

open science

## **Presentation 1**: Landscape overview of data stewardship in the Netherlands

## **Presentation 2**: Community building among data stewards: the Dutch experience

Mijke Jetten Community Manager Data Stewardship (DTL) Programme Manager FAIR Data (Health-RI) March 7, 2022



## Outline

- Getting to know each other
- Learning goals
- Data stewardship roles and responsibilities
- Data stewardship training
- Local, disciplinary, national and international communities
- Takeaways

Slides: <u>https://doi.org/10.5281/zenodo.6334297</u>



## Getting to know each other

### You are a data professional

The ideal trainee:

TRAINING DATA STEWARD

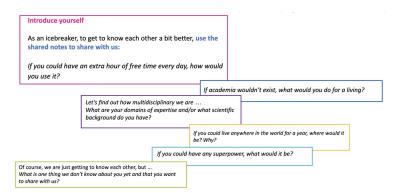
has a PhD in Life Sciences

- has a permanent position, ideally with an interest in data
- is committed to take the whole training program
- after this program, will be involved in the support to the data management, FAIR data and open science policies in her/his organizations.

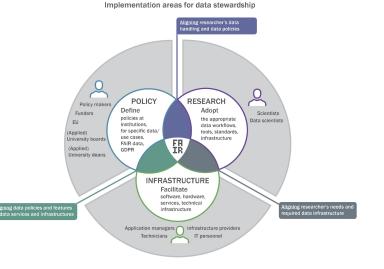


By the end of this course you will be able to make significant contribution in:

- Writing data management plans in support to grant applications • Ensuring continuity and sharing of data over projects • Ensuring data quality
- Supporting the elaboration of institutional data management guidelines
   Supporting compliance with institution data management policy
   Contributing to knowledge exchange with regard to data management



- Do you recognize your current and/or future role in the diagram?
   Where are you in the diagram?
- Where are your colleagues in the diagram? Does your current and/or future team cover all the roles needed?
- What is your biggest challenge?



## Getting to know each other

- 1. Where are you in the diagram?
- Current:
  - Mainly researchers and/or research data stewards (workflow/pipelines, data analysis, stimulate data management, project management, data deposit, metadata)
  - Some (also) infrastructure data stewards (hardware purchases, technical support, data management plans)
  - Some (also) policy data stewards (writing policy, promote FAIR and Open Science)
  - Some are in the centre of the diagram, having all three roles (in one person)
- Future: wish/need to perform other roles as well

### 2. Where are your colleagues in the diagram? Does your current and/or future team cover all the roles needed?

- Mainly researchers and/or research data stewards (disadvantage: not all roles covered in team)
- Combination of researchers/research data stewards and policy data stewards
- Combination of researchers/research data stewards and infrastructure data stewards
- Particularly lack of policy data stewards
- All roles covered but too small team (to truly promote/realise FAIR data and Open Science among researchers)



## Getting to know each other

- 3. What is your biggest challenge?
- Awareness of the importance of Open Data for scientific research (progress is slow)
- Effective data management in the institution (align that with the needs of each researcher)
- Engage researchers (time constraints; 'administrative tasks')
- Identify people with the required competences (data steward as a profession; too many tasks/roles for one person; lack of resources)
- From policy to implementation (implementing FAIR and Open Science)
- Structured approach to data management
- Alignment with policy makers and funders (DMPs)



## Learning goals

After this presentation ...

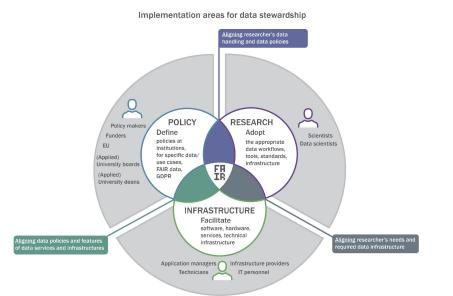
- you are able to position yourself in the data steward landscape
- you are aware of tools with information on the required knowledge, skills and abilities
- you are aware of training resources that help you acquire this knowledge, skills and abilities
- you understand how community building may help you to become a better data steward
- and you might even be inspired to start/participate in similar communities yourself

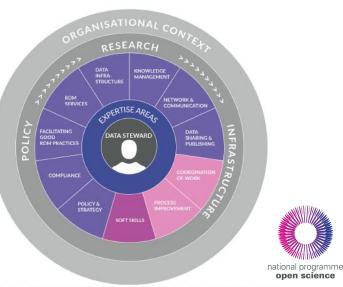


# Data stewardship roles and responsibilities



## Data stewardship roles and responsibilities





ZonMw/ELIXIR-NL funded project "Towards FAIR Data Steward as profession for the Life Sciences"

- All project output: https://zenodo.org/communities/nl-ds-pd-ls .
- Final report (2019): https://doi.org/10.5281/zenodo.3471707 .
- Matrices: https://doi.org/10.5281/zenodo.3239079
- Since 2019 also in the https://competency.ebi.ac.uk .

Figure 4.2 Basic job profile components of a data steward

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



### umcg Radboud University



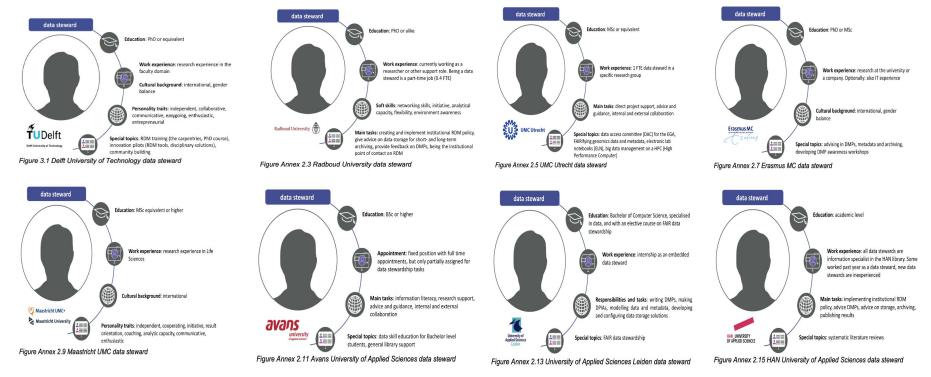
DUTCH TECHCENTRE FOR LIFE SCIENCES





open science

### What does a data steward do?





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

## Data steward personas



- As a policy oriented data steward, I can help my organisation, so that my organisation can strengthen its RDM policy, including efficient workflows, cost handling, FAIR data and Open Science

