

DCC voor FAIR

LCRDM netwerkdag, 04-02-2020, Laurents Sesink



**Universiteit
Leiden**
The Netherlands

FAIR principles

- 2016. Publicatie van de FAIR principes voor wetenschappelijke data en data stewardship.
- Verbeteren van de vindbaarheid, toegang, interoperabiliteit en hergebruik van digitale objecten.
- De principes benadrukken **machine leesbaarheid** omdat mensen in toenemende mate gebruik maken van algoritmes om zodoende gebruik te kunnen maken van de beschikbare hoeveelheid wetenschappelijke data.
- **Machine leesbaarheid** is nodig vanwege de toevloed van digitale data, de complexiteit van de data en de snelheid waarmee nieuwe data geproduceerd wordt.



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

Implementatie van de FAIR principes

Onderzoeksorganisaties hebben beleid of reglementen waarin de omgang met onderzoeksdata beschreven wordt. Voorbeeld Universiteit Leiden.

Voor

✓ Write a data management plan (DMP)

Tijdens

✓ Secure storage : integrity, availability, confidentiality

Na

- ✓ Data must remain available for **10 years**
- ✓ Data should be 'FAIR'
- ✓ Provide documentation, metadata, software for reuse
- ✓ Archive data in a certified archive

Onderzoeksorganisaties hebben een meerjaren implementatie programma RDM

Eerste prioriteit: **FAIR** -- data management

Doel: Voldoen aan wet (AVG en auteursrechten), en regelgeving (gedragscode wetenschappelijke integriteit), eisen van financiers (data delen) en bewustwording over vindbaarheid (**F**) en toegankelijkheid (**A**).

(Beginnend met mensen)

Diensten

- DMP tool, Affordable, secure, shared storage, archief, preserving

Ondersteuning

- Informatievoorziening, vraagafhandeling, advies, training en hands-on support
- DMP workshops, DPIA,

Kennis delen en verwerven door lokale experts voornamelijk op national niveau

- LCRDM, RDNL, NFU, DTL, UKB, etc



Ambitie: **FAIR** - data stewardship

Doel: Van data delen (Vindbaar **(F)** en Toegankelijk **(A)**)

naar gebruik van data (Interoperabel **(I)** en (her)bruikbaar **(R)**)

Kennis delen en verwerven door lokale experts voornamelijk op internationaal niveau.

- GOFAIR, RDA, EOSC, DARIAH, CLARIN, Elixir, CODATA, etc.)



CLARIN



Ondersteuning

- Informatievoorziening, vraagafhandeling, advies, training en hands-on support
- Top 10 FAIR Things (Bewustwording FAIR), Proof of Concepts met onderzoekers (Bring Your Own Data)
- Pilots met instituten, Advies aan faculteiten en instituten over FAIR protocollen, GOFAIR implementation network Matrix

Diensten?

Van FAIR naar FAIR

Vindbaar en Toegankelijk

- Generieke diensten.
- Generieke ondersteuning.

Interoperable en (her)bruikbaar

- Domein specifieke diensten.
- Domein specifieke ondersteuning.



Technical infrastructure (generic operations)
Data / social agreements (domain-specific content)

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

Volgende stap: FAIR data tijdens onderzoek



FAIR data by design

Ondersteuning bij:

- Data management
- Data stewardship
- Curation/wrangling of data

Een Leids voorbeeld

Leiden University Centre for Linguistics (LUCL)

