DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES

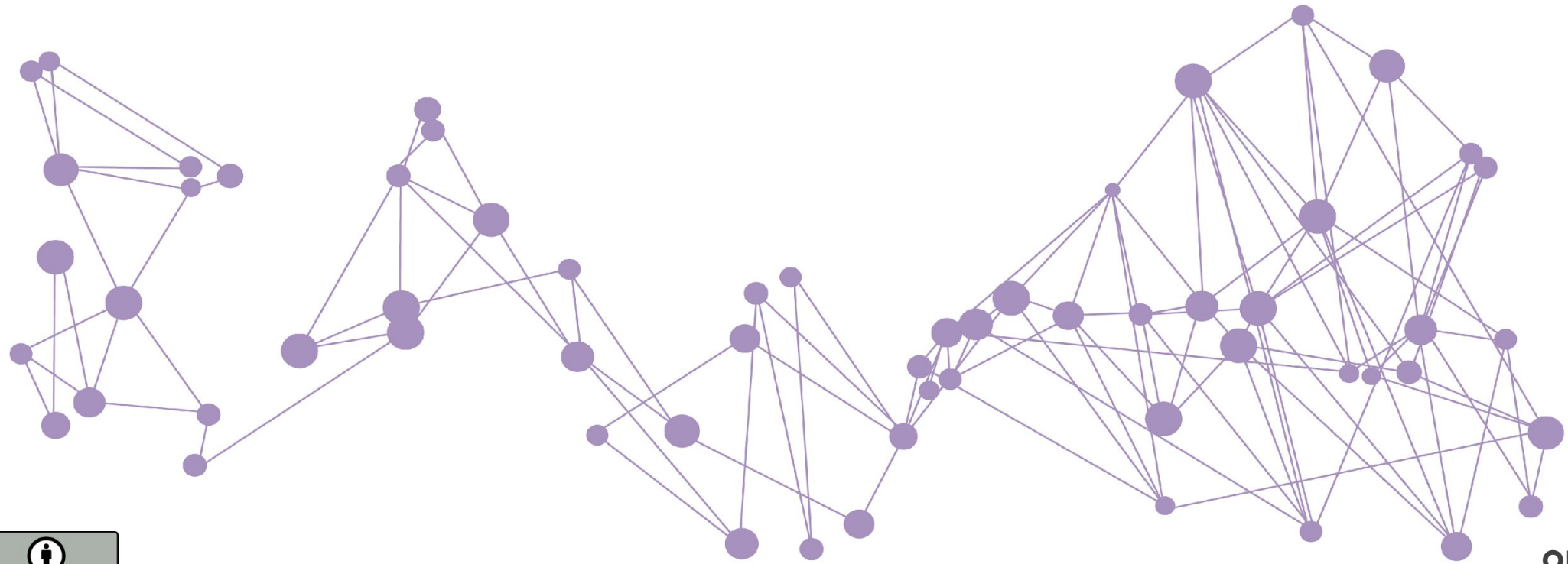
health RI
enabling data driven health

nationaal programma open science

DEVELOPING A FAIR-BASED RESEARCH ECOSYSTEM IN NL

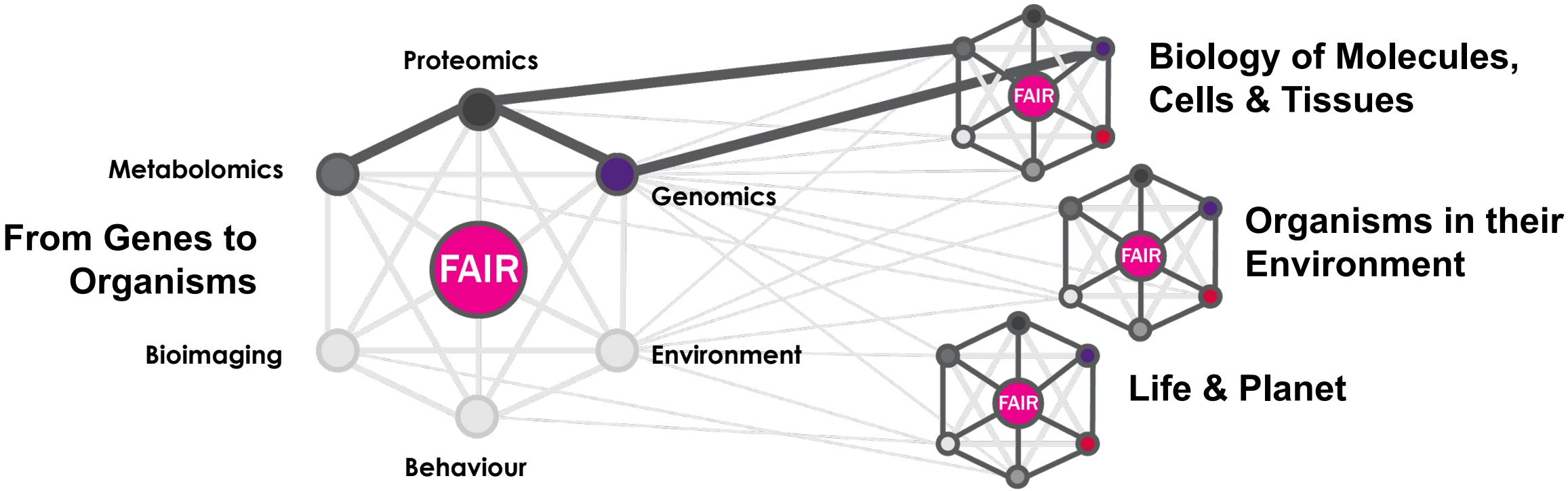
EOSC Policy Event Strassbourg, May 2nd 2022

Ruben Kok



dtls.nl
health-ri.nl
openscience.nl

INTERDISCIPLINARY LIFE SCIENCES RESEARCH NEEDS INTEGRATING DATA



Similar experimentation

Similar data integration challenges

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

NON-REPRODUCIBLE LIFE SCIENCES RESEARCH (!) TOWARDS GOOD DATA STEWARDSHIP

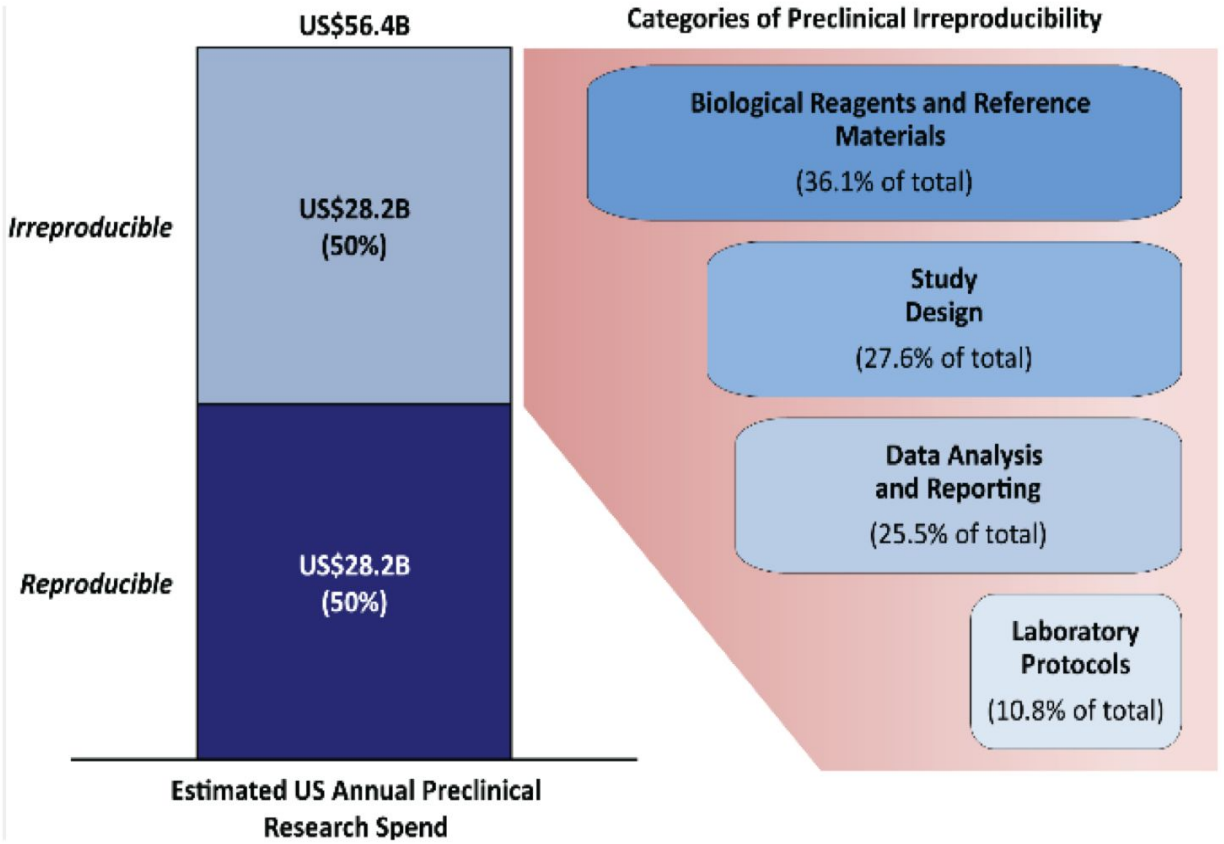


Figure 5: Irreproducibility of research and contributing factors

Needed:

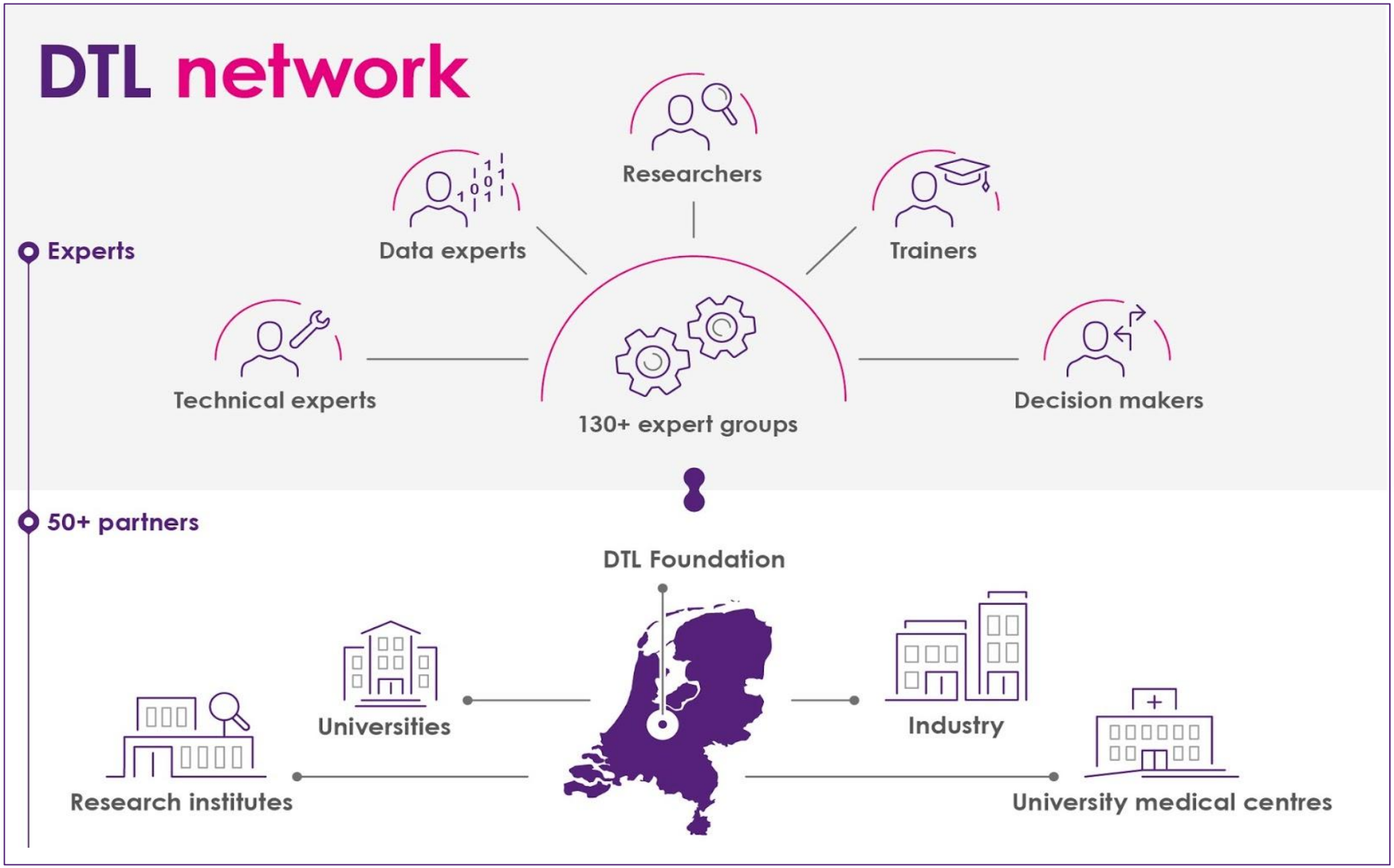
High-quality

- **experimental design**
- **data generation**
- **data stewardship**

Freedman et al PLoS Biol 2015
doi:10.1371/journal.pbio.1002165

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

DTL network



DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES

SCIENCE & INNOVATION DATA MUST BE FAIR

Jan 2014

Findable

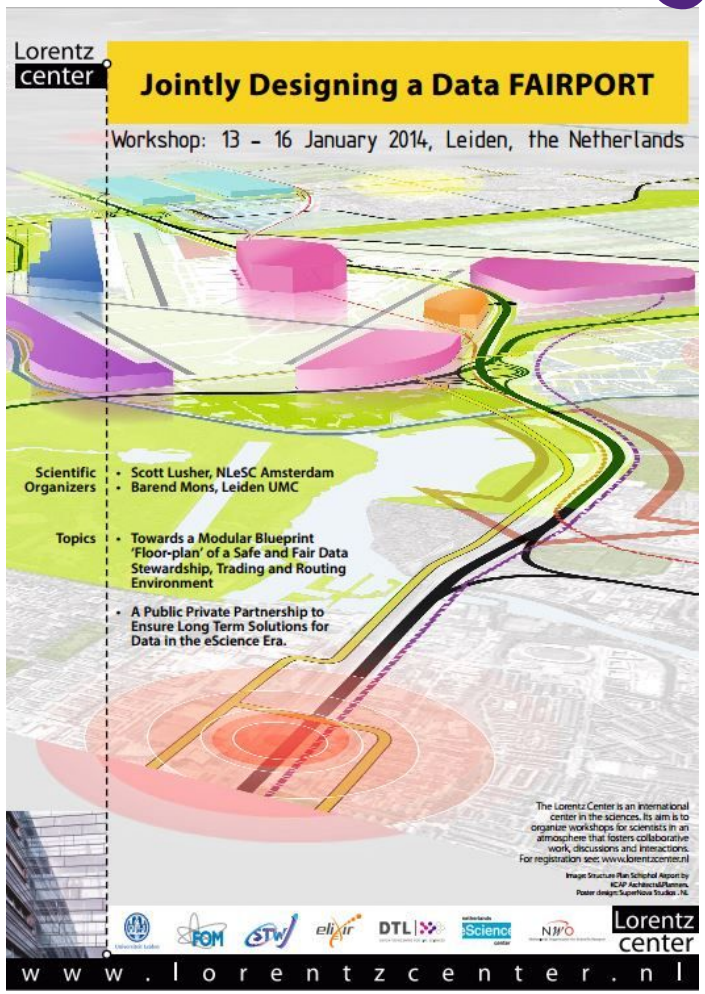
Accessible*

Interoperable

Re-usable

…… for both people and computers

*) NB: accessible ≠ open!
(e.g. proprietary data, privacy-sensitive data)



DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

elixir NETHERLANDS

netherlands eScience center

FAIR GUIDING PRINCIPLES PUBLISHED

March 2016

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2. the protocol allows for an authentication and authorization procedure, where necessary;

A2. metadata are accessible, even when the data are no longer available;

Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles;

I3. (meta)data include qualified references to other (meta)data;

Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes;

R1.1. (meta)data are released with a clear and accessible data usage license;

R1.2. (meta)data are associated with detailed provenance;

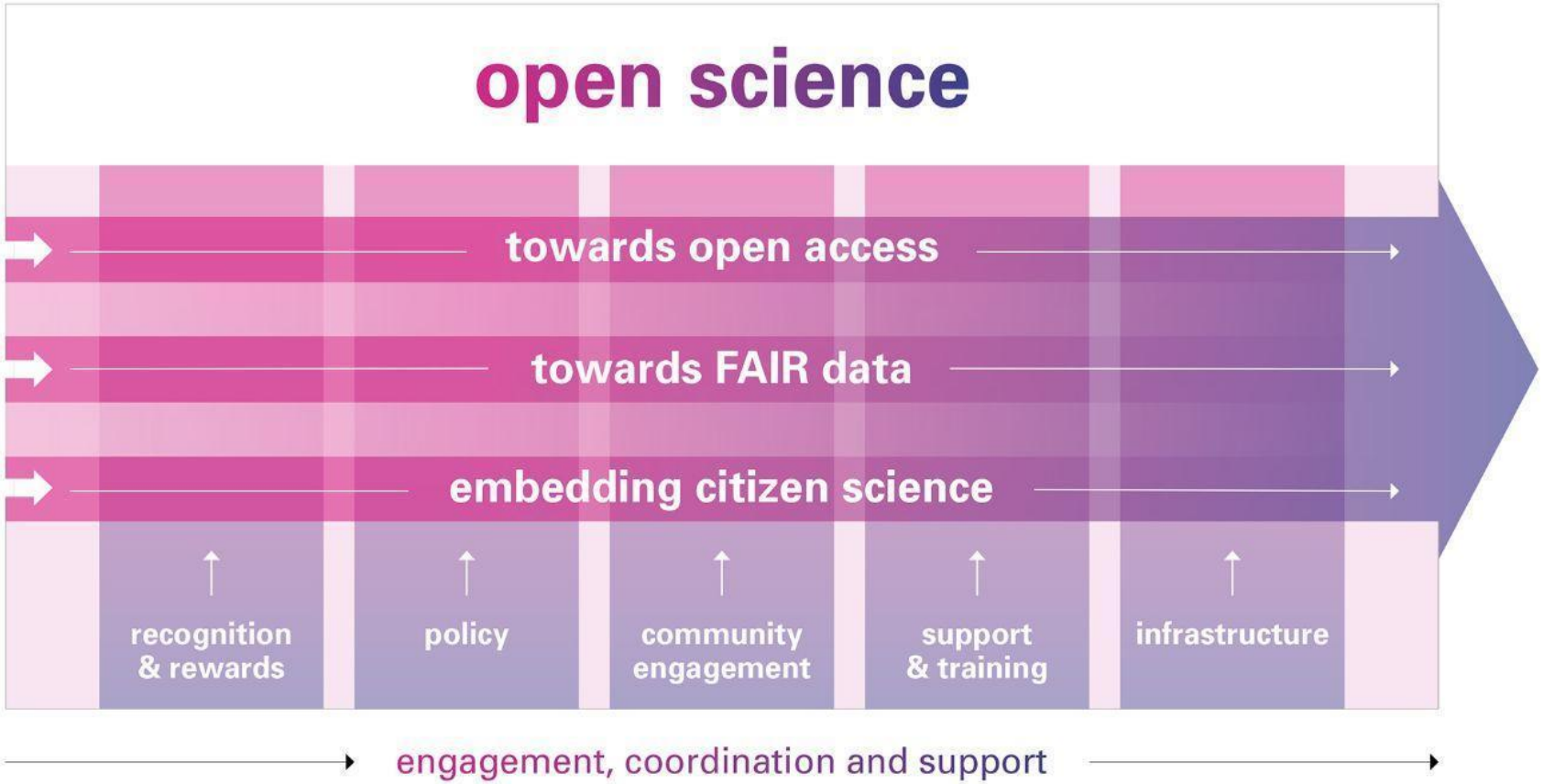
R1.3. (meta)data meet domain-relevant community standards;

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

Wilkinson et al. & Mons 2016 Nature Scientific Data
doi:10.1038/sdata.2016.18

NATIONAL PROGRAMME OPEN SCIENCE

June 2017



nationaal programma open science

National Programme Open Science What is Open Science? Documentation

Home

Project E: Exploring the Dutch data landscape

On 18 June 2020 the Steering Group of the National Programme Open Science (NPOS) gratefully accepted the final report "Exploring and optimising the Dutch data landscape". The report describes both the process of the landscape analysis performed - with input from many experts and stakeholders in the field – and the results of this analysis. The report provides an orientation on the direction to move forward towards a coordinated national approach to implement FAIR data in the Netherlands. The report serves to seed further discussions guided by the NPOS Steering Group, to define the national policy to realise a strong FAIR data programme, including the organizational steps to move forward.