Figure Annex 12.1 Policy data steward persona (individual perspective)



Work experience and background: research experience in university or business; speaks the language of the researcher; aware of researchers' needs; embedded in specific projects

Education and gualification: Ma or PhD in a relevant scientific domain

Responsibilities and tasks: efficient and quick in handling data; analytical focus; responsible for data processing and analysing; manages big data; takes a consultancy role

#### User story (example) for tool usage:

- As a research oriented data steward, I can effectively support researchers, so that data management is an intrinsic part of their project

- As a research oriented data steward, I want to become more skilled in computing data, so I can help my researchers with managing data in a project - As a research oriented data steward. I want to be able to identify my current knowledge of FAIR data management, so I know what training to take to improve my FAIR data competences

Figure Annex 12.2 Research data steward persona (individual perspective)

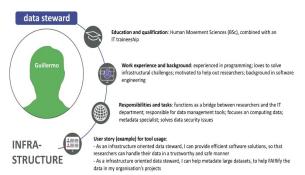
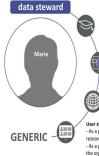


Figure Annex 12.3 Infrastructure data steward persona (individual perspective)



Education and qualification: Information management (BSc: Ma)

Work experience and background: has worked as librarian at the university; has experience with research support: experience with infrastructural and policy developments in the organisation: focuses at matching researchers' needs to organisational developments, including policies

Responsibilities and tasks: supports researchers; advises the organisation; focus at FAIR data, including metadata, persistent identifiers and archiving services; connects people, strong networking skills; matches needs to development processes and policies

#### User story (example) for tool usage:

- As a generic data steward, I can explain my organisational data management policy to the researchers, so that the researchers can meet the requirements - As a generic data steward. I can explain the research needs to the infrastructure people, so that the organisational infrastructure optimally supports research - As a librarian, I have the knowledge to support researchers with archiving, so data is findable

Figure Annex 12.4 Generic data steward persona (organisational perspective)



Education and qualification: Computer Science (BSc), or Biomedical Sciences (MSc)

Work experience and background: has learned via internship or an additional course on FAIR data stewardship; familiar with a specific research domain; hands-on experienced with FAIR data; enjoys working with large datasets; wants to contribute to the quality of research in an organisation

Responsibilities and tasks: supports researchers: writes DMPs for projects: carries out DPIAs; monitors data capture; validates data; focuses at FAIR data; matches researchers needs to development processes

#### User story (example) for tool usage:



- As an embedded data steward, I can implement up-to-date knowledge, to enhance the quality of the data in the projects that I monitor

- As an embedded data steward, I can coordinate the data in a project, so it easier for my researchers to meet the requirements of the organisational data policy - As an embedded data steward, I can use the tool, so I can assess my level of knowledge, improve my career prospects and identify potentially interesting job profiles for my future

Figure Annex 12.5 Embedded data steward persona (organisational perspective)



### https://doi.org/10.5281/zenodo.4320504

## Data steward competences

## Network: Obtaining and maintaining a network of aligned expertise areas and relevant organisations by the institute

> [Policy Oriented Data Steward] Responsible for obtaining and maintaining a network of aligned expertise areas and relevant departments and organisations inside and outside the institute with regard to research data management

 [Research Oriented Data Steward] Responsible for liaison and alignment of research data management within the department or project group and with relevant stakeholders outside the department or project group

#### Activities and tasks

- Refers researchers to other RDM related facilities and services (legal, financial or operational), inside and outside the department or project
- · Liaises with RDM-related experts inside and outside the department or project
- Maintains a network with colleagues and other relevant departments and projects

#### Knowledge, skills and abilities (KSAs)

- Knowledge about where to find department or project stakeholders and researchers, including relevant networks
- · Liaise with department or project stakeholders, researchers and other data stewards and establish an active network
- Communicate with a diverse range of stakeholders

#### Learning objectives (LOs)

- Explain where to find department or project stakeholders and researchers (remembering)
- Liaise with department or project stakeholders and researchers (applying)
- Establish an active network in which regular consultations are taking place (evaluating)



https://competency.ebi.ac.uk

#### 1 More details

## Data steward competences

### Policy and Strategy: Development, implementation and monitoring of the research data management policy and strategy of the institute

> [Policy Oriented Data Steward] Responsible for advice on and development, implementation and monitoring of a research data management policy and strategy for the 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. The policy is the basis for (project) data management plans

✓ [Research Oriented Data Steward] Responsible for the development and implementation of a data management plan for departments, projects or data collections within the institute that is in alignment with the research requirements, specifications and practices, and is in line with the institute's research data management policy and supports FAIR data and Open Science

#### Activities and tasks

- Develops DMP templates tailored for the departments, projects or data collections within the institute
- Writes and/or supports researchers in writing a DMP for departments, projects and data collections, in line with the institute's RDM policy
- Implements RDM as a regular aspect of doing research