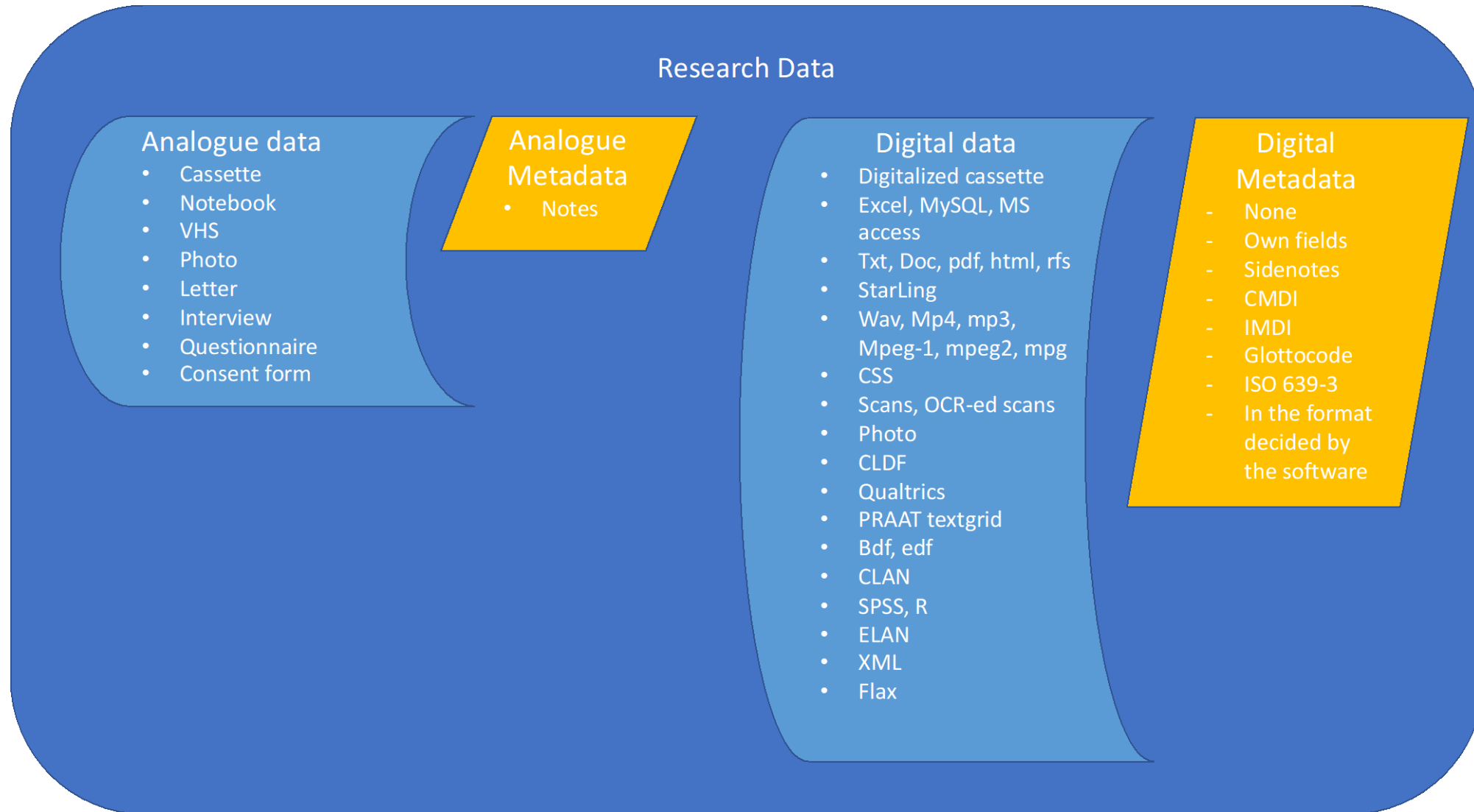
- Stap 1: Inventarisatie van data in het instituut.
- Stap 2: Vindbaar **(F)** en toegankelijk **(A)** maken van de data.
- Stap 3: Begrijpelijk **(I)** en herbruikbaar **(R)** maken van de data.

Stap 1: Inventarisatie LUCL

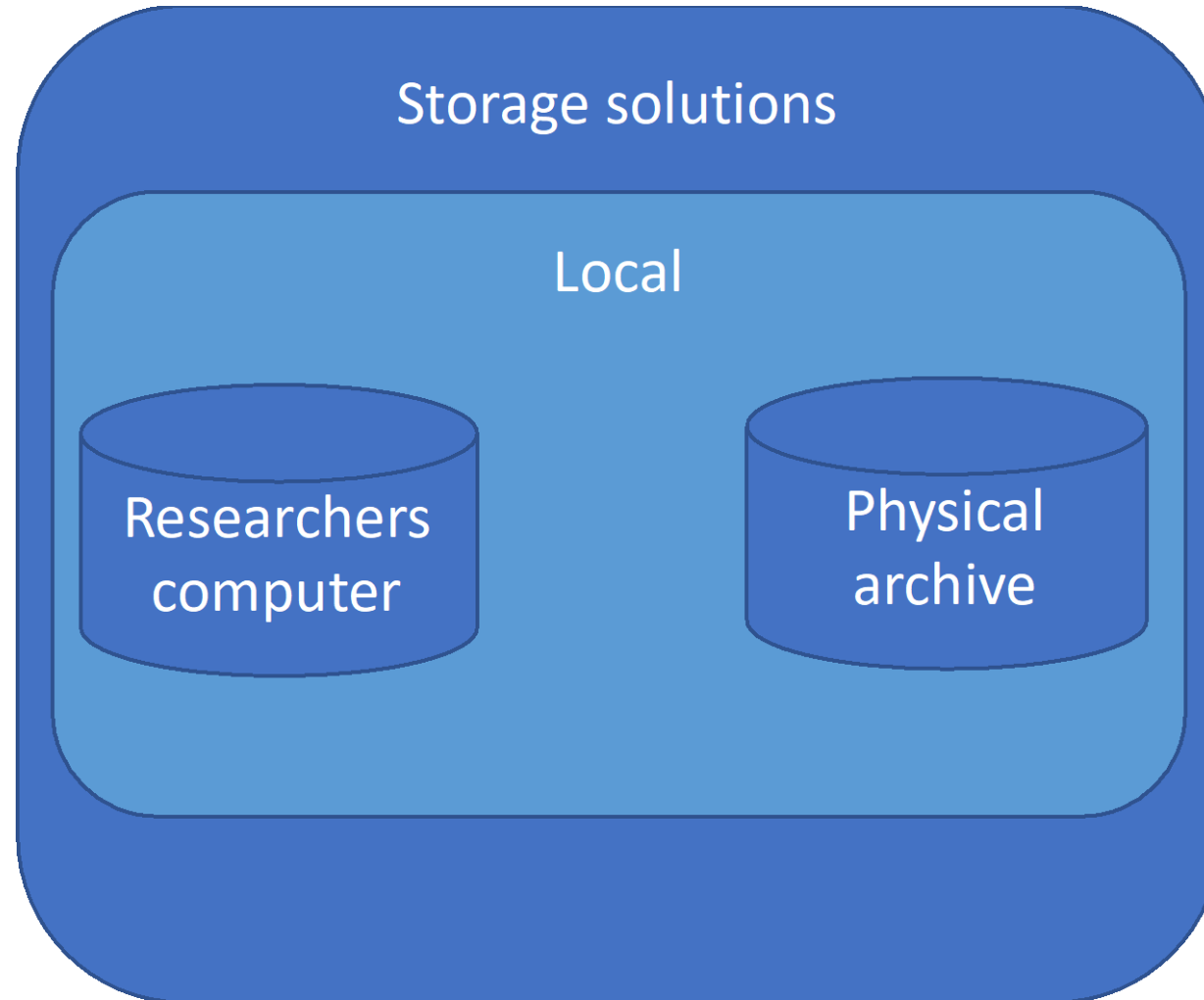
- Naam onderzoeker, taal, datatype, dataformaat, grootte, software tools, metadatastandaard, rechten, toegang, curatie nodig?