Final report Exploring and optimising the Dutch data landscape

NPOS (2020) Final report Exploring and optimising the Dutch data landscape

Download English, PDF (1 MB)

https://www.openscience.nl/en/projects/project-e-exploring-the-dutch-data-landscape

NPOS FAIR Roundtable

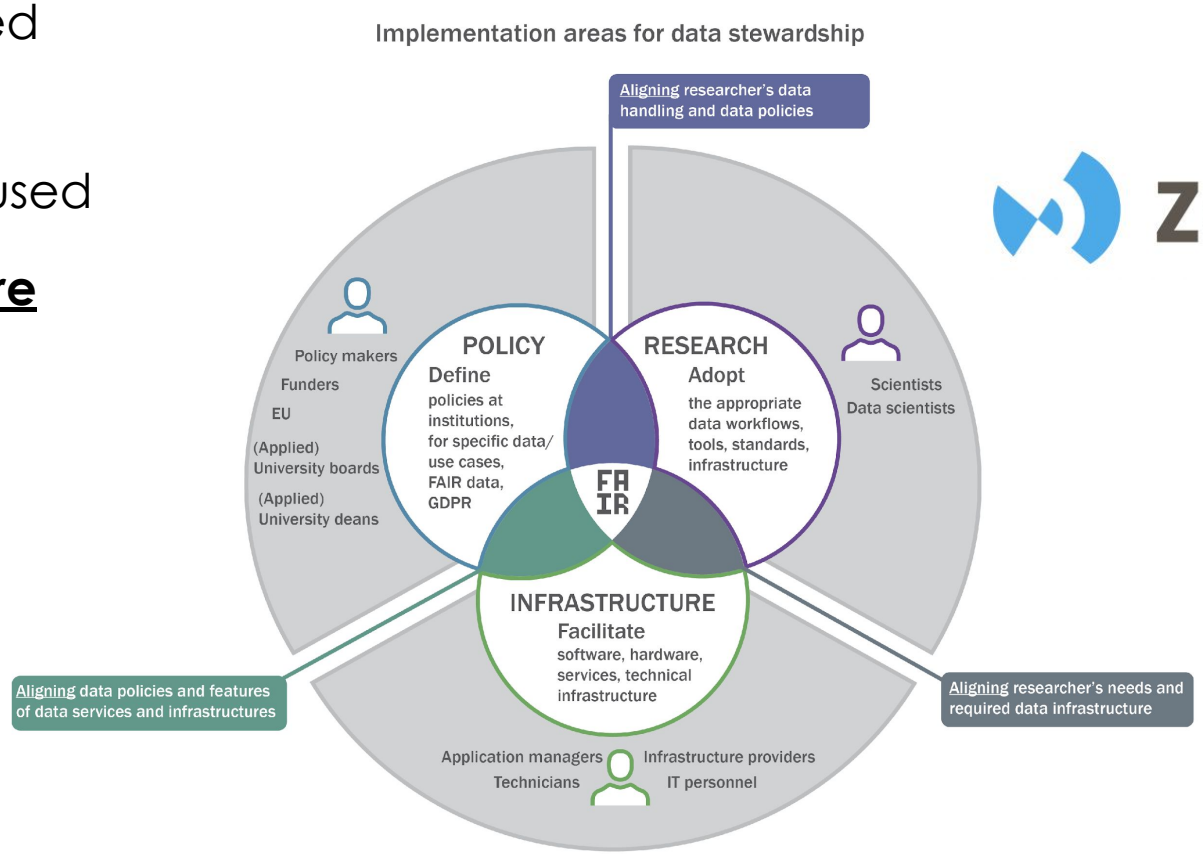
- **Strong basis to work from**
  - Many strong players and initiatives
- **Too little alignment among initiatives**
  - National coordination needed

DEFINE ROLES OF DATA STEWARDS IN THE FAIR DATA LANDSCAPE

- **3 data stewards roles**
  - Data Steward **Policy**
    institute and policy focused
  - Data Steward **Research**
    project and research focused
  - Data Steward **Infrastructure**
    data and e-infrastructure focussed

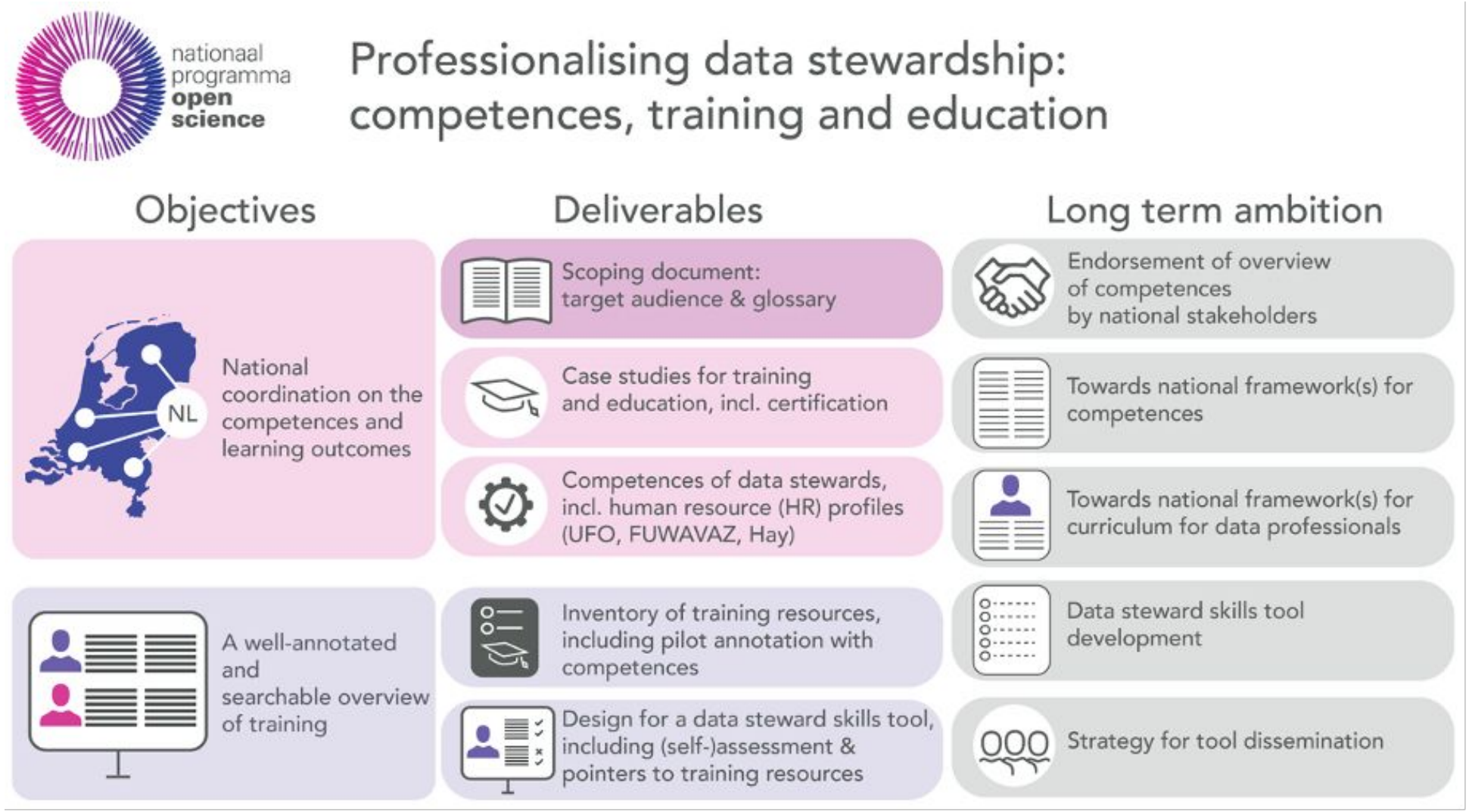
All project output
https://zenodo.org/communities/nl-ds-pd-ls/

Final report, Oct 2019
https://doi.org/10.5281/zenodo.3471707



ZonMw

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES



From NPOS-project report. ***Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship.*** Zenodo. https://doi.org/10.5281/zenodo.4320504

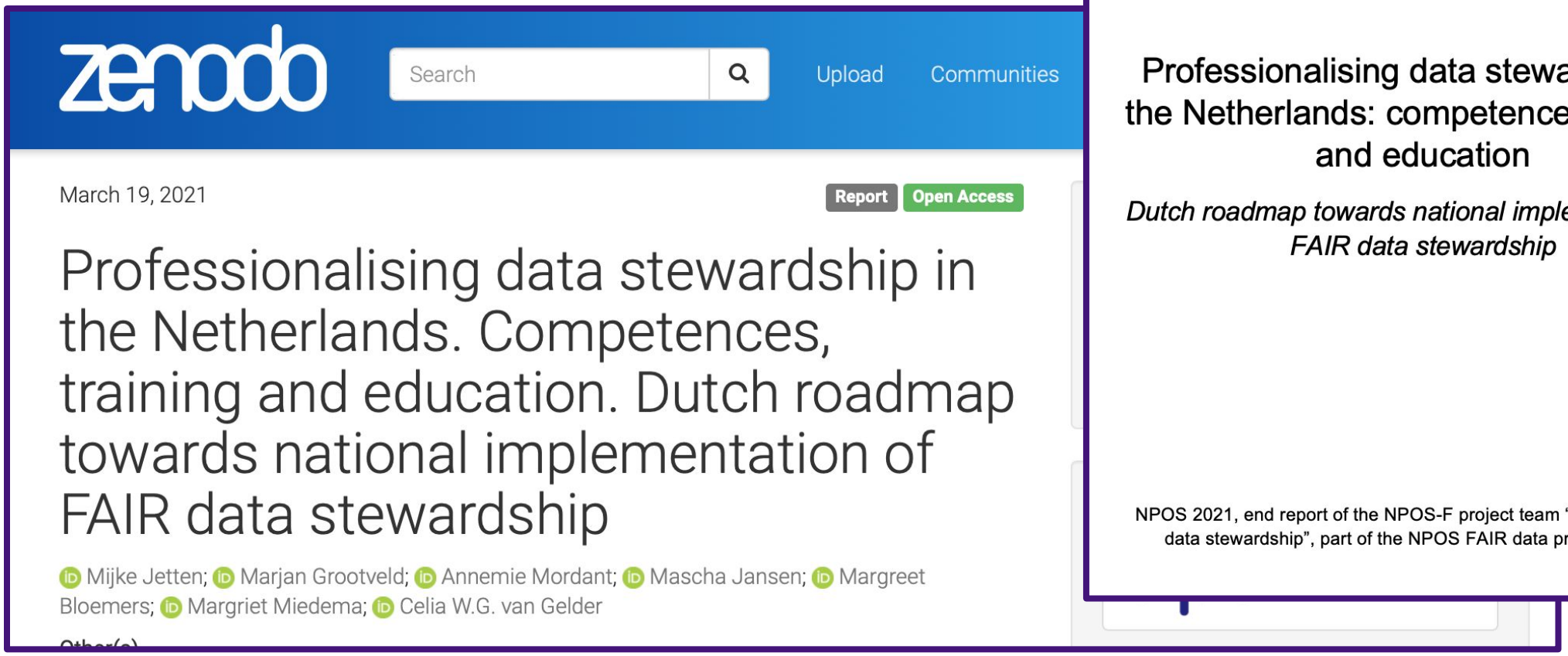
Formal job profiles for data stewards part of function ordering systems at RPOs

Universiteiten *van* Nederland

NFU NEDERLANDSE FEDERATIE VAN UNIVERSITAIR MEDISCHE CENTRA

NPOS REPORT PUBLISHED

- **Recommendations to build a professional community of data stewards**



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

research data netherlands

Essentials 4 Data Support

ABOUT THE COURSE ▸ START THE COURSE ▸ LOGIN ▸

Essentials 4 Data Support is an introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data.

