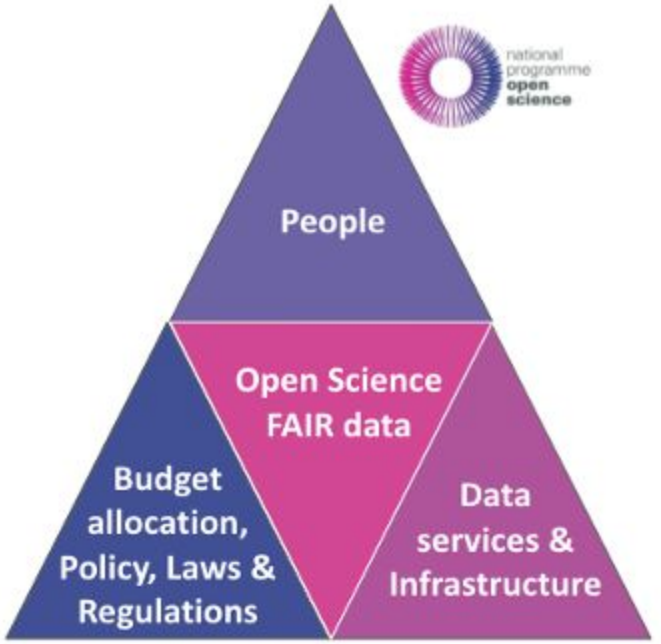
- … the number of data stewards
- … where in the organisation
- … and with what competences
- … including training

Realising data steward capacity implies

- Changes in research-performing organisations, including HR management
- National funds for science and institutional budgets
- Coordinated action of research institutes, policy makers and research-funding organisations for the required changes

Dutch National Programme Open Science (NPOS)

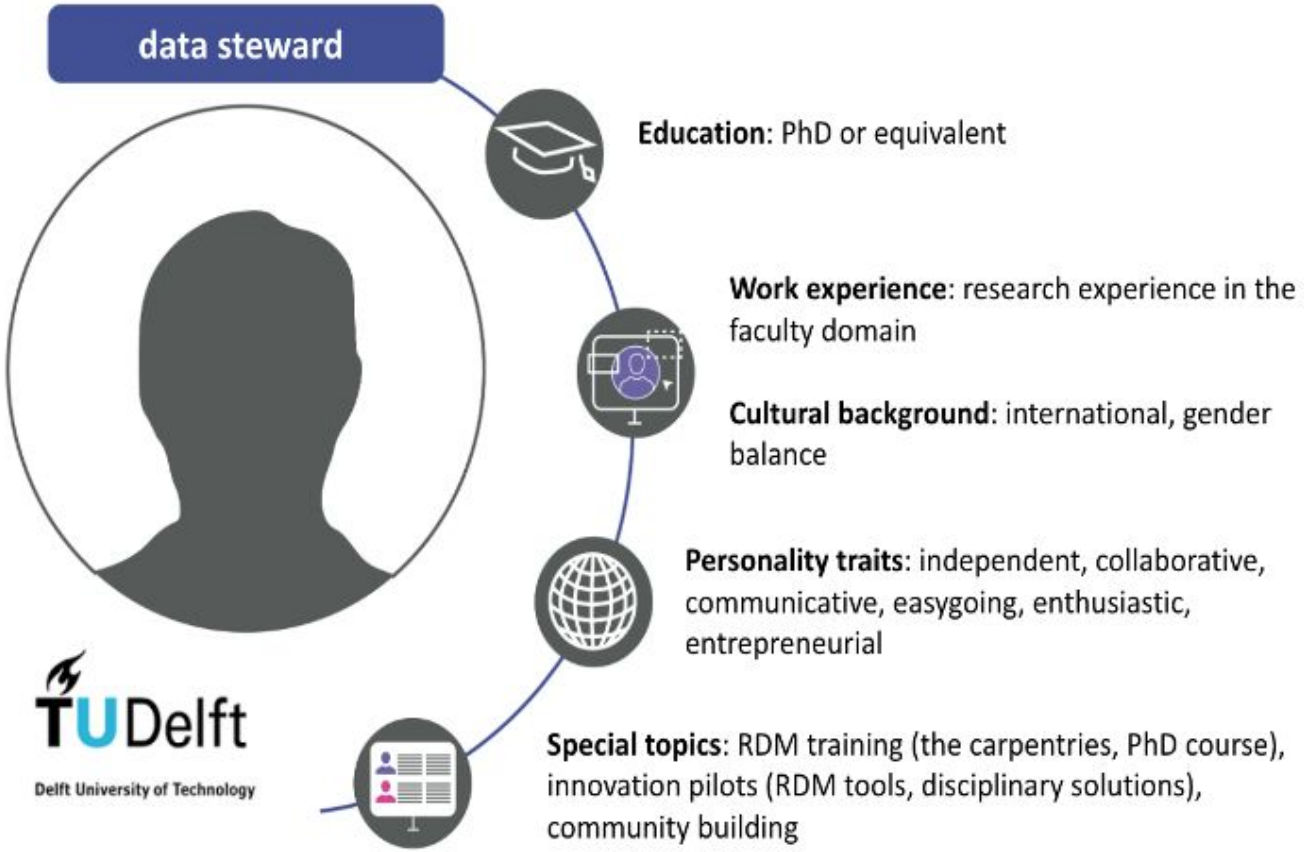
- Three key areas: 100% Open Access publishing, optimal reuse (FAIR) of research data, and corresponding evaluation and valuation systems
- The data stewardship report (NPOS F) links to the second key area, together with a report on the Dutch data infrastructure and services landscape (NPOS E)
- NPOS F: for a quick overview, we advise to read the preambule, executive summary and Chapter 7