1	Researcher Name	Language(s)	Data type	Data format	Size	Software tools	Metadata stand.	Resource rights	Access	Curation need(s)
	Iraqw	Audio	Digital;	Unknown	No				N/A	Conversion to database
	Sereer	Audio	Digital;	Unknown	No				N/A	Conversion to database
	Yaaku	Audio	Digital; Digitalized cassette	4 cassettes	No				N/A	Conversion to database
	Konso	Audio	Digital; Digitalized cassette	9 cassettes	No				N/A	Conversion to database
	Mbugu	Audio	Digital; Digitalized cassette	5 cassettes	No				N/A	Conversion to database
	Seme	Audio	Digital; Digitalized cassette	3 cassettes	No				N/A	Conversion to database
	Toussian	Audio	Digital; Digitalized cassette	2 cassettes	No				N/A	Conversion to database
	Bété	Audio	Analogue; Cassette	7 cassettes	No				N/A	Digitalization; Conversion to database
	Mbugu	Audio	Analogue; Cassette	4 cassettes	No				N/A	Digitalization; Conversion to database
	Zigula	Audio	Analogue; Cassette	2 cassettes	No				N/A	Digitalization; Conversion to database
	Datooga	Audio	Analogue; Cassette	1 cassette	No				N/A	Digitalization; Conversion to database
	Tunen	Audio	Analogue; Cassette	5 cassettes	No				N/A	Digitalization; Conversion to database
	Minkana	Audio	Analogue; Cassette	9 cassettes	No				N/A	Digitalization; Conversion to database
	Iraqw	Written	Analogue; Notebook	around 60 notebooks	No				N/A	Scanning; Conversion to database
	Seme	Written	Analogue; Notebook	2 notebooks	No				N/A	Digitalization; Scanning; Conversion to database
	Alagwa	Written	Analogue; Notebook	10 notebooks	No				N/A	Digitalization; Scanning; Conversion to database
	Hieroglyphic Luwian	Written	Digital; Excel	48kb	No			All forms come from c	N/A	None
	Cuneiform Luwian	Written	Digital; Excel	207kb	Yes			All items taken from N	N/A	None
	Germanic languages	Written	Digital; StarLing	135 KB	No				Restricted	None
	Indo-European Language	Written	Digital; MySQL	8887 word entries	Yes			When relevant; refere	Open	None
	(Classical) Armenian	Written	Digital; Excel	around 200 KB	No				Other:	None
	Various	Written	Digital; .txt	7655 wordlists; 40 items each	Yes		Glottocode; ISO 639-3; etc.		Open	None
	Hieroglyphic Luwian	Written	Digital; Excel	212kb	Yes			All data taken from H	N/A	None
	Latin	Written	Digital; Excel	A few kilobytes	No				N/A	Conversion to database
	Old Khotanese	Written	Digital; Excel	364 entries	No				N/A	None
	Kelantan Malay	Written; Audio	Digital; .wav; .mp4	16GB	No				N/A	None