Essentials 4 Data Support is a product of Research Data Netherlands.

SURF

Data Archiving and Networked Services DANS

4TU.ResearchData SCIENCE • ENGINEERING • DESIGN

DTL DIGITAL LIFE SCIENCES

https://datasupport.researchdata.nl

Oct 2019

Integrale aanpak voor digitalisering in wetenschap

Inleiding

Op alle terreinen is de impact van digitalisering van de wetenschap merkbaar. De manier van inwinning, analyse, deling en presentatie van onderzoeksdata- en bronnen is sterk aan het veranderen: de omvang en complexiteit van datasets is fors toegenomen, en internationale en discipline-doorkruisende samenwerkingen zijn belangrijker en meer gemeengoed dan ooit.

Met de steeds luidere roep om FAIR (Findable, Accessible, Interoperable, Reusable) in het kader van de principes van Open Science ligt er nu een sterke nadruk op het delen en vindbaar maken van data. Onderzoeksdata vormen echter maar één facet van de digitalisering. Hetzelfde geldt voor de enorme groei in de behoefte van rekencapaciteit. Enerzijds is er een steeds verder groeiende behoefte aan snellere supercomputers en anderzijds —aangejaagd door het beschikbaar komen van grote volumes aan metingen— neemt de vraag naar rekenkracht voor nieuwe analysevormen, zoals Machine Learning, vanuit bijna alle wetenschapsgebieden sterk toe. Dataverwerking is vaak afhankelijk van specialistische software. Veel van die software wordt binnen projecten ontwikkeld en gaat na afloop van die projecten verloren. Software en data zijn onlosmakelijk met elkaar verbonden; verduurzaming van software is essentieel om het delen van data een succes te maken.

DIGITAL COMPETENCE CENTRES

Sept 2020

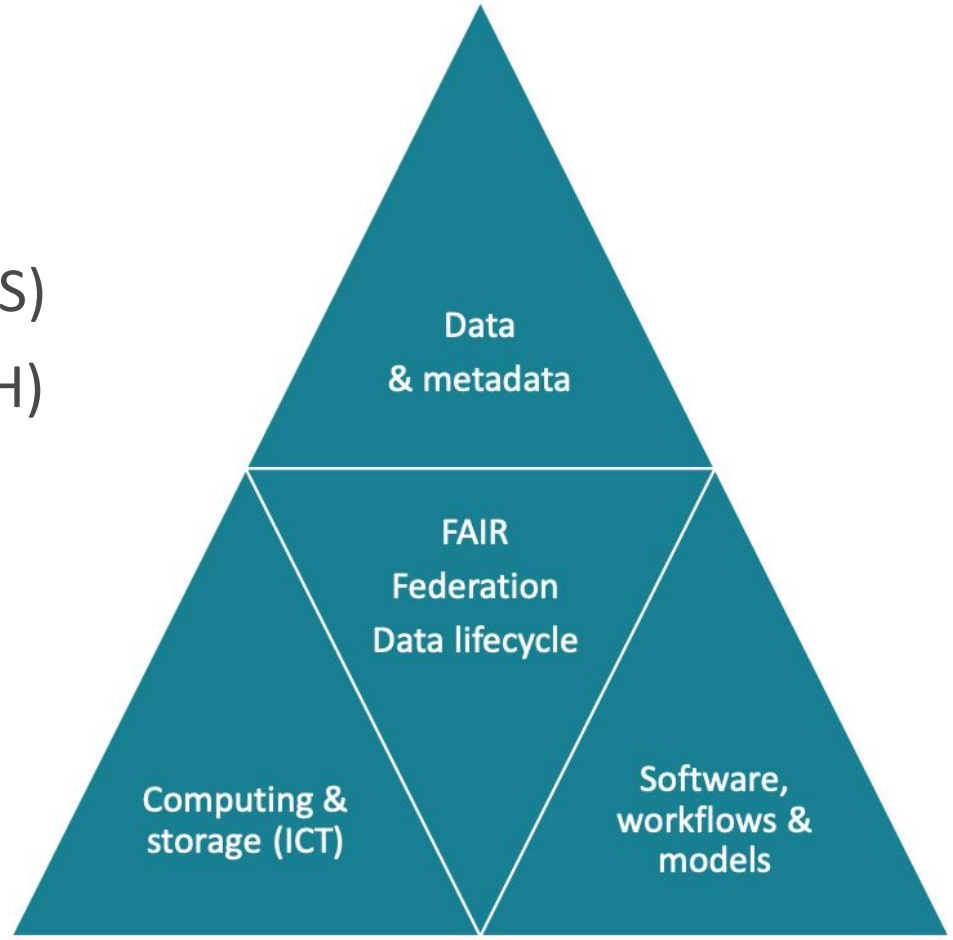
- **NWO funding for network of DCCs**
  - 4.5M€ (impuls) for *local* DCCs at RPOs (30/70)
  - 2.5M€/yr for domain-specific *Thematic* DCCs (30/70)
  - 2M€/yr for SURF to support and interconnect DCCs

- **Local DCCs now established in many RPOs in the Netherlands:**
  - "DCCs largely employ Data Stewards en Research Software Engineers who support scientists within their organisation"
  - Expertise and consulting centre for FAIR data and software
  - Expertise and consulting centre for local ICT infrastructure
  - Connector in federated network of data, compute and expertise

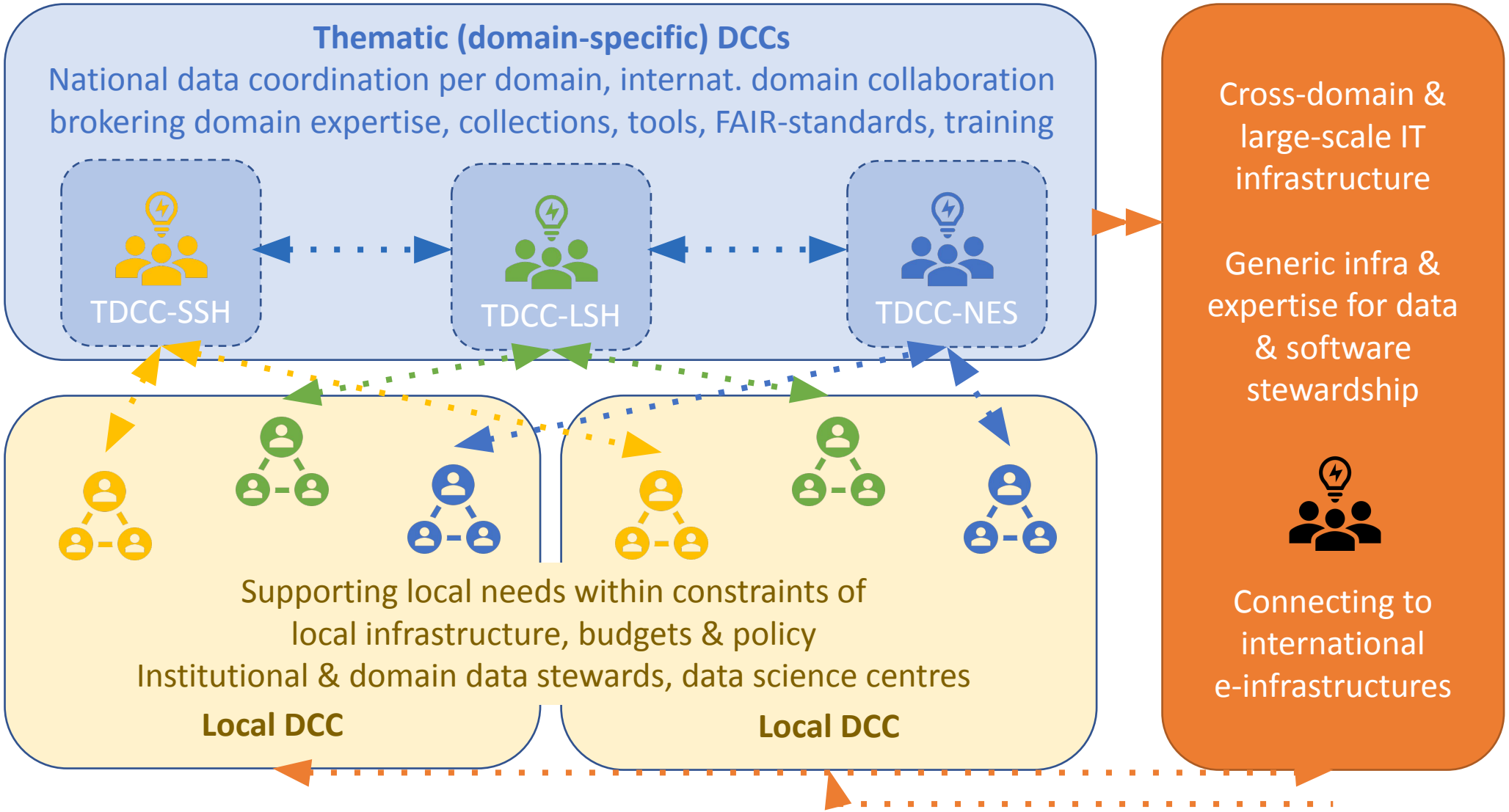
3 Thematic Digital Competence Centres are now being established

- **Per science domain**
  - Life Science & Health (LSH)
  - Natural & Engineering Sciences (NES)
  - Social Sciences and Humanities (SSH)

- **Build on existing networks**
  - LSH: DTL/Health-RI
  - NES: 4TU-Research.data
  - SSH: DANS



3 DOMAIN-SPECIFIC DIGITAL COMPETENCE CENTRES



National health data infrastructure for research and innovation supported by National Growth Fund (69M€)

April 2021

1

Afsprakenstelsel

A. Zeggenschap burger/patiënt
B. Interoperabiliteit & koppelstandaarden
C. Data governance processen & overeenkomsten
D. Publiek-private samenwerkingsmodellen