#### Knowledge, skills and abilities (KSAs)

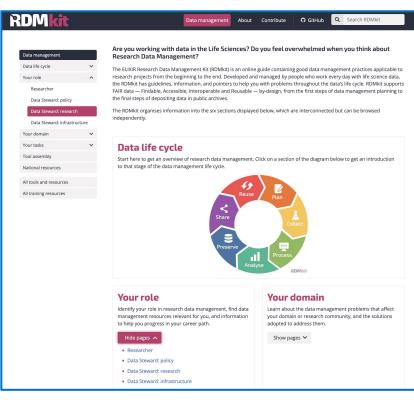
- Knowledge about the general research process and all aspects of RDM and the research data life cycle
- Knowledge about the content of a DMPs, including knowledge about the purpose and how to use it within the research process
- Develop tailored DMP templates together with stakeholders that are understood and can be used by researchers and support staff within the institute
- Translate RDM policies and DMPs to practical implications and guidelines that researchers can understand
- Monitor the use of DMPs by researchers
- Design and draft DMPs for research projects
- Review DMPs written by researchers and help adjusting and refining DMPs over time
- Identify services which support the researcher in putting the DMP into action
- · Monitor research projects with regard to data management
- Knowledge about relevant stakeholders and how to contact them

#### More details



https://competency.ebi.ac.uk

### Data steward tools



### Yourrole Data Steward: research > >



#### Description

As a research data steward, I support and work in close collaboration with the main data producers and users in a cademia: the researchers, ranging from undergraduate students to full professors. I advise researchers, make sure data is handled in a manner compliant with the institute's policy and may also perform hands-on work in a project.

My work focuses on implementing the institute's data guidelines and translating them into domain and project specific procedures, for example by managing a database or reviewing data management plans. My responsibilities and tasks focus on translating the researcher needs on data into infravrutural and service requirements.

#### **Focus**

- Develop and implement data management plans for projects and data collections and align Data Managements Plans (DMP) with the FAIR (Findable, Accessible, Interoperable, Reusable) data principles and the principles of Open Science
- Advise projects and data collections on compliance with codes of conduct, regulations and field specific legal and ethical standards
- Provide adequate research data management (RDM) support to researchers. This involves, for example, supporting
  researchers in improving the reproducibility of their computational analyses or directing researchers to appropriate data
  management and archival solutions
- · Monitor a project's needs regarding data-infrastructure and tools for RDM
- · Determine the adequate level of knowledge and skills of researchers on RDM
- · Identify the requirements of adequate support and data infrastructure for FAIR and long-term archiving of data of a project

#### **Learning path**

Institutes across Europe have started hiring professional data stewards. A research oriented data steward is expected to be competent in the following areas:

- Create awareness and communicate about RDM and the FAIR data principles and translate RDM policies into guidelines for researchers
- · Transform discipline specific research data into FAIR data with help of available services and tools
- · Advise and assist researchers on short and long term actions for RDM
- · Assess RDM knowledge and skills, identify gaps among researchers and take action when needed
- Understand the purpose and use of a DMP in a project and have the skills to utilise the available tools and templates to
  produce a DMP
- · Assist researchers in developing a DMP, review DMPs, and support researchers in putting DMPs into action
- Liaise with the surrounding environment (department, project, national stakeholders and international network) and continuously follow the field to gain knowledge of relevant facilities, tools and emerging standards available for RDM

#### https://rdmkit.elixir-europe.org



### Data steward tools

#### **Related pages**

#### Your tasks

Compliance monitoring & measurement Measure compliance to data management regulations and standards.

Data management plan How to write a data management plan (dmp).

Data organisation Best practices to name and organise research data.

Licensing How to license research data.

Documentation and metadata How to document and describe your data.

Data protection How to make research data compliant to gdpr.

Data publication Prepare data and find repositories for publication.

Data quality Ensure high quality research data.

Data transfer How to transfer data files.

Identifiers How to use identifiers for research data.

### Data Steward: research

- Description
- Focus
- Learning path
- Related pages
- More information
- Relevant tools and resources

**More information** 

Training	
elixir	TeSS - ELIXIR's training portal
Ľ	RDNL - Essentials for Data Support
Ľ	Mantra - RDM training
Ľ	GO FAIR starter kit
Ľ	Data Carpentry lessons
Ľ	RDNL & DCC - Delivering RDM Services
Ľ	NPOS/ELIXIR data steward competency framework
Ľ	ELIXIR Data Management Network

#### **Relevant tools and resources**

Tool or resource 🛈	Description	Related pages	Registry
Argos	Plan and follow your data. Bring your Data Management Plans closer to where data are generated, analysed and stored.	Data management plan Researcher	
Atlas	Free, publicly available web-based, open- source software application developed by the OHDSI community to support the design and execution of observational analyses to generate real world evidence from patient level observational data.	Researcher TransMed	Tool info Training
BBMRI-ERIC's ELSI Knowledge Base	The ELSI Knowledge Base is an open-access resource platform that aims at providing practical know-how for responsible research.	Data protection Sensitive data Data Steward: policy Human data	
Beacon	The Beacon protocol defines an open standard for genomics data discovery.	Researcher Data Steward: infrastructure Human data	Tool info Training
BisQue	Resource for management and analysis of 5D biological images	Data organisation Data analysis Bioimaging data	Tool info
Bitbucket	Git based code hosting and collaboration tool, built for teams.	Data organisation Data Steward: infrastructure	Standards/Databases
Bulk Rename Utility	File renaming software for Windows	Data organisation Researcher	
Castor	Castor is an EDC system for researchers and institutions. With Castor, you can create and customize your own database in no time. Without any prior technical knowledge, you can build a study in just a few clicks using our intuitive Form Builder. Simply define your data points and start collecting high quality data, all you need is a web browser.	Identifiers Data Steward: infrastructure	Tool info
CEDAR	CEDAR is making data submission smarter and faster, so that scientific researchers and analysts can create and use better metadata.	Documentation and metadata Machine actionability Researcher	Tool info Standards/Databases