Inventarisatie LUCL

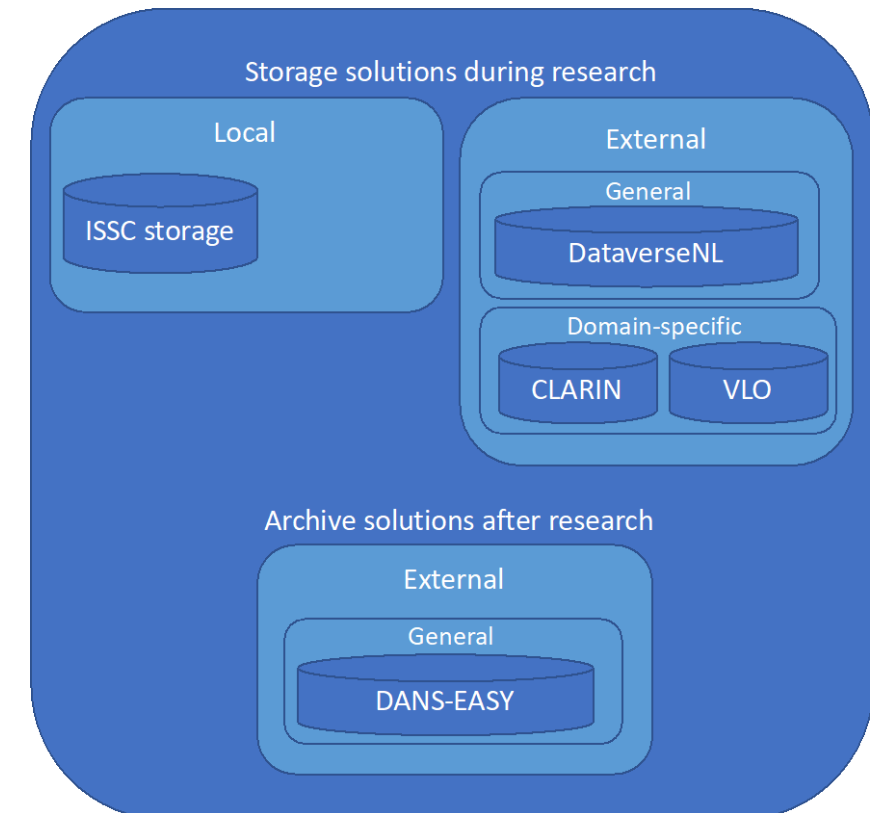


Inventarisatie LUCL



Stap 2: Vindbaar en toegankelijk maken

- Data vindbaar en toegankelijk binnen het LUCL tijdens/kort na onderzoek
 - **Minimale eisen.**
 - Alle data op ISSC storage.
 - Basis metadata, alle dataformaten geaccepteerd, eigenaar van de data, bewaartermijn, rechten.
- Data vindbaar en toegankelijk buiten het LUCL tijdens/kort na onderzoek
 - **Voorkeur voor het gebruik van standaarden.**
 - Voorkeurs dataformaten, DataverseNL metadata en persistent identifier
 - CLARIN Metadata en persistent identifier in CLARIN portaal
(Virtual Language Observatory)
- Data vindbaar en toegankelijk buiten het LUCL na afloop onderzoek
 - **Aansluiten bij discipline specifieke standaarden.**
 - Duurzame dataformaten, DANS-EASY metadata en persistent identifier
 - NARCIS en CLARIN metadata en persistent identifier in NARCIS, OpenAire, VLO



Stap 3: Begrijp- en herbruikbaar maken voor mens en machine

- Welke best practices/standaarden worden er binnen een wetenschappelijke discipline gebruikt?
- **UBL/CDS** werkt samen met **Go FAIR** voor het geven van advies over te gebruiken discipline specifieke standaarden en het opstellen van werkprocessen voor het begrijpelijk (**I**) en herbruikbaar (**R**) maken van de data.





Technical infrastructure (generic operations)
Data / social agreements (domain-specific content)

Box 2 | The FAIR Guiding Principles

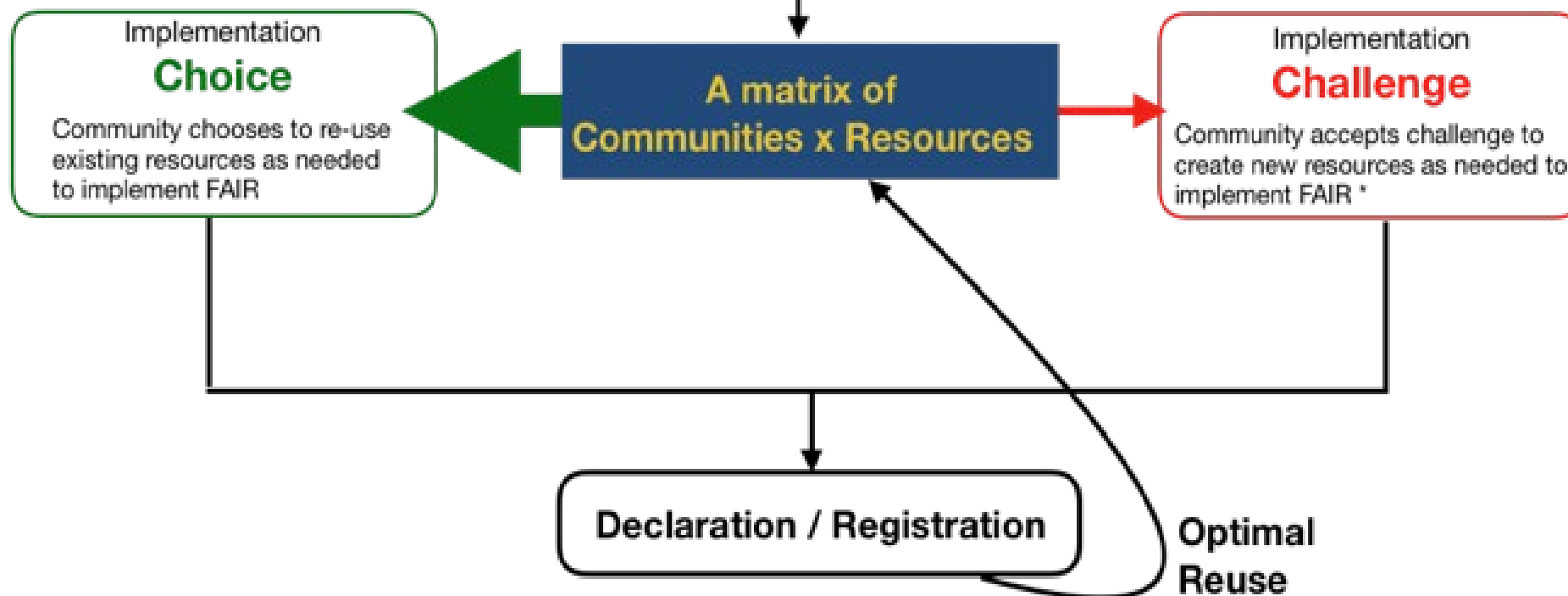
To be Findable:
 F1. (meta)data are assigned a globally unique and persistent identifier
 F2. data are described with rich metadata (defined by R3 below)
 F3. metadata clearly and explicitly include the identifier of the data it describes
 F4. (meta)data are registered or indexed in a searchable resource