2

FAIR gezondheidsdata

A. Inrichting regionale knooppunten
B. Inrichting centrale hub en internationale aansluiting
C. Faciliteren gefedereerde analyse en gefedereerd leren

3

Een loket

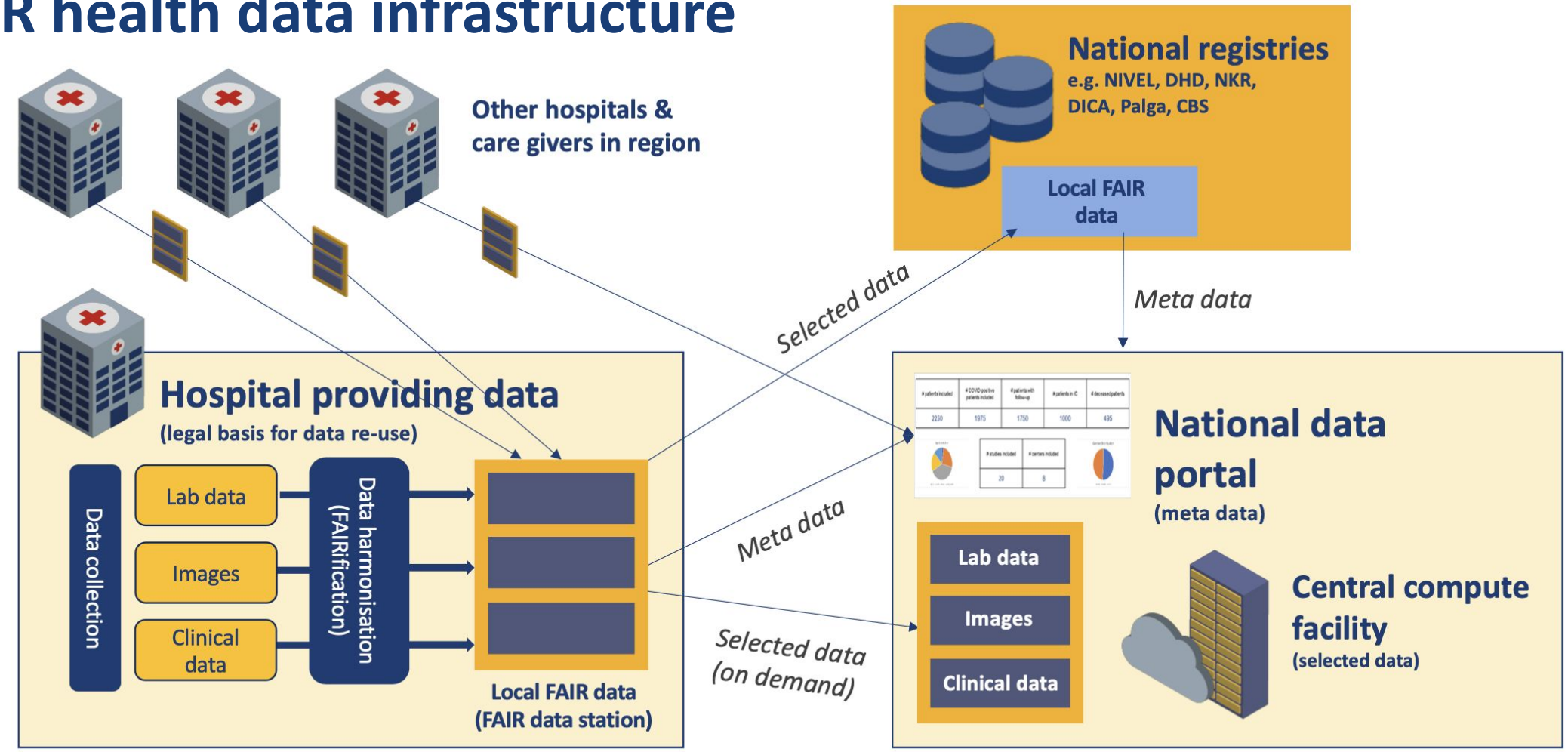
A. Centrale FAIR data catalogus
B. Ontwikkelen one-stop shop voor data-gedreven onderzoek en innovatie
C. Analysesoftware & digitale werkomgeving
D. Support, training & dissemination

0

Management & organisatie

Communicatie, Management, Financiën

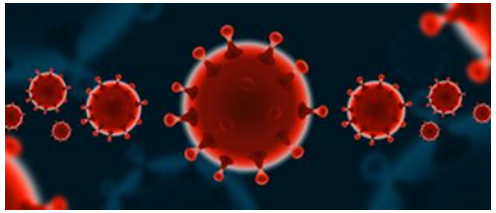
Multi-stakeholder-organisation to design & build the national FAIR health data infrastructure



ZonMw

Starting March 2020 …

Next steps in FAIRifying data in ZonMw-projects



- The urgency of Corona — **Virus Outbreak Data Network (VODAN)**

WHO eCRF data model (semantic, contr.vocabs, machine readable)

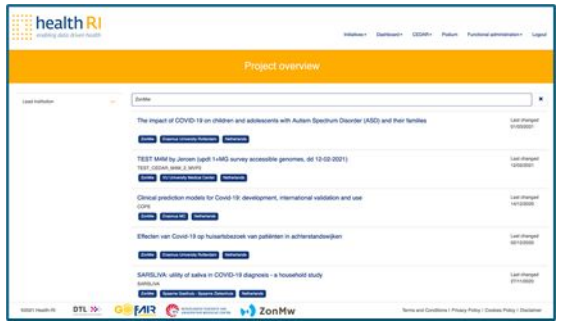
- **Followed by FAIR data services in the COVID-programme (GO FAIR Foundation, DTL, Health-RI)**

3-point FAIRification Framework

Some important features:

- FAIR is for 90% about metadata
- Machine actionable > the computer can find, understand, use
- Domain specific, relevant for community >> ZonMw promotes community agreement!

- Metadata for machines (M4M)
- specific for COVID-19-research (all subdisciplines)
- Health data portal
- Support & community building



National COVID-19 Data Portal
Find&Access portal

Slide: Margreet Bloemers, ZonMw

DTL | GO FAIR | health RI enabling data driven health

Who do you need?

Datastewards take a central position!

- ❑ They help PIs to FAIRify their data
- ❑ ZonMw requires grantees to involve data stewards in the projects
- ❑ Data stewards obtain specialised knowledge at data services & research infrastructures

Funding agencies

(ZonMw, NWO, Horizon Europe, …)

researcher

Data steward
at univ/umc/hs

Experts at FAIR data services & research infrastructures

health RI
enabling data driven health

DTL

GO FAIR

Landelijk Coördinatiepunt Research Data Management

ODISSEI
Open Data Infrastructure for Social Science and Economic Innovations

Local Digital Compentence Centres

NWO

Slide: Margreet Bloemers, ZonMw

ZonMw

COVID-19 Program
62 Projects

Brainstorm controlled lists

M4M.7	M4M.8	M4M.9	M4M.10	M4M.11	M4M.12
13 projects	10 projects	6 projects	7 projects	9 projects	14 projects
21 participants	17 participants	12 participants	13 participants	16 participants	21 participants
1. Data	1. Data	1. Data	1. Data	1. Data	1. Data
2. Images	2. Images	2. Images	2. Images	2. Images	2. Images
3. Biomaterials	3. Biomaterials	3. Biomaterials	3. Biomaterials	3. Biomaterials	3. Biomaterials
4. Services	4. Services	4. Services	4. Services	4. Services	4. Services
5. Standards	5. Standards	5. Standards	5. Standards	5. Standards	5. Standards

COVID-19 Program controlled list
1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

The research community agrees:
- ❑ What information is needed to be able to reuse the resource?
- ❑ What standards for our data do we prefer?

health RI research infrastructure

Build & register
ZonMW COVID-19 Program Vocab

Project Admin Metadata

Project Content Metadata

Generic Dataset Metadata

M4M Support Team

CEDAR

FAIR Data Collective

Exposed on data portal

Researchers create Metadata instances (Routine)

FAIR Data Points

COVID-19 Program Metadata input forms

GO FAIR M4M

Creating FAIR Implementation Profiles (FIPs)

FAIR Principles related questions

A FIP is a collection of machine-readable human agreements addressing each of the FAIR Principles

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

- Choices
- Challenges

GFF

Quality controlled FAIR Enabling Resources (FERs): identifiers, vocabularies, schemas

Slide: Erik Schultes & Barbara Magagna (GO FAIR)

← Home

Data reuse obstacles removal traject

Collaborative commitment to remove obstacles that block the (re)use of health data for research and innovation

Ministerie van Economische Zaken en Klimaat

Ministerie van Onderwijs, Cultuur en Wetenschap

Ministerie van Volksgezondheid, Welzijn en Sport

The Health-RI board and management, the National Growth Fund Committee, the relevant ministries and the field parties involved realize that creating the right conditions, and removing obstacles, is essential to the success of this project. Therefore, the ministries of EZK, VWS and OCW involved, and Health-RI have committed to jointly identify and then remove the current obstacles. This is summarized in the following ambition:

*Remove obstacles and create preconditions in the entire process chain of the reuse of healthcare data, for the benefit of a learning healthcare system. Realize an infrastructure in the Netherlands based on FAIR principles to collect and process data in a **safe** and **standardized** way, and, while guaranteeing **privacy**, make it **findable** and **accessible** for reuse by healthcare professionals, researchers and companies.*

https://www.health-ri.nl/data-reuse-obstacles-removal-traject

Collectively working towards removing obstacles for data reuse

Principes / Draagvlak / Communicatie (alle 4 in samenspraak en parallel)	Kader: Juridisch / Ethisch / Maatschappelijk / Privacy / Kennisveiligheid	Technisch / Logistiek / Organisatie / Services (eerste 3 in samenspraak en parallel)
Cluster 1: "Gezamenlijk waarom"	***Cluster 2: "Spelregels"***	***Cluster 3: "Praktijk"***
4. Patiënten/burgers zijn niet goed op de hoogte van het belang van hergebruik van zorgdata	**3. (Gepercipieerde) Juridische / ethische / sociale barrières (leidend)**	1/2. Zorgdata zijn niet vindbaar / Zorgdata zijn versnipperd
8. Onduidelijkheid over wie zeggenschap heeft over data	**14. Zorgdata kunnen niet veilig en exact gekoppeld worden (leidend)**	18. Zorgdata worden gewijzigd en/of geschrapt (door administratieve last)
9. Angst om erkenning mist te lopen bij organisatie die de data genereren (link met 7, toetsing)	5. Waarborgen privacy (volgend)	12/13/15/16. Zorg(meta)data zijn niet gestructureerd volgens een standaard
10. Toegang tot gezondheidsdata voor/van bedrijven kan extra lastig zijn **(link met 3)**	6. Waarborgen kennisveiligheid (volgend)	11. Technisch / logistiek / organisatorisch (compliance-by-design, volgend op 3,5,6,7,14,17)
	7. Gestapelde toetsing (volgend) (link met 9)	
	17. Zorgdata kunnen niet internationaal gedeeld worden (volgend)	