**RDMkit** 

### Data steward tools

Output

National Coordination Point Research Data Management

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

23 Things for Outo Stewards	https://do
An overview of precised resources and hole that you are hep to say looky to recepte the research data management into your data second big practice.	mups.//uu
Destants Public Development	and the COLOR
Date interagreement Plane	Dagarnerit Carlo
compliance pero redecision & contracth	
Learning Responses	Epidem via DMP-rafine, the MR-TH-I-CHEMISTER FIRS
Community of Dualina	the de cara ren ren
Averadora	ATTON Y
cogets reversion a separations. to way data sharends separa with policy, research and in hadrochare oriented data hadrochars in	
Instanti dia mangement.	bits management plans
Policy Development.	
con newark develop, implement and monitor the responsibility management policy and money for the mail and department or property including version.	
en un serent education beine anone	putanism on building maps
1. Use the users now sookit to check rokew the relevant domonts of a successful research data	
management policy,	wither care within protein
3. Use the respected of bill and bills in learn how in develop and debut effective data	
management services, also review SCROM Put into Practice to ensure that appropriate local inflastructures and provides are implicents support researchers,	ges, for data protection and
which where and solvers are impact to support researchers, where the TTLA where the solvers A where the test and	400.
A results on facilitates the optimal and of intervisions, make, facilities, leadsh data and samples in	
<ol> <li>Health in facilitates the optimal are of knowledge, tools, facilities, health data and samples in personalised medicine and health, while the CEVIDS bats Management Expert Guide might help</li> </ol>	
tiefe slewenk workey with source scenes date.	
eda KUQVA & the All (2007	
Date Management Flam	at in the partition,
sato deviatib help researchers write and inglement effective data management plant in the	
extinute, department or project	a, get begaled by the
4. Understand the life of respects data with examples of data He cyclic at and earse,	
eda algunde & eda algund	
	and the lot is being and the lot of the lot
	LE COMMUNICIO,
A	angement is Baschever
Describe Resources Data classeris help people in the restricte, department or proper	
pillo en rocarch data management.	In the second
1	
udiences	
uliences	
-	17 1 17
Researcher	<ul> <li>IT staff</li> </ul>
o	
Student	<ul> <li>Research Software Engineer</li> </ul>
oludoin	- Roodaron Conward Engineer
Librarian	
	<ul> <li>Policy maker</li> </ul>
Libranan	<ul> <li>Policy maker</li> </ul>
	<ul> <li>Policy maker</li> </ul>
Data steward	<ul> <li>Policy maker</li> </ul>

Set and inconsection from (ISS)



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

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



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

## Data stewardship training



### Helis course FAIR data stewardship

## 3rd edition of *FAIR data stewardship (for the life sciences)*

Did you ever wonder how to improve your data handling in your daily research practice? Want to improve the FAIRness of your data?

Did you recently start a job as data steward? Are you thinking of starting a career as a data steward in the life sciences in academia or industry?

### Join this introductory course to FAIR data stewardship in the life sciences! Course topics

We will pass the stages of the data life cycle in more detail in the training modules of the course. The following topics will be discussed at an introductory level:

- Data stewardship competency framework & the FAIR data stewardship landscape
- (Reviewing) data management plans
- Informed consent procedures
- Data discovery and capturing data
- Preregistration
- Data security and privacy
- Infrastructure for storing and sharing data
- Tools for processing and analysing data
- Organising, versioning and documenting data
- Data and software carpentry
- Archiving data
- Data rights
- FAIR data





https://www.aanmelder.nl/fair-data-stewardship-2021/part\_program https://tess.elixir-europe.org/events/helis-course-fair-data-stewardship

#### Related training materials:

•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 1	, Introduction Landscape Wrap up
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 1	, Institute requirements
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 1	, Funder requirements
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 1	, Publisher requirements
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 1	, (Reviewing) Data Management Plans
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Introduction Wrap up
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Informed consent procedures
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Finding and capturing data part 1
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Finding and capturing data part 2
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Data security and privacy
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Infrastructure for storing and sharing data
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 2	, Tools for processing and analysing data
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 3	, Introduction Wrap up
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 3	, Electronic Lab Notebooks (ELN)
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 3	, Software carpentry/Versioning
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 4	, Introduction Wrap up
• 1	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 4	, Software carpentry
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 4	, Data carpentry
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 5	, Introduction Wrap up
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 5	i, Archiving data 🛛 🔪 🚚
•	Helis Acade	my course	e FAIR data	a stewardship	2021,	Day 5	, Data rights
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 5	, Preregistration
•	Helis Acade	my course	e FAIR data	a stewardship	2021,	Day 6	, Introduction Wrap up 🛛 🔑
•	Helis Acade	my course	e FAIR data	a stewardship	2021,	Day 6	, FAIR Data 🛛 🚺 Tess
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 6	, Metadata
•	Helis Acade	my course	FAIR data	a stewardship	2021,	Day 6	, Metadata FAIR Data Point

Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer

### Welcome and introduction



### https://doi.org/10.5281/zenodo.5704716

### Your needs

Via the assignments, you have provided the content of today's session. We invite you to participate actively, ask questions and use examples so we can tailor the session to your needs. This session is a success if you feel that it is time well spent!

### Objectives

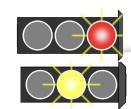
- By the end of this session, you will be able to recognise the basics of FAIR data stewardship in the various stages of the data life cycle
- And you will be able to evaluate what actions need to be taken to solve FAIR data stewardship issues in the proEVlifecycle research project 3-2-1 Waterfall

### Content

- 1. FAIR data stewardship generics (slides 4-9)
- 2. How to be FAIR aware (slides 10-25)
- 3. Recap: group efforts on FAIR data stewardship (slides 26-36)
- 4. Recap: *ten practices*: 'prepare well to prevent data disaster' (slides 37-47)
- 5. Next steps: FAIR awareness and group actions (slides 48-50)



 Ask a question
 Students answer in the chot, but <u>do not</u> click send
 Teacher calls out "3-2-1 Waterfall"
 Students all press send together for a cascade of answers!



Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer

### FAIR awareness

FAIR starts right at the beginning of your research, in the planning phase. Use the FAIR Aware tool to reflect on your (future) FAIR practices. Keep your project's (future) dataset in mind. What are already good practices/what is feasible and what are the challenges or bottlenecks?

