



Preconference

WHAT IS METADATA? COMMON STANDARDS AND PROPERTIES

9 November | 13:30-17:00 CET



BERLIN | 9-12 NOVEMBER 2022



Pétronille BOGAERT
Sciensano



Truls KORSGAARD
*Norwegian
Directorate for e-
Health*



Hanna TOLONEN
*Finnish Institute for
Health and Welfare*



Irini KESSISOGLU
Sciensano



**Susanna-Assunta
SANSONE**
ELIXIR



Pascal DERYCKE
Sciensano



PHIRI

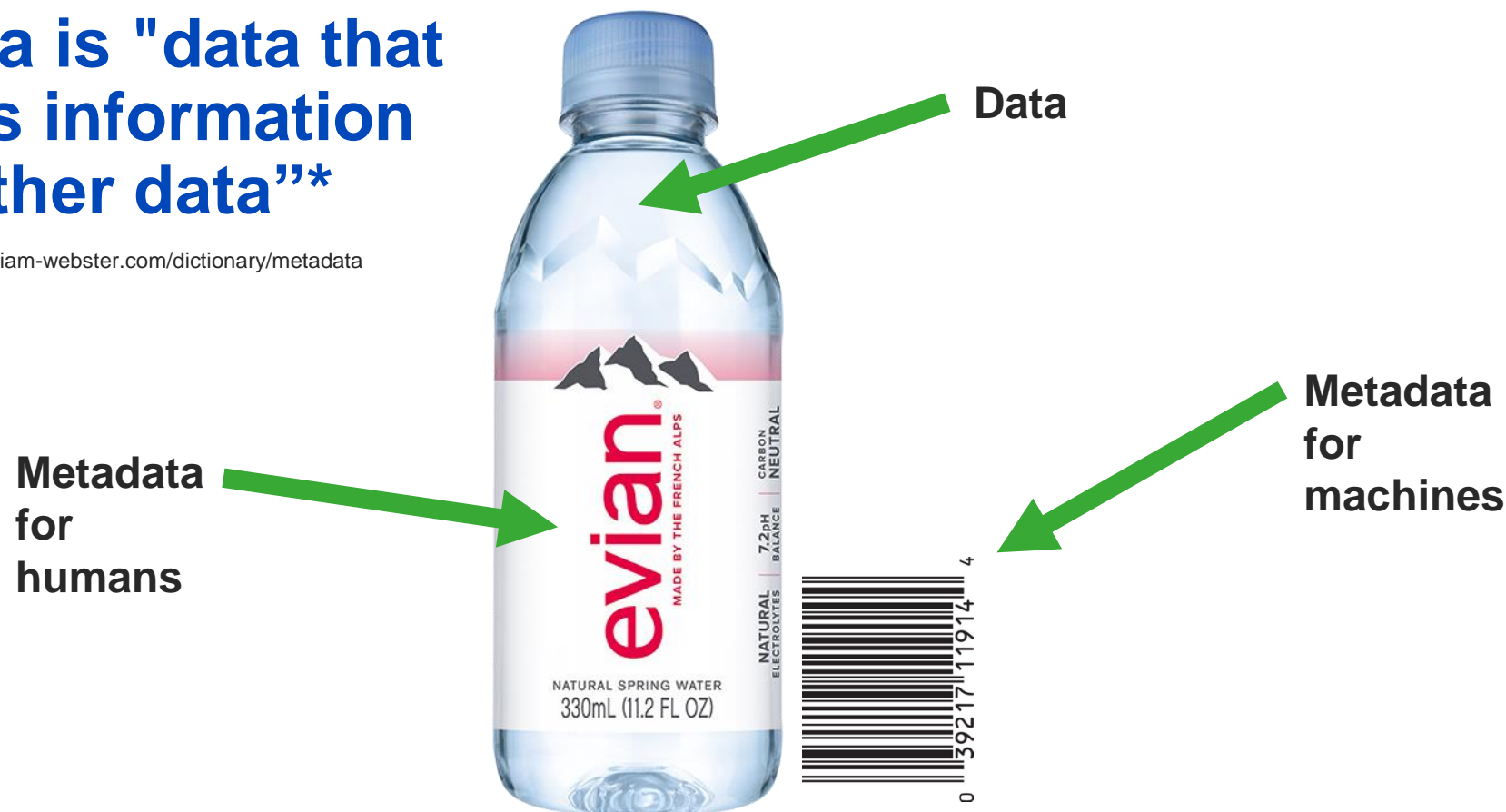
Population Health Information
Research Infrastructure



What is metadata?

Metadata is "data that provides information about other data"*

* Source: <https://www.merriam-webster.com/dictionary/metadata>



Standards for dummies



Standards used to structure metadata

Descriptive metadata standards:

Descriptive information about a resource. Used for discovery and identification. It includes elements such as title, abstract, author, and keywords.

- **DCAT-AP**

TABLE 3.15 : PATIENT-DISCHARGES-MEAN-AND-MEDIAN-LENGTH-OF-STAY-DAYS-BY-PRINCIPAL-PROCEDURE-SEX-AND-AGE-GROUP-2017

Presents in-patient mean and median length of stay for principal procedure by sex and age group. This measure includes pre-operative and post-operative length of stay. It should also be noted that this analysis by mean length of stay does not take into account the status of the patient on discharge. Activity in Acute Public Hospitals in Ireland Annual Report 2017, is a report on in-patient and day patient discharges from acute public hospitals participating in the Hospital In-Patient Enquiry (HPIE) scheme in 2017. Discharge activity is examined by type of patient (day patient/in-patient), admission type (elective/emergency/ambulatory) and hospital group, and by demographic parameters (such as age and sex). Particular issues of relevance to the Irish health care system covered in the report relate to the composition of discharges by medical card and public/private status. Discharges are also analysed by diagnoses, procedures, major diagnostic categories, and diagnosis related groups. The analysis is presented at the national level. In 2017 HPIE discharges were coded using ICD-10-AMACHACS 8th Edition and grouped into AHS-DRG Version 8.0. See the complete Activity in Acute Public Hospitals in Ireland Annual Report 2017 at <http://www.hse.ie/healthcare/annual-report-2017>

DATA AND RESOURCES

patient-discharges-mean-and-median-length-of-stay-2017

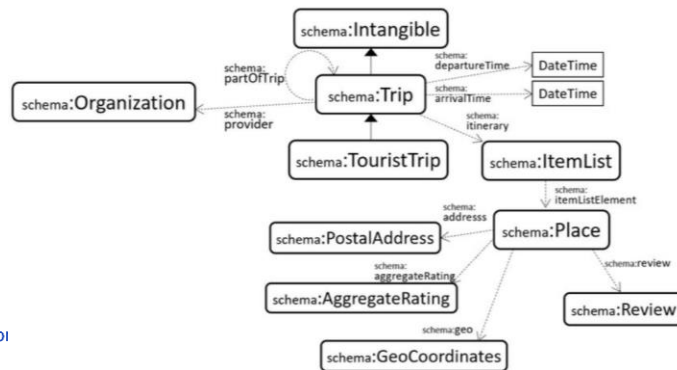
activity acute a-drugs casemix day patient diagnoses discharge discharges hpi hospital hospitals icd-10-am in-patient principal diagnosis principal procedure procedures achi

Explore

Annotations:

- @prefix dcat: <http://www.w3.org/ns/dcat#>
- @prefix dct: <http://purl.org/dc/terms/>
- dct:title (@datasets)
- dct:description
- dct:distribution
- dct:title
- dct:accessURL
- dct:identifier
- dct:description
- dct:format
- dct:issued

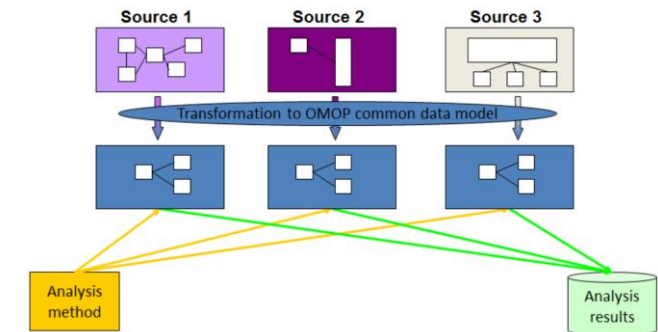
- **Schema.org**



Structural metadata standards:

- **OMOP**

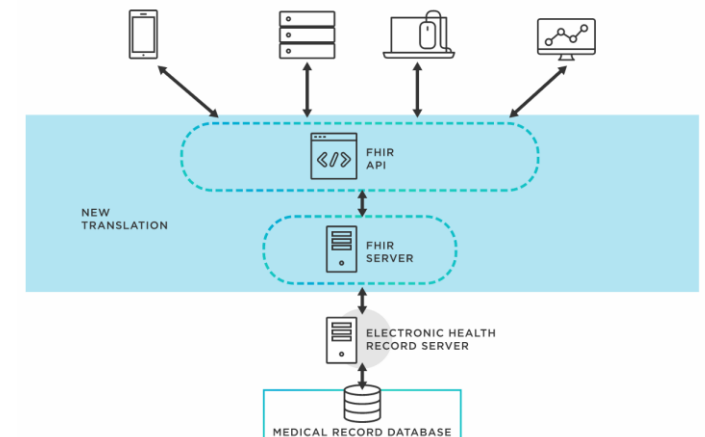
A common way to describe datasets at variable level



Standards used for data exchange:

- **HL7 FHIR**

To enable health data, including clinical and administrative data, to be quickly and efficiently exchanged electronically.



Standards used to structure data

Standards used for semantic interoperability/ontology:

- ICD9-11 (international classification of diseases)
- SNOMED-CT
- LOINC
- ORPHACODE: standardising ontology for rare diseases

Standards used to structure images:

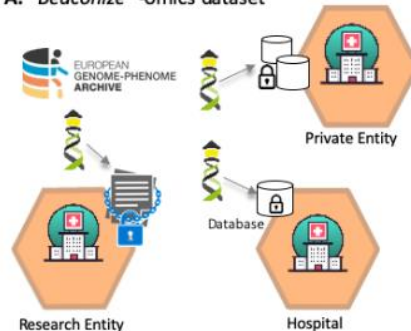
- DICOM



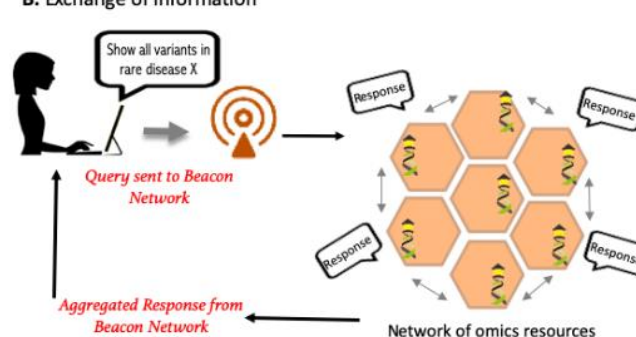
Standards used to make genomic data discoverable:

- BEACON

A. "Beaconize" -omics dataset



B. Exchange of information



Agenda

Setting the scene and Inspirational cases

- **13.40 - 14.20:** **What is metadata? What are common standards and their properties? Why is metadata important for researchers?**
 - **Truls Korsgaard** Norwegian Directorate for e-Health, Oslo, Norway
- **14.20 - 14.40:** **Why we need metadata for public health data sources?**
Illustrated with the European Health Information Portal and the ambition of the European Health Data Space (EHDS2 pilot).
 - **Hanna Tolonen**, Challenges and Solutions on Population Health and Welfare, Finish Institute for Health and Welfare, Helsinki, Finland
 - **Petronille Bogaert**, EU Health Information System Unit, Sciensano, Brussels, Belgium
- **14.40 - 15.10:** **Break**

Agenda

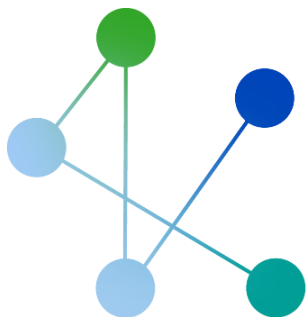
Setting the scene and Inspirational cases

- **15.10 - 15.30:** **Assessing FAIRness levels of health data within HealthyCloud**
 - **Irini Kesisoglou**, EU Health Information System Unit, Sciensano, Brussels, Belgium
- **15.30 - 15.45:** **FAIRsharing and the FAIR Cookbook: helping you choose and use metadata standards**
 - **Susanna-Assunta Sansone**, ELIXIR

Agenda

Practical exercise

- **15.45 - 16.50:** **Create your own metadata documentation using DCAT standard**
 - **Group work facilitated by: Pascal Derycke**, EU Health Information System Unit, Sciensano, Brussels, Belgium
- **16.50 - 17.00:** **Wrap up and key messages**
 - **Petronille Bogaert**, EU Health Information System Unit, Sciensano, Brussels, Belgium



PHIRI

Population Health Information
Research Infrastructure

Contact us: PHIRI.coordination@sciensano.be

Follow us on Twitter:  @PHIRI4EU
 @EU_HIS_unit
 @PetronilleBo
 @PHMRsection



www.phiri.eu



BERLIN | 9-12 NOVEMBER 2022



This project has received
funding from the European
Union's Horizon 2020
research and innovation
programme under grant
agreement No 101018317

What is metadata?

Why is metadata important for researchers?

Why is common standards important?



What is metadata?

Why is metadata important for researchers?

Why is common standards important?



What is metadata?

- Metadata is information about for example a health registry.
For example:
 - Title
 - Description
 - Contact point
 - Keywords
 - Coverage
- Metadata helps you to find the registry you are looking for
 - Filtering
 - Browsing
- And helps you to discover more information about the registry

Search gave 73 hit(s) among datasets, 3 hit(s) among APIs, 46 hit(s) among concepts 5 hit(s) among information models

Norwegian Cause of Death Registry

All (127) Datasets (73) APIs (3) Concepts (46) Information models (5) Services and events (0)

Sorted by relevance Sort by last published

Theme

- ☐ Building and property (5)
- ☐ Business (5)
- ☐ Family and children (3)
- ☐ Work (2)
- ☐ Democracy and citizen rights (1)
- [Show more](#)

EU-theme

- ☐ Science and technology (34)
- ☐ Government and public sector (19)
- ☐ Health (9)
- ☐ Regions and cities (8)
- ☐ Population and society (4)
- [Show more](#)

Access

- ☐ Open data (48)
- ☐ Public (56)
- ☐ Non public (11)
- ☐ Restricted (5)
- ☐ Unknown (1)

Owner

Search for owner

- ☐ Stat (70)
- ☐ Kommune (2)
- ☐ Privat (1)

Norwegian Cause of Death Registry
Dataset

Owner: Direktoratet for e-helse nb en

The Norwegian Cause of Death Registry covers deaths amongst those registered as inhabitants of Norway, and include history from 1951. Deaths of non-citizens is included after medio-2012. The registry is the source of the official cause of death statistics for Norway.

Health

csv

Brønnbaner
Dataset

Owner: Oljedirektoratet no

Dataset is public

Brønnbaner

Energy Regions and cities

vnd.ahp

NorNE - Norwegian Named Entities
Dataset

Owner: Nasjonalbiblioteket nb en

Dataset is public

NorNE (Norwegian Named Entities) is a text corpus composed of the same texts as the Norwegian Dependency Treebank (NDT), but is in addition tagged with named entities. The corpus contains approx. 300,000 words of running text for Norwegian Bokmål and Norwegian Nynorsk, respectively.

[Show full description](#)

How is metadata?

- In our context metadata is information about a datasource not the micro-data itself
- Metadata may be structured or not structured, or a combination of both
 - Like a health record
- Metadata may be standardized or not, or partly
- Metadata may be machine-readable or not
 - Excel, CSV, PDF files for downloading is not machine-readable
 - Standardized JSON-files provided through an API-interface or an URL are machine-readable
- **FAIR metadata have to be based on common standards, controlled vocabularies and be machine-readable**
 - **Or there will be too much work.....**

Description

The Norwegian Cause of Death Registry covers deaths amongst those registered as inhabitants of Norway, and include history from 1951. Deaths of non-citizens is included after medio-2012. The registry is the source of the official cause of death statistics for Norway.

Distributions (1)

Norwegian Cause of Death Registry	
Format	csv
AccessURL	http://helsedata.no
License	https://lovdata.no/dokument/NL/lov/2014-06-20-43
Description	Helsedata.no provides guidance and tools to access high quality health data for research and innovation purposes.

Dataset usage

[More information about this dataset](#)

Provenance

Update frequency	annual
------------------	--------

Issue date 01.01.1928

Keywords

Health, Health registry, Health data

Restrictions

Spatial <https://publications.europa.eu/resource/authority/country/NOR>

Contact information

Contact point	https://helsedata.no/no/forvaltere/folkehelseinstituttet/dsarsaksregisteret/
---------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Data Community

This dataset has no mentions in our data community [🔗](#)

Do you have questions or comments regarding the dataset?

The comment will be visible for others and may be answered here.

[Log in to post a question](#)

dct:keyword

Property	URI	Range	ReqLevel	Card
contact point	dcat:contactPoint	vcard:Kind	M	1..n
description	dct:description	rdfs:Literal	M	1..n
identifier	dct:identifier	rdfs:Literal	M	1..n
publisher	dct:publisher	foaf:Agent	M	1..1
Title	dct:title	rdfs:Literal	M	1..n
dataset distribution	dcat:distribution	dcat:Distribution	R	0..n
keyword/tag	dcat:keyword	rdfs:Literal	R	0..n
landing page	dcat:landingPage	foaf:Document	R	0..n
release date	dct:issued	rdfs:Literal (typed as as xsd:date , xsd:dateTime , xsd:gYear Or xsd:gYearMonth)	R	0..1
spatial/ geographical coverage	dct:spatial	dct:Location	R	0..n
temporal coverage	dct:temporal	dct:PeriodOfTime	R	0..n
theme/category	dcat:theme	skos:Concept	R	0..n
update/ modification date	dct:modified	rdfs:Literal (typed as as xsd:date , xsd:dateTime , xsd:gYear Or xsd:gYearMonth)	R	0..1
access rights	dct:accessRights	dct:RightsStatement	O	0..1
conforms to	dct:conformsTo	dct:Standard	O	0..n
documentation	foaf:page	foaf:Document	O	0..n
frequency	dct:accrualPeriodicity	dct:Frequency	O	0..1
image	schema:image	schema:url Or schema:ImageObject	O	0..3
is referenced by	dct:isReferencedBy	rdfs:Resource	O	0..n
language	dct:language	dct:LinguisticSystem	O	0..n
qualified attribution	prov:qualifiedAttribution	prov:Attribution	O	0..n
qualified relation	dcat:qualifiedRelation	dcat:Relationship	O	0..n
related resource	dct:relation	rdfs:Resource	O	0..n

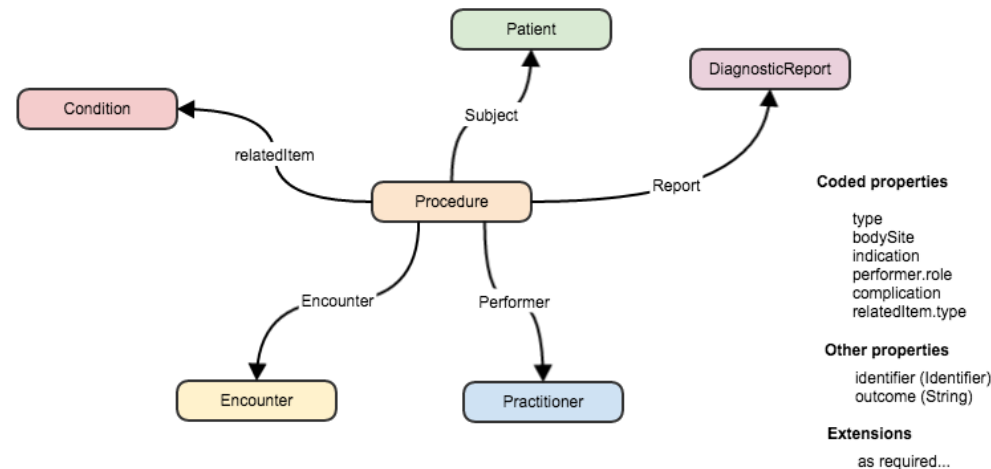
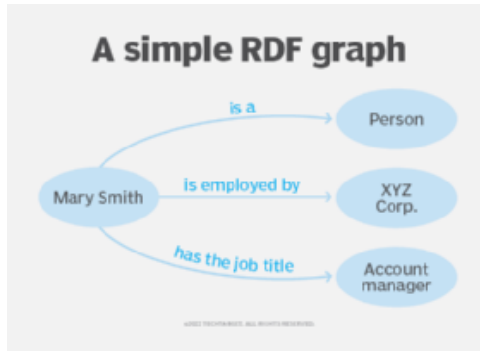
What is DCAT?

Data Catalog Vocabulary (DCAT) - Version 2

W3C Recommendation 04 February 2020

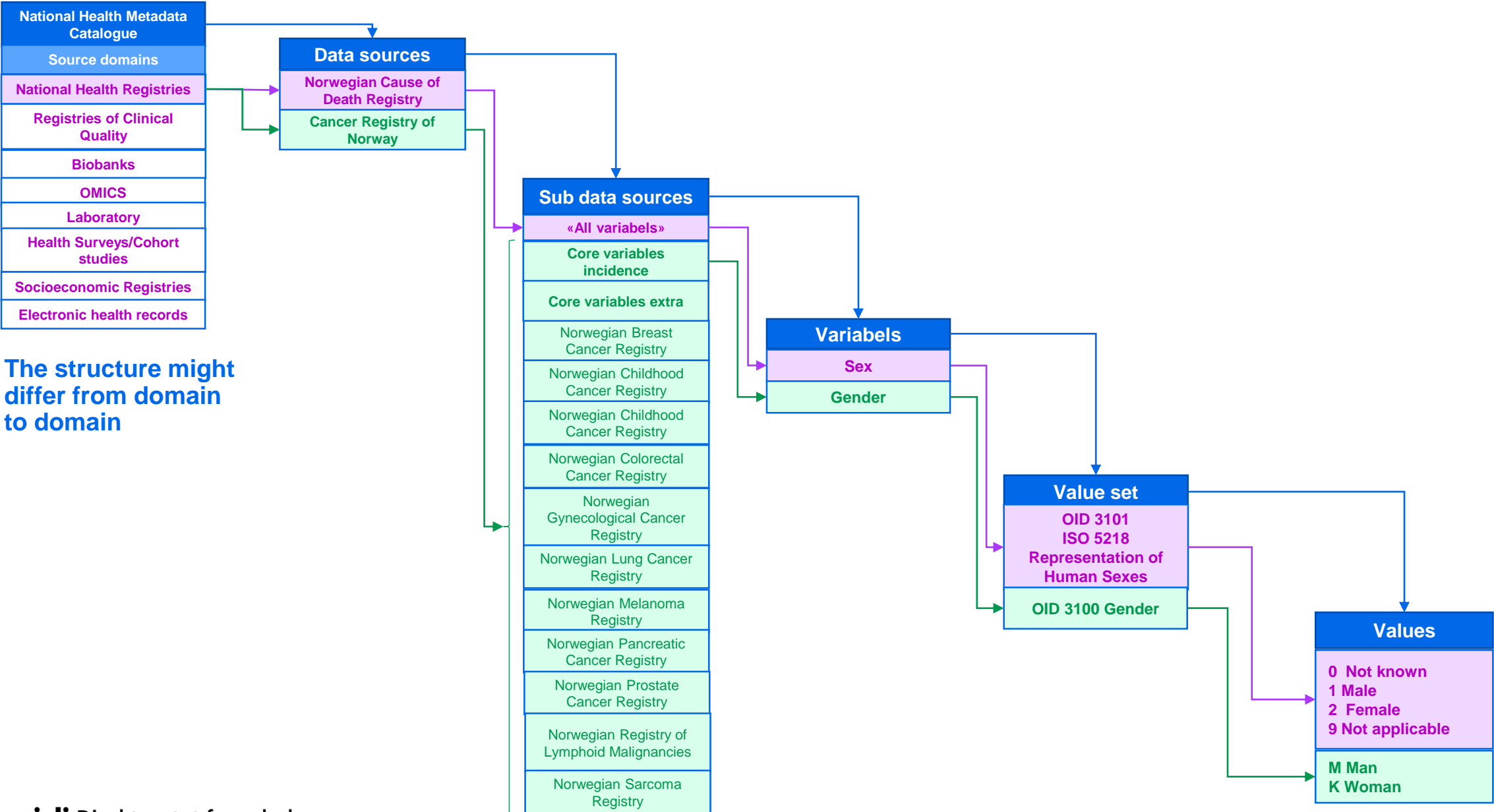
DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. This document defines the schema and provides examples for its use.

DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs. This can increase the discoverability of datasets and data services. It also makes it possible to have a decentralized approach to publishing data catalogs and makes federated search for datasets across catalogs in multiple sites possible using the same query mechanism and structure. Aggregated DCAT metadata can serve as a manifest file as part of the digital preservation process.



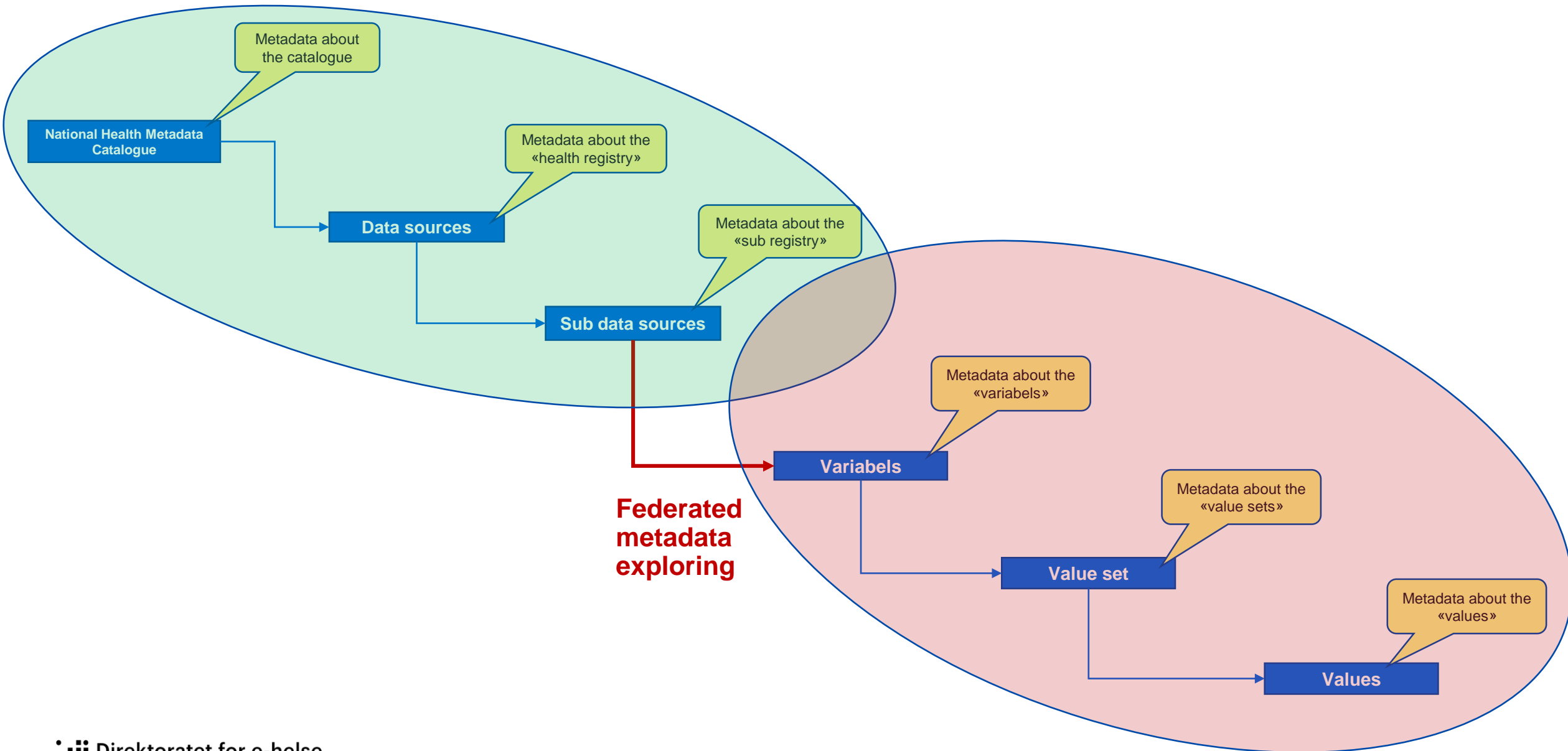
- **RDF vocabulary**
- Interoperability between catalogs
- Describe datasets and data services
- Vocabulary (controlled)
- Decentralized approach and federated research across catalogs
- Using the same query mechanism and structure
- **This can increase the discoverability**


Common understanding of the break down structure of data sources



The structure might differ from domain to domain

DCAT covers (so far) metadata on register and sub-register level (about datasets)





Norwegian Cause of Death Registry

Dataset description published in Felles datakatalog on 08.06.2022

Owner: Direktoratet for e-helse

Metadata quality: 49 %

Non public data Health

Description

The Norwegian Cause of Death Registry covers deaths amongst those registered as inhabitants of Norway, and include history from 1951. Deaths of non-citizens is included after medio-2012. The registry is the source of the official cause of death statistics for Norway.

Distributions (1)

Norwegian Cause of Death Registry	
Format	csv
AccessURL	http://helsedata.no
License	https://lovdata.no/dokument/NL/lov/2014-06-20-43
Description	Helsedata.no provides guidance and tools to access high quality health data for research and innovation purposes.

Dataset usage

[More information about this dataset](#)

Provenance

Update frequency	annual
Issue date	01.01.1928

Keywords

[Health](#), [Health registry](#), [Health data](#)

Restrictions

Spacial	https://publications.europa.eu/resource/authority/country/NOR
---------	-------------------------------------------------------------------------------------------------------------------------------------------

Contact information

Contact point	https://helsedata.no/no/forvaltere/folkkehelseinstituttet/dsarsaksregisteret/
---------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Data Community

This dataset has no mentions in our data community

Do you have questions or comments regarding the dataset?

The comment will be visible for others and may be answered here.

[Log in to post a question](#)

Norwegian Cause of Death Registry

[Norwegian Institute of Public Health](#) • [National health registry](#) • 1951 -


The Norwegian Cause of Death Registry contains information on deaths and causes of death in Norway from 1951 until today. The registry is an important source of information about the state of health and mortality in the population, and about the changes in causes of death over time.


Other Physical health

Contents

- Variables
- Criteria for data access
- Apply for access to data
- Contact

Variables

 Go to the variable overview to create variable lists and to see detailed information about the variables in this data source.

[View variables from data source](#) 


Criteria for data access

The data sources have different purposes and are regulated by different laws and regulations. In order to access information from the data sources, what you plan to use the information for must be in accordance with the purpose of the data source.

The types of approvals and documents you must submit to access information depend on what you are applying for, what you are going to use the information for and how you are going to process it. We therefore recommend that you take the time to familiarize yourself with the application guides before starting the application process:


[Application guide for anonymous, aggregated data \(statistical data\)](#)
[Application guide for personally identifiable data](#)

Apply for access to data


 You can apply for access to data from the registry via the application forms at [helsedata.no](#).

[+ Create new application](#)

Contact

 datatilgang@hi.no

LINES

[Data source homepage](#) 

Variables

Here you can create variable lists, download them and use them in applications. Information about variables is currently only available in Norwegian.

[Reset all filters](#)

Type of data source

Data sources

- ☐ The whole data register (11)
- ☒ The cause of death register (39)
 - ☐ All variables (39)
 - ☒ Deceased (8)
 - ☐ Data source (2)
 - ☐ Deaths (19)
 - ☐ Geography (7)
 - ☐ Autopsy (3)
- ☐ The Norwegian Armed Forces' health register (8)
- ☐ The health archive register (16)
- ☐ The Heart and Vascular Register (81)

Find variables (8 hits)

[Log in to save variables](#)

Variable	Data source	Data collection	Variable group
Identity	The cause of death register	All variables	The deceased
Resident in Norway at the time of death	The cause of death register	All variables	The deceased
Identifier type for identity	The cause of death register	All variables	The deceased
Sex	The cause of death register	All variables	The deceased
Date of death	The cause of death register	All variables	The deceased
Date of birth	The cause of death register	All variables	The deceased
Age in years	The cause of death register	All variables	The deceased

What is metadata?

Why is metadata important for researchers?

Why is common standards important?

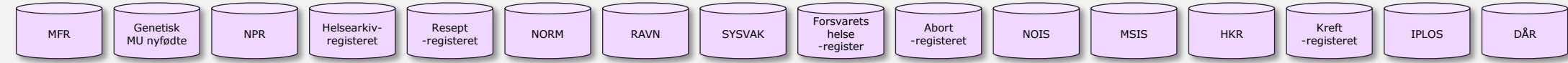


Makes it easier to find the right data source for your purpose

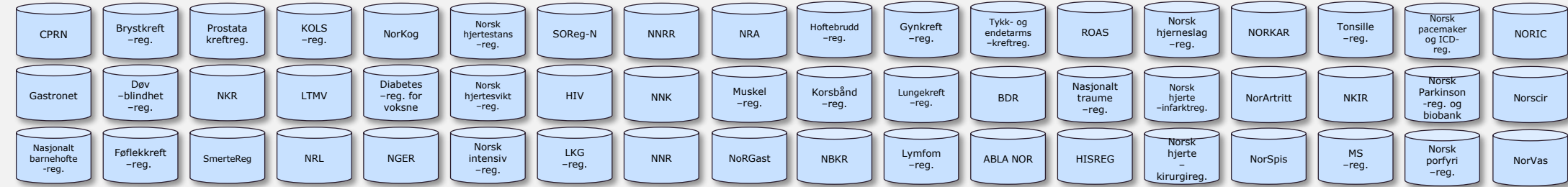
Patient journals



National/Central health registers



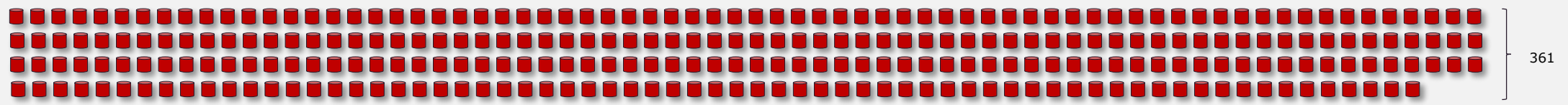
National medical quality registers



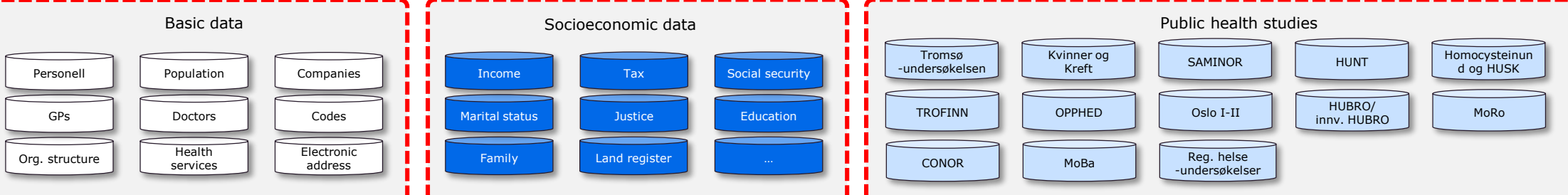
Other medical quality registers



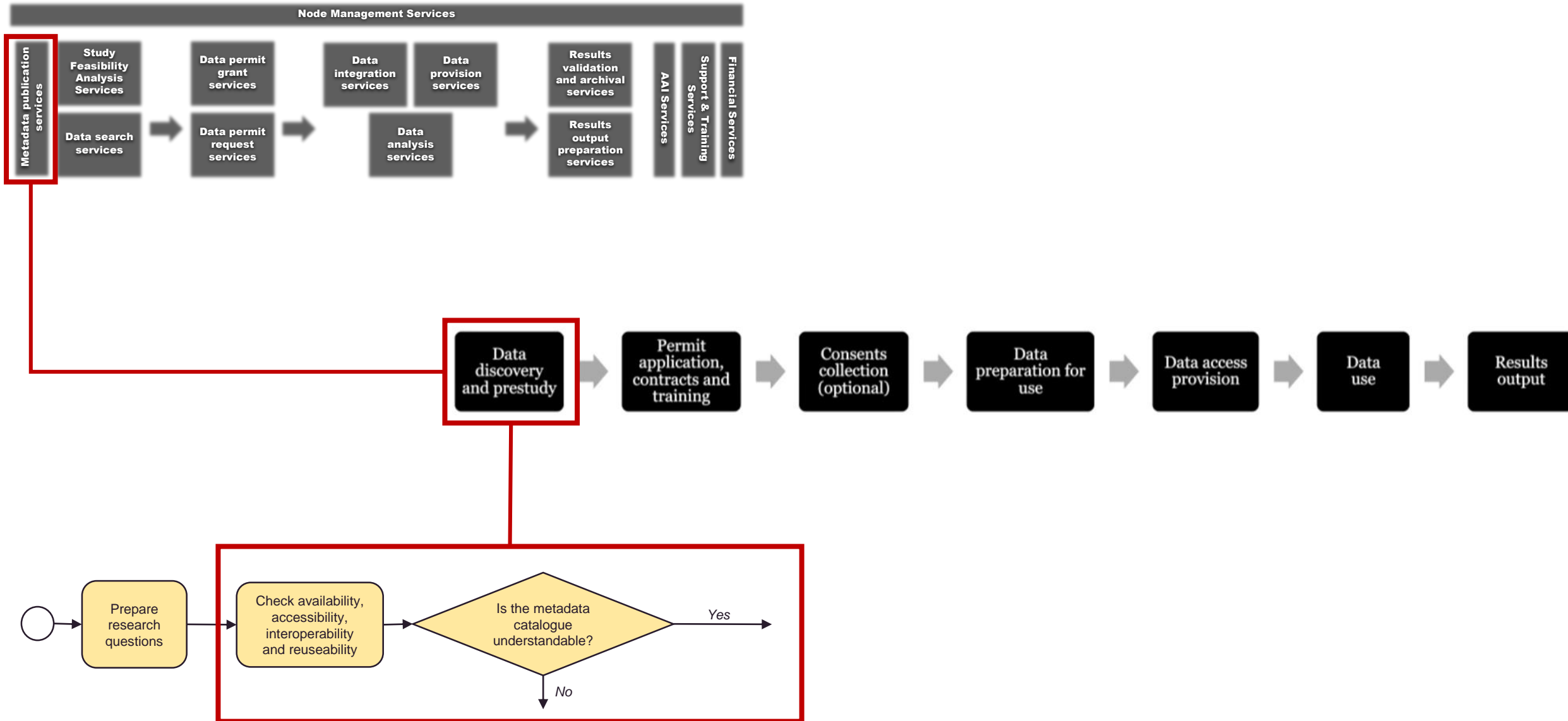
Biobanks



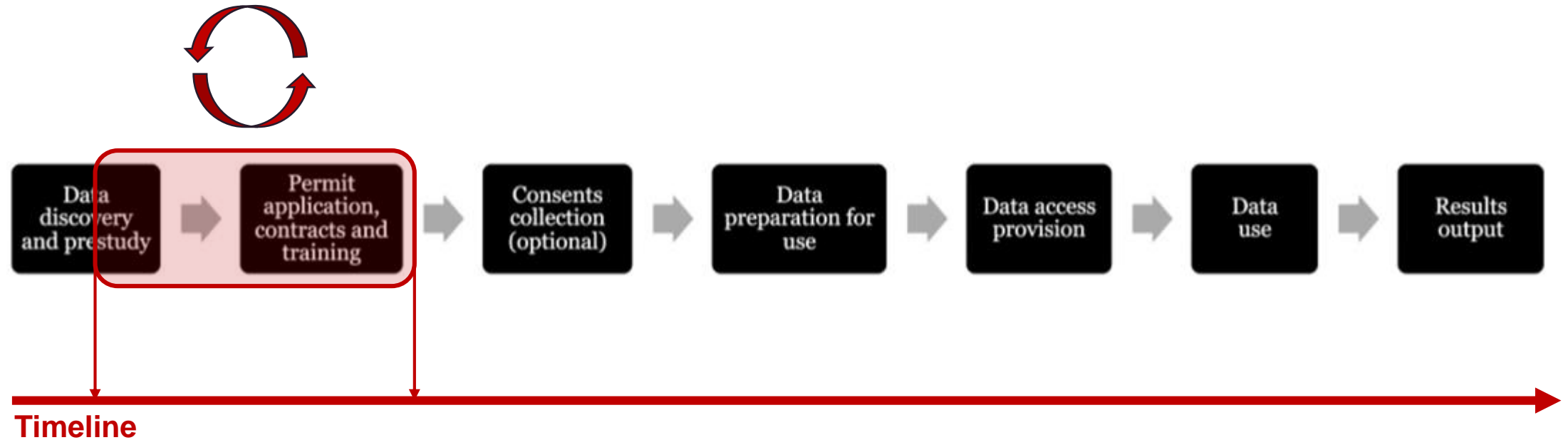
Other



The first step in a users journey



Reducing time to find and choose the right variabels



Create, share, discuss and the variabel list to your application (www.helsedata.no)

helsedata

Search

Language

Sign In

Menu

Variables

28 Variables from 14 data sources. Last modified: 18.07.22, 10:06

Download list

Share

Empty list

Change list

Variable

Kjønn

Kjønn

Kjønn

Pasientens kjønn

Kjønn

Pasientens kjønn

Pasientens kjønn

Pasient Kjønn

Kjønn

Barnets kjønn

Z-score basert på barnets vekt, svangerskapslengde og kjønn

Pasientens kjønn slik den er definert av folkeregisteret

Kjønn

PatientGender

PatientGender

PatientGender

Harmonization_Sex, gender and kjønn

28 Variables from 14 data sources. Last modified: 18.07.22, 10:06

Download list

Share

Empty list

Change list

Variable

Kjønn

Kjønn

Kjønn

Pasientens kjønn

Kjønn

Pasientens kjønn

Pasientens kjønn

Pasient Kjønn

Kjønn

Barnets kjønn

Z-score basert på barnets vekt, svangerskapslengde og kjønn

Pasientens kjønn slik den er definert av folkeregisteret

Kjønn

PatientGender

PatientGender

PatientGender

Period for data collection

Not defined

Variable group

Helse- og omsorgstjenester (IPLOS)

Alle variabler

Kjernevariabler - insidens

Svangerskap fødsler og nyfødte

Svangerskap fødsler og nyfødte

Personopplysninger

Alle variabler

Oppfølgingsskjema for Norsk hjerneslagregister

Hovedskjema for Norsk hjerneslagregister

Trombektomiskjema for Norsk hjerneslagregister

Variable group

Pasientjournal

Pasienter

Pasient

Pasient

Enkeltregning

Person

Om barnet

Om svangerskap og mors helse

Personopplysninger

Vaksinand

Mrskjerne

Mrskjerne

Mrskjerne

Find variables (28 hits)

Log in to save variables

Variable

Kjønn

Kjønn

Kjønn

Pasientens kjønn

Kjønn

Pasientens kjønn

Pasientens kjønn

Pasient Kjønn

Kjønn

Barnets kjønn

Z-score basert på barnets vekt, svangerskapslengde og kjønn

Pasientens kjønn slik den er definert av folkeregisteret

Kjønn

PatientGender

PatientGender

PatientGender

Data source

Bivirkningsregisteret

Dødsårsaksregisteret

Forsvarets helseregister

Helsearkivregisteret

Hjerte- og karregisteret

Kommunalt pasient- og brukerregister

Kommunalt pasient- og brukerregister

Kontroll og utbetaling av helseerfusjoner

Kreftregisteret

Medisinsk fødselsregister

Medisinsk fødselsregister

Nasjonalt kvalitetsregister for ryggkirurgi

Nasjonalt vaksinasjonsregister

Norsk hjerneslagsregister

Norsk hjerneslagsregister

Norsk hjerneslagsregister

sex gender kjønn

Find variables (28 hits)

Log in to save variables

Sign in to store variables

Then you can create lists, save all the variables you need, and retrieve them later.

Close

Sign In

VELG ELEKTRONISK ID

MiniID

BANKID

BUYPASS ID

COMMFIDES

Slik skaffer du deg elektronisk ID

Logg på HelseID

VELG ELEKTRONISK ID

MiniID

BANKID

BUYPASS ID

COMMFIDES

Slik skaffer du deg elektronisk ID

What is metadata?

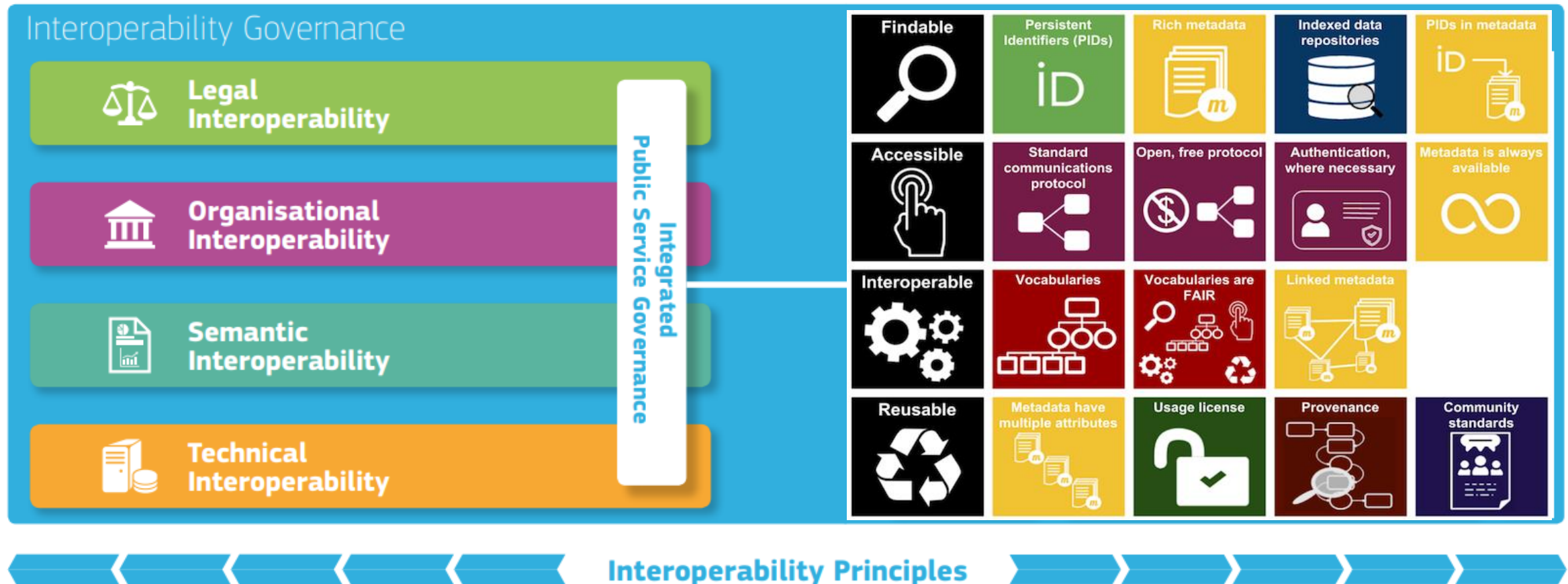
Why is metadata important for researchers?

Why is common standards important?



FAIR operationalizes the European Interoperability Framework

A common language between, architects, technicians and researchers



DCAT operationalize the FAIR criterias

...and makes it possible for machines to share and validate the quality of metadata....
among others



TO BE FINDABLE:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

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.

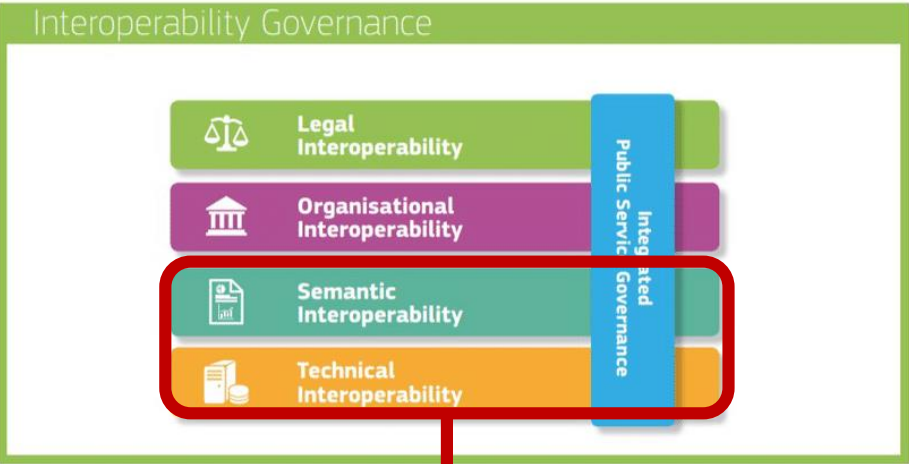
TO BE RE-USABLE:

- R1. meta(data) have 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 their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

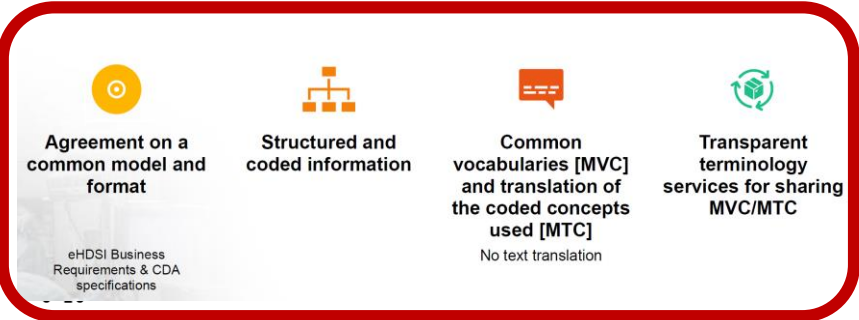
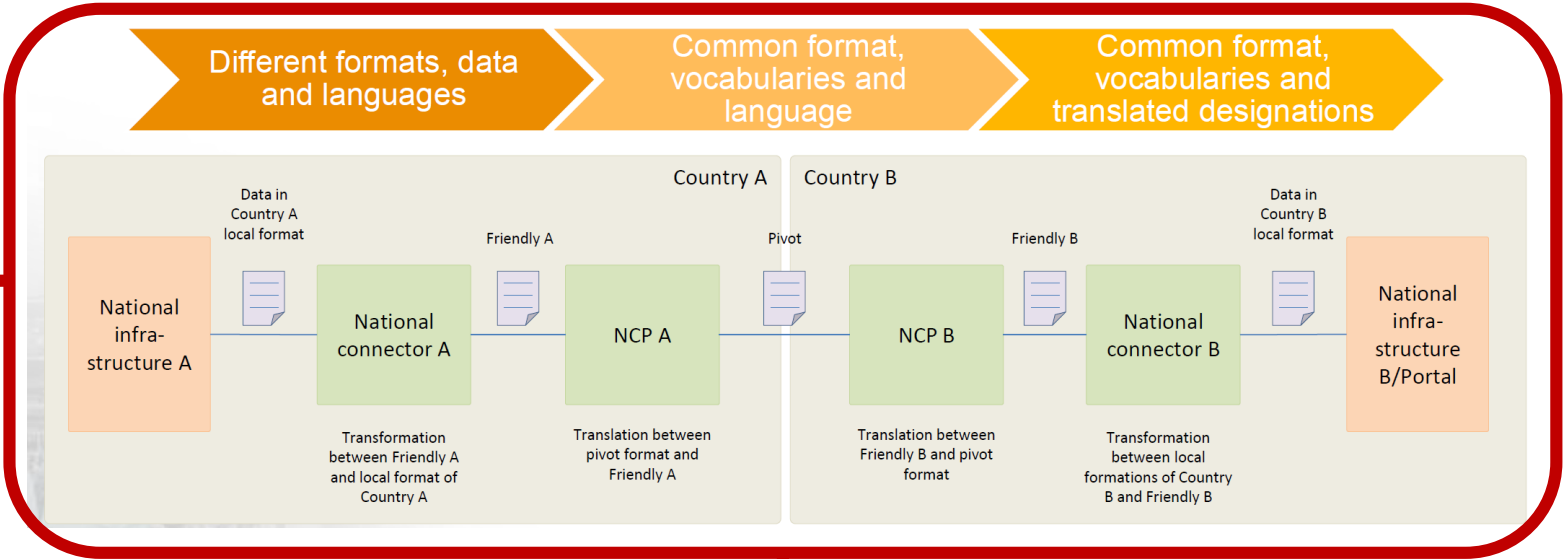
Rating evolution ✓ Sufficient	Findability ① Time based search 12% ① Geo search 66% ① Keyword usage 100% ① Categories 61%	Accessibility ① Most frequent accessURL ... 1100 ① Download URL 100% ① Most frequent downloadUR... 200
Interoperability ① Non-proprietary ?? ① Format / Media type from v... 0% ① Machine readable ?? ① DCAT-AP compliance 0% ① Media type 0% ① Format 75%	Contextuality ① Date of issue 100% ① Modification date 0% ① File size 15% ① Rights 19%	Reusability ① Contact point 100% ① License information 37% ① License vocabulary 58% ① Access restrictions 0% ① Publisher 100% ① Access restrictions vocabulary 0%

Common framework for interoperability and standards

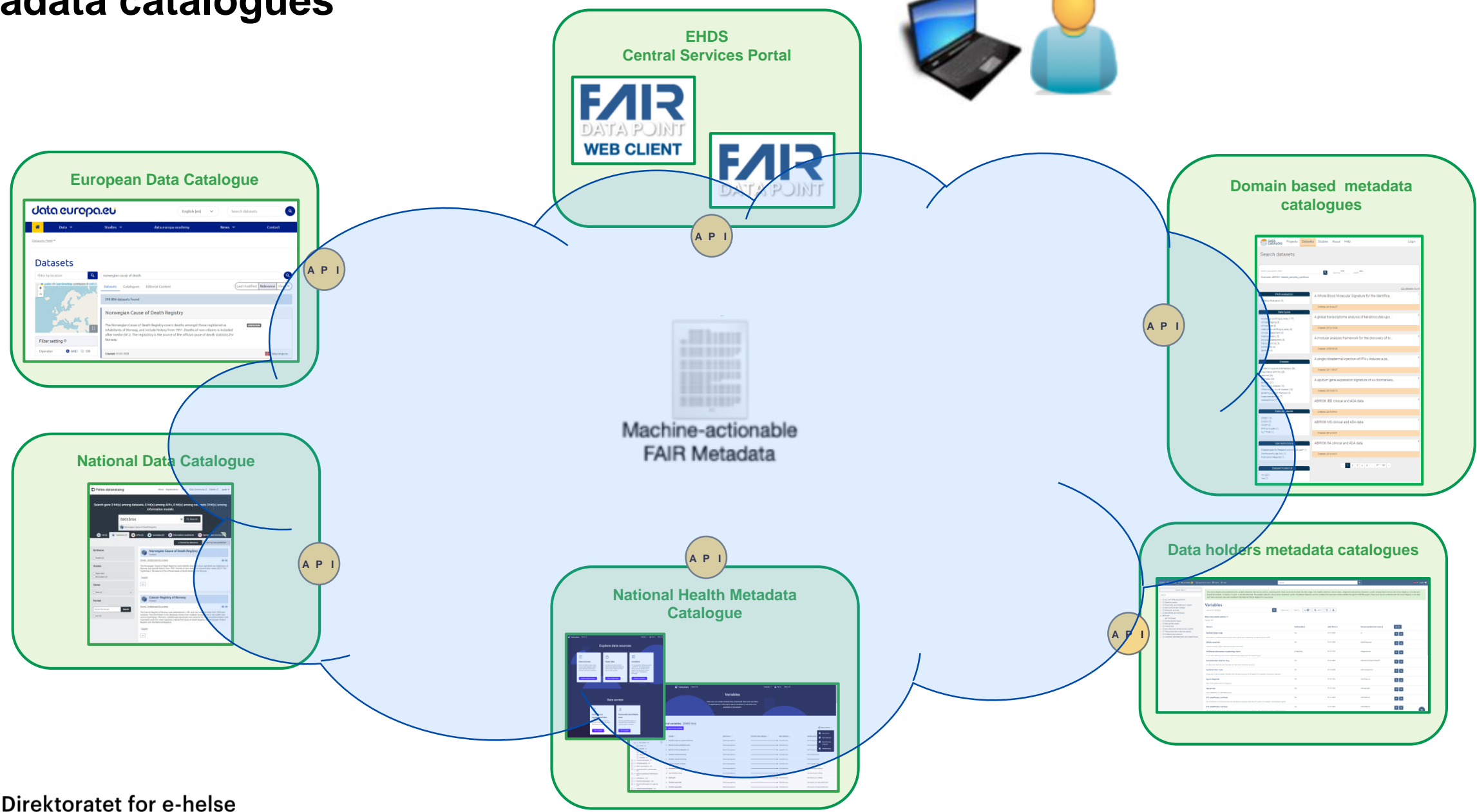
European Interoperability Framework (EIF)



NB! Data Quality



Standardization is prerequisites for an efficient federated ecosystem of metadata catalogues



The jungle of standards

Descriptions	Content	Examples	Typologi	Utility	Domain	Ex. Standards
Frameworks and "high level" standards and terminologies	How to describe data and concepts used for descriptions	Principles, ConceptSystem, Concepts, Structure				FAIR-principles ISO 11179 (Metadata) GSIM (Metadata) HL7 FHIR (Interoperability)
"High level" metadata standards	Attributes/properties to describe the dataset and more	Title, Description, Owner, Publisher, Identifier				DataCite DDI (Document, Discover and Interoperate) Dublin Core DCAT-AP SKOS (Simple Knowledge Organization System) DQV (Data Quality Vocabulary) PHIRI (Population health) ECRIN (CMRD)/ERIC)(Clinical research data)
Domain specific metadata standards	What should be described and details on how.					MIABIS (Biobanks) ISO 19115-1 (Geographic information Metadata) ORPHANET (Rare diseases) DICOM SPOR (ISO IDMP) (Medical products, drugs and medical substances) PHIRI (Population health) ECRIN (CMRD)/ERIC)(Clinical research data)
Common Data Models and semantics	Concepts and terms to define meaning and context for humans and computers.					OMOP (Observational Common Data Model) ORPHANET (Rare diseeseases) CONTSYS (A system of concepts for the continuity of care) HL7 FHIR SnomedCT Mesh Loinc
Persistent Identifiers	Unique keys for metadata and data	Persistent Identifiers for researchers, Data Sets				DOI (OID) ORCHID

We need standardized, structured and machine-readable metadata about standards for categorization and sorting

Health Data Source domains		Common standards	Domain specific standards	Clinical classifications and terminologies
	National Health Registries			
	Registrars of Clinical Quality			
	Biobanks			
	OMICS data			
	Laboratory data			
	Health Surveys/Cohort studies			
	Socioeconomic Registers			
	Electronic health records			

Ongoing work in TEHDAS WP6.1

Name	Typology	Utility	Domain	URL	Familiar with?	Widely adopted?	Barriers and challenges	Comments
ISO 8000 1(10)	Metadata standard	Conformance messaging (processing)	Any master data file	https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en	No	No		
(ISO 23494)	Meta-data standard	Data provenance	Bio samples	In development here: https://www.iso.org/standard/80715.html	No			This ISO is currently under development with the ISCO/TC 276 biotechnology Technical Committee.
BEACON	Metadata standard	Discoverability	Genomics, clinical data	https://beacon-project.io	No+	No, but someone in Oslo is actively promoting using Beacon.	Variant-Disease associations collected from curated resources and the literature. Beacon is an API that allows for data discovery of genomic and phenoclinic data.	I have heard FEGA (https://ega.elixir.no/), which covers whole sequences compared to variants only (Beacon), is about to be more prominently used.
SNOMED-CT	Ontology	Data provenance	EHR, clinical data, claims	https://www.snomed.org/	Yes	No+/Yes-	Health regions, communities and registries have different plans, ambitions and prerequisites to use Snomed CT. Also challenging to understand joint use of ICD-10/11 and SCT for diagnostic information.	Under implementation in one of four health regions, at some areas in other regions, and in selected registries including the cancer registry.
SPOR (ISO-IDMP)	Ontology	Data provenance	Medical products (drugs and medical substances)		Yes	Not yet, but will be.* The Norwegian Directorate of e-health recommends the use of IDMP for describing product-specific information. The use of Medicinal Product Identifier is recommended throughout the value chain for medicinal products.	Not yet, but will be.* The Norwegian Directorate of e-health recommends the use of IDMP for describing product-specific information. The use of Medicinal Product Identifier is recommended throughout the value chain for medicinal products.	*The Norwegian Medical agency (NoMA) has an ongoing project for developing a new drug database in accordance with IDMP and with a portal to SPOR.
HL7-FHIR	Meta-data standard	Conformance messaging	EHR output, clinical data, registries,	https://www.hl7.org/fhir/ ; See also https://art-decor.org/mediawiki/index.php/Main_Page	Yes	No, but increasing. The Norwegian Directorate of e-health has issued a high-level recommendation to use HL7 FHIR for integrations based on data sharing in the healthcare sector in 2019. The Directorate also recommends using SMART on FHIR for integration of applications to EHRs.	The use of a selection of national core profiles is a recommended standard in Norway. However, there is a lack of coordination on a national level for development and maintenance of use-case specific FHIR-profiles, implementation guides etc. leading to unwanted variations and inconsistent use. There are also challenges linked to discovery and re-use of existing profiles/implementation guides.	
DICOM	Meta-data standard	Conformance messaging	Medical image	https://www.dicomstandard.org/	Yes	This is the major standard for digital medical applications handling medical imaging and was established in 1992. All hospitals in Norway use DICOM for medical image communication.		

Overview clinical classifications and terminologies per 02.09.2022

National experts

Domain	Sub domain	International	Finland	Norway	Denmark	Sweden	Iceland	NC preferred	TEHDAS preferred	1+MG	OMOP	Expert (Norway)	Expert (Finland)	Expert (Sweden)	Expert (Denmark)	Expert (Iceland)
Diagnoses	Morbidity	ICD	ICD-10 (National version, Finland)	ICD-10 (National version, Norway)	ICD-10, SKS (National version, Denmark)	ICD-10-SE (national version, Sweden) (ICD7-ICD11)	ICD-10 (National version, Iceland)					Marie.Vikdal@ehelse.n				
	Mortality	ICD	ICD-10 International version	ICD-10 International version	ICD-10 International version	ICD-10 International version	ICD-10 International version					Marie.Vikdal@ehelse.n				
Diagnose groups/Casemix	Casemix	DRG (Under development in WHO)	NordDRG	NordDRG	NordDRG (Danish adaptation)	DKDRG	NordDRG					Marie.Vikdal@ehelse.n				
Health Interventions	Health Interventions (collected files)	ICHI (Under finalization in WHO)	Finnish code system (Toimenpideluokitus) that is based on NCSP	NKPK (consist of all the three mentioned under)	ICHI?	KVÅ			ICHI			Marie.Vikdal@ehelse.n				
	Health Interventions - Surgical Procedures	ICHI (Under finalization in WHO)	Finnish code system (Toimenpideluokitus) that is based on NCSP	NCSP (National version, Norway)	NCSP-DK (National version, Denmark)	KVÅ	NCSP (National version, Iceland)	NCSP				Marie.Vikdal@ehelse.n				
	Health Interventions - Medical Procedures	ICHI (Under finalization in WHO)	Finnish code system (Toimenpideluokitus) that is based on NCSP	NCMP (National version, Norway)	ICD-10, SKS	KVÅ						Marie.Vikdal@ehelse.n				
	Classification of Radiological Procedures and nuclear medicine	ICHI (Under finalization in WHO)	Finnish code system (Toimenpideluokitus) that is based on NCSP	NCRP (National version, Norway)	ICD-U/LUX, SKS (Classification regarding clinical physiology and nuclear medicine, a classification for results and one for additional coding)	KVÅ						Marie.Vikdal@ehelse.n				
	Classification of other health interventions (i.e. environment, participation, public health etc.)	ICHI (Under finalization in WHO)		Not classified in Norway		Classification of social services', interventions and activities (KSI)						Marie.Vikdal@ehelse.n				
Cancer	Cancer	ICD	ICD-10, ICD-O-3	ICD-10 (Int.) and ICD-O-3	ICD-10, SKS	ICD-O/3.2						Marie.Vikdal@ehelse.n				
Drugs	Drugs	ATC	ATC	ATC+ other?	ATC	ATC (+ NPL and NSL ?)			SPOR/IDMP	SPOR/IDMP		Marie.Vikdal@ehelse.n				
Laboratory	Laboratory	LOINC	LOINC	NLK (NPU)	IUPAC, NPU	NPU	LOINC		LOINC			Marie.Vikdal@ehelse.n				
Laboratory	Anatomical pathology			NORPAT (somewhat aligned with ICD-O3.2 and early SNOMED for pathology)		ICD-O/3 ?						Marie.Vikdal@ehelse.n				
Functioning	Functioning	ICF	ICF	ICF (CY)	ICF	ICF	ICF	ICF				Marie.Vikdal@ehelse.n				
Primary care diagnoses	Primary care diagnosis		ICPC-2 and ICD-10	ICPC-2	ICPC-2-DK	ICD-10-SE and a national primary care version of ICD-10-SE (KSH97-P)	ICD-10 (National version, Iceland)					Marie.Vikdal@ehelse.n				
PROMS/PREMS			In quality registries				ICNP					Marie.Vikdal@ehelse.n				
Quality Of Life										SF-12, SF-36 or EORTC-QLQ-C30.						
Psykisk helsevern for barn og unge		ICD		PHBU		DSM, ICD-10-SE						Marie.Vikdal@ehelse.n				
Nursing				ICNP/Snomed CT	National classification for nursing/care and treatment,	KVÅ (ongoing: ICNP/Snomed CT ?)	ICNP					Marie.Vikdal@ehelse.n				
Smoking	GATS Lifetime smoking status Pack-yers										GATS Lifetime smoking status Pack-yers					
Physical aktivitet	IPAQ										IPAQ					
Obesity	BMI Waist circumference										BMI Waist circumference					

Metadata standards

Metadata Specifications

The core objective of the workshop will be to investigate and advance alignment between the cross-disciplinary and domain-specific metadata standards, and to bridge from standards focusing on collection-level to variable-level metadata.

Metadata standards that may be considered include⁴:

- Study- or collection-level: DCAT, Dublin Core, ISO 19115-1, DDI 4
- Variable and dimension level
 - Microdata: DDI 4, W3C SSN, FHIR-HL7, CDISC, EML, SensorML, GSIM
 - Aggregate data: W3C DataCube, ISO 19123, Frictionless data
- Provenance: W3C PROV-O, ISO 19115-2
- Workflows/data transformation: DDI 4

Data transformations to prepare data for analysis may be described in machine-actionable form. DDI 4 uses some patterns of BPMN to achieve this, and CSV on the Web addresses transformation of tabular data into semantic form.

Reuse existing vocabularies for providing metadata to your resources

General purpose standards and specifications:

- **Dublin Core** for published material (text, images), <http://dublincore.org/documents/dcmi-terms/>
- **FOAF** for people and organisations, <http://xmlns.com/foaf/spec/>
- **SKOS** for concept collections, <http://www.w3.org/TR/skos-reference>
- **ADMS** for interoperability assets, <http://www.w3.org/TR/vocab-adms/>

Specific standard for datasets:

- **Data Catalog Vocabulary DCAT**, <http://www.w3.org/TR/vocab-dcat/>

Specific usage of DCAT and other vocabularies to support interoperability of data portals across Europe:

- **DCAT application profile for data portals in Europe**, http://joinup.ec.europa.eu/asset/dcat_application_profile/description

DCAT Application Profile for data portals in Europe (DCAT-AP) reuses terms from DCAT, Dublin Core, FOAF, SKOS, ADMS and others.

Contents [\[hide\]](#)

- 1 DCAT in Context of Standards and related Work
 - 1.1 Standards and related Work
 - 1.1.1 Dublin Core
 - 1.1.2 Asset Description Metadata Schema (ADMS)
 - 1.1.3 Comprehensive Knowledge Archive Network (CKAN)
 - 1.1.4 Data Tag Suite (DATS)
 - 1.1.5 DDI-RDF Discovery Vocabulary (Disco)
 - 1.1.6 RDF Data Cube Vocabulary (DQ)
 - 1.1.7 Geographic Information – Metadata (ISO 19115)
 - 1.1.8 Dataset Descriptions: Community Profile (HCLS)
 - 1.1.9 Schema.org
 - 1.1.10 Vocabulary of Interlinked Datasets (VoID)
 - 1.1.11 DataCite
 - 1.1.12 Research Data Alliance
 - 1.2 Comparative analysis of the "Catalog" concept
 - 1.2.1 Geographic Information – Metadata (ISO 19115)
 - 1.3 Comparative analysis of the "Dataset" concept
 - 1.3.1 Dublin Core
 - 1.3.2 ADMS
 - 1.3.3 CKAN
 - 1.3.4 DATS
 - 1.3.5 DQ
 - 1.3.6 Disco
 - 1.3.7 HCLS
 - 1.3.8 Schema.org
 - 1.3.9 VoID
 - 1.3.10 Geographic Information – Metadata (ISO 19115)
 - 1.3.11 DataCite
 - 1.4 Comparative analysis of the "Distribution" concept
 - 1.4.1 ADMS
 - 1.4.2 Disco
 - 1.4.3 Geographic Information – Metadata (ISO 19115)
 - 1.4.4 DataCite

Preparing for federated analyses

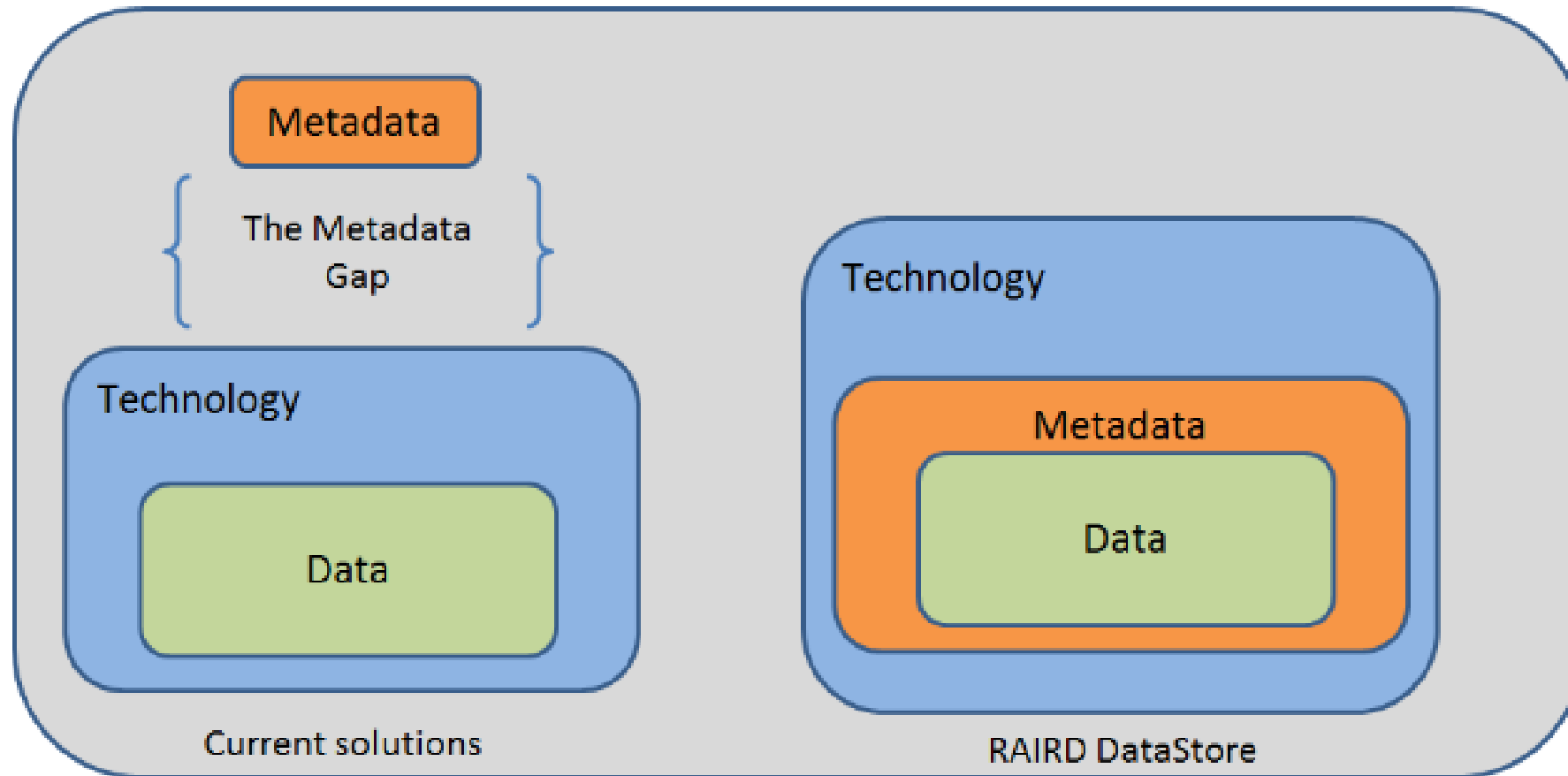



Figure 5 - The Metadata Gap and the RAIRD DataStore

Clinical Epidemiology **Dovepress**
open access to scientific and medical research

 Open Access Full Text Article

REVIEW

Nordic Health Registry-Based Research: A Review of Health Care Systems and Key Registries

Variables
Personal identity number
Sex
Date of birth
Country of birth
Name
Address, including date of address changes
Immigration and date
Emigration and date
Civil status, including changes in and personal identity number of spouse or registered partner
Kinship, including personal identity number of parents, siblings, and children
Date of death

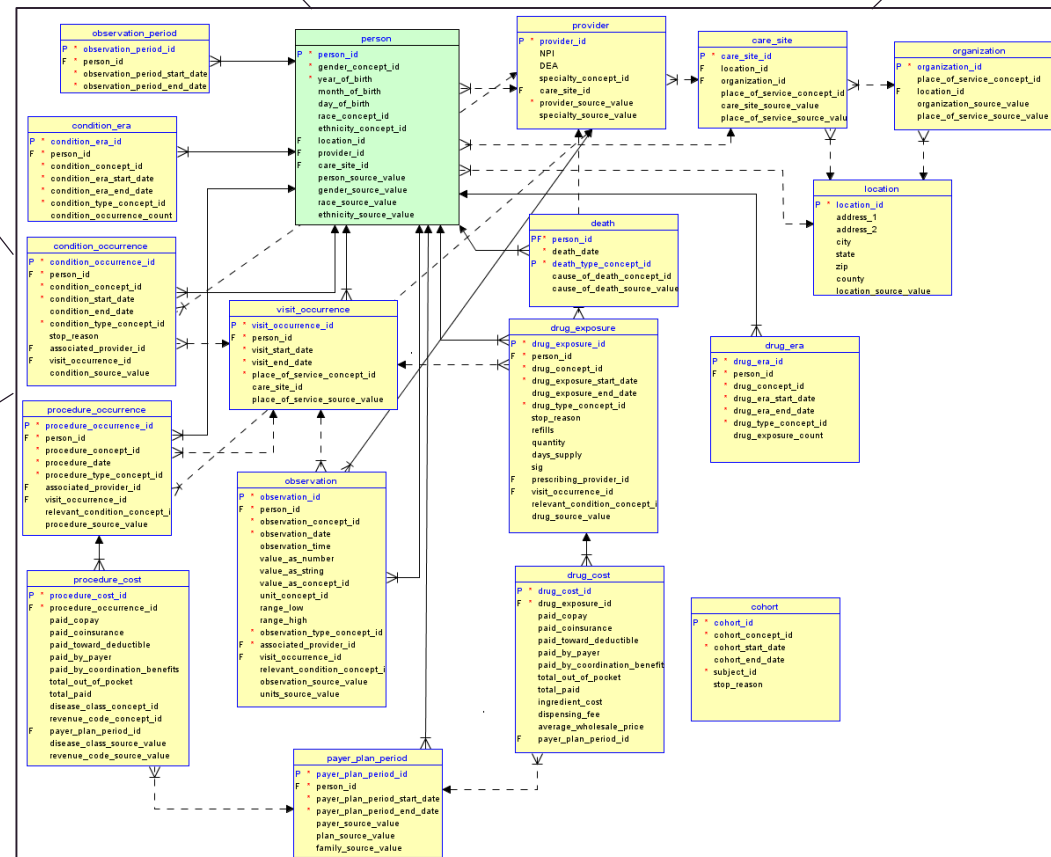
Information Type	Variables
Basic information	The personal identity number of the infant
	The personal identity number of the mother The personal identity number of the father ^a
Delivery information	Date of birth
	Place of birth
	Presentation at birth (eg, cephalic, breech, or shoulder presentation)
	Method of delivery
	Delivery complications
	Procedures around delivery
Maternal characteristics	Age at birth
	Height, weight, body mass index ^b
	Smoking status ^c
	Parity
	Diagnoses ^d and complications during pregnancy or delivery
	Number of previous pregnancies and deliveries
Infant characteristics	Single or multiple birth
	Sex
	Gestational age at birth
	Birth weight
	Length
	Head circumference
	Live or stillborn ^e
	Health status of the child (Apgar score, infant diagnoses, and treatment)
	Congenital malformations at birth ^f

Information Type	Variables
The patient	Personal identity number Area of residence
Hospital and department	Hospital code Department code/specialty
Admission	Admission date Discharge date Admission type (acute, non-acute, etc.) Patient contact type (eg, inpatient, outpatient) ICD diagnoses (primary and secondary/ additional codes) Surgical and medical procedure codes

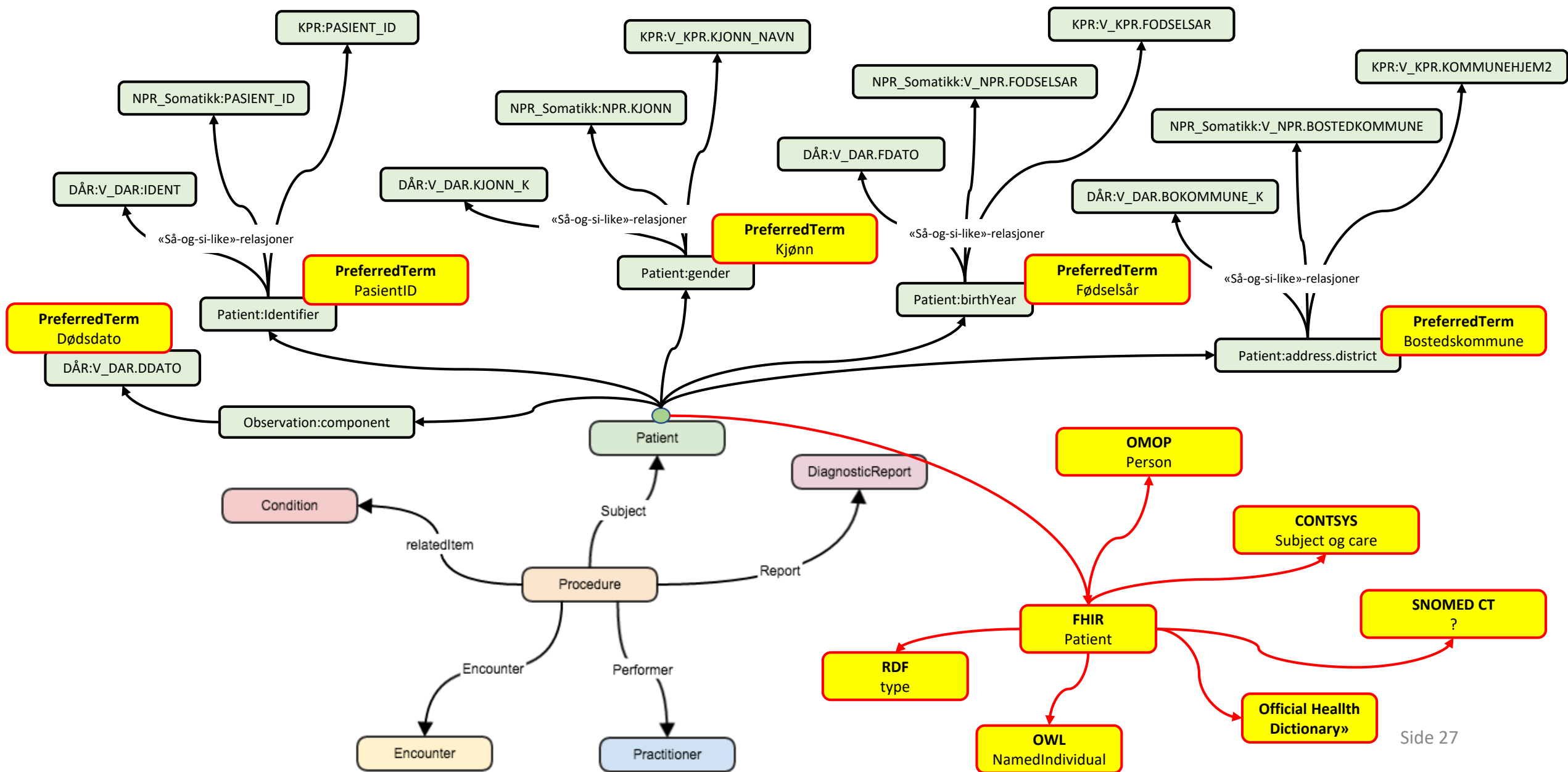
Information Type	Variables
The person	Personal identity number Place of residence
The death	Date of death (or date of discovery if found dead) Manner of death (natural, accident, violence, suicide, uncertain) Underlying cause of death The immediate or direct cause of death The contributing causes of death Place of death (private residence, nursing home, hospital) Autopsy performed (yes/no) Type of autopsy (clinical, medico-legal)
Physician issuing the death certificate	Hospital physician, GP health officer, etc

Information Type	Variables
The patient	Personal identity number (or pseudonymized number) Sex Date of birth
The prescriber	Prescriber type (eg, GP, hospital physician, or private physician)
The pharmacy	Identifier Location of the pharmacy ^a
The drug	Date of dispensing Nordic article number (unique identifier) Anatomical Therapeutic Chemical classification (ATC) code Number of packages dispensed Number of tablets in one package Tablet strengths The defined daily dose (DDD) Formulation of the drug Drug reimbursement

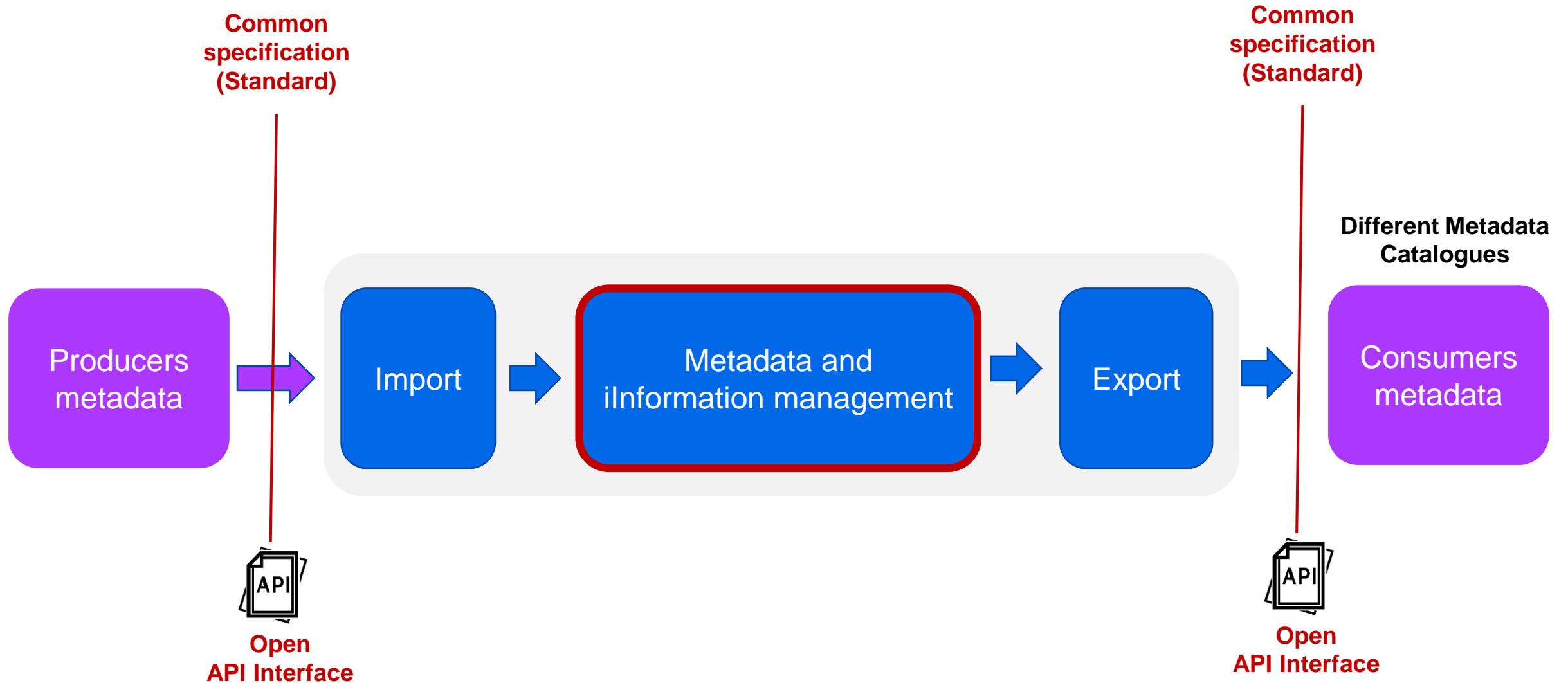
Information Type	Variables
The patient	Personal identity number Date of birth Sex Place of residence (unit) Vital status Date of death
Tumor characteristics	Date of diagnosis Topography (primary site) Morphology/histology Tumor stage or grade ^a Method of confirmation Behavior (malignant, premalignant, and borderline behavior)



Understanding the importance of information management

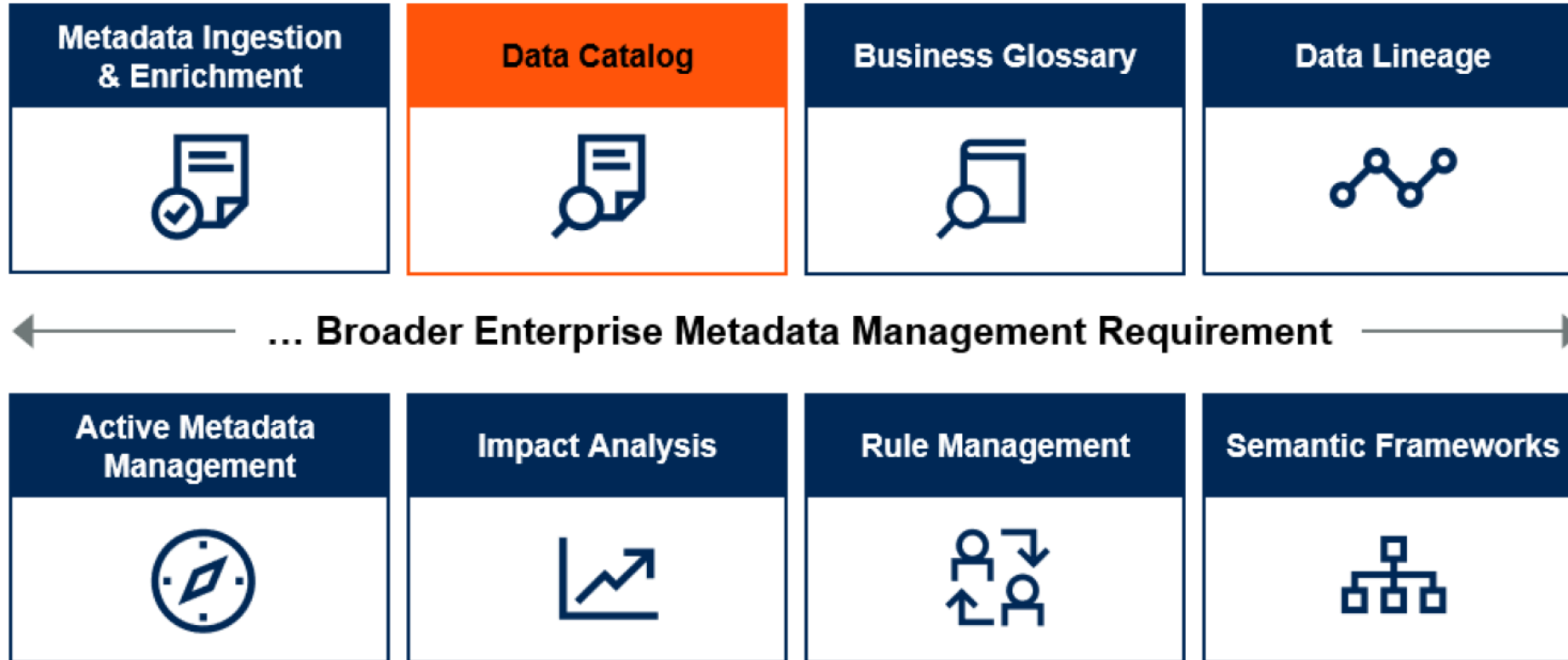


The Metadata Management Process



Broader Metadata Management is required!

Data Catalog Is One Key Component of the ...



Source: Gartner
ID: 394570

The road from «my» metadata to standardized, structured, machine-readable and FAIR metadata

AS - IS

Metadata about my data source

Based on any metadata standard?

Based on controlled vocabularies?

Which format are they stored in?

Where can I find the metadata about your data source?

Can I download it?


How can we get there together?

Machine to machine communication – m2m

```
<rdf:RDF>
  <dc:Dataset rdf:about="https://fair-v1.healthdata.be/dataset/68f53c1a-6109-4e8b-b7a5-a7f7484f6e8b">
    <dc:title>
      Registry Next-Generation-Sequencing (NGS) tests for oncology
    </dc:title>
    <dc:description></dc:description>
    <dc:keyword>Federal Public Service Health</dc:keyword>
    <dc:keyword>Food Chain Safety and Environment</dc:keyword>
    <dc:keyword>Laboratories: clinical biology</dc:keyword>
    <dc:keyword>Laboratory Workers</dc:keyword>
    <dc:keyword>
      National Institute for Health and Disability Insurance
    </dc:keyword>
    <dc:keyword>Physicians</dc:keyword>
    <dc:identifier>68f53c1a-6109-4e8b-b7a5-a7f7484f6e8b</dc:identifier>
    <dc:issued>2018-05-08T23:22:19+02:00</dc:issued>
    <dc:modified>2021-10-31T22:35:01+01:00</dc:modified>
    <dc:publisher>Sciensano</dc:publisher>
    <dc:distribution rdf:resource="https://fair-v1.healthdata.be/resource/27bb2003-e8a2-428c-b265-5e5b7708c924"/>
  </dc:Dataset>
  <dc:Distribution rdf:about="https://fair-v1.healthdata.be/resource/27bb2003-e8a2-428c-b265-5e5b7708c924">
    <dc:title>
      Registry Next-Generation-Sequencing (NGS) tests for oncology
    </dc:title>
    <dc:description>
      <p>Machine-readable representation of this dataset.</p>
    </dc:description>
    <dc:issued>2018-05-08T23:22:19+02:00</dc:issued>
    <dc:modified>2021-10-31T22:35:01+01:00</dc:modified>
    <dc:accessURL rdf:resource=""/>
    <dc:downloadURL rdf:resource="https://fair-v1.healthdata.be/sites/default/files/68f53c1a-6109-4e8b-b7a5-a7f7484f6e8b.json"/>
    <dc:mediaType>application/json</dc:mediaType>
    <dc:format>json</dc:format>
    <dc:byteSize>14823</dc:byteSize>
    <foaf:page>
      https://fair-v1.healthdata.be/resource/27bb2003-e8a2-428c-b265-5e5b7708c924
    </foaf:page>
    <dc:Distribution>
    </dc:Distribution>
    <foaf:Agent rdf:about="https://fair-v1.healthdata.be/publisher/n0">
      <foaf:name>DKAN</foaf:name>
      <foaf:homepage>https://fair-v1.healthdata.be</foaf:homepage>
      <dc:type rdf:resource="http://purl.org/adms/publishertype/NonProfitOrganisation"/>
    </foaf:Agent>
  </dc:Distribution>
</rdf:RDF>
```


Thank you!

www.helsedata.no

 Direktoratet for e-helse

Truls Korsgaard
Senior advisor
The Norwegian Directorate of eHealth
truls.Korsgaard@ehelse.no



PHIRI

Population Health Information
Research Infrastructure

Why do we need metadata for public health data sources?

Hanna Tolonen, Finnish Institute for Health and
Welfare (THL), Finland

Petronille Bogaert, Sciensano, Belgium

www.phiri.eu

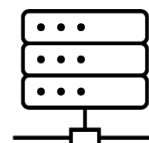


This project has received
funding from the European
Union's Horizon 2020
research and innovation
programme under grant
agreement No 101018317

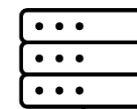
Why metadata?

For my research, I would need data about prevalence of NCDs in EU MSs. Where to find it?

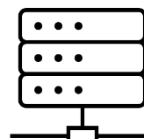
Data **without** metadata



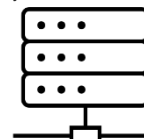
Hospital Discharge Data, NL



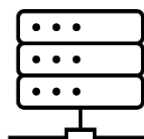
Care Register for Health Care, FI



Social Health Insurance Data, AT



Morbidity data, RO



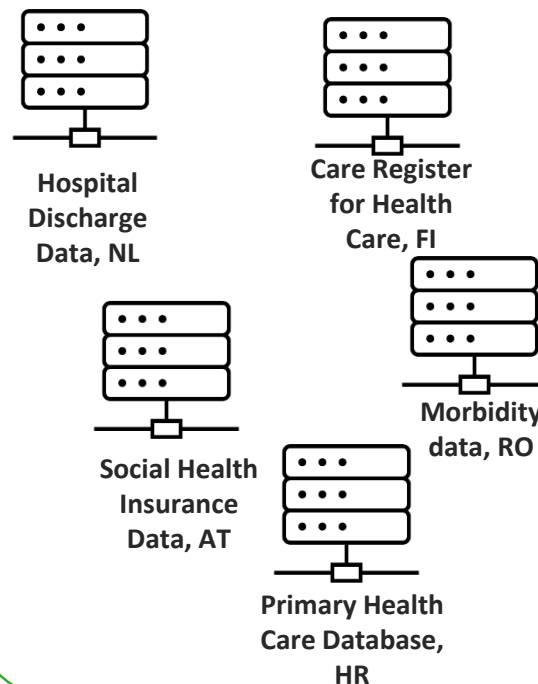
Primary Health Care Database, HR

- What type of data are these?
- What age groups they covers?
- What are their geographical coverage?
- What type of variables they have?
- How do I get access to these data sources?

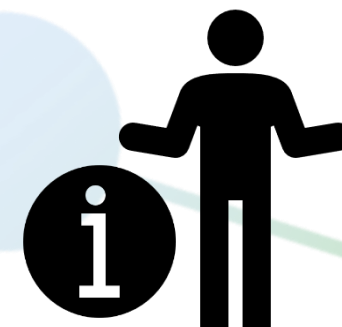
Why metadata?

For my research, I would need data about prevalence of NCDs in EU MSs. Where to find it?

Data **with** metadata



Through my networks and internet search I was able to identify following potential data sources



In structured format

- Data on hospitalizations
- Covers all age groups
- National coverage
- Basic background variables on sex, age and diagnostic information
- Contact details for access requests

Welcome to the **one-stop shop** that facilitates access to population health and health care data, information and expertise across Europe

FIND DATA

Fulltext Search

Country

- None - ▾

SEARCH



CARE REGISTER FOR HEALTH CARE

CONTACT

Registry data

[Link to Care Register for Health Care](#)

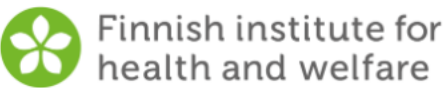
The purpose of the register is to collect data on the activities of health centres, hospitals and other institutions providing inpatient care and on the clients treated in them as well as on home-nursing clients for the purposes of statistics, research and planning.

More details about collected data at: <https://aineistokatalogi.fi/catalog/studies/7567e45d-72b7-428b-be9e-510440336edf>

Based on the Act on the National Institute of Health and Welfare (668/2008 <https://www.finlex.fi/en/laki/kaannokset/2008/en20080668?search%5Btype%5D=pika&search%5Bkieli%5D%5B0%5D=en&search%5Bpika%5D=668%2F2008>) and the Act digital handling of record from the social and health care customers (159/2007)

ICD-10; disease; Hospital admission

Health status » Morbidity/disability » Accidents & injuries; Health status » Morbidity/disability » Communicable diseases;



[CONTACT](#) / [FOR MEDIA](#)

Search... 

- [WHAT'S NEW](#) ▾
- [TOPICS](#) ▾
- [SERVICES](#) ▾
- [RESEARCH AND DEVELOPMENT](#) ▾
- [STATISTICS AND DATA](#)** ▾
- [PUBLICATIONS](#) ▾
- [ABOUT US](#) ▾

[THL.FI](#) ▸ [Statistics and data](#) ▸ [Data and services](#) ▸ [Register descriptions](#) ▸ [Care Register for Health Care](#)

- [STATISTICS AND DATA](#)
- [Data and services](#)
- [Register descriptions](#)
- [Care Register for Health Care](#)**
- [Medical Birth Register](#)
- [Register of Congenital malformations](#)
- [Register of Induced Abortions](#)

Care Register for Health Care

Register description

Register description, in accordance with the Personal Data Act (523/1999)

- [Controller](#)
- [Person responsible for the register](#)
- [Other persons maintaining the register](#)
- [Name of the register](#)

[Grounds for maintaining the register](#)

[ON OUR WEBSITE](#)
[All THL's register descriptions](#)

CARE REGISTER FOR HEALTH CARE

- OVERVIEW
- DEMOGRAPHICS
- DATASET
- CONTACT

Country(ies)	Finland
GEO coverage	Nuts 3
Target Population	General population
Age range	From 0 years
Sex	Both



CARE REGISTER FOR HEALTH CARE

- OVERVIEW
- DEMOGRAPHICS
- DATASET**
- CONTACT

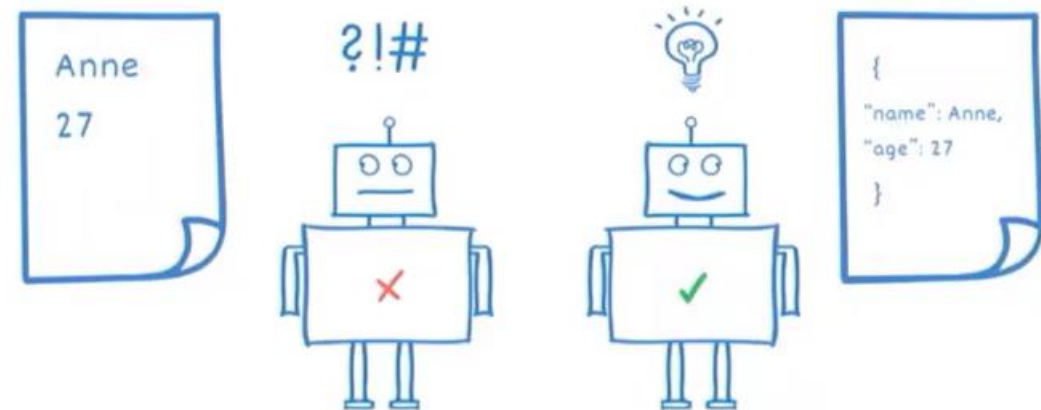
Data Collection Period	From 01 January 1969 to 14 March 2022
Language(s)	Finnish, Swedish
Personal Identifier	National identifier
Level of aggregation	Individual
Terms of data access	Access can be provide through Findata (findata.fir)
Linkage possible	Yes

Schema.org & DCAT metadata Standards

Why Schema.org

Schema.org is a metadata standard for the indexation of web pages

- Improves machine to machine communication
- Increases the discoverability of datasets



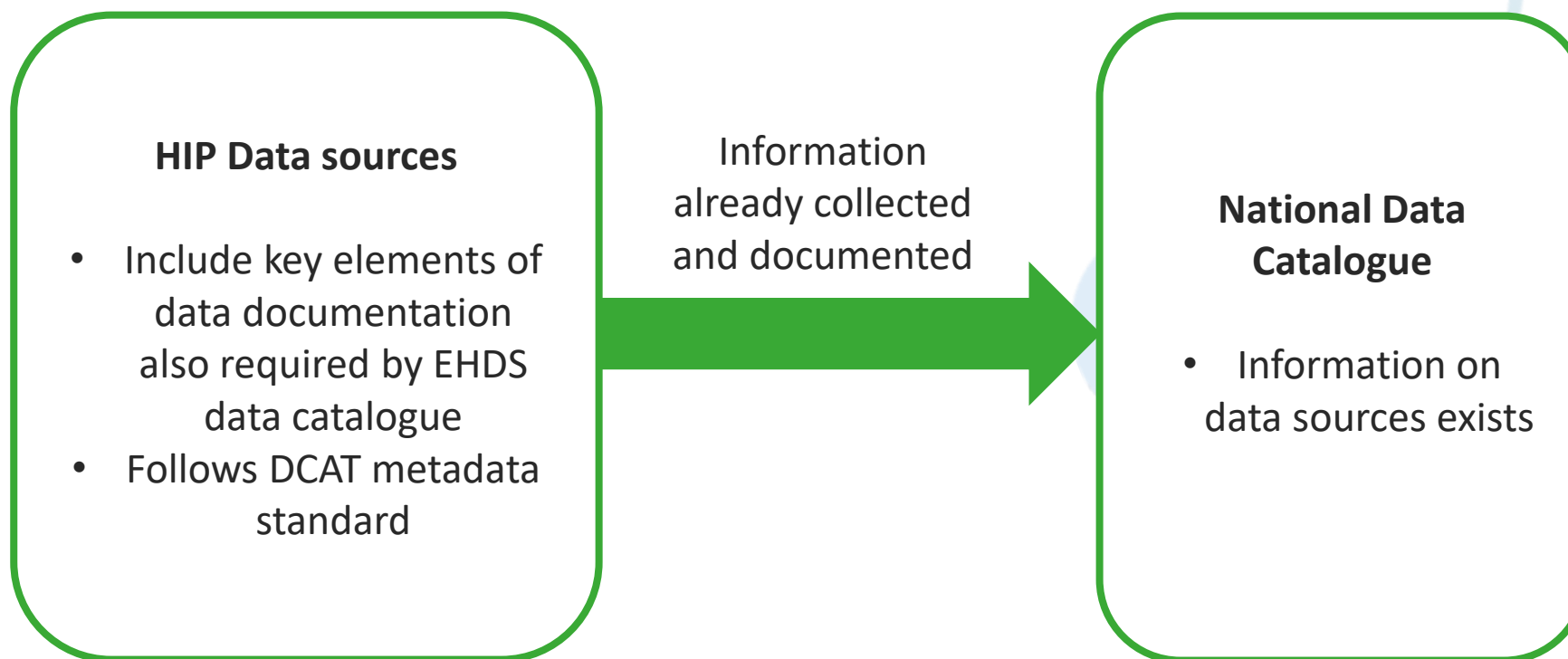
Why DCAT

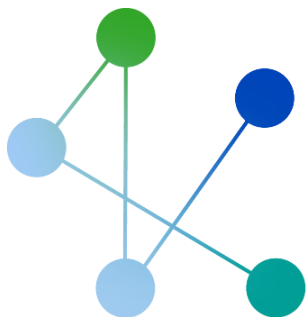
DCAT is a vocabulary for publishing data catalogs on the web

- Facilitates sharing of data & interoperability between data catalogs
- Increases the discoverability of datasets
- Allows federated search for datasets across catalogs in multiple sites

National data catalogue and HIP data sources

- EHDS regulation proposal: each country should have a national data catalogue (Article 55)





PHIRI

Population Health Information
Research Infrastructure

Thank you for your attention!

 [@PHIRI4EU](https://twitter.com/PHIRI4EU)
 [/company/phiri](https://www.linkedin.com/company/phiri)

www.phiri.eu



This project has received
funding from the European
Union's Horizon 2020
research and innovation
programme under grant
agreement No 101018317

The ambition of the European Health Data Space for Secondary Use (EHDS2) Pilot project

Petronille Bogaert – European Public Health Conference 2022 Berlin

European Health Data Space

Top priority of the European Commission is the creation of a **European Health Data Space (EHDS)**.

Secondary use of health data: the use of health data for a different purpose than the one they were initially collected for, e.g. research and policy making



EHDS pilot

creates and tests a first
version of European Health
Data Space network

16 partners

10 countries

5 million € grant



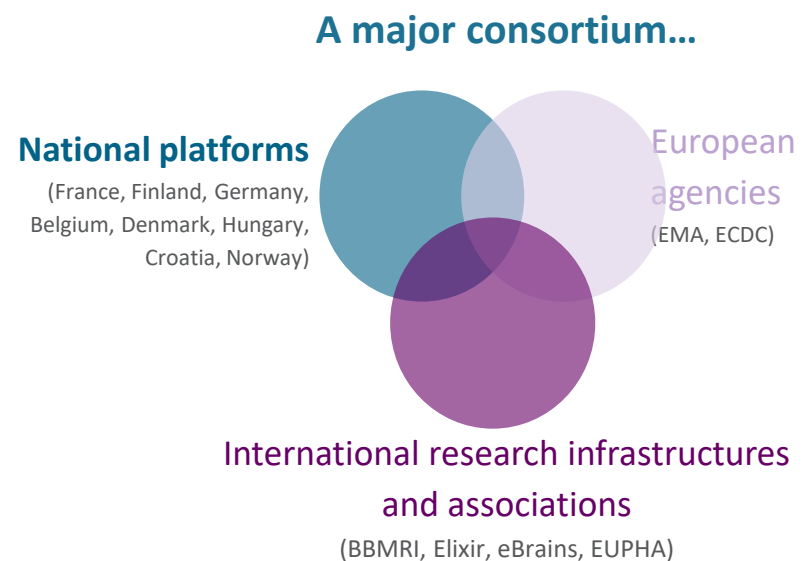
European Health Data Space: metadata catalogue requirement

Proposal for a regulation - The European Health Data Space published on May 3rd by the European Commission.

Section 5 – Article 55.1

*“The health data access bodies shall inform the data users about the available datasets and their characteristics through a **metadata catalogue**. Each dataset shall include information concerning the source, the scope, the main characteristics, nature of electronic health data and conditions for making electronic health data available”*

EHDS2 pilot: The consortium



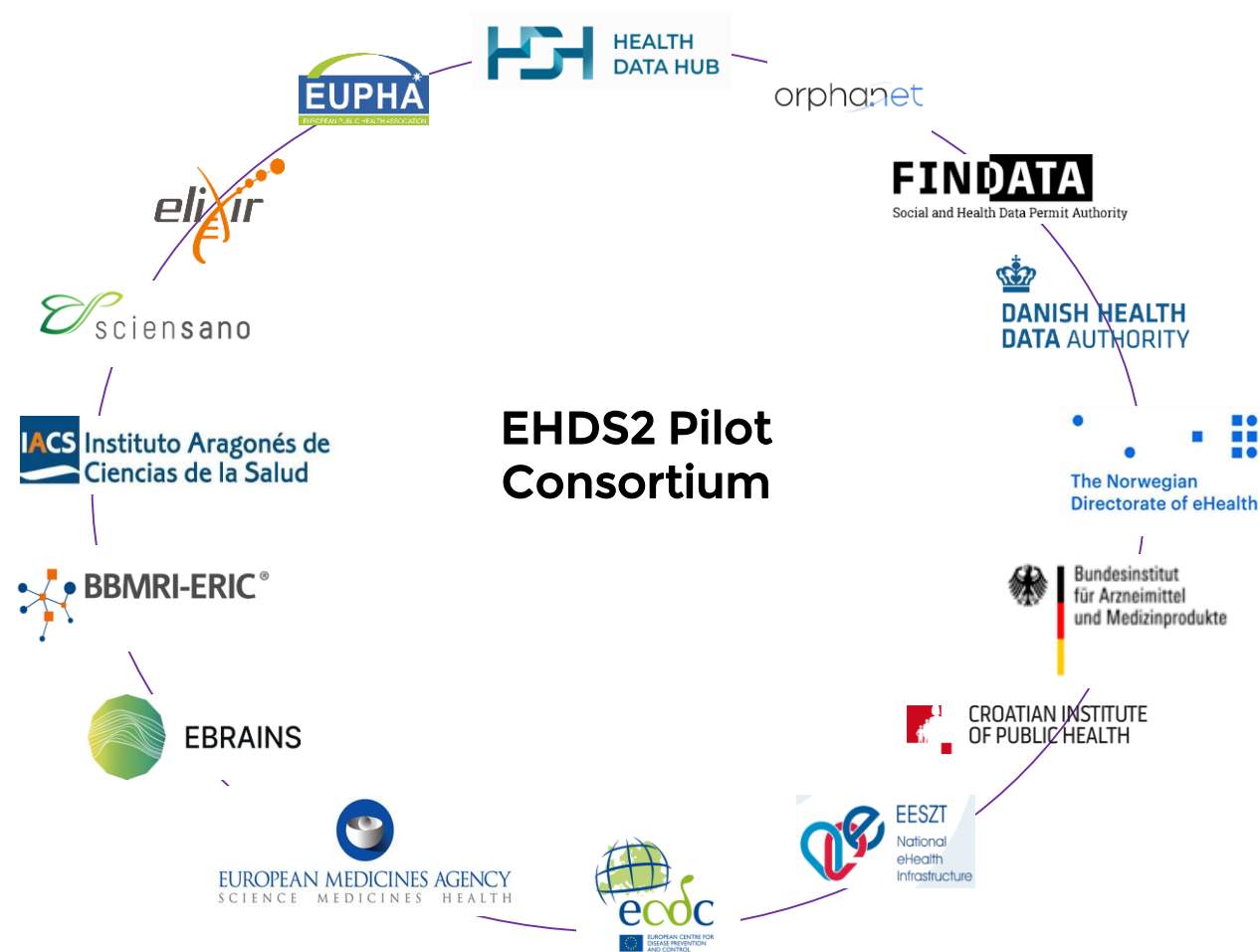
... to concretely test the EHDS in practice

A **network of data platforms** at European level

Cross-border research **use cases**

Recommendations for the future EHDS

A consortium of 16 major partners

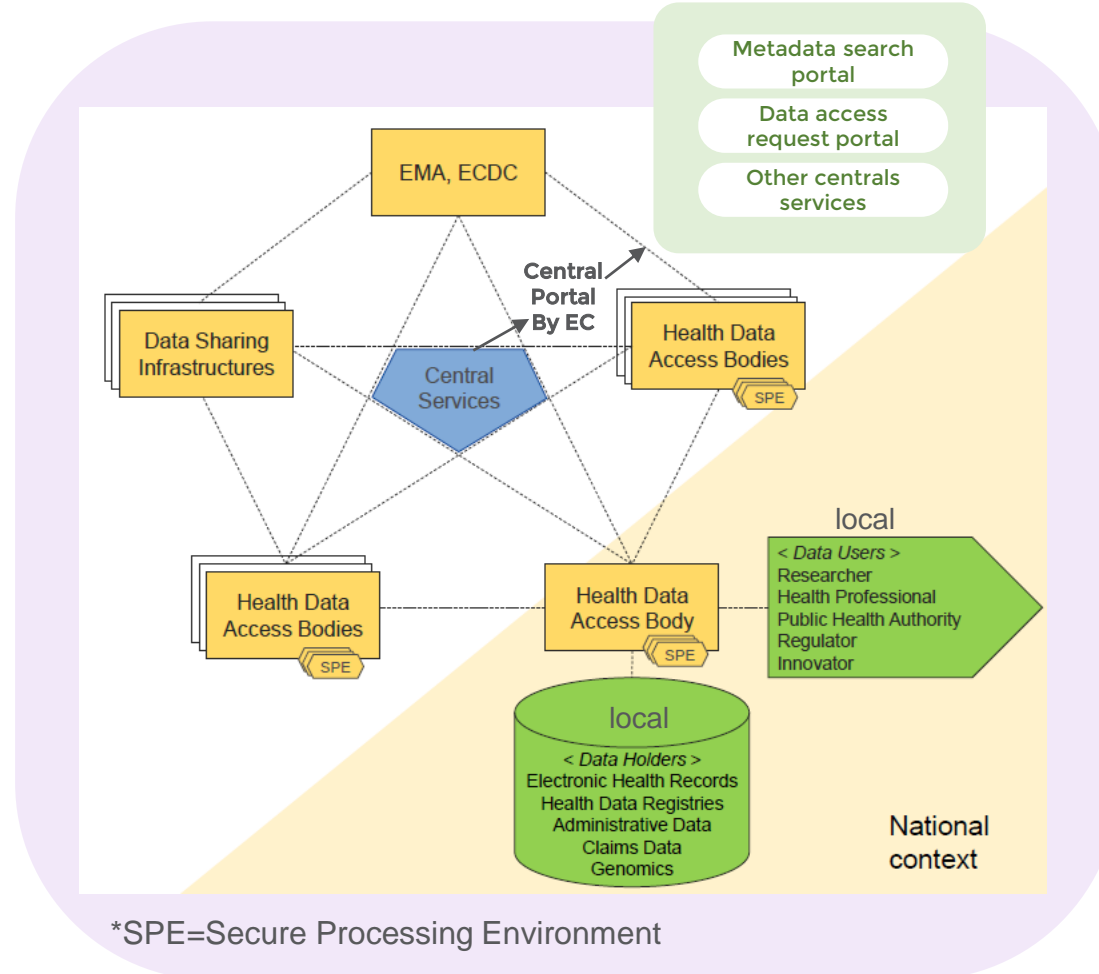


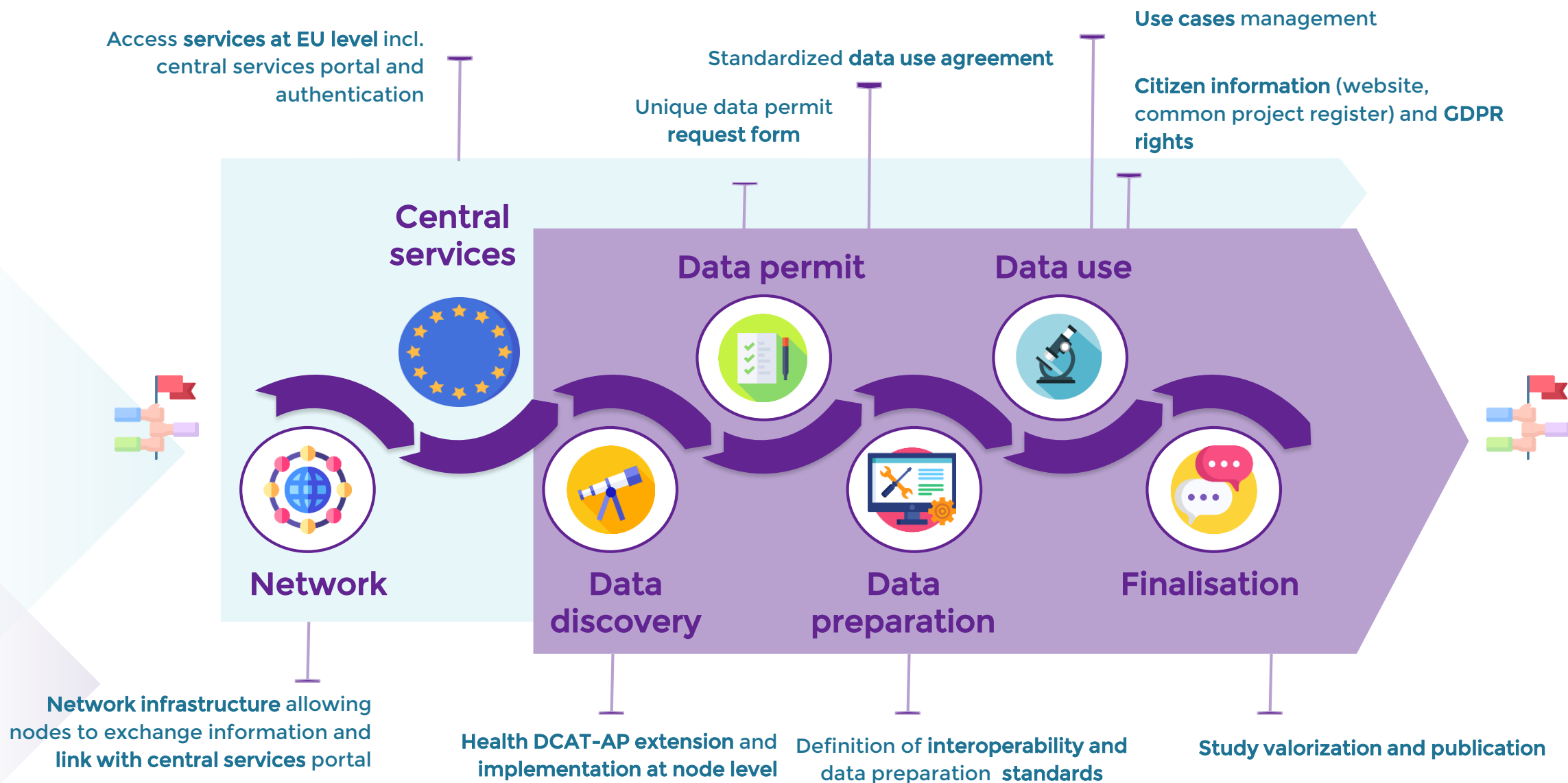
The EHDS2 Pilot project: the first European network to facilitate the secondary use of health data

Within this network of nodes researchers will be able to:

Query metadata catalogs and discover all available data collections in Europe relevant to health

Ask for data access in several nodes with a single data application form





Technical WPs to build and test the user journey :

WP5	IT INFRASTRUCTURE Building an IT infrastructure connecting the nodes for information exchange and allowing services discovery
WP6	METADATA STANDARDS Develop and implement a standardised descriptive metadata for health data portals respecting the FAIR principles and user needs of the EHDS with an Health-DCAT extension. Implement service that exposes on each node the Metadata catalogue of the node
WP7	REGULATORY AND LEGAL COMPLIANCE Providing guidelines for harmonizing legal, ethics data access procedures, security requirements and GDPR compliances in order to build a unique data access application for the EHDS and allow cross-border use of data. Implement a portal allowing form filling up and dispatch to different nodes
WP8	DATA INTEROPERABILITY, QUALITY, AND PROTECTION Providing guidelines for data standards including data quality assurance and solutions for query and semantic interoperability. Definition of security prerequisite for transfer. Implementation of Data transfer protocol for use cases.
WP9	USE CASES MANAGEMENT Ensure project management for implementing standards on chosen data use cases, provide use cases management (completion, opening of HDW...), and produces recommendations on use cases management procedures and governance.



Health Data Lab (*co-leader*)



Health Data Hub (*co-leader*)



Sciensano (*co-leader*)



Norwegian Directorate of eHealth (*co-leader*)



BBMRI (*leader*)



FinData (*Task leader: data application*)



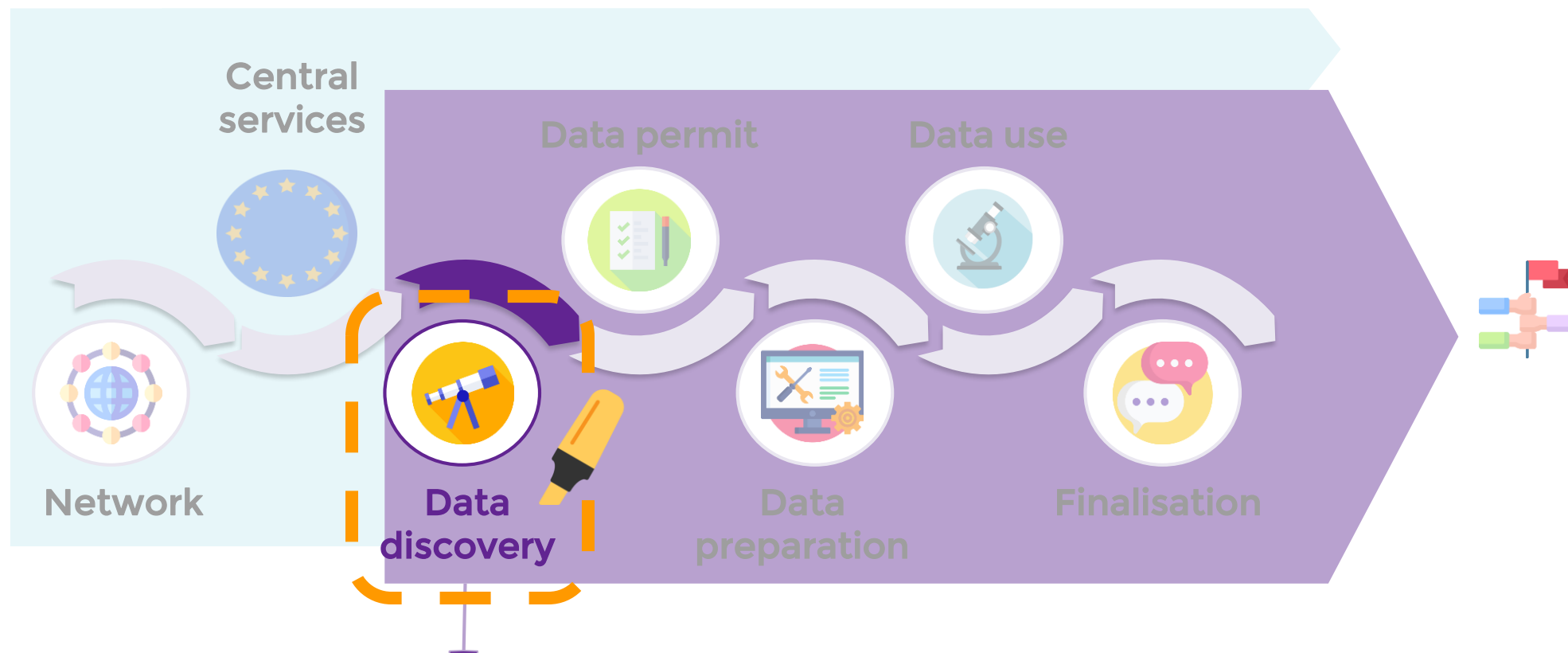
BBMRI (*leader*)



Health Data Hub (*co-leader*)



Danish Health Data Authority (*co-leader*)



Health DCAT-AP extension and
implementation at node level

Objectives of WP6



To develop and implement a standardized descriptive metadata template for health data portals respecting the FAIR principles and user needs of the EHDS.



**Landscape analysis
of descriptive
metadata catalogue
templates used by
different MSs/nodes**



**Design the Health
DCAT-AP extension**



**Implementation and
assessment of the
descriptive metadata
templates based on
the Health DCAT-AP
extension**



**Management of
development of
central search
portal**



**Recommendations on
further development
and deployment**

Contact

**EU health information system unit
SD Epidemiology and Public Health**

Email: irene.kesisoglou@sciensano.be
Petronille.bogaert@sciensano.be



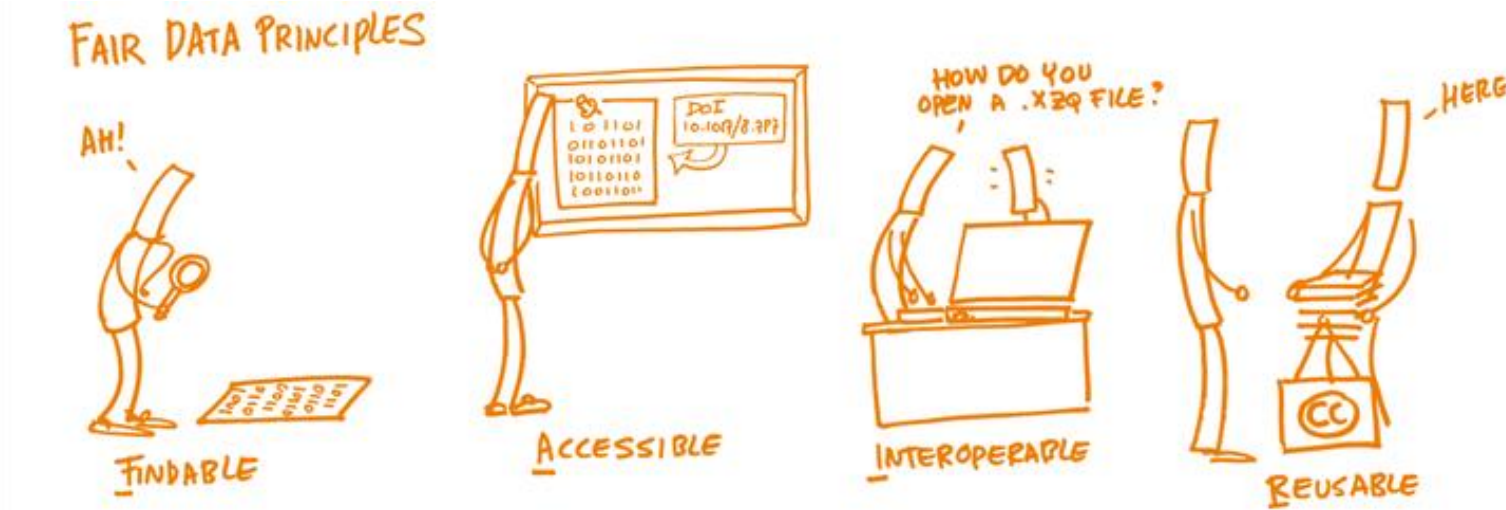
HEALTHYCLOUD

Health Research & Innovation Cloud

Assessing the FAIRness level of health-related data collections

Introduction to the FAIR principles

To enable and enhance the reuse of data by both humans and machines data collections need to comply with the following principles:



F.A.I.R.

- **Findable:** Your data can be discovered by others
 - Rich metadata & persistent identifier (e.g. DOI)
- **Accessible:** Your data can be made accessible by others
 - Protocol to access the data openly available
- **Interoperable:** Your data can be integrated/linked with other data
 - Structured data and implementation of internationally recognized standards
- **Reusable:** Your data can be reused by others
 - Context of the data & license

Landscape analysis

In HealthyCloud we carried out a landscape analysis of available health related data collections for secondary use in Europe and mapped their level of FAIRness (compliance with the FAIR principles)

- ❖ A health-related catalogue matrix was designed using a survey, asking key information from 26 data collections regarding the data quality and compliance with the FAIR principles.
 - **Administrative aspects:** Title, URL, contact details of data controller etc.
 - **Information about the data:** Storage, level of aggregation, domain of data etc.
 - **Data completeness and quality aspects:** geo coverage, timeline, upgrading periodicity, quality checks etc.
 - **Metadata:** existence of metadata on the datasets publicly available etc.
 - **Compliance with the FAIR principles...**

FAIRness evaluation tool

- ❖ After analysing the responses to the survey we evaluated their level of compliance with the FAIR principles using an adapted FAIRness evaluation tool:
- This tool is based on the Australian Research Data Commons tool (ARDC) and was adapted according to the questions of the survey
- The tool was then published in Zenodo in open access. [HealthyCloud FAIRness assessment tool | Zenodo](#)

HEALTHYCLOUD FAIRness evaluation tool

Document produced by HealthyCloud

2022-11-04 16:54:14

1 Data collection

Give a title to your data collection FAIR assessment

This FAIRness evaluation tool has been created by members of the [HealthyCloud consortium](#), specifically Work Package 3 (WP3). The tool is based on the Australian Research Data Commons (ARDC) [FAIR data self-assessment tool](#).

This 2-in-1 tool (survey form and reporting) is designed to allow assessment of the FAIRness of health-related data infrastructures, and includes questions to assess each of the FAIR principles. These questions are based on a survey developed by members of the consortium with the aim of carrying out a landscape analysis of health-related data infrastructures.

There is an option to select 'I don't know' under each question. However, we would encourage users to consider if there is someone else in their organisation who does know the answer to that question, to increase the accuracy of the assessment. You would need to share the FAIRness report within your organisation. At each updating step, a new csv file can be produced and used to generate an updated FAIRness report.

Step by step guide for users:

- Input your answers for each question in the tool
- Click 'Download' and save the csv file
- Re-run the script*, specifying the location at which the csv file has been saved
- This creates a FAIRness report, including pie charts demonstrating the percentage scores for each principle as well as an overall score.
- Save and share the FAIRness report as an HTML document.

[*] The tool can be downloaded on [ZENODO.org](#) (R studio is required) or accessed on [MyBinder](#).

2 FAIR survey

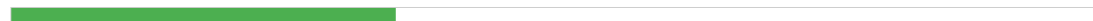
Findable

Accessible

Interoperable

Reusable

FAIRNESS score 35%



Findable

Do you have a unique identifier for your datasets?

Yes (5)

Do you have a unique identifier for the metadata?

Yes (4)

Do you produce or collect metadata for all your data (e.g. handbook, guide for users, description, keywords, timestamp, spatial coverage etc.)?

Brief title and description (2)

Do you have a public metadata catalogue service?

I don't know (0)

Accessible

Do you provide access to individual and/or aggregated data (for third party users)?

Publicly accessible aggregated (3)

Is it possible to extract the data from the data infrastructure (e.g. download) or do they have to stay in the data infrastructure?

Standard web service API (e.g. OGC) (4)

Do third party users have to register to the data infrastructure and have an account in order to access the data?

I don't know (0)

Are the conditions of access published?

Yes (1)

Interoperable

What is the format(s) for distributing data?

In a structured, open standard, non-machine-readable format (pdf) (2)

Which community-recognised vocabularies, standards or methodologies are used for metadata and data to facilitate interoperability?

Other (1)

Do you have a metadata record API endpoint (m2m) in place?

I don't know (0)

Reusable

Is there a clear procedure for third party users to request (the license) for data re-use?

I don't know (0)

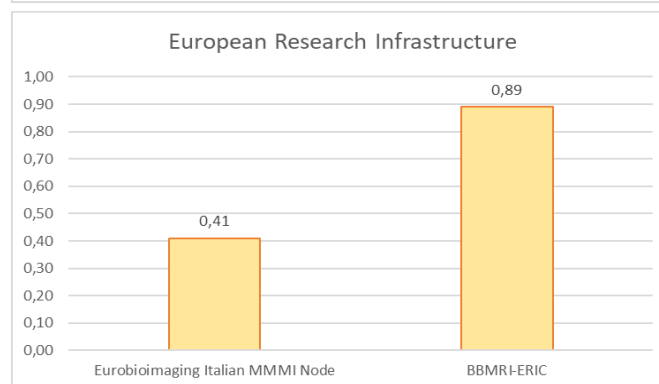
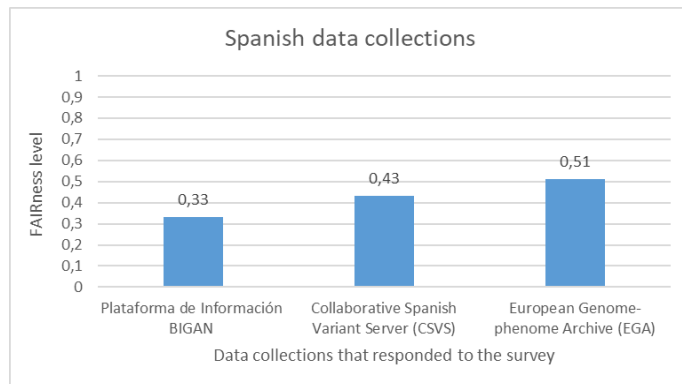
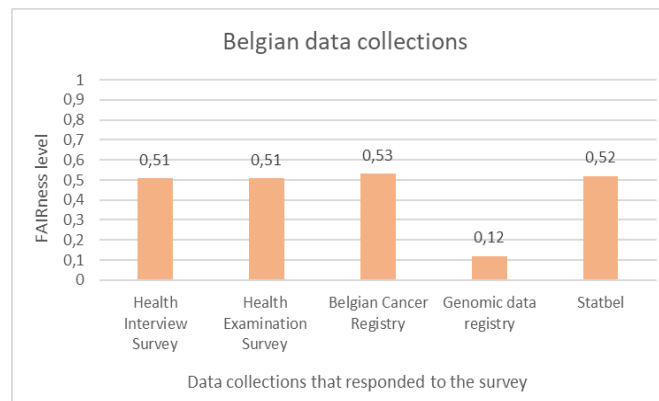
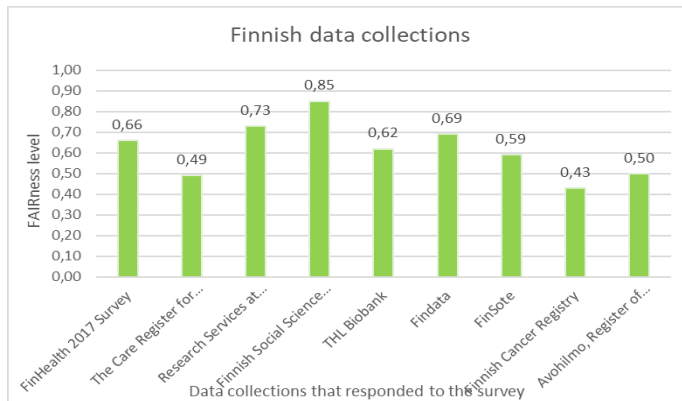
Have you placed the metadata related to your data infrastructure (that is, the above information provided in this survey) in another available source already?

I don't know (0)

Is it possible for third party users to access the data and re-use it for more than one purpose/project?

I don't know (0)

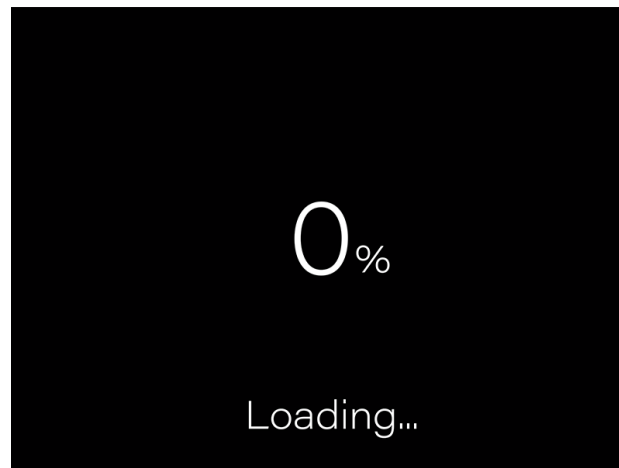
Results



FAIR is a *spectrum*

Add your data in domain specific
repositories to start your

FAIR data journey!



Thank you

Any questions?

Follow us on social media:

www.healthycloud.eu

 @HealthyCloudEU

 www.linkedin.com/company/healthycloudeu



HEALTHYCLOUD
Health Research & Innovation Cloud



The HealthyCloud project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement N° 965345.



FAIRsharing and the FAIR Cookbook: Helping you choose and use metadata standards

Susanna-Assunta Sansone, PhD

ORCID: 0000-0001-5306-5690

Twitter: @SusannaASansone

ELIXIR

Interoperability Platform Co-Lead



elixir-europe.org

Professor of Data Readiness

Associate Director, Oxford e-Research Centre



Group: datareadiness.eng.ox.ac.uk



Founding

Academic Editor

scientific **data**
SPRINGER NATURE

nature.com/sdata

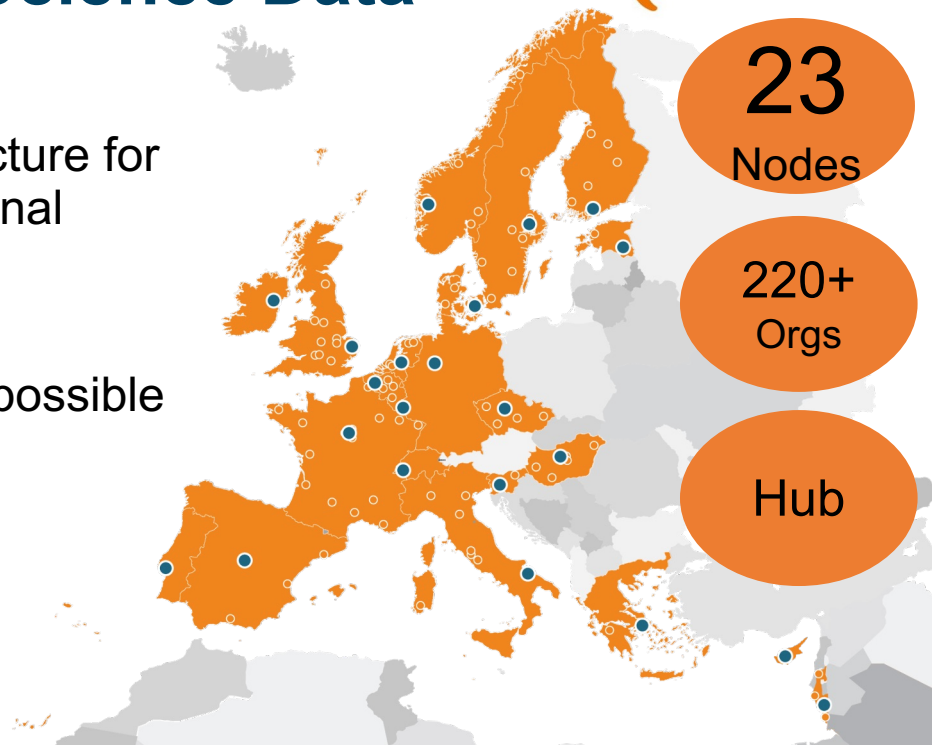
ELIXIR European Research Infrastructure for Life Science Data



Towards a federated digital infrastructure for Life Science Data, coordinating national capabilities

Data & software **FAIR and open** as possible
Transnational **access and analysis**

Gateway Communities of Practice,
European and Global initiatives,
Standards Bodies



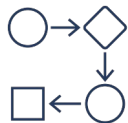
<https://elixir-europe.org>

The ELIXIR interoperability platform



FAIR services & resources

Registries, standards, ontologies, identifiers, data management platforms, stewardship tools, templates.



FAIR data techniques

Workflows, reproducible processing, transparent reporting and provenance, FAIR assessment and evaluation, FAIRification methods.

Leadership



Susanna-Assunta Sansone
(ELIXIR UK)



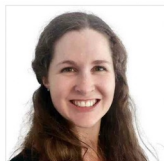
Chris Evelo
(ELIXIR Netherlands)



Tony Burdett
(EMBL-EBI)



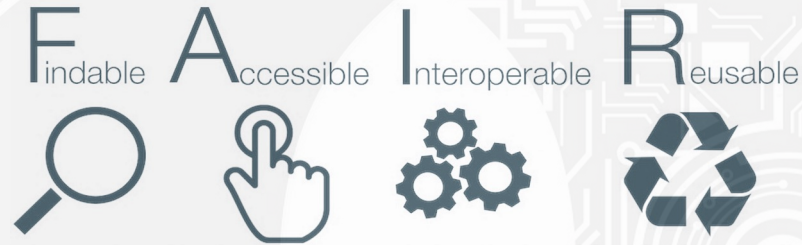
Peter MacCallum
(ad interim Platform
Coordinator, ELIXIR Hub)



Clare Garrard
(Platform Officer, ELIXIR
Hub)



Metadata make data count



Globally unique and
persistent **identifiers**

Detailed
provenance

Community defined
terminologies

Terms of **access**

Terms of
use

Community defined
descriptive **metadata**

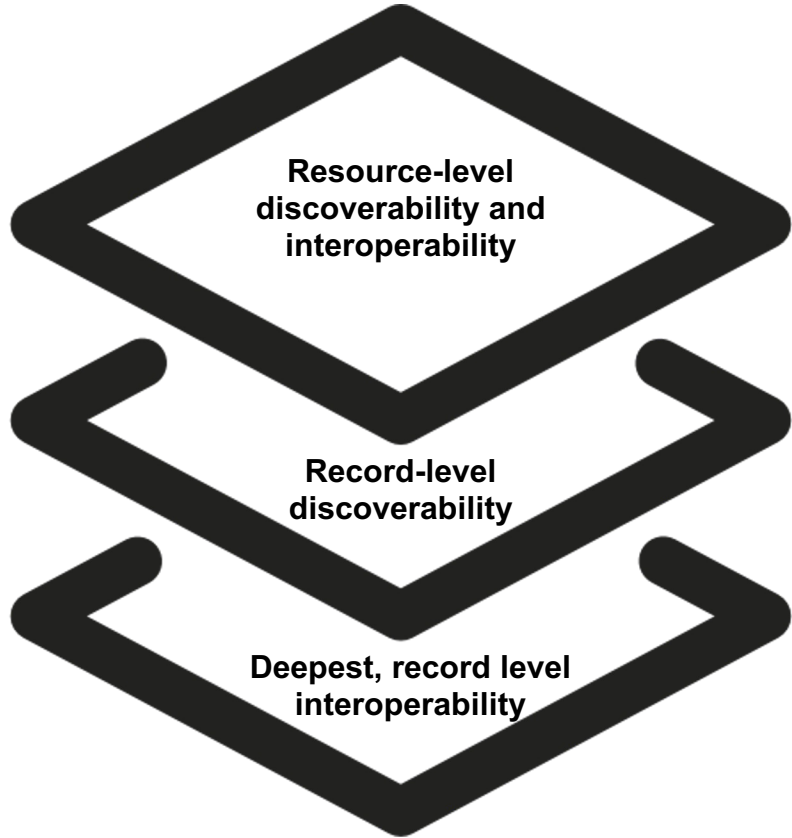
A continuum of features, attributes and behaviours



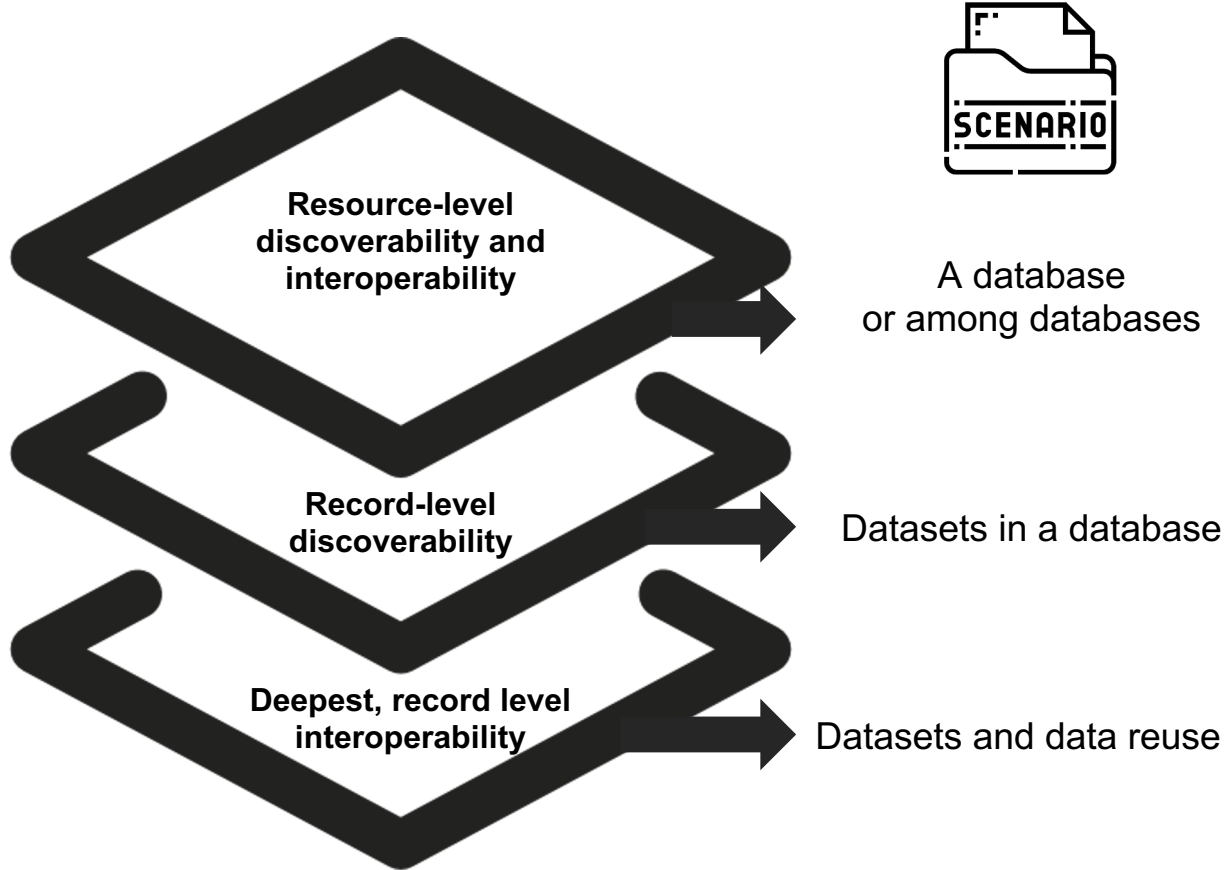
Fairness does not mean everyone gets the same. Fairness means everyone gets what they need.

Rick Riordan

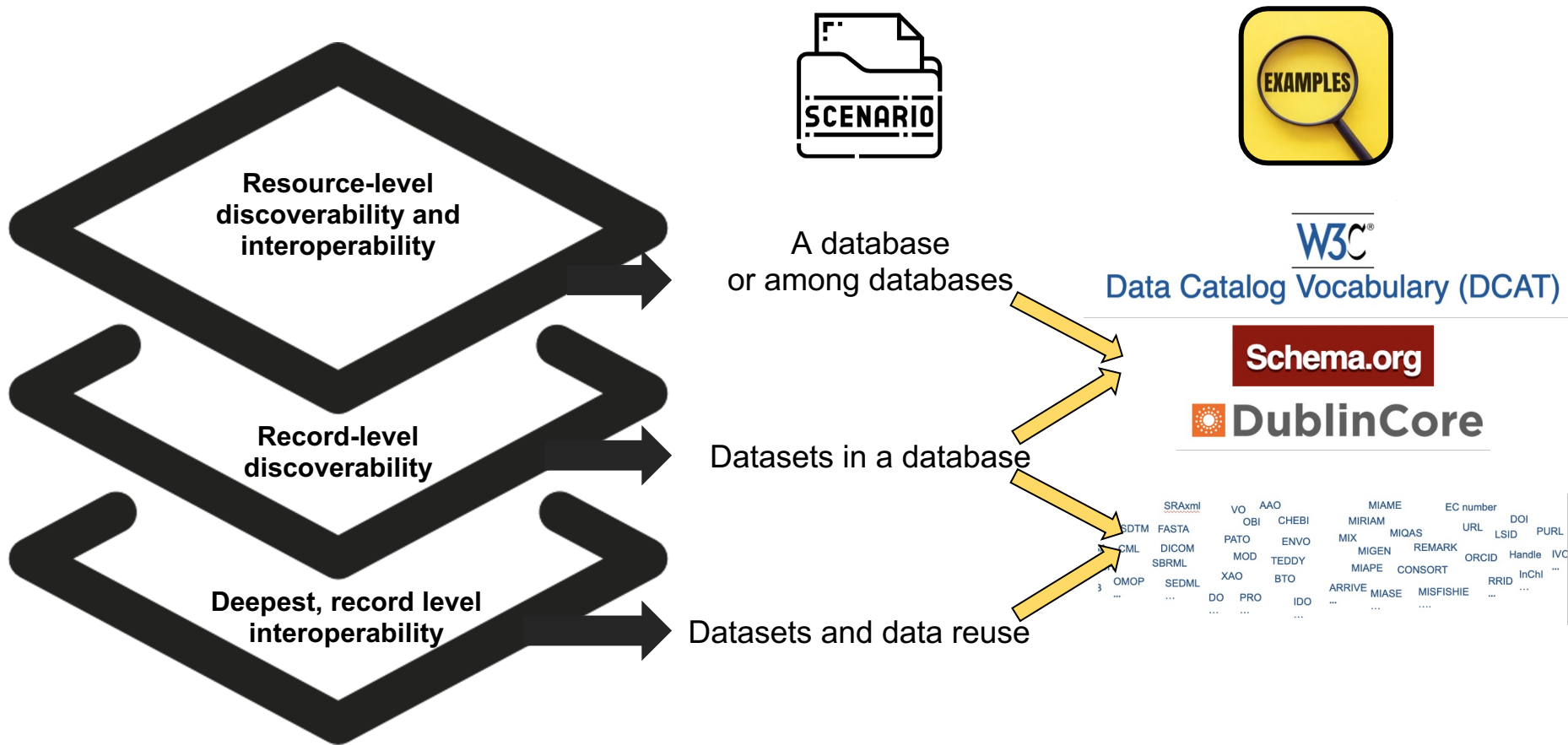
Metadata standards for different purposes



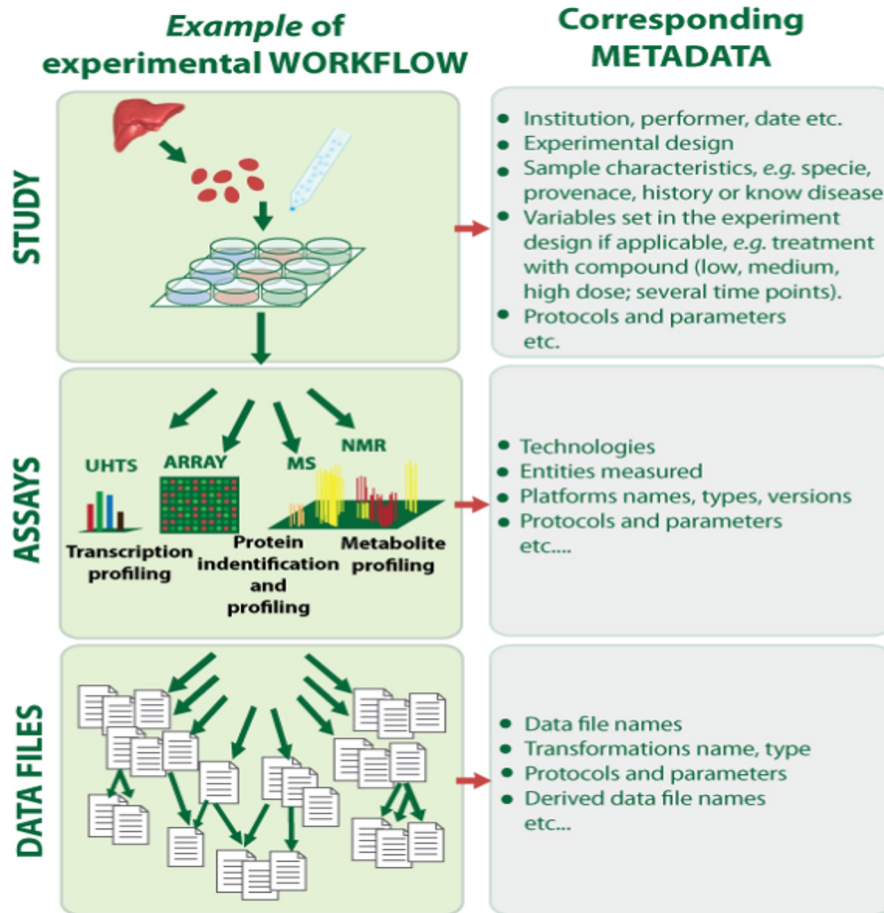
Metadata standards for different purposes



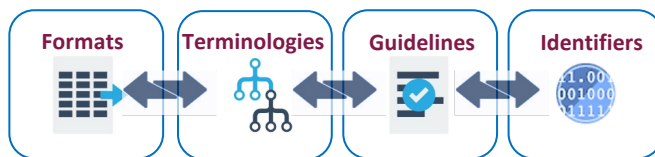
Metadata standards for different purposes



Standards to report metadata at dataset level



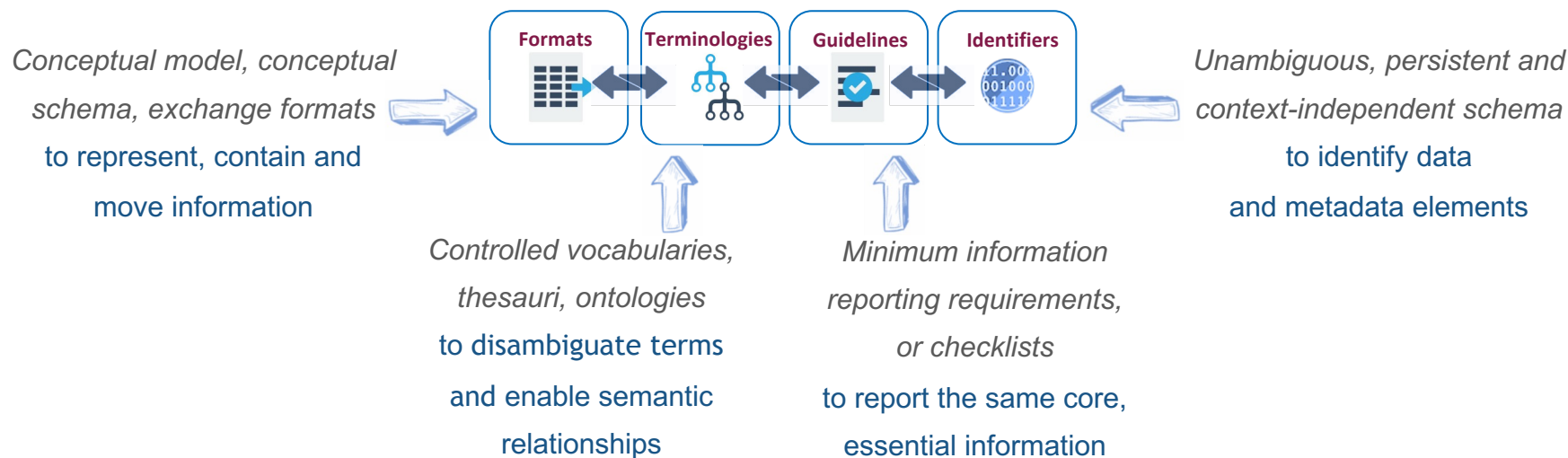
Standards to report metadata at dataset level



Source:

[FAIRsharing.org](https://fairsharing.org)

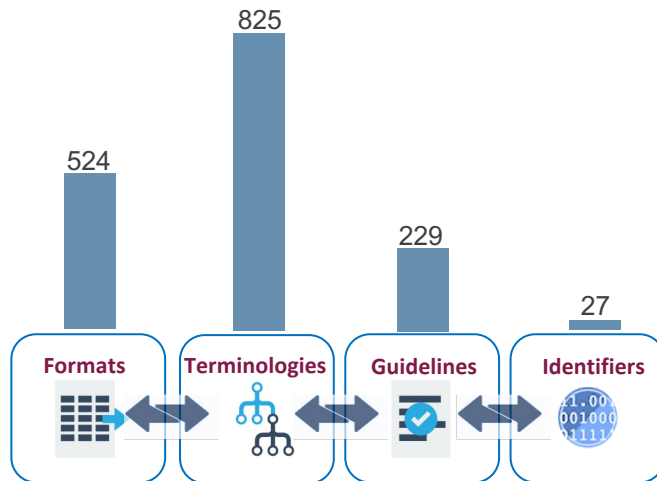
Standards to report metadata at dataset level



Source:

[FAIRsharing.org](https://fairsharing.org)

Natural, engineering, humanities & social sciences



More than 1600 data and metadata standards

Source:

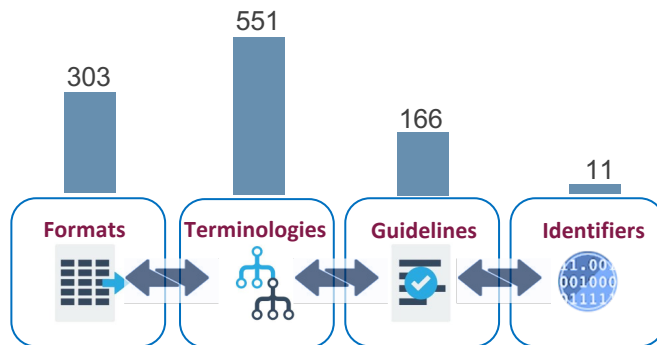
[FAIRsharing.org](https://fairsharing.org)

Life and biomedical sciences

Standard organizations, e.g.:



Grass-roots groups, e.g.:



More than 1000 data and metadata standards

SRAxml VO AAO MIAME EC number
 OBI CHEBI MIRIAM URL DOI
 SDTM FASTA PATO ENVO MIX MIQAS LSID PURL
 ISA CML DICOM MOD TEDDY MIGEN REMARK ORCID Handle IVOA ID
 CDASH SBRML XAO BTO MIAPE CONSORT RRID InChI ...
 MITAB OMOP SEDML DO PRO IDO ARRIVE MIASE MISFISHIE ...

Source:

[FAIRsharing.org](https://fairsharing.org)

Understanding their life cycle and landscape

Standard organizations, e.g.:

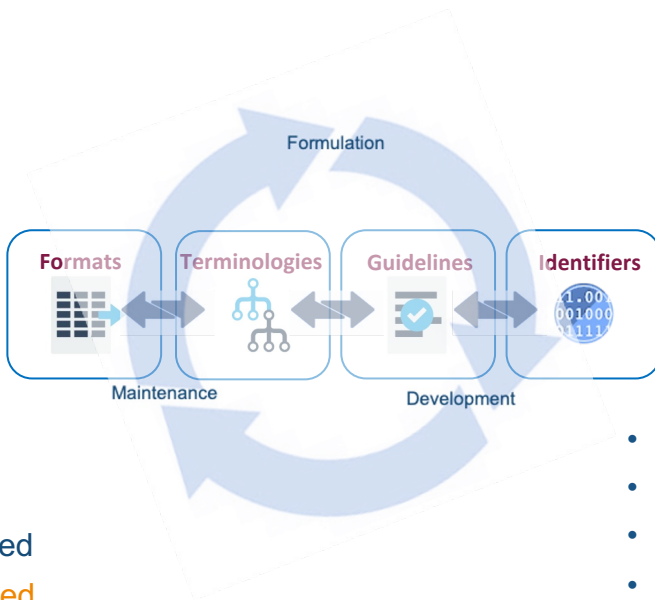


- **Industry-level standards**
- **Mostly regulators-driven**
- Participation is often regulated
- **Standards are sold or licenced**
- Formal development process, often less flexible, could be lengthy
- **Charges apply to advanced training or programmatic access**

Grass-roots groups, e.g.:



- **Mostly research-level standards**
- **Open to any interested party**
- Volunteering efforts
- **Standards are free for use**
- Development process varies, more flexible and adaptable to changes
- **Minimal or little funds for carry out the work, let alone provide training**

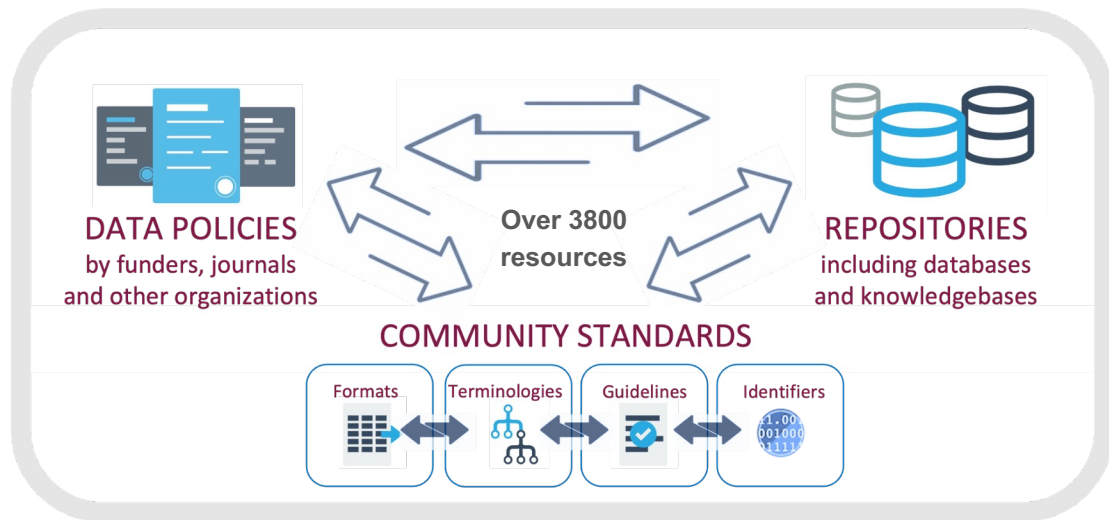


Source:

[FAIRsharing.org](https://fairsharing.org)

Informative and educational resource

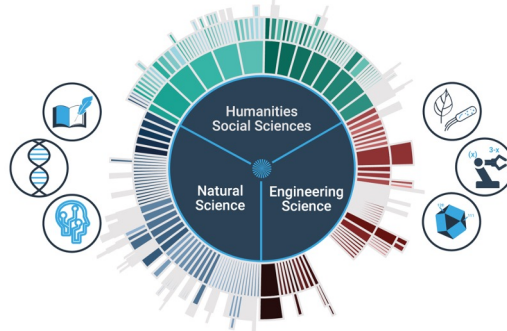
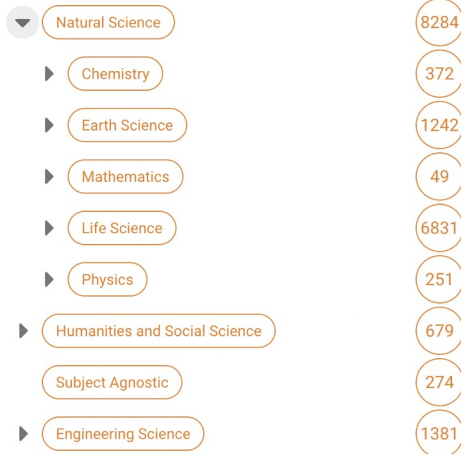
FAIRsharing.org



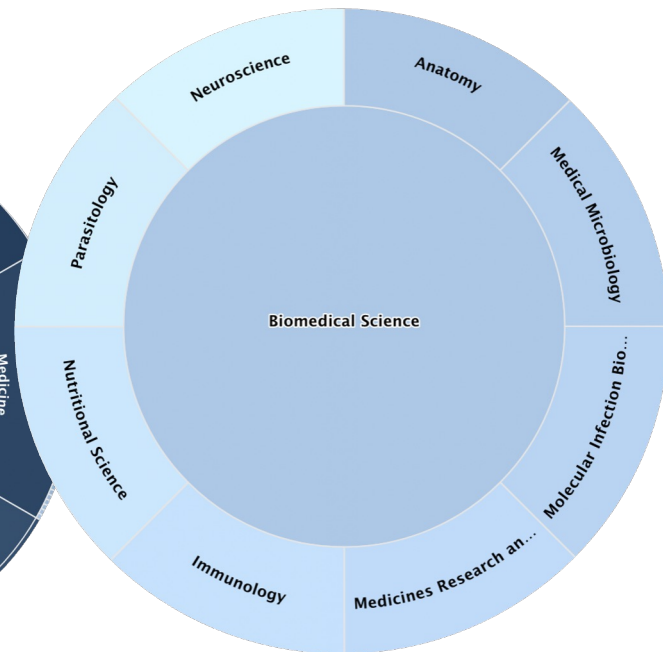
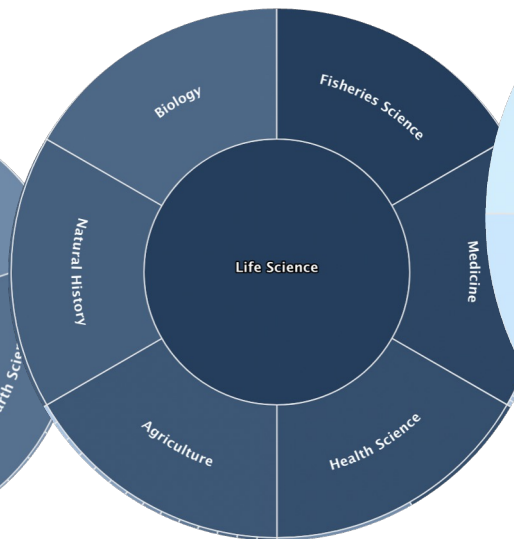
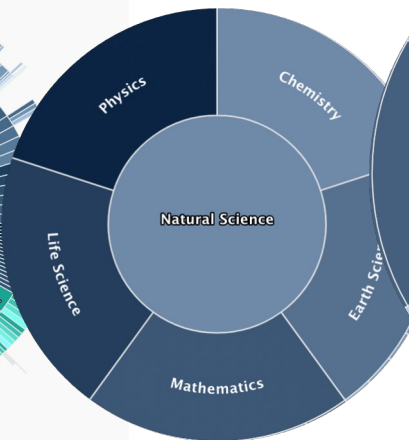
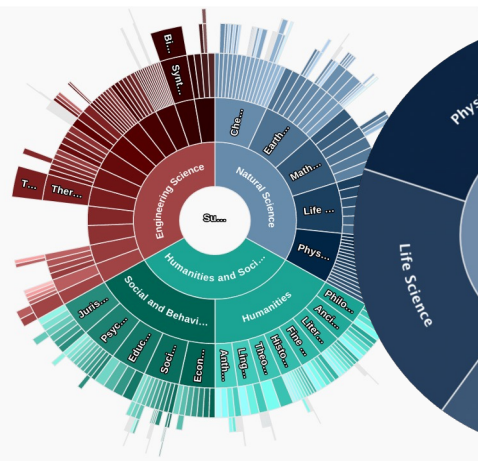
Guides **consumers** to *discover*, *select* and *use* these resources with confidence

Helps **producers** to make their resources more *visible*, more widely *adopted* and *cited*

FAIRsharing.org



Browse by subject



Track their evolution

Search through current results.

MATCH ALL TERMS

MATCH ANY TERM

MAINTAINED

NOT MAINTAINED

RECOMMENDED

NOT RECOMMENDED

READY

DEPRECATED

UNCERTAIN

IN DEV.

Record Status

ready

deprecated

uncertain

In Development

Clear All

Registry: Standard

Output status: deprecated

< 1 2 3 4 5 >

Displaying 1 to 30 of 140.

MAT

Minimal Anatomical Terminology

Material Element component of the SWEET ontology.

Material Element component of SWEET ontology. SWEET is a highly modular ontology suite with many separate ontologies covering Earth system scienc...

Earth Scien...

Not applic...

Related Standards

Implementing Databases

Endorsing Policies

SAO

Subcellular Anatomy Ontology

SAO describes structures from the dimensional range encompassing cellular and subcellular structure, supracellular domains, and macromolecules.

Anatomy

Cell

Homo sapi...

+2 more tags

Related Standards

HC

Habronattus Courtship Ontology

A demonstration of ontology construction as a general technique for coding ethograms and other descriptions of behavior into machine understandable forms. ...

Life Science

Behavior

Habronatta...

Related Standards

GENERAL INFORMATION

This record was deprecated on 2014-04-01 for the following reason(s): This standard has been revised and is superceded by ISO 19115-1:2014. However, many other resources still reference this particular version.

This record is replaced by:

[ISO 19115-1:2014 Geographic information -- Metadata -- Part 1: Fundamentals](#)



ISO 19115:2003 Geographic information -- Metadata (ISO 19115:2003)

Awaiting DOI



Type

Model and format


Registry

Standard

Description

ISO 19115:2003 defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

[illegible]



ISO 91: 07: 2019

ISO/DIS 20691

Biotechnology – Requirements for data formatting and description in the life sciences

ISO/TC 276
Biotechnology

BUY THIS STANDARD

FORMAT LANGUAGE

☒ PDF ☐ English ☐

☐ HTML ☐ English ☐

Only 58

BUY

GENERAL INFORMATION ☐ PREVIEW

Status : In order development

This document is in this draft international standard stage, indicating your national committee.

Editor(s) : Number of pages : 161

Technical Committee : ISO/TC 276, Biotechnology

ICS : 67.080 Biotechnology, Biotechnology

● Standard

Collects

Related To

Deprecates

Deprecated by

Implements

Implemented by

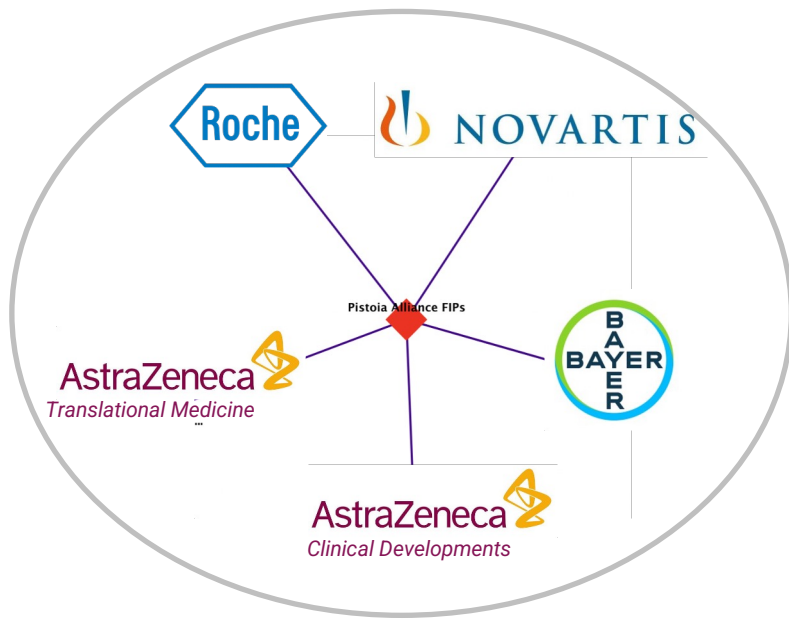
Shares Data With

Collected by

Extended by

Extends


A collaboration with their FAIR Implementation WG





URL: <https://fairsharing.org/3519>
(work in progress!)

Building and comparing “FAIR profiles”

CPT Current Procedural Terminology <p>Current Procedural Terminology is a medical nomenclature used to report medical procedures and services under public and private health insurance programs.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	LOINC Logical Observation Identifier Names and Codes <p>LOINC is a common language for clinical, research, and public health information. It is a coding system for observations, including laboratory tests, clinical measures, the vital signs and anthropometric measures, the identified codes, and the codes for the results of the tests.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	ATC Anatomical Therapeutic Chemical Classification <p>The Anatomical Therapeutic Chemical (ATC) Classification System is used for the classification of active ingredients of drugs according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. It is a hierarchical classification system.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases
RxNorm RxNorm <p>RxNorm provides standardized names for clinical drugs and their uses. It is a common language for clinical, research, and public health information. It is a coding system for observations, including laboratory tests, clinical measures, the vital signs and anthropometric measures, the identified codes, and the codes for the results of the tests.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	ICD-10 CM International Classification of Diseases, Tenth Revision, Clinical Modification <p>The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10 CM) is a worldwide classification for identifying diagnoses and reasons for visits in all intensive health care settings. The ICD-10 CM is based on the ICD-10, the official classification of diseases published by the World Health Organization (WHO) which replaced ICD-9 in 1989.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	ISO 3166-1:2013 ISO 3166-1:2013 Codes for the representation of names of countries and their subdivisions - Part 1: Country codes <p>ISO 3166-1:2013 is intended to be used in any application requiring the representation of current country names in coded form. ISO 3166-1 also includes basic guidelines for its implementation and maintenance.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases
NCD NCD Thesaurus <p>NCD Thesaurus (NCD) provides evidence-based terminology for non-communicable diseases (NCD) and other systems. It is a common language for clinical, research, and public health information. It is a coding system for observations, including laboratory tests, clinical measures, the vital signs and anthropometric measures, the identified codes, and the codes for the results of the tests.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	MeSH Medical Subject Headings <p>MeSH is the National Library of Medicine's controlled vocabulary thesaurus. It consists of a list of terms, having associated with them hierarchical structure that permits searching at various levels of specificity. MeSH thesaurus are arranged in both an alphabetical and a hierarchical structure. The hierarchical structure is the MeSH thesaurus structure. The alphabetical structure is the MeSH thesaurus structure.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	JSON-LD JavaScript Object Notation for Linking Data <p>JSON-LD is a JSON-based format for linking data. It is a common language for clinical, research, and public health information. It is a coding system for observations, including laboratory tests, clinical measures, the vital signs and anthropometric measures, the identified codes, and the codes for the results of the tests.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases
DC Dublin Core Metadata Element Set <p>The Dublin Core Metadata Element Set, which is often called Dublin Core (DC), is a metadata vocabulary scheme for description of any kind of resource such as documents in electronic and non-electronic form, digital materials such as video, sound, images, and other computer media for web access. Dublin Core Metadata may be used for metadata projects, from simple research.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	RSFS Resource Description Framework Schema <p>RSFS (RDF Schema) is the W3C vocabulary description language. RSFS defines classes and properties that may be used to describe objects, properties and other resources.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	DOI Digital Object Identifier <p>The Digital Object Identifier (DOI) system, registered in a joint initiative of three trade associations in the publishing industry (International Publishers Association, International Association of Scientific, Technical and Medical Publishers Association of America Publishers). The system was introduced in the first half of the 1990s. The International DOI Foundation (IDF) is the organization responsible for the system.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases
DCAT Data Catalog Vocabulary <p>DCAT (Data Catalog Vocabulary) is a W3C vocabulary description language. DCAT defines classes and properties that may be used to describe objects, properties and other resources.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	CDISC CDISC <p>CDISC (Clinical Data Interchange Standards Consortium) is a global community of experts in developing and advancing data standards of the highest quality. CDISC creates standards in data exchange, data collection, data management, and data analysis. CDISC standards are used by pharmaceutical, biotechnology, and medical device companies to ensure consistency and interoperability of data for more meaningful and efficient analysis of clinical research data.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases 	HTTPS Hypertext Transfer Protocol Secure <p>Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP). It is used for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS) or, formerly, Secure Sockets Layer (SSL). The content is therefore sent as H₂.</p> <p>Related Standards</p> <ul style="list-style-type: none"> International Classification of Diseases International Classification of Diseases International Classification of Diseases



Monarch Disease Ontology (MONDO)

 [10.25504/FAIRsharing.b2979t](https://doi.org/10.25504/FAIRsharing.b2979t) 

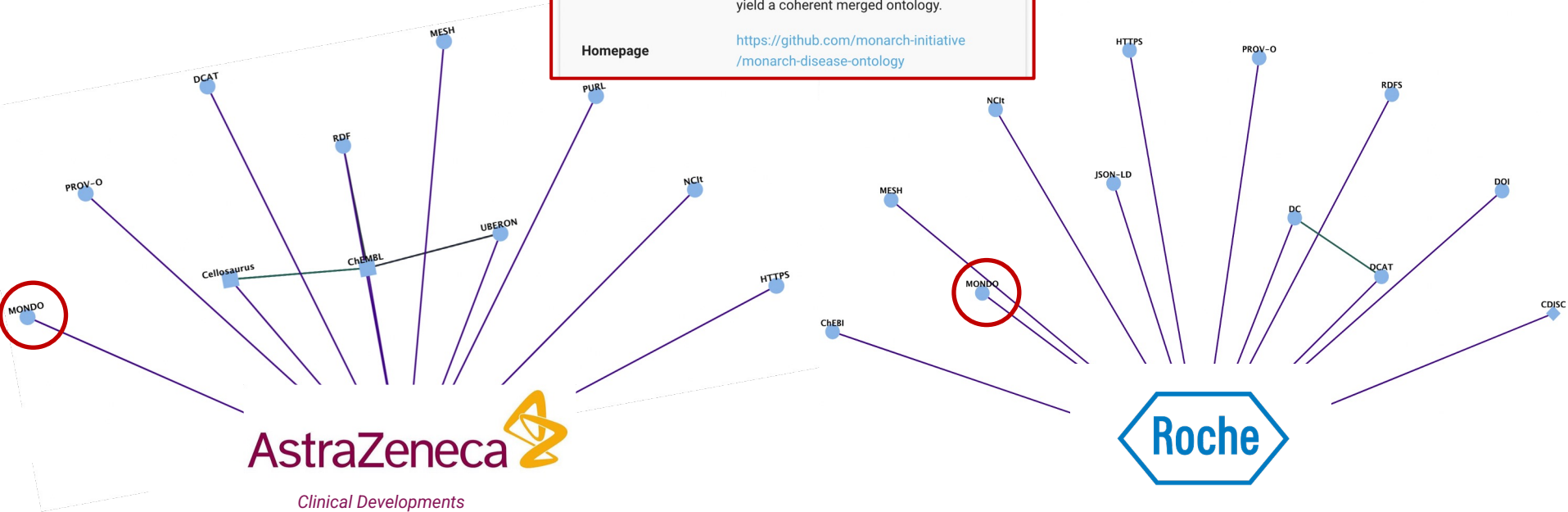
Type Terminology artefact

Registry Standard

Description MonDO (Monarch Disease Ontology) is a semi-automatically constructed ontology that merges in multiple disease resources to yield a coherent merged ontology.

Homepage <https://github.com/monarch-initiative/monarch-disease-ontology>

Snapshot of the semantic and syntactic standards used



FAIR cookbook: from knowledge to recipes

The screenshot shows the FAIR Cookbook website. At the top, the logo 'FAIR cookbook' is displayed. Below it is a search bar with the text 'Search this book...'. On the left side, there is a navigation menu with categories: 'FAIR Cookbook', 'FOREWORD' (Introduction, Ethical values of FAIR, Glossary), 'RECIPES' (Findability, Accessibility, Interoperability, Reusability, Infrastructure, Assessment, Applied examples), and 'ABOUT' (Community, Contribute, Platform). The main content area is titled 'The recipes' and contains a paragraph explaining the organization of recipes. Below this, there are four recipe categories: Findability (F), Accessibility (A), Interoperability (I), and Reusability (R). Each category lists exemplar recipes and a link to 'More about' the category. At the bottom, there are three more categories: Infrastructure, Applied Examples, and Assessment.

FAIR cookbook

Search this book...

FAIR Cookbook

FOREWORD

- Introduction
- Ethical values of FAIR
- Glossary

RECIPES

- Findability
- Accessibility
- Interoperability
- Reusability
- Infrastructure
- Assessment
- Applied examples

ABOUT

- Community
- Contribute
- Platform

Powered by **Jupyter Book**

The recipes

The FAIR Cookbook organizes the recipes according to the FAIR elements, audience type (your role), reading time, and level of difficulty. The FAIR Cookbook is a 'live resource'; recipes are added and improved, iteratively, in an open manner, therefore bear with us if several sections are work in progress! Below there are links to some key recipes, click on them to explore their content; otherwise use the main menu on the left hand side to browse all the current recipes.

F Findability

Exemplar recipes:

- Unique, persistent identifiers
- Search engine optimization

→ More about Findability

A Accessibility

Exemplar recipes:

- Transferring data with SFTP
- Downloading data with Aspera

→ More about Accessibility

I Interoperability

Exemplar recipes:

- Selecting terminologies and ontologies
- Creating a metadata profile

→ More about Interoperability

R Reusability

Exemplar recipes:

- Data licenses
- Declaring data's permitted uses

→ More about Reusability

Infrastructure

Applied Examples

Assessment

A collection of recipes that cover the operational steps of FAIR data management.
Example:

The example recipe card is titled 'Selecting terminologies and ontologies'. It features a 'Recipe Overview' section on the left with icons for 'Reading Time' (15 minutes), 'Executable Code' (No), and 'Difficulty' (4 stars). The main section on the right is titled 'Selecting terminologies and ontologies' and includes a 'Recipe Type' (Guidance) and an 'Audience' (Principal Investigator, Data Manager, Terminology Manager, Data Scientist, Ontologist). At the bottom right, it says 'Cite me with FCB020'.

Recipe Overview

- Reading Time: 15 minutes
- Executable Code: No
- Difficulty: 4 stars

Selecting terminologies and ontologies

Recipe Type: Guidance

Audience: Principal Investigator, Data Manager, Terminology Manager, Data Scientist, Ontologist

Cite me with FCB020

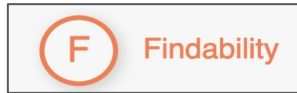
Authored by almost 100 data professionals from industry and academia, including:



New! Publication pre-print: <https://doi.org/10.5281/zenodo.7156792>

Define what your needs are

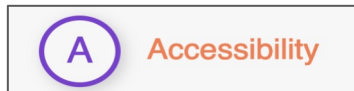
Goal: **improving visibility of content**



Goal: **semantic integration of datasets from multiple sources**

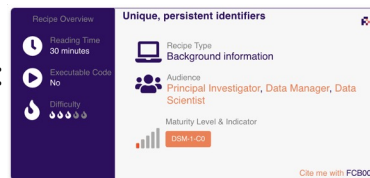
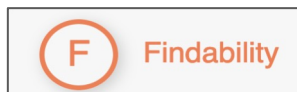


Goal: **security compliance and with regulators**

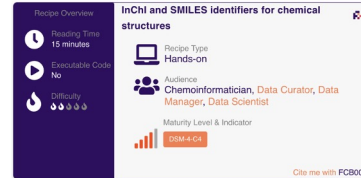


Define what your needs are

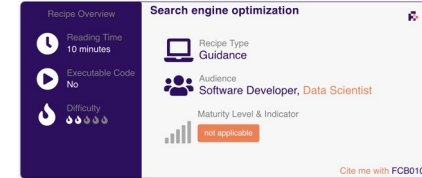
Goal: **improving visibility of content**, e.g.:



<https://w3id.org/faircookbook/FCB006>

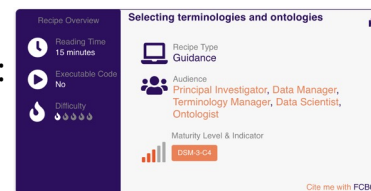
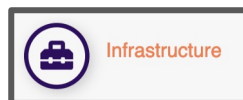


<https://w3id.org/faircookbook/FCB007>

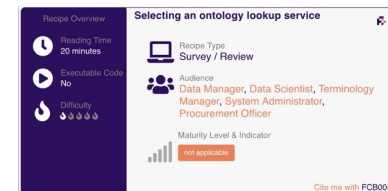


<https://w3id.org/faircookbook/FCB010>

Goal: **semantic integration of datasets from multiple sources**, e.g.:

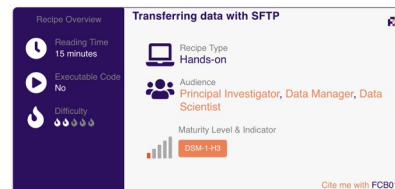


<https://w3id.org/faircookbook/FCB020>

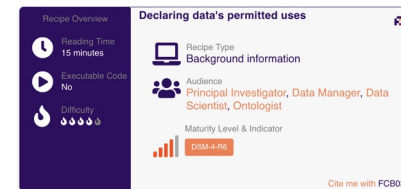


<https://w3id.org/faircookbook/FCB004>

Goal: **security compliance and with regulators**, e.g.:

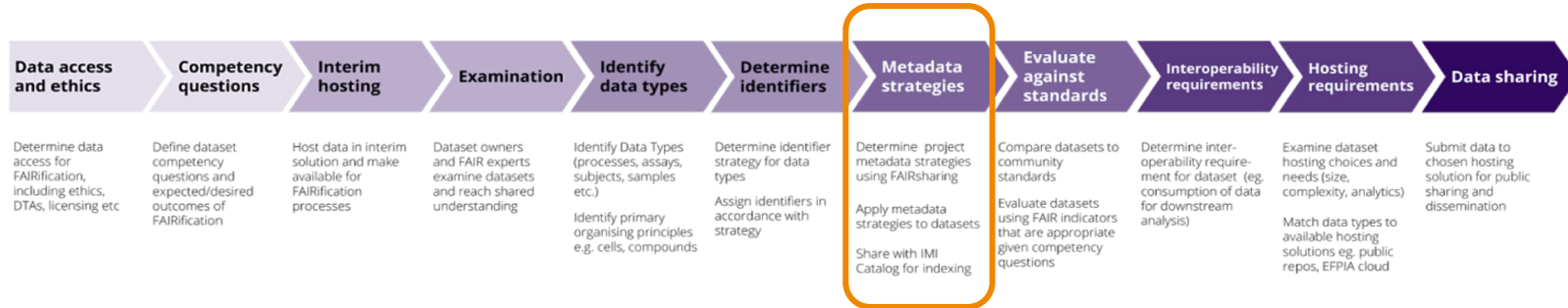


<https://w3id.org/faircookbook/FCB014>



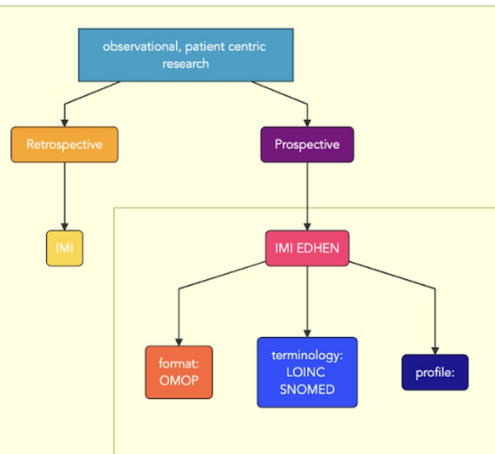
<https://w3id.org/faircookbook/FCB035>

FAIRification paths: one size does not fit all

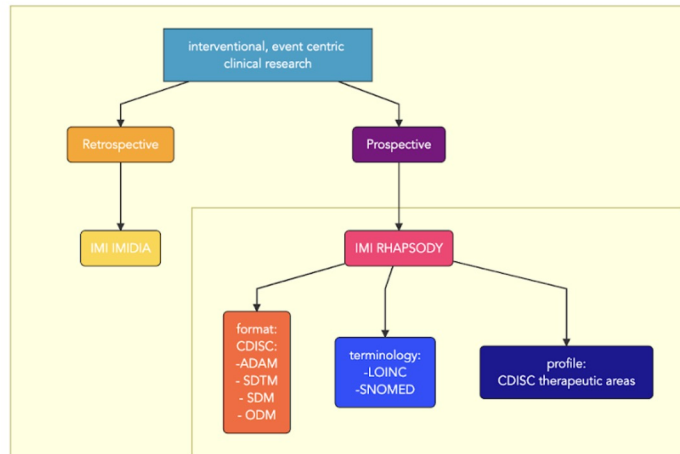


Different contexts mandate different metadata strategies

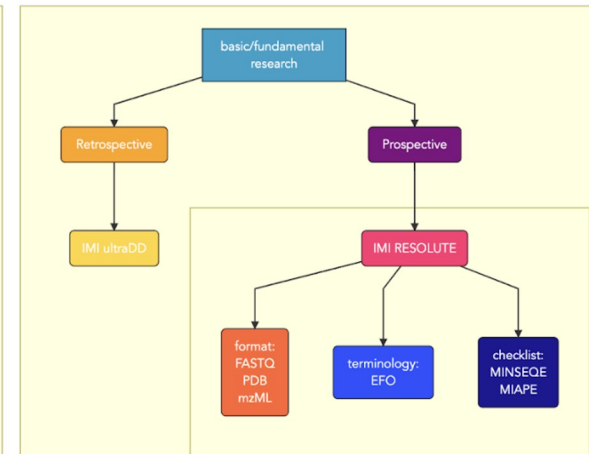
Clinical (observation based)



Clinical trial (event based) data

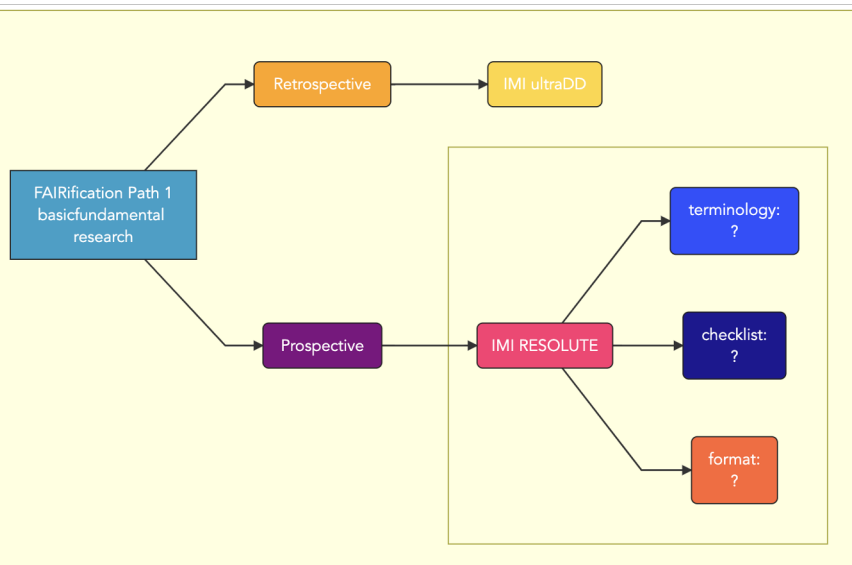


Molecular data



Selecting a 'standard stacks' for the FAIRification

Molecular data



Terminologies



FAIRsharing.org

R

FAIRsharing.org



EFO

Experimental Factor
Ontology
TERMINOLOGY ARTIFACT

standards > terminology artifact >
doi:10.25504/fairsharing.1gr4tz

Guidelines



R



Minimal Information about a
high throughput
SEQUENCING Experiment
REPORTING GUIDELINE

standards > reporting guideline >
doi:10.25504/fairsharing.a55z32

R



Minimum Information
About a Proteomics
Experiment
REPORTING GUIDELINE

standards > reporting guideline >
doi:10.25504/fairsharing.8vv5fc

Formats



R



FASTQ Sequence and
Sequence Quality Format
MODEL/FORMAT

standards > model/format >
doi:10.25504/fairsharing.r2ts5t

R



Protein Data Bank Format
MODEL/FORMAT

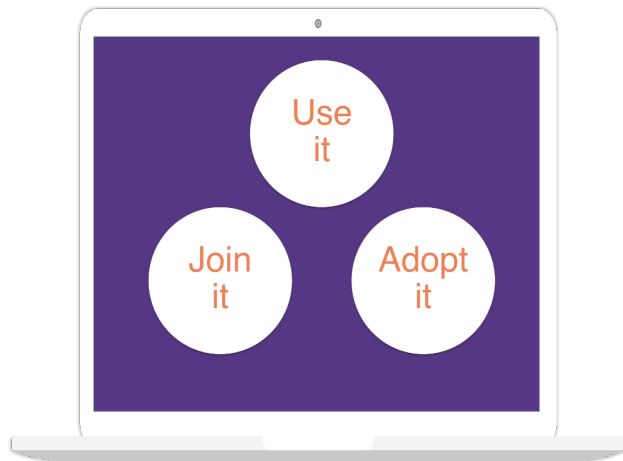
standards > model/format >
doi:10.25504/fairsharing.9y4cqW

R

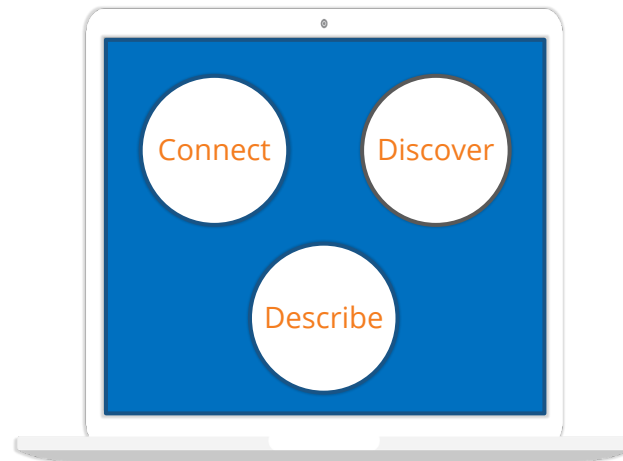


mzML
mz Markup Language
MODEL/FORMAT

standards > model/format >
doi:10.25504/fairsharing.26dmba



faircookbook.elixir-europe.org
fairplus-cookbook@elixir-europe.org



fairsharing.org
contact@fairsharing.org



PHIRI

Population Health Information
Research Infrastructure



BERLIN | 9-12 NOVEMBER 2022

What is metadata?

**Practical exercise: Create your own
metadata documentation
using DCAT standard**

Pascal Derycke

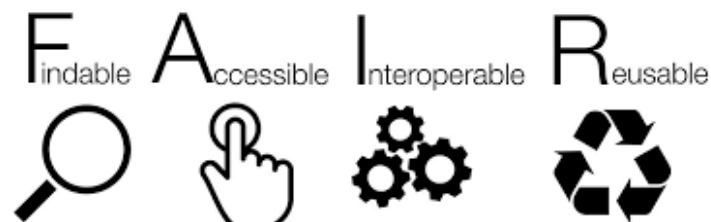
Pre-conference 09-11-2022



The Data Catalog Vocabulary (DCAT) is a W3C metadata recommendation for publishing data on the Web. DCAT is defined in RDF and reuses the Dublin Core Metadata standard.

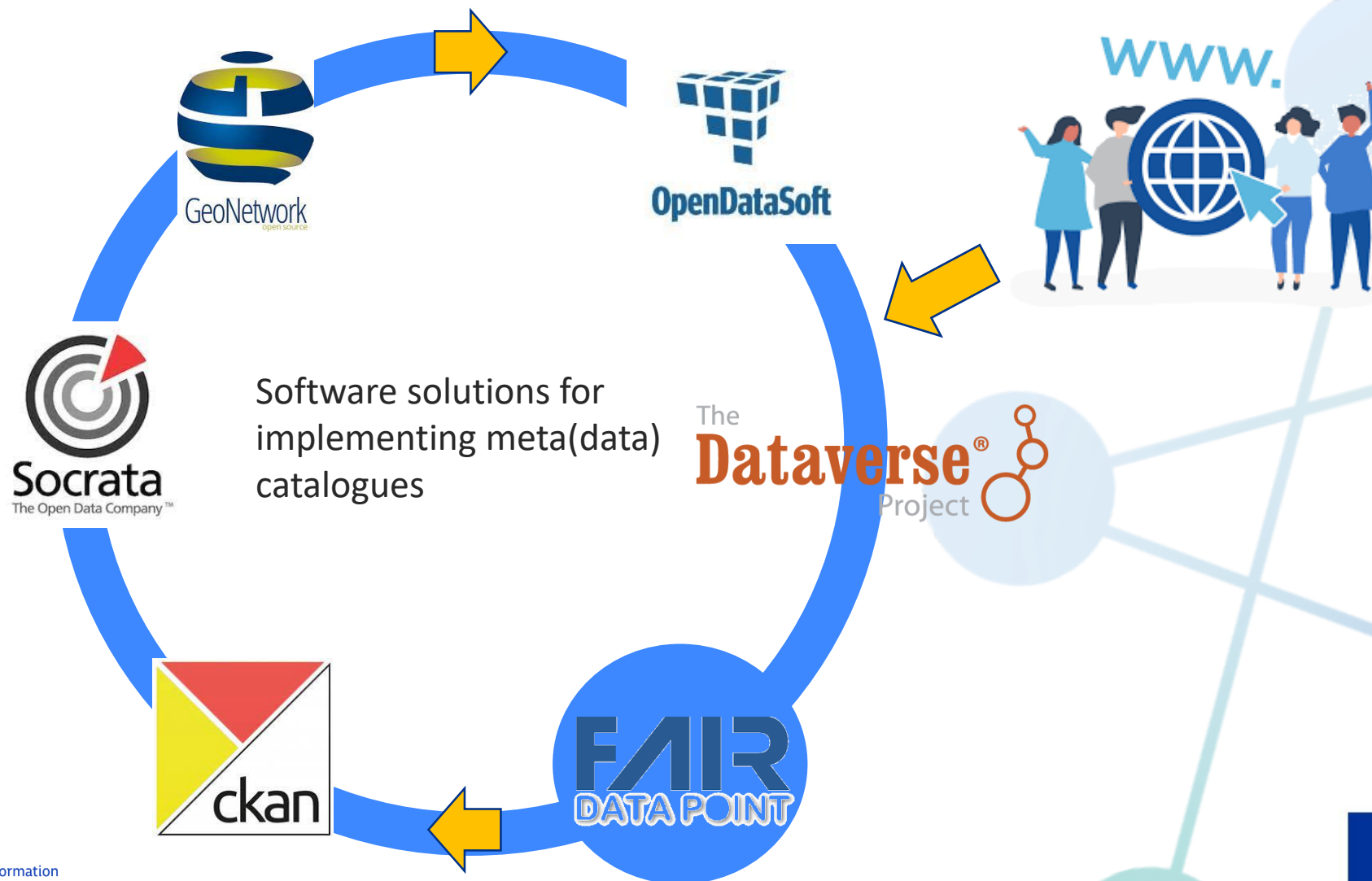
DCAT is designed to facilitate interoperability between data catalogs published on the Web.

By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications to easily consume metadata from multiple catalogs. It further enables decentralized publishing of catalogs and