NPOS FAIR RESEARCH ROUNDTABLE

Universiteiten van Nederland

NWO

ZonMw

SURF

NFU Nederlandse Federatie van Universitair Medische Centra

Koninklijke Nederlandse Akademie van Wetenschappen

Data Archiving and Networked Services DANS

Vereniging Hogescholen

netherlands eScience center

Universiteitsbibliotheken & Nationale Bibliotheek

4TU.ResearchData Science • Engineering • Design

KB nationale bibliotheek

rdnl research data netherlands

Landelijk Coördinatiepunt Research Data Management

DATA SHARING COALITION

health RI enabling data driven health

NL AI Coalition

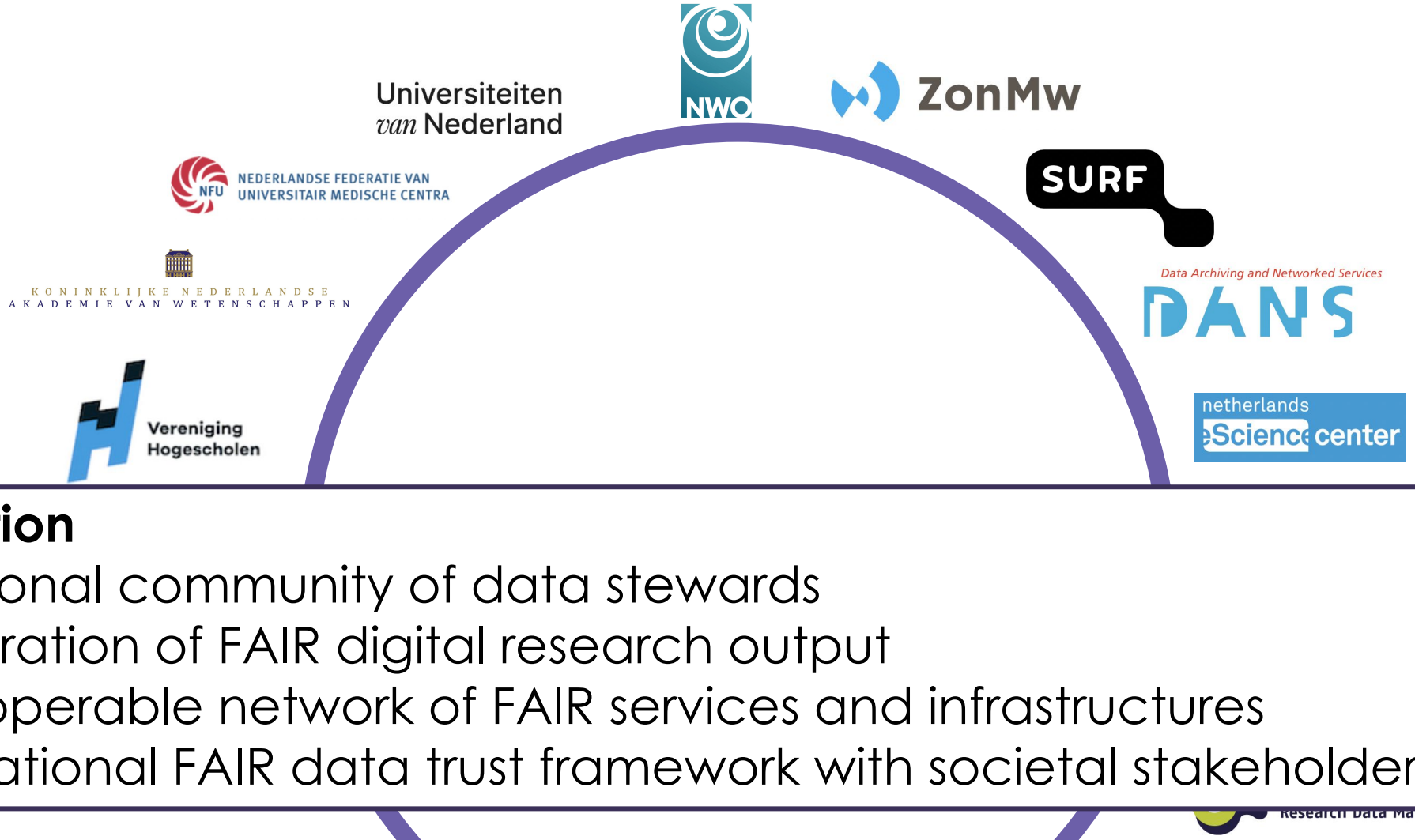
RDA Research Data Alliance

GO FAIR

DTL Digital Life Sciences

nationaal programma open science

NPOS FAIR RESEARCH ROUNDTABLE



Key lines of action

- Build professional community of data stewards
- Enable generation of FAIR digital research output
- Realise interoperable network of FAIR services and infrastructures
- Develop a national FAIR data trust framework with societal stakeholders