ESR Projects

The 10 proEVLifeCycle ESR projects

**FAIR** Aware Awareness ("Yes, I am aware") Persistent Identifier 90% Discovery metadata 80% Metdata for humans and machines 50% Access control metadata 70% Persistance of metadata 60% Controlled vocabularies 50% Provenance information 90% Community-endorsed metadata standards 60% Preferred data format 90% Digital curation and 90% preservation 0% 25% 50% 75% 100%

Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer

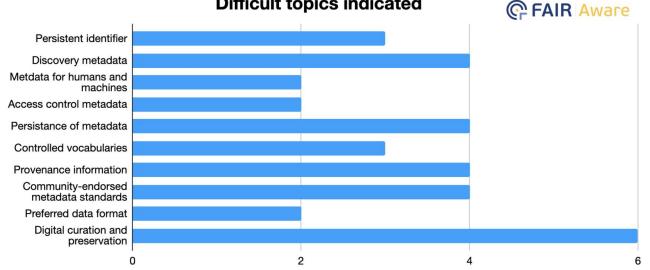
### **FAIR difficulties**

https://fairaware.dans.knaw.nl/ https://doi.org/10.5281/zenodo.5704716

### **Difficult topics indicated**

**ESR Projects** 

The 10 proEVLifeCycle ESR projects



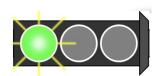
Count

Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer



### Findable data (cf. FAIR Aware)

- 1. Your dataset should be assigned a globally unique persistent and resolvable identifier
  - $\checkmark$  So your dataset can be located unambiguously by humans or machines
  - $\checkmark$  Identifiers (such as a DOI) are usually assigned by data repositories
  - ✓ Persistent identifiers (PID) remain stable and direct users to the object consistently over time
  - ✓ Not all data will need a PID; in general, those that underpin published findings or have longer term value are worth assigning a PID
  - ✓ PIDs ideally point to an online page that contains metadata for context and the link to access the actual data, or details about how to request access





#### https://fairaware.dans.knaw.nl/ https://doi.org/10.5281/zenodo.5704716

### proEVLifeCycle

Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer

### ESR Projects The 10 proEVLifeCycle ESR projects

### Findable data (cf. FAIR Aware)

- 2. You will need to provide details (discovery metadata) to make the data findable, understandable and reusable
  - ✓ Metadata: descriptive information about the data object (e.g. creator, title, publisher, creation date, publication date, summary, keywords, PIDs, license)
  - ✓ Data content: an accurate reflection of the data deposited (e.g. resource type, format, size, variables, methods)
  - ✓ Other research outputs: include links to other research output to increase reuse (e.g. prior version, other datasets, publications, data source, data creators/collectors, funders, host institution)
  - ✓ A discipline-specific repository will likely be using common metadata standards (see for instance <u>Re3data</u> or <u>FAIRsharing</u>)
  - ✓ It pays to spend time on providing a good description rather than just the minimum information required, so your data is clearly understood 3-2-1 Waterfall!

You have indicated this as a difficult topic





 Ask a question
 Students aswer in the chot, but <u>do not</u> click send
 Teacher calls out '3-2-1 Waterfall'
 Students all press send together for a cascade of answers!



Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer

### ESR Projects The 10 proEVLifeCycle ESR projects

### 4. Recap: ten practices: 'prepare well to prevent data disaster'

Reread last session's <u>article</u>: Briney KA, Coates H, Goben A (2020) Foundational Practices of Research Data Management. Research Ideas and Outcomes 6: e56508

Reread last session's <u>slide deck</u> for example best practices from your own team, and details of the recommended actions

Check the Data Management Plan (D6.3) and EV database report (D4.3)

What steps have you taken in the past year to improve FAIR data stewardship?



Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer



### 3. Recap: group efforts on FAIR data stewardship



Create standards and templates (for documentation, organisation/structures of files, file names, versioning, SOPs, change management)



Use a version control system such as Git (and also use it for corrections and annotations)



Explore your institution's storage solutions, such as the DRE, and commit to the (automated) backup procedures of your institution



Develop a secure and feasible research workflow (including working towards future data integration)



Regularly structure and organise outputs (including taking snapshots at key points in the project)



Discuss with the group to use external repositories (including making plans: when, what and how, aimed at reproducibility and reusability)



Create a DMP if you didn't do so yet, per project or as joint effort, and periodically discuss and update it



Formally appoint roles and responsibilities



Knowledge exchange and training (including exploring tools together, learn how to prepare for data archiving, learn from best practices, and check the understandability of each others data activities)



Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer

### ESR Projects The 10 proEVLifeCycle ESR projects

### Practice 1. Keep sufficient documentation

Easy steps to take

- Create documentation standards/templates to ensure recording of the same information (group effort)
- Take the 'outsider' perspective: can others understand my data activities? (group effort)
- Document so that research is reproducible (document more and more)
- Don't wait: document your data right away

### How does this help you?

- Producing documentation in the course of your research ensures that data can be
  - $\checkmark$  properly interpreted as relevant context is available
  - $\checkmark$  verifiable and reproducible
  - $\checkmark$  reusable (by you or by others)
- It helps to explain
  - $\checkmark$  the content of the dataset, at the data level (codebook)
  - $\checkmark$  the context of the dataset, and how the research was done (methodology section)
  - $\checkmark$  the structure of the dataset (readme.txt file with the structure of the dataset)
- Optimising the usability, reusability and reproducibility of the resulting data

Innovative, multidisciplinary research and training program on extracellular vesicles in prostate cancer



### Practice 2. Organise files and name them consistently

Easy steps to take

- Record the structure you choose, and create standards/templates to ensure organising in the same way (group effort)
- Separate raw, analysed, processed data
- Separate ongoing from closed work (milestone versions)
- If you use abbreviations, make sure to explain them
- Keeping folder and file names as short as possible
- Apply the same to physical data (samples)
- Don't wait: organise your data right away

### How does this help you?

- Optimising the usability, reusability and reproducibility of the resulting data
- In 3 years time, would you know what these are?





#### ZonMw and FAIR metadata



### The M4M Workshop concept

### Making it easy for humans to make metadata for machines

Metadata for Machines (M4M) workshops are agile, hackathon-style events that bring together domain experts (who are able and willing to represent a domain community) with FAIR metadata experts (data stewards) who guide a discussion leading to the metadata requirements that meet the FAIR data needs of that domain community. M4M Workshops are lightweight, fast-track (often 1-day) events where policy and domain experts can build new, or make informed choices regarding the reuse of already existing metadata schema. Although M4M Workshops can serve many purposes, they are usually intended to kick-start FAIRification efforts with minimally viable metadata components that are modular, and can be later extended as needed.



#### **ZonMw and FAIR metadata**



### The M4M Workshop concept

### ZonMw COVID-19 Community

#### **Guidelines for completing CEDAR forms**

Link to the guidelines document

#### CEDAR forms - to be filled out

Project Admin form [CEDAR account required] Project Content form [CEDAR account required] Data Catalogue form [CEDAR account required] Dataset form [CEDAR account required] Data Distribution form [CEDAR account required]

#### CEDAR forms in OpenView - easy way to view/share forms

Project Admin form in OpenView [no CEDAR account required] Project Content form in OpenView [no CEDAR account required] Data Catalogue form in OpenView [no CEDAR account required] Dataset form in OpenView [no CEDAR account required] Data Distribution form in OpenView [no CEDAR account required]

**BioPortal vocabularies** 

ZonMw Generic Terms ZonMw COVID-19 Vocabulary

### **Useful links**

- · ZonMw's news item on their approach to optimize reuse of COVID-19 related data
- ZonMw's new item on FAIR metadata about the COVID-19-projects available on COVID-19 Data Portal of Health-RI
- Health-RI's Workshops on delivering FAIR metadata for COVID-19 data portal
- Health-RI's information on the COVID-19 data portal
- Health-RI direct link to the COVID-19 data portal
- Information on the workshops can also be found on the GO FAIR Foundation website on M4Ms.

#### ZonMw and FAIR metadata

https://covid19initiatives.health-ri.nl/p/ProjectOverview https://www.gofairfoundation.org/m4m/





### The M4M Workshop concept

variable

Which categ	ory of diseases	, if any, is releva	nt for your projec	t? Multiple answ	ers are allowed.
If you ticked	human diseasi	, what specific c	lisease is relevar	t for your project	1?
If you ticked	animal disease	, what specific d	isease is relevan	t for your project	?
If you ticked	plant disease,	what specific dis	ease is relevant	or your project?	
Bloace ente	the disease if	not found in the	dron down lint		

What type of biomate	rial, if any, is relevant for	your project? Multiple a	nswers are allowed.
What organisms are t	the target of your analysi	is?	
Please enter the orga	inism if not found in the	drop-down list.	
If you ticked substand	ce, what kind of substance	ce is relevant for your pr	oject?
If you ticked specime	n, what kind of specimer	is relevant for your proj	ect?
If you ticked macrome	plecule, what kind of mar	cromolecules is relevant	for your project?
If you ticked other, wh	nich ones would you like	to add?	

 Temporal Scope (1.3)
 ■
 ■
 ■
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●
 ●

### ZonMW COVID-19

Last uploaded: January 1, 2022

Summary Classes Properties Jump to:	health enabling data driven he		
<ul> <li>biomaterial</li> <li>constraint</li> <li>data</li> <li>audiovisual data</li> <li>diagnostic imaging</li> </ul>			Project overview
<ul> <li>economic data</li> <li>environmental data</li> <li>genomics data</li> <li>geographical data</li> </ul>	Request data	Reset	Q Search
<ul> <li>health data</li> <li>personal information data</li> <li>political data</li> </ul>	Lead institution	~	Changes in the use and organization of care in general pr the COVID-19 pandemic
<ul> <li>qualitative data</li> <li>social data</li> <li>socio-economic data</li> </ul>	Data availability	~	COVID-GP ZonMw - Netherlands Organisation for Health Research and Development NIVE
└─ survey data ─ deprecated ● focus area	ZonMw focus area	~	
<ul> <li>population group</li> <li>COVID-19 patient grouping</li> <li>COVID-19 vaccination grouping</li> </ul>	Type of provided assets	~	Aanhoudende Klachten na COVID-19: perspectief vanuit
<ul> <li>people by age</li> <li>people by COVID-19 risk factors</li> <li>people by education</li> </ul>	Provided data	~	ZonMw - Netherlands Organisation for Health Research and Development Universe
<ul> <li>people by employment</li> <li>people by gender</li> <li>people by sex</li> </ul>	Type of provided biomaterial	~	COVID-NL cohort RadboudUMC RadboudUMC COVID-NL cohort
<ul> <li>people with migration background</li> <li>target group of policy interventions</li> </ul>	Type of provided service	~	other Radboud University Nijmegen Medical Centre
property     service     spatial scope     standard     subject     temporal scope		<u>http</u>	s://www.gofairfoundation.org/m4m/

## Data stewardship community building



## Four levels of community building in the Netherlands

- Local (universities, universities of applied sciences, university medical centres)
- Disciplinary (for example Health-RI, Odissei, Clariah)
- National (for example LCRDM, DSIG, GO FAIR)
- International (for instance RDA)



### Implementation Plan Investments Digital Research Infrastructure

With the extra structural investments and the kickstarter funding, deployed as an integral coherent plan, NWO wants to stimulate the following aspects:

- Data sharing according to the principles of FAIR (Findable, Accessible, Interoperable, Reusable) and open science.
- Federated digital infrastructure both locally and at the (inter)national level.
- Making research data easier to analyse by using more powerful computer capacity, larger storage capacity and new technological developments.
- Making more use of software already developed because this is better known and more accessible.

### Subprogrammes

There are four funding lines within this programme: Local DCCs, thematic DCCs, Investments in eScience, Computing facilities (supercomputer and computing time). The PC-LRI subcommittee ICT is in charge of the programme.

Local DCCs	(+)
Thematic DCCs	+
Investments in eScience	$(\pm)$
Computer facilities	$(\pm)$
ICT subcommittee of the PC-LRI	+



#### 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.

IN PROGRESS	Leids Digital Competence Centre (LDCC)	DCC NWO-I
De impulsfinanciering gaat binnen Tilburg University gebruikt worden om een volwaardig Digital Compe	Deze aanvraag behelst de realisatie van het Leids Digital Competence Centre (LDCC) voor de Universit	Het NWO-I vraagt middels dit project impulsfinanciering aan voor het starten van een Digital Compete
*	*	
DEC Utrecht: omgaan met gevoelige data m reconsta Met deze impulsfinanciering beogen we de beschikbere informatie over het omgaan met gevoelige data L	Radboudumc aanvraag DCC er moantear Het Radboudumc vraagt voor haar lokale DCC een generieke data steward aan.	Het Digitaal Competentiecentrum van de Universiteit van Amsterdam wernoenees Wetenschap wordt steeds digitaler. De manier waarop resultaten verkregen, geanalyweert, gedeeld en g
→	→	-
Aanvraag impulsfinanciering Lokaal DCC KNAW w recontrast Impulsfinanciering voor het opzetten van een Digitaal Competentie Centrum binnen de KNAW, ter verbre	Aanvraag NWO impulsfinanciering voor Digital Competence Centre Maastricht University w moontse In het licht van de strategische UM deelstellingen ten anzien van Open Science, FAIR en ROM werkt d	Oprichting Digital Competence Center Erasmus MC W monester Hete Erasmus MC vraagt €246.880 subsidie aan voor de inrichting van een lokaal Digital Competence Cen
•	<b>→</b>	-
Oprichting van een lokaal Digital Competence Center (DCC) in het UMC	Lokaal Digitaal Competentie centrum aan de Universiteit Twente (UT-DCC)	Opzetten EUR Digitaal Competentie Centrum (DCC)
Utrecht nr Prodettass In bijgevoegd voorstel wordt beschreven hoe we NWO's Impulstinanciering voor oprichting	BI PROFILES Bijgaande aanvraag is gericht op de verdere ontwikkeling van het DCC van de Universiteit Twente. Ορ	N PROGRESS Eraamus Universiteit Rotterdam (EUR) heeft in "Strategie 2024' zeven strategische prioriteiten benoe
Utrecht IN PROGRESS In bijgevoegd voorstel wordt beschreven hoe we NWO's Impulstinanciering voor oprichting	Bijgaande aanvraag is gericht op de verdere ontwikkeling van het DCC van de Universiteit	IN PROGRESS Erasmus Universiteit Rotterdam (EUR) heeft in 'Strategie 2024' zeven strategische
Utrecht In Prodettast In bijgevoegd voorstel wordt beschreven hoe we NWO's impuisfinanciering voor oprichting van een loka	Biggande anvæg is gericht op de verdere ontwikkeling van het DCC van de Universiteit Twente. Op <b>Verming DCC Radboud Universiteit</b> Anstanting De Roboud Universiteit wil met het voortiggende financieringsinstrument het huidige RDM aupport tea	er PROGRESS Eraamus Universiteit Rotterdam (EUR) heeft in "Strategie 2024 zeven strategische prioriteiten benoe
Utrecht In worktes In bijgevegd voorstel wordt beschreven hoe worktes wan een loka <b>UUMC-DCC: hoogwaardige data voor</b> <b>Internetseller</b> <b>Internetseller</b> <b>Internetseller</b> Her LUMC migreert naar een data gedreven organisatiet. Het dev van het LUMC-DCC is	Biggande aervrag is gelcht op de verdere ontwikkeling van het DCC van de Universiteit Twente. Op	Erramus Universiteit Rotterdam (EUR) heeft in Strategie 2024 zeven strategische prioriteiten Benoe Groningen Digital Competence Center er recontes Met behult van deze Impulafinanciering gaan de Rijkseniversiteit Groningen en hat

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

### Implementation Plan Investments Digital Research Infrastructure

### **Subprogrammes**

There are four funding lines within this programme: Local DCCs, thematic DCCs, Investments in eScience, Computing facilities (supercomputer and computing time). The PC-LRI subcommittee ICT is in charge of the programme.

Local DCCs	$(\neq)$
Thematic DCCs	$(\pm)$
Investments in eScience	$(\pm)$
Computer facilities	$(\pm)$
ICT subcommittee of the PC-LRI	$(\pm)$





Thematic DCCs

In the spring of 2022, the thematic DCC (TDCC) network organisations will start. NWO will subsequently publish a call for proposals each year for the realisation of these networks.

The local DCCs provide generic support to researchers within a single institution. However, many researchers collaborate across institutions on specific research themes, and they could benefit from joint support for their research theme. For example, this could concern how you make agreements to better disclose and share data, or about the development of software that makes data analysis within that research theme easier.

To support this, the implementation plan provides for the formation of TDDCs. These network organisations will be established at the start of 2022 with funding from NWO. There will be three TDCC networks that will bring together researchers in three 'domains' (just like the domains in the National Roadmap for Large-scale Research Facilities):

- Life Sciences and Health (LSH)
- Natural and Engineering Sciences (NES)
- Social Sciences and Humanities (SSH)

Besides researchers and the aforementioned local DCCs, other parties will also be important within these three domains:

- Institutions that have expertise in supporting researchers (such as DANS, NLeSC);
- Research infrastructures (such as Health-RI, NPOS-NL, ODISSEI);
- Partnerships (for example, 4TU.ResearchData, DTL);
- Strategic initiatives and councils.

All of these parties will be involved in the networks. Over the next ten years, NWO will invest about 2.4 million euros per year. This will give a boost to the establishment and support of the TDCC organisations, and the funding can be used to support concrete projects within the digital needs that have been identified.

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

## enabling data driven health

Health-RI is the Dutch non-profit foundation supporting a public private partnership of organizations that want to realize a national health-data infrastructure. More than 70 organizations involved in health research and care endorse our efforts.

Health-RI is the Dutch national initiative to facilitate and stimulate an integrated health data infrastructure accessible for researchers, citizens, care providers and industry. It will enable optimal use of health data, samples and images, a learning healthcare system and accelerate personalized health.



### How we operate

The strategy of Health-RI follows three lines of action

- 1. Collective action: optimizing the conditions for building and maintaining a national health data infrastructure;
- 2. Building a national health data infrastructure: fostering and facilitating initiatives and collaborations directed at developing health data infrastructure;
- 3. Providing mature services: supporting researchers and data managers by making infrastructure services, tools and data easy to locate, access and use;

### Our mission

Build an integrated health data research infrastructure accessible for researchers, citizens and care providers.

Facilitate and foster the optimal use of knowledge, tools, facilities, health data and samples to enable a learning healthcare system and accelerate sustainable and affordable personalized medicine and health.

#### ← Communities

### Data Stewardship Community

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

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

### DUTCH NATIONAL PLATFORM IN INTEGRATED DIGITAL LIFE SCIENCES

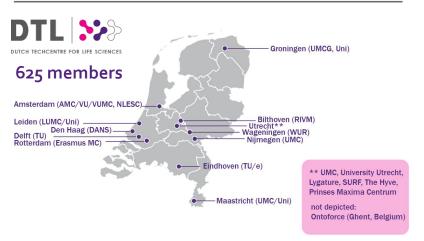
We assemble public and private partners to jointly establish a digital competence centre on biological and molecular data in support of the Dutch health and biosciences communities.

We connect data generation facilities, data and computational infrastructure, data collections, software and models, data stewardship practices and FAIR-compliant standards.

The platform connects experts in digital life sciences and enables easy access to relevant data and tools, and supports advanced computational analysis and modelling via bioinformatics and computational modeling techniques, including machine learning and artificial intelligence.

### **Data Stewardship Interest Group**

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



### **DIGITAL LIFE SCIENCE COMMUNITIES**



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 <u>Data Stewards Interest Group</u> (DSIG), with regular meetings (next one in September) and a vibrant (<u>slack channel</u>) 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

### Dutch roadmap towards professionalising data stewardship

### Creating FAIR data implies

Making well informed choices about

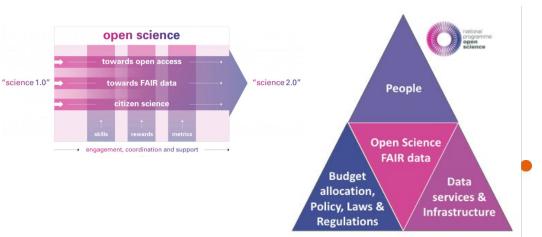
- ... the number of data stewards
- ... where in the organisation
- ... and with what competences
- ... including training

### 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 <u>data stewardship report</u> (NPOS F) links to the second key area, together with a <u>report on the Dutch data infrastructure</u> <u>and services landscape</u> (NPOS E)
- NPOS F: for a quick overview, we advise to read the preambule, executive summary and Chapter 7





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



Glossarv



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'

Pitch and team

Digital consent

enabling.

Policy Inventory of trans-Recommendations on

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

- E-

Anonymisation



A legal and ethical framework for obtaining digitized informed consent.



a DPIA per research

scenario, so that a

scenario can be

selected for new

research proposals.





V2019

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 aroup members is: "I

also have to do something with my subject in my own working environment'

Read more about pitching new task groups.





Data Stewardship



"23 Things" Grant V2019

Working groups 2015

2017

## **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

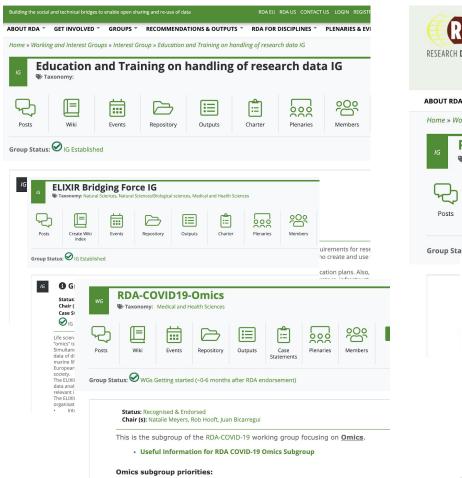
from 13:00-17:00 (CEST) via Zoom



DCC implementation network

## Local Digital Compentence Centres





- A set of guideline documents, highlighting the primary data [and software/ci resources in Omics research, addressing different data types and cross-cutti 2. Resource [data and software/code] List(s) in Omics.
- 3. A Decision Tree tool to facilitate navigation to specific Omics Resources.
- Members: 10956 **O&A Members** 63 MEMBERSHIP R Active Organisational & Affiliate Becoming a member of RDA is simple and Di open to both individuals and organizations G members RESEARCH DATA ALLIANCE 01 **Register now** ABOUT RDA \* GET INVOLVED \* GROUPS \* **RECOMMENDATIONS & OUTPUTS** RDA FOR DISCIPLINES \* **PLENARIES & EVI** Home » Working and Interest Groups » Interest Group » Professionalising Data Stewardship IG **Professionalising Data Stewardship IG** > Taxonomy: 4 Ξ 200 |=  $\sim$ := 000 ññň Create Wiki Events Repository Outputs Charter Plenaries Members index Group Status: 🕀 Not vet endorsed Expected outputs by task group RDA and interdependencies RESEARCH DATA ALLIANCE 8. Certification 1. Business case 7. Networking Create self-assessment tool Stakeholder maps Recommend existing certification Create networking Business case studies opportunities 2. Terminology 9. Webinars 6. Career tracks Personas Minimal common terminology Career case studies LOD vocabularies Library of job profiles Matrix of similarities 5. Training Collection of data and difference stewardship models Existing training library 3. Models Training pathways library 4. Job profiles Gap analysis for missing training

## Takeaways

After this presentation ...

- take time to dive into the resources in this presentation: look at the reports, tools and training resources
- keep discussing your current/future roles, responsibilities and tasks in your local teams and organisations
- start/participate in data steward communities yourself



### 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