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

Communities



Convergence Matrix & FAIR Implementation Profiles

		Communities									
		A	B	C	D	E	F	G	H	I	J
Resources	F	Resource 1	0	1	0	0	0	0	0	0	0
	F	Resource 2	1	1	1	0	1	1	1	1	0
	A	Resource 3	1	1	0	0	0	0	0	1	0
	A	Resource 4	0	0	1	1	1	1	1	1	1
	I	Resource 5	1	0	0	0	0	0	0	1	0
	I	Resource 6	0	1	1	0	1	1	1	1	0
	R	Resource 7	1	0	0	0	0	0	0	0	0
	R	Resource 8	1	0	1	0	1	1	1	1	0

FIP for community C = {0,1,0,1,0,1,0,1}

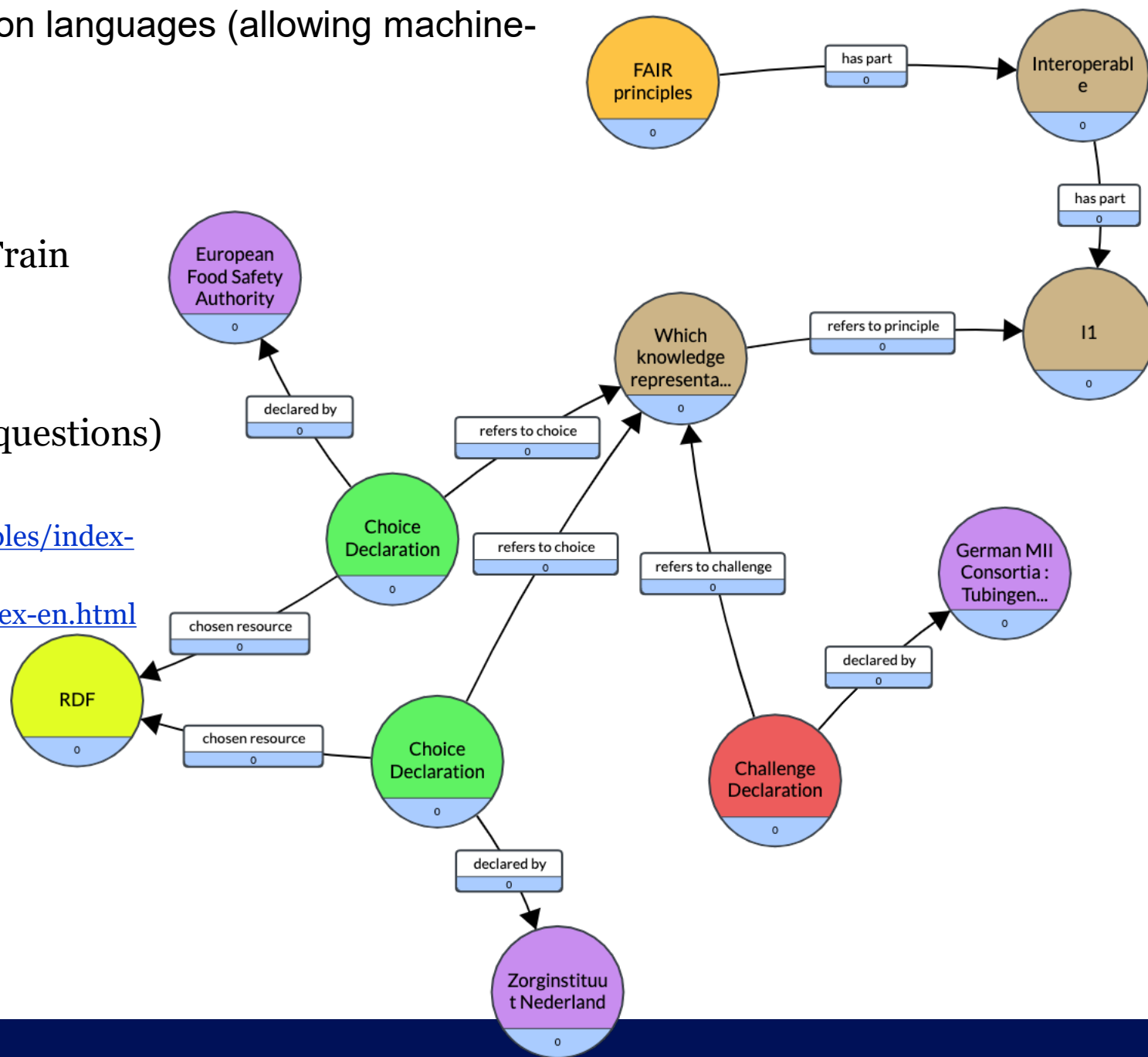
“Which knowledge representation languages (allowing machine-interoperation) do you use?”

- FAIR Data Trains
- Resource chosen by more than one Train
- Choice declaration
- Challenge declaration
- FAIR principles concept
- I principles and related choices (i.e. questions)

Ontologies used:

<https://peta-pico.github.io/FAIR-nanopubs/principles/index-en.html>

<https://peta-pico.github.io/FAIR-nanopubs/icc/index-en.html>

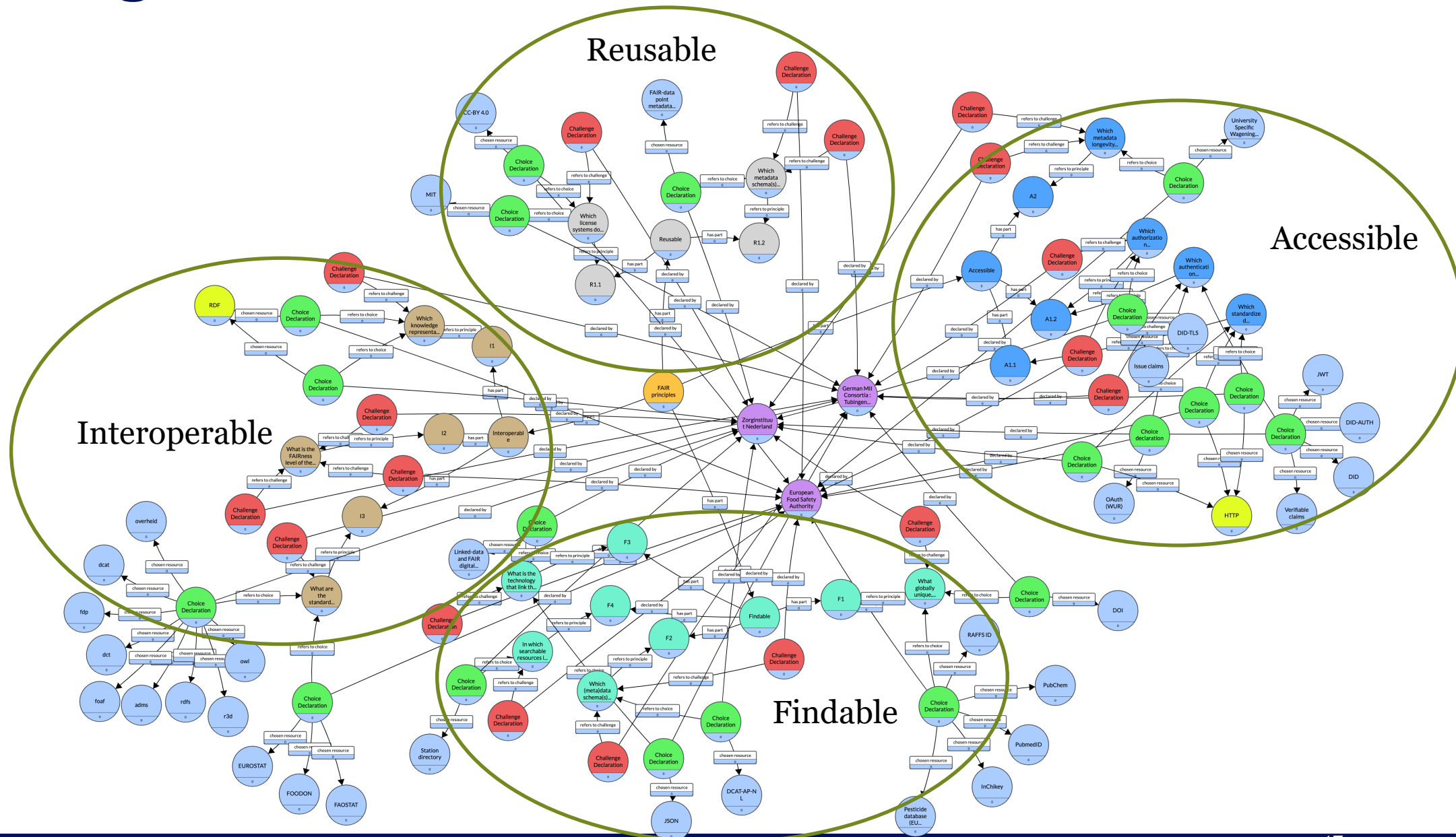


Made with:



<https://gra.fo>

Convergentie van FAIR Data INs



Uitdagingen FAIR implementatie

Onderzoeksorganisaties bieden onderdak aan veel verschillende wetenschappelijke disciplines, studenten, promovendi, Early career researcher, hoogleraren.



Faculteiten



Studenten



PhDs



Hoogleraren



Jaarlijkse
NWO subsidies



Jaarlijkse ERC
subsidies

Datamanagement/data stewardship is onderdeel van een pallet aan onderzoeks ondersteuning.

- Subsidieadvies
 - Financieel projectbeheer
 - Projectmanagement
 - Juridische ondersteuning
 - Impact / Valorisatie
 - Datamanagement
- Ethiek
 - Informatiebeveiliging en privacy
 - Advisering Publiceren en Open Access
 - Ondersteuning lobby en beïnvloeding financiers
 - Wetenschapscommunicatie
 - Apparatuur en fac. specifieke ondersteuning en overige ICT



Onderzoeksorganisaties zijn complexe bestuurlijke organisaties.

Faculty				
Faculty secretary	Institute manager	Information manager	Ethics committee	Leiden Centre of Data Science

Administration and Central Services					
IM	SCM	Procurement	Data protection officer	Security officer	HR

Expertise Centres							
Luris			University Library			ICT Shared Service Centre	
Knowledge partnership	Grant development	Legal affairs	UBL	AIP	CDS	ISSC	

Leiden Data Competence Center

Universitaire Bibliotheken Leiden - CDS

- FAIR data management/stewardship
- Netwerk van data-managers/stewards

Bestuurs bureau - Information Management

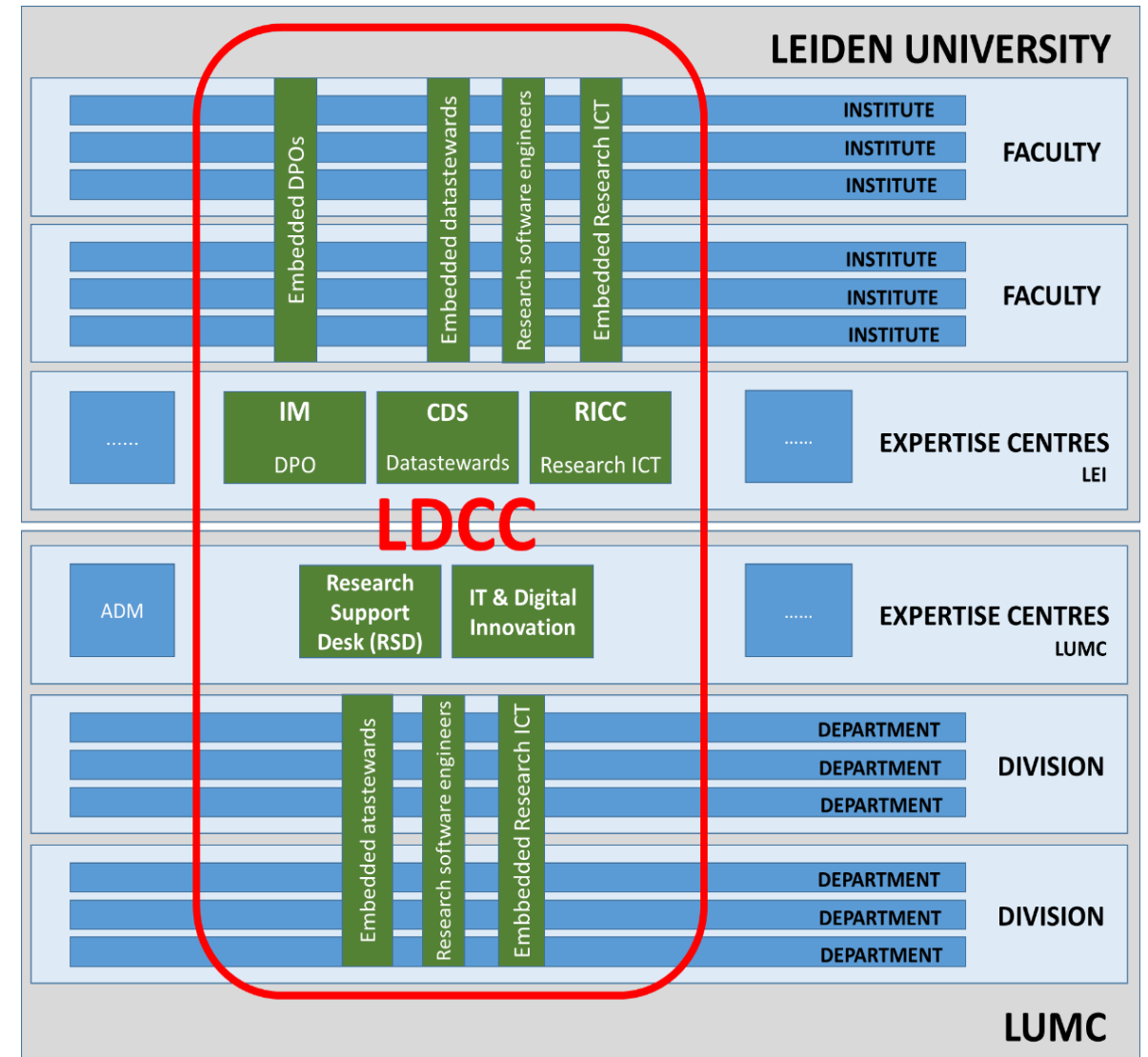
- Informatie beveiliging (CISO, DPO)
- Netwerk van DPOs

ISCC- Research IT Competence Centre

- ICT-diensten voor onderzoek (opslag en HPC)
- Netwerk van Research ICT

Faculteiten en instituten

- Embedded DPO, datasteward, research software engineers, research ICT



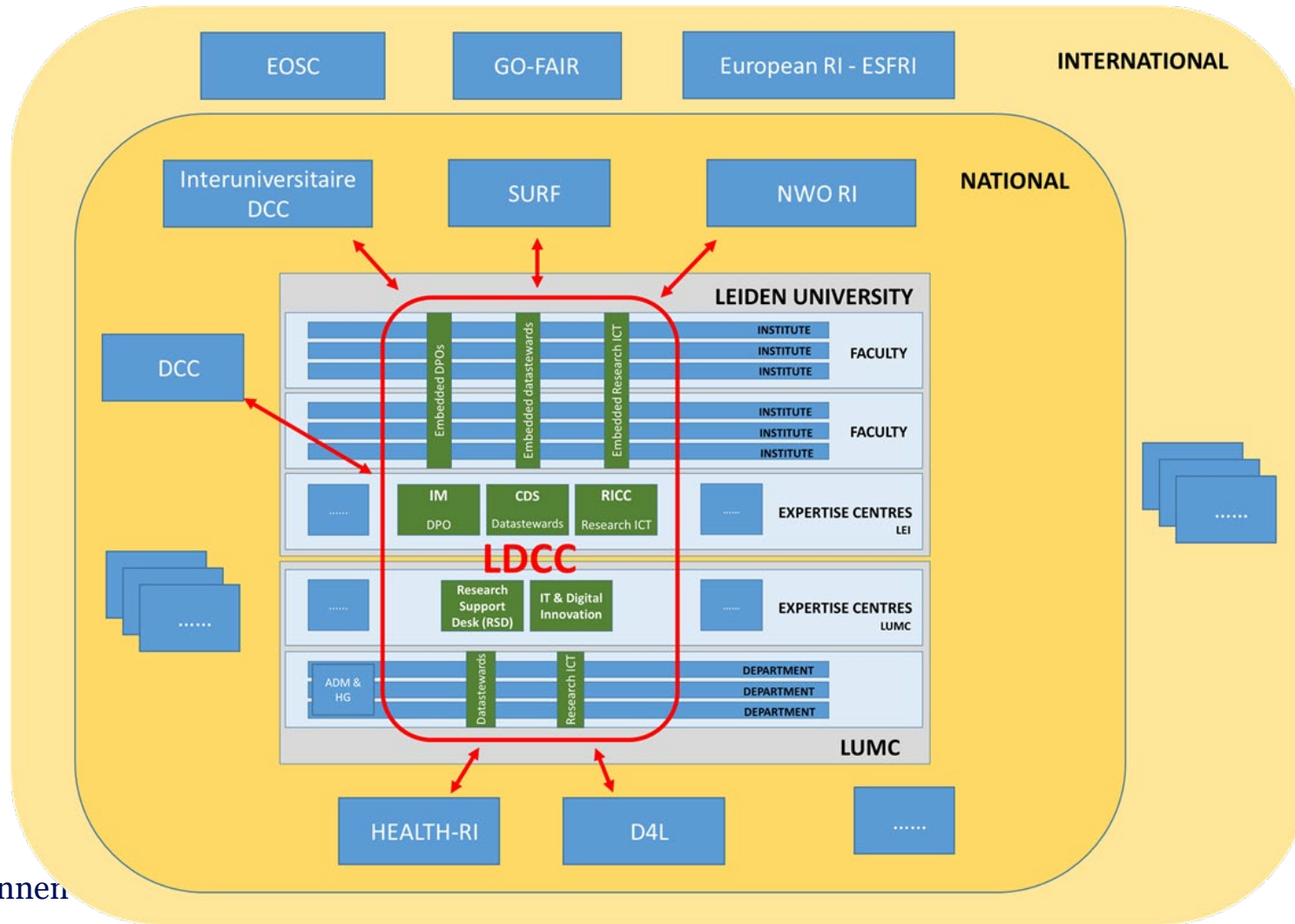
Leiden DCC & inter university DCCs

Door middel van de inrichting van voorzieningen onderzoekers in staat stellen om data te beheren zodat voldaan wordt aan de **FAIR** principes van het Open Science programma.

Zorgplicht **FAIR**
databeheer
Gedragscode
Wetenschappelijk
Integriteit.

Het organiseren van de Data Stewardship organisatie in de instelling.

Aanspreekpunt voor alle vragen mbt **FAIR** data binnen de lokale organisatie.



Uitwisselen van kennis mbt **FAIR** data ontwikkelingen en Data infrastructuur binnen het netwerk.

Het delen van succesvolle strategieën en ontwikkelingen die leiden tot gedragsveranderingen in de transitie naar **FAIR** data.

Aanspreekpunt voor externe partijen die op het gebied van data met de instelling zou willen samenwerken en deze samenwerking bevorderen.

Questions?



Universiteit
Leiden
The Netherlands