DATA SHARING COALITION

health RI
enabling data driven health

NL AI Coalition

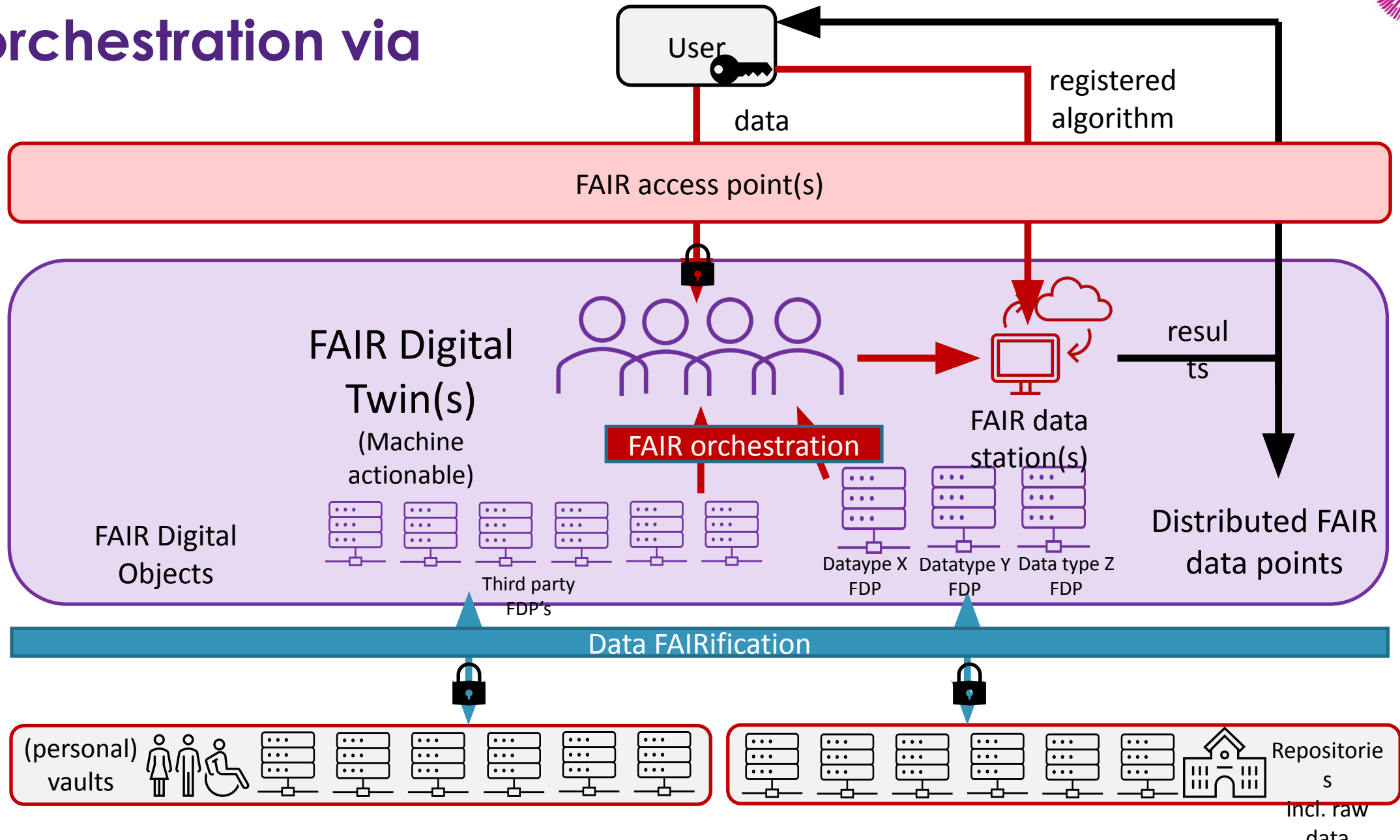
nationaal programma open science

RDA RESEARCH DATA ALLIANCE

GO FAIR

DTL DIGITAL LIFE SCIENCES

FAIR-based data orchestration via



Hub & nodes model

National FAIR Coordination

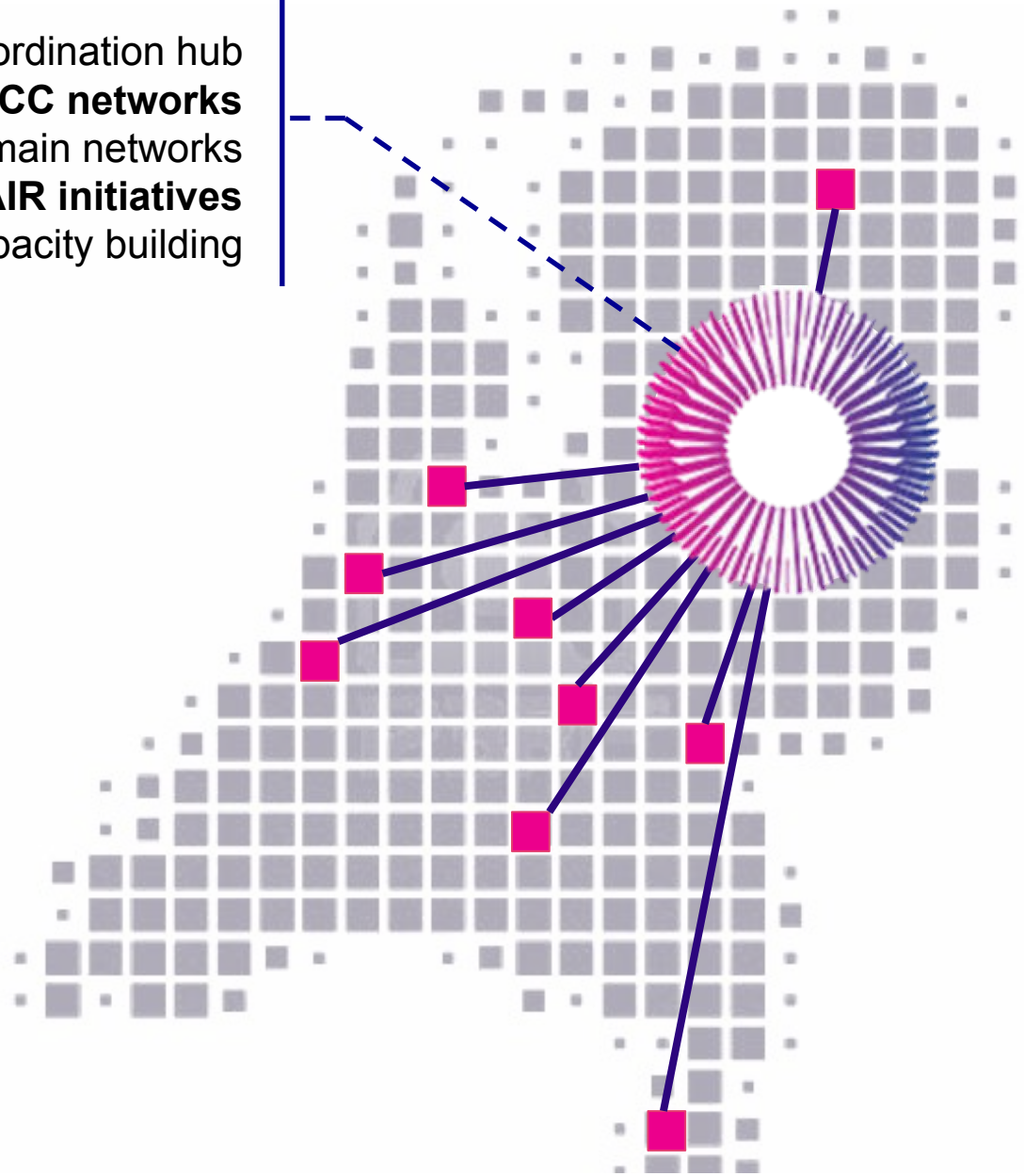
NL-FAIR coordination hub

National coordination via DCC networks

LDCCs at RPOs, and TDCCs connecting domain networks

Connect to EOSC & international FAIR initiatives

Drive and coordinate capacity building



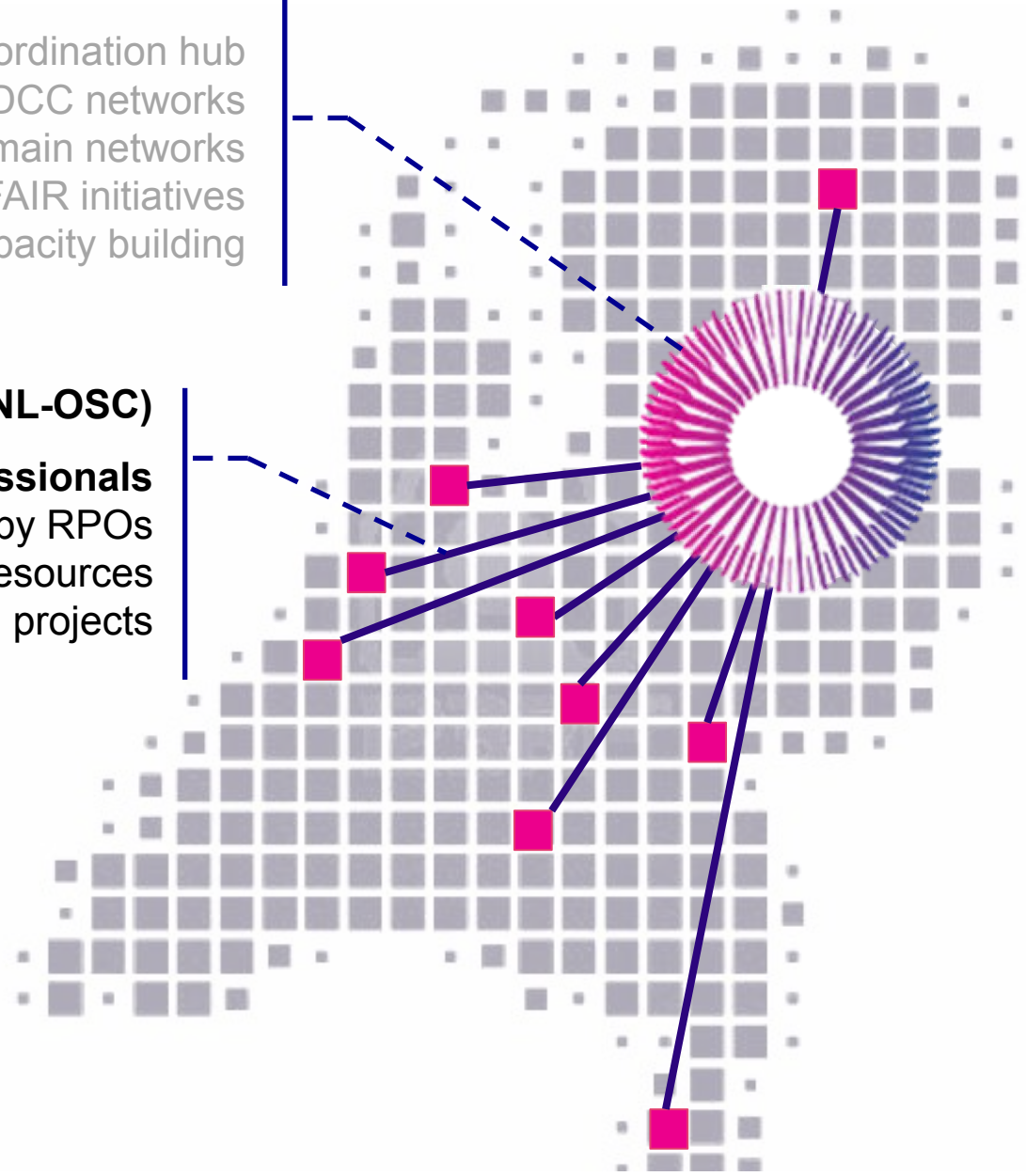
nationaal programma open science

National FAIR Coordination

NL-FAIR coordination hub
National coordination via DCC networks
LDCCs at RPOs, and TDCCs connecting domain networks
Connect to EOSC & international FAIR initiatives
Drive and coordinate capacity building

National ‘web’ of local FAIR research resources (NL-OSC)

Build community of FAIR professionals
FAIR data implementation by RPOs
Domain-specific FAIR data resources
Central & federated data analysis in research projects



nationaal programma open science

National FAIR Coordination

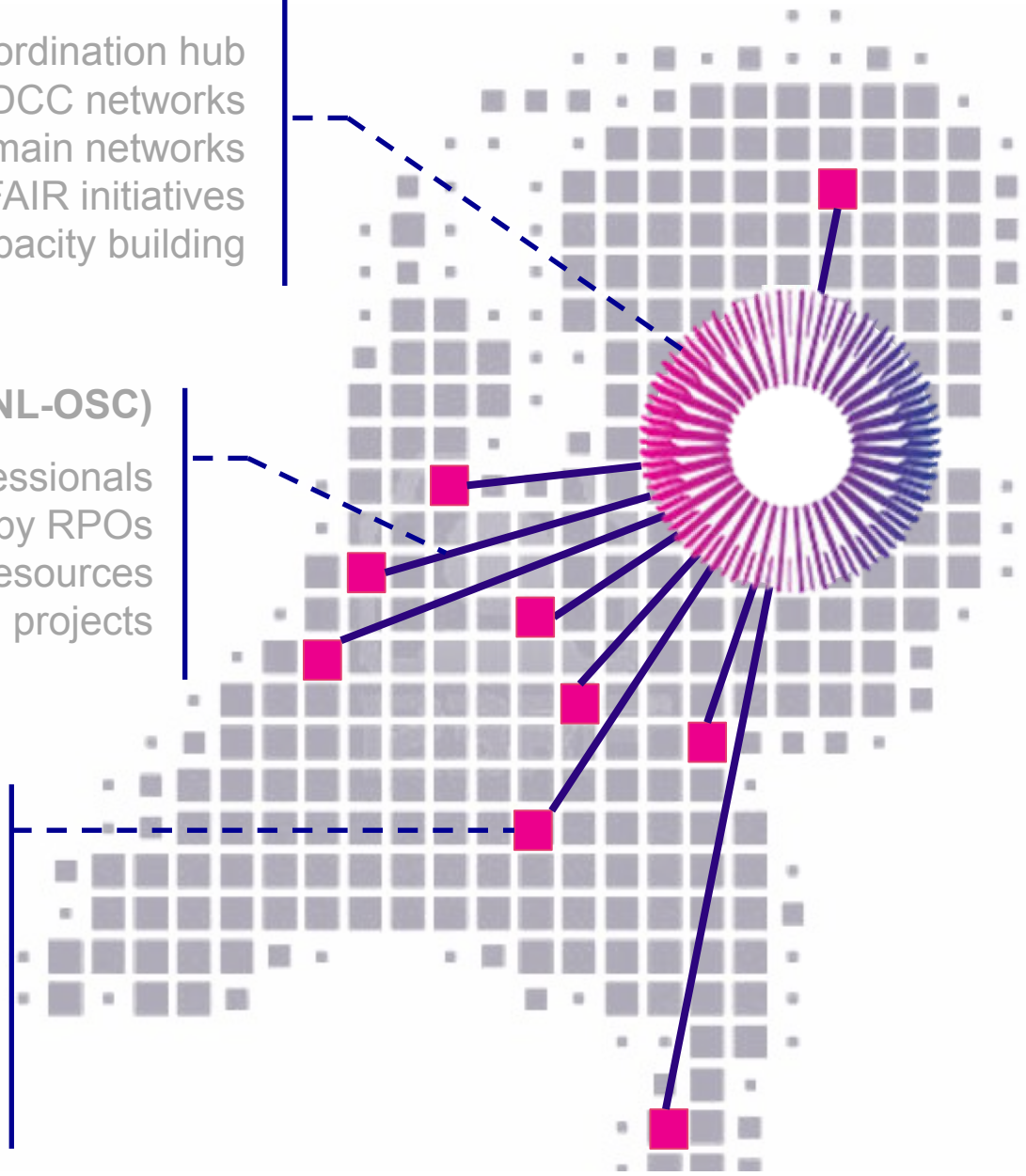
NL-FAIR coordination hub
National coordination via DCC networks
LDCCs at RPOs, and TDCCs connecting domain networks
Connect to EOSC & international FAIR initiatives
Drive and coordinate capacity building