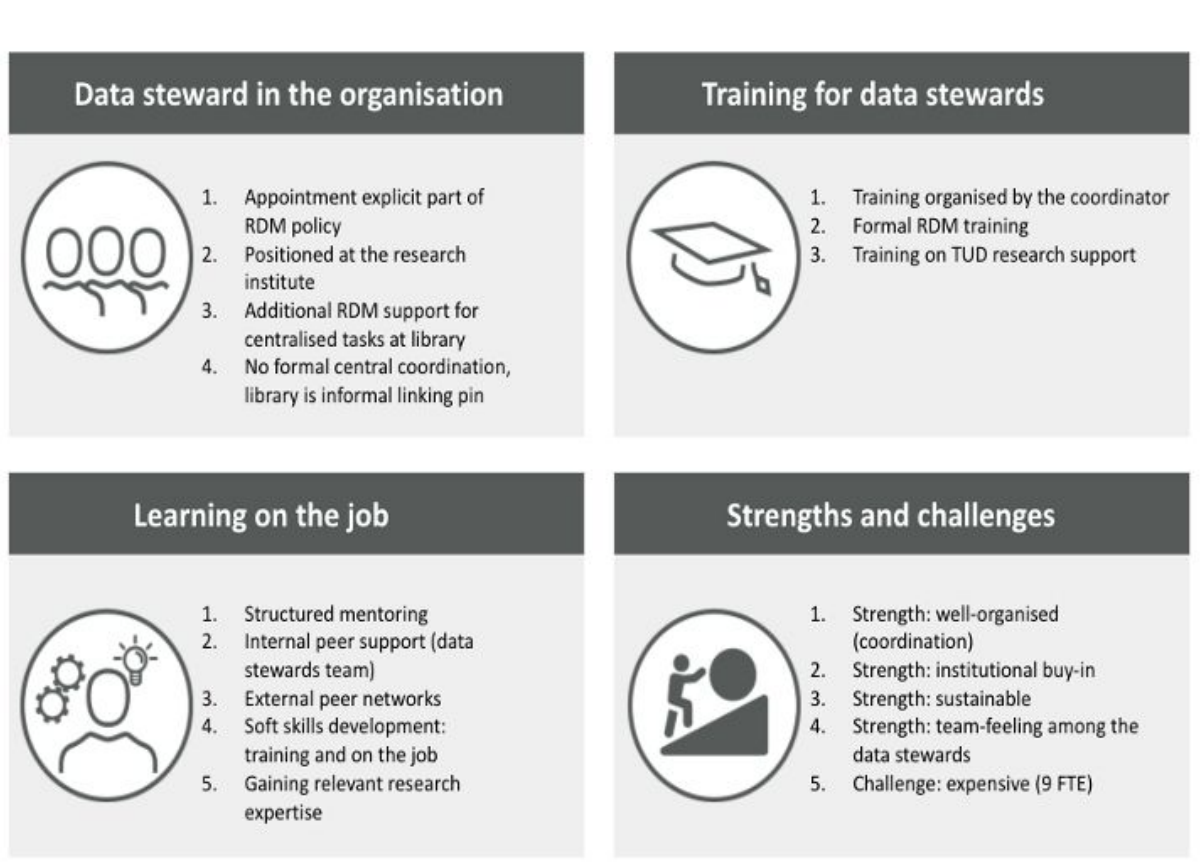


CC BY
DTL DUTCH TECHCENTRE FOR LIFE SCIENCES
health RI enabling data driven health
elixir NETHERLANDS
nationaal programma open science

NPOS F: Landscape analysis on training and education

8 case studies (universities, UASs & UMCs) as a first fact check on data stewardship training and education in the Netherlands. Example case study: Delft University of Technology





CC BY
DTL DUTCH TECHCENTRE FOR LIFE SCIENCES
health RI enabling data driven health
elixir NETHERLANDS
nationaal programma open science

NPOS F: Basic data steward job profile components

Various annexes with detailed information

- Domain areas, responsibilities, tasks, and competences of a data steward
- Basic components of the data steward job profile - for the UFO (universities) job classification system, as a first step
- Three local university data steward job profiles for the FUWAVAZ (UMCs) job classification system
- Proposal for a job profile for data stewards in UASs
- Context of the Digital Competence Centers: additionally, the basic components of the research software engineer job profile

CC BY
DTL DUTCH TECHCENTRE FOR LIFE SCIENCES
health RI enabling data driven health
elixir NETHERLANDS
nationaal programma open science

NPOS F: Training, education and certification

- Inventory of training resources and pilot annotation of courses based on competences
- Inventory of existing certification mechanism, for different categories: courses, trainees, trainers, and organisations
- Certification for data stewardship is still in its early days and needs to be done in alignment with similar activities in Europe and beyond

NPOS F: Data stewards skills tool

- Sketch for a data steward skills tool as a single point of reference for data steward competences, training and education
- 5 data steward personas, to show how stewards could benefit from the tool, and possible pathways for these personas in such a tool
- In collaboration with ELIXIR, as a pilot, content was added to the EBI Competency Hub: expertise areas, responsibilities, tasks, KSAs, and learning objectives. The Competency Hub is interested to expand their tool for our use case and discuss adaptations needed

CC BY

Data stewardship training

Helis Academy course FAIR data stewardship (for life sciences), materials available for reuse

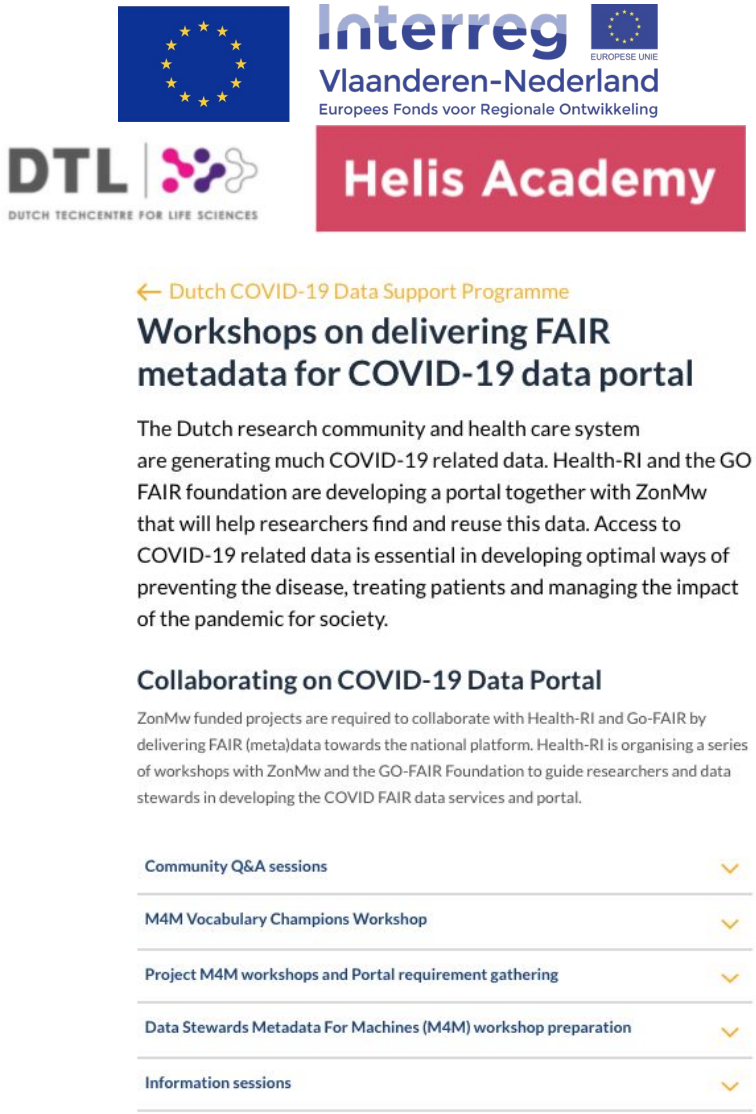
- First edition ('19)
- Second edition ('19)
- Third edition ('21)

ELIXIR-NL Technical Data Stewardship Course 2019

Data stewardship training for Marie Curie Early Career Researchers

- COSMIC (materials) ('20)
- ProEVLifeCycle (materials) ('20)

Various national data stewardship webinars and metadata workshops were given for data stewards and project leads of the ZonMw granted Covid-19 projects, as a collaborative action of GO FAIR, Health-RI and DTL, as part of the ZonMw FAIR data requirements



CC BY
DTL DUTCH TECHCENTRE FOR LIFE SCIENCES
health RI enabling data driven health
elixir NETHERLANDS
nationaal programma open science

2021 DCC Spring Training Days

May 20 — **Engagement: connecting researchers and data stewards**

June 3 — **Organising your data and software with a reproducible workflow**

June 17 — **The role of good RDM in accelerating scientific progress**

from 13:00-17:00 (CEST)
via Zoom

Local Digital Competence Centres

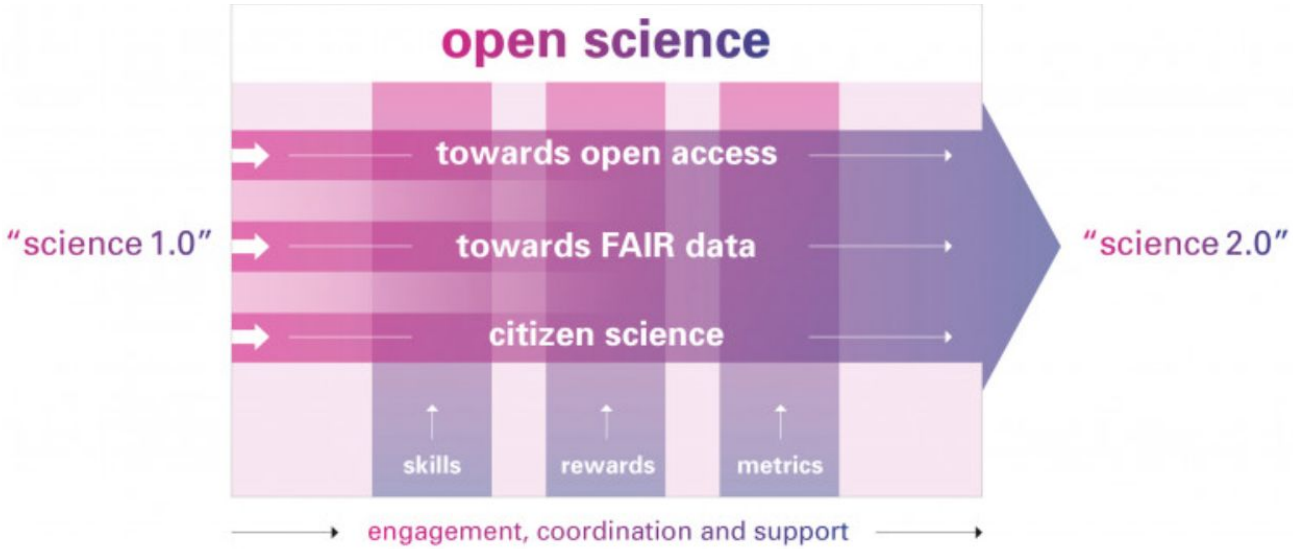
DTL Dutch Techcentre for Life Sciences · umcg · health RI enabling data driven health · nationaal programma open science · DCC implementation network · Landelijk Coördinatiepunt Research Data Management · SURF · Universiteit Leiden · netherlands eScience center · Hogeschool van Amsterdam · Universiteit van Amsterdam · UMC Utrecht · VU Vrije Universiteit Amsterdam · Saxion Hogeschool · TU/e Eindhoven University of Technology · TU Delft

Digital Competence Centres

Calls › Local Digital Competence Centres

Local Digital Competence Centres

Universities, academic medical centres and the NWO and KNAW institute organisation can apply for one-off impulse funding for the establishment or further development of a locally based Digital Competence Centre (DCC). The funding is intended to stimulate research institutions to further develop an existing centralised DCC, or to set up a DCC at a central level within the institution.



Thematic Digital Competence Centres

TU Delft

TU/e

University of Twente.

Wageningen University & Research

SURF

CLARIAH
Common Lab Research Infrastructure for the Arts and Humanities

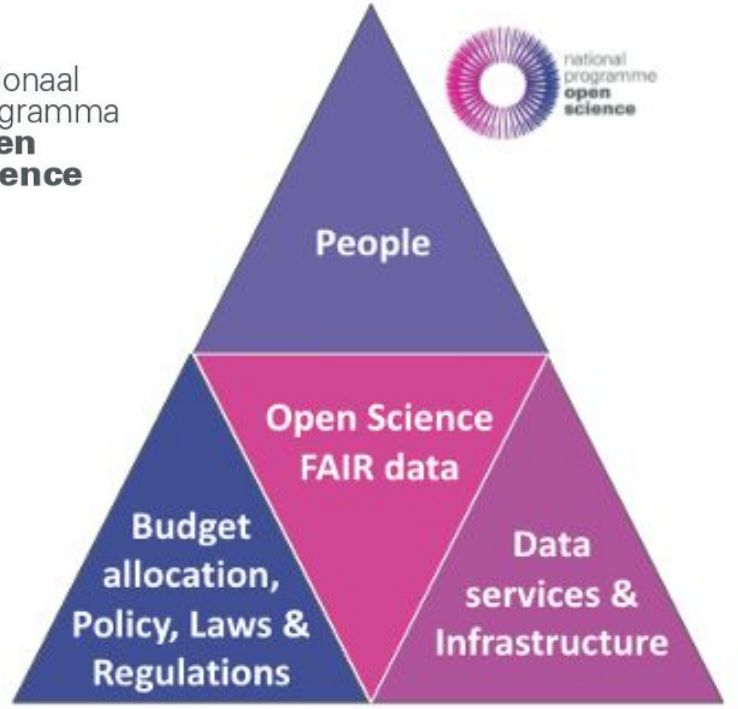
ODISSEI

DCC implementation network

Landelijk Coördinatiepunt Research Data Management

nationaal programma open science

national programme open science



CC BY

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

health RI enabling data driven health

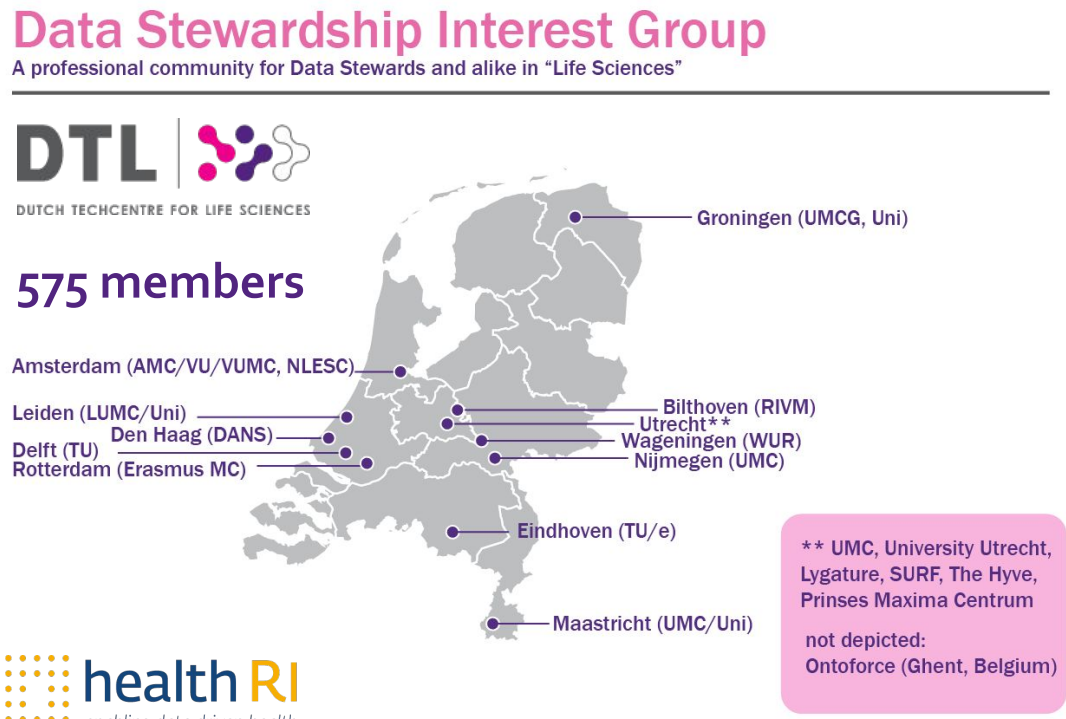
Data stewardship community building

Facilitating communities is an essential element of professionalising data stewardship and capacity building

- Exchange experiences and good practices
- Jointly tackle data challenges

For many years, DTL facilitates the Data Stewards Interest Group (DSIG), with regular meetings (next one in September) and a vibrant (slack channel) community for data stewards and like-minded in the Netherlands **and beyond (welcome to join!)** to share experiences and foster the (Dutch) national implementation of data stewardship

In Health-RI, the Data Stewardship Community (DSC) unites healthcare data stewards in national collaborations, so each institute does not have to reinvent the wheel locally



CC BY

Landelijk Coördinatiepunt 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.

https://www.lcrdm.nl/en

Pool of Experts

The LCDRM 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.

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.

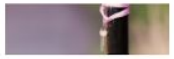
Pitch and team

"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

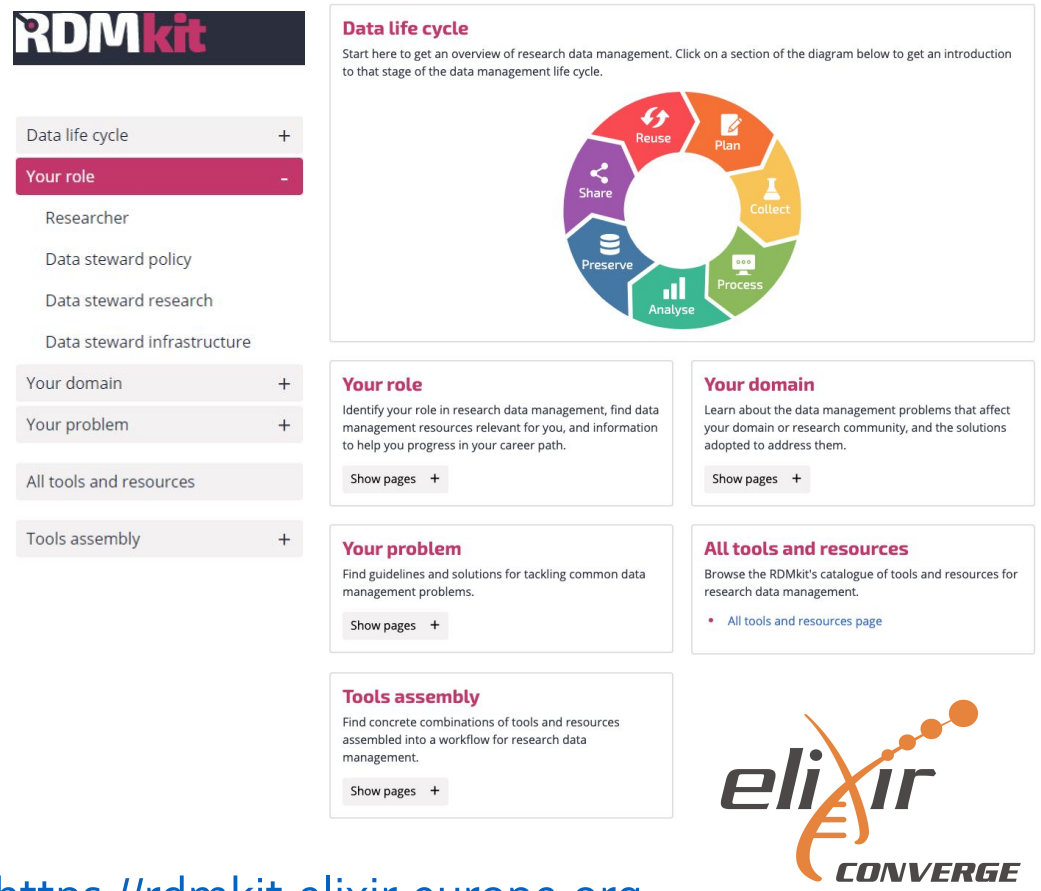
Working groups 2015 - 2017

Part 3

International initiatives

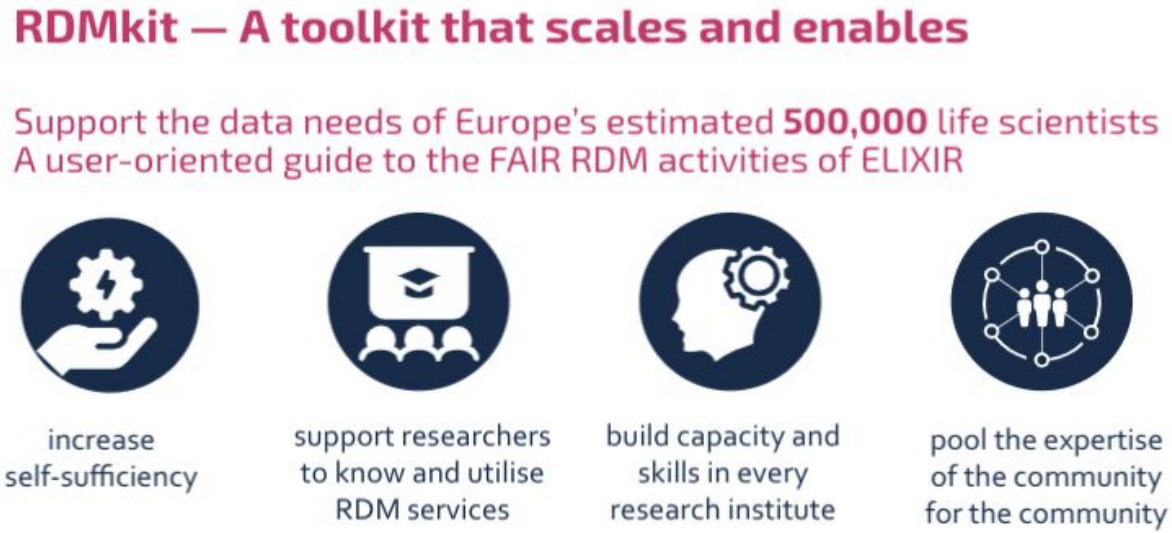
CC BY

ELIXIR-CONVERGE: Toolkit



elixir CONVERGE

https://rdmkit.elixir-europe.org



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!

RDMkit in numbers

87	266	69
Contributors The force behind RDMkit	**Tools & resources** Explained in the context of real world problems	**Pages** Helping you with data management

CC BY

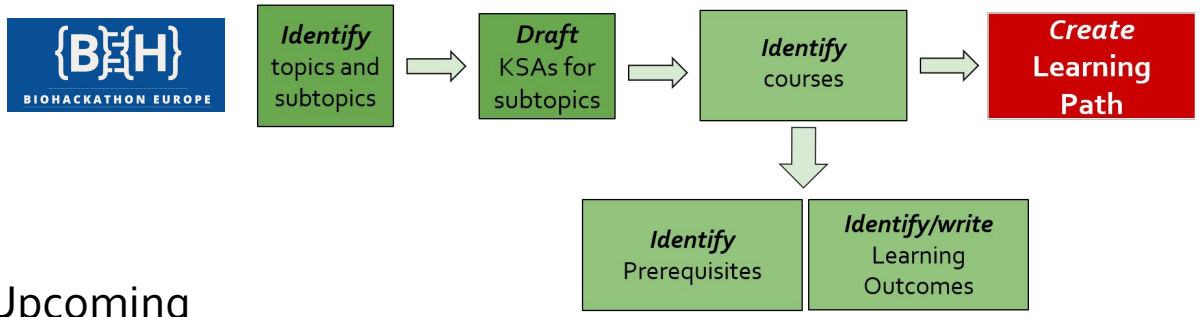
DTL DUTCH TECHCENTRE FOR LIFE SCIENCES | health RI enabling data driven health | elixir NETHERLANDS | nationaal programma open science

ELIXIR-CONVERGE: Training

- Design courses, develop course materials (a.o. through hackathons) and deliver courses
- Audiences: data stewards, researchers, tool developers and trainers
- In alignment with node priorities and/or national strategies

Actions

- First trainings have been delivered
- Training gap analysis has been performed
- Defining topics, KSAs and learning paths at biohackathon 2020 (#29) & 2021 (#13)



ELIXIR-CONVERGE
Data Management Training and Impact - WP2

Training Events

24
Training events
from Feb 2020 to April 2021

11
Countries
organised training in DM

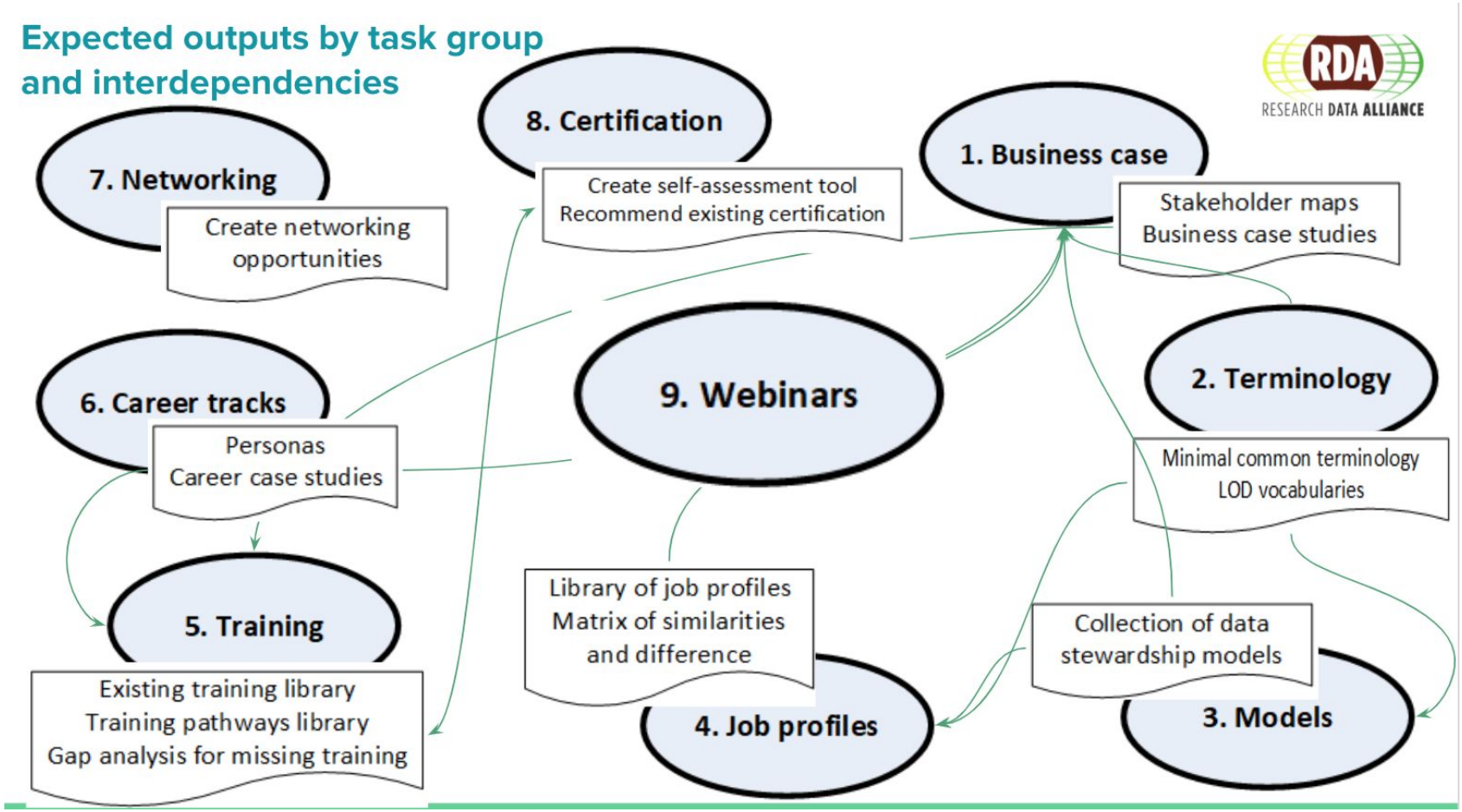
Data Management and Stewardship	ELIXIR Data Management Toolkit	Reproducibility
Data Management Plans	Policies	Ethics
Capturing/collecting/creating data	Finding data/Data discovery (Reusing existing data)	infrastructure for storing and sharing data
Documenting & Organizing data	Processing, analysing data	FAIR
Meta-data	Publishing, sharing data & archiving data	

Upcoming

- Training material generation hackathons related to the gap analysis topics
- Train-the-data-stewardship-trainer courses

CC BY

RDA Professionalising Data Stewardship IG: topics



We need you to make professionalising data stewardship a real community effort, by and for the community!

Make sure to check into our slack community for regular communication

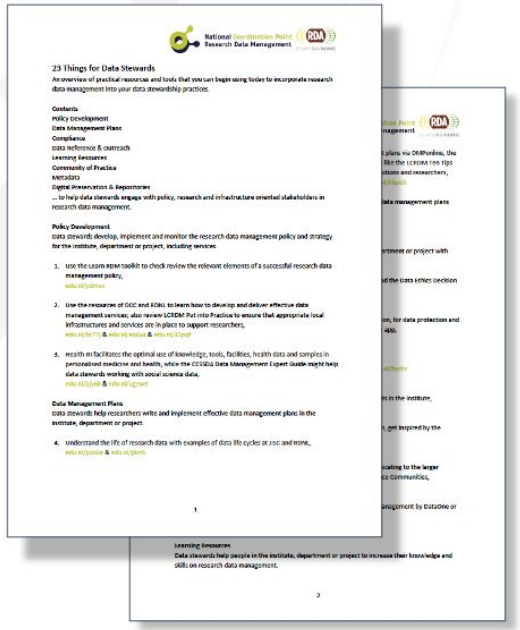
See PDS IG session notes with full slide deck and relevant links in the VP17 notes

CC BY

RDA Libraries 4 Research Data IG: 23 things for data stewards

RDA Output

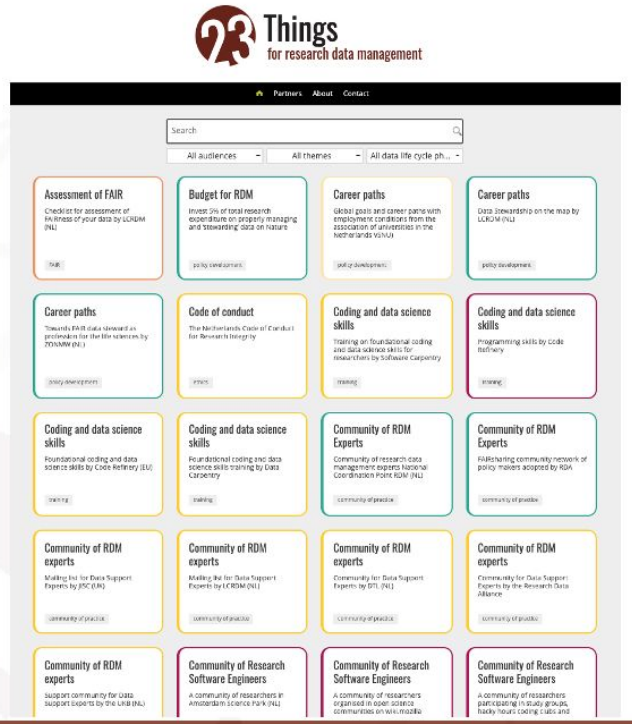
National Coordination Point Research Data Management



Digital sheets for training purposes
https://doi.org/10.5281/zenodo.3773663

Audiences

- Researcher
- Student
- Librarian
- Data steward
- IT staff
- Research Software Engineer
- Policy maker



Toolkit (beta version)
https://23things.sites.uu.nl

Filter by
Audience (7 options)
Theme (17 options)
Data life cycle phase (6 options)

Work in progress

CC BY SA

Acknowledgements

- The ELIXIR-NL, DTL and Health-RI team, particularly Celia van Gelder
- ELIXIR-CONVERGE members
- NPOS-F team
- DTL Data Stewards Interest Group (DSIG) and the Health-RI Data Stewardship Community (DSC)
- RDA Professionalising Data Stewardship Interest Group

Thank you for listening!

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

CC BY

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

health RI enabling data driven health

elixir NETHERLANDS

nationaal programma open science