National 'web' of local FAIR reserarch resources (NL-OSC)

Build community of FAIR professionals
FAIR data implementation by RPOs
Domain-specific FAIR data resources
Central & federated data analysis in research projects

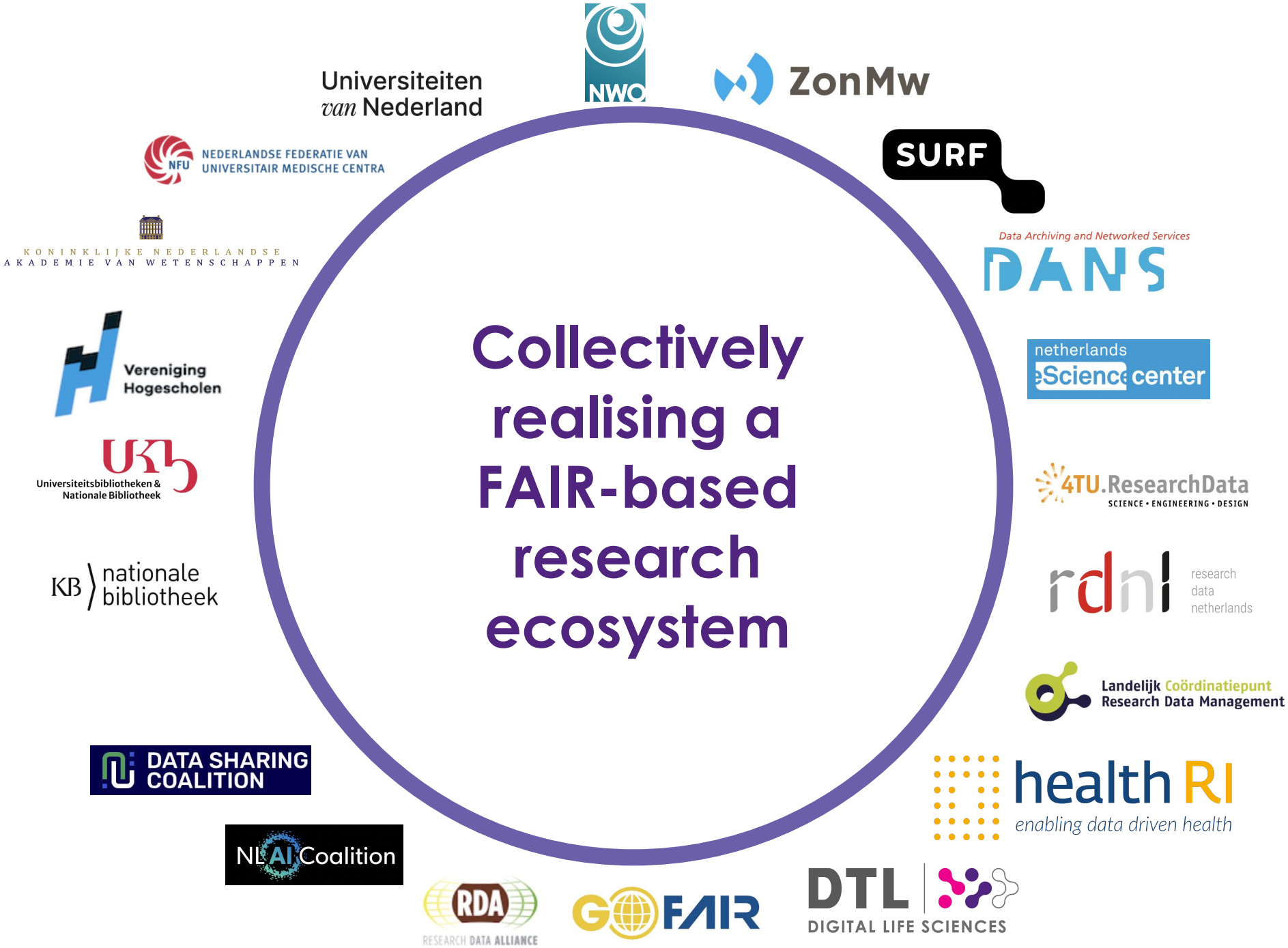
National FAIR Services Infrastructure Network

National FAIR architecture team
Network of interoperable data & IT centres & libraries at RPOs and RSOs
Domain-specific services & infrastructures
Train-the-trainer to support capacity building
National registries of FAIR research output



nationaal programma open science

NPOS FAIR RESEARCH ROUNDTABLE

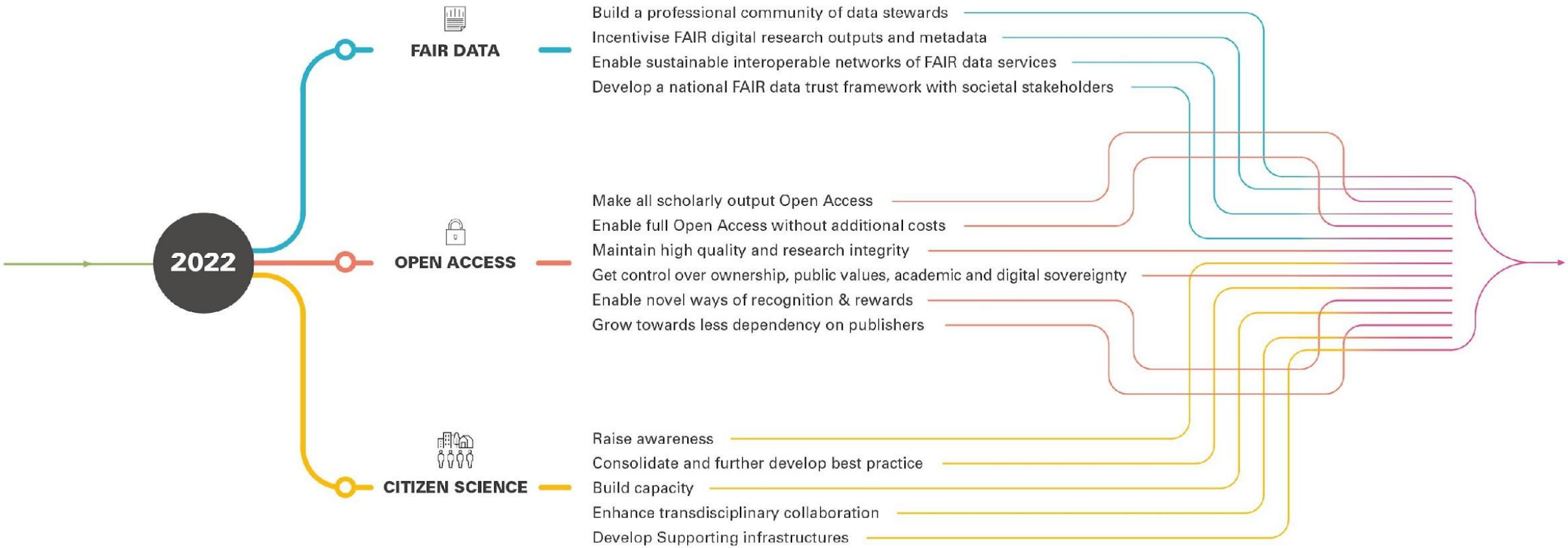


Collectively realising a FAIR-based research ecosystem

NPOS Ambition 2030 2/3

2022 - 2030

key lines of action



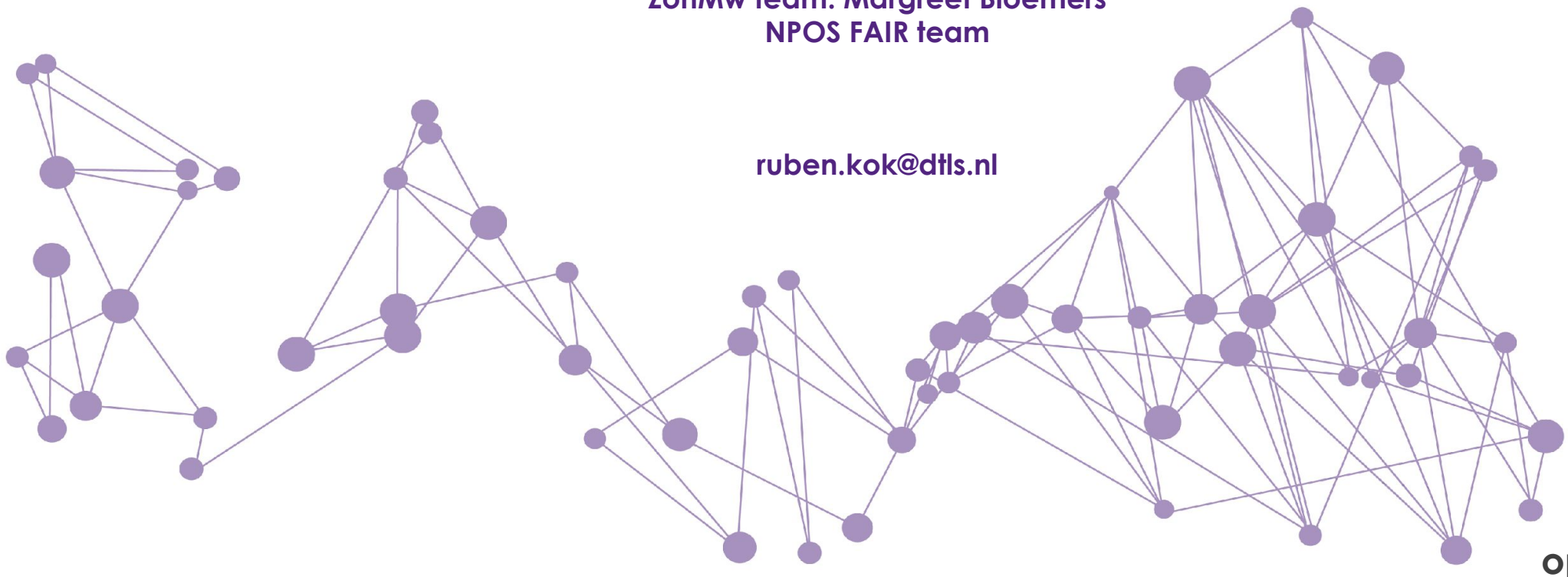
engagement, coordination and support

DEVELOPING A FAIR-BASED RESEARCH ECOSYSTEM IN NL

Acknowledgements

DTL/Health-RI team: Mijke Jetten, Celia van Gelder, Rob Hooft, Mascha Jansen
GO FAIR team: Barend Mons, Barbara Magagna
ZonMw team: Margreet Bloemers
NPOS FAIR team

ruben.kok@dtls.nl



dtls.nl
health-ri.nl
openscience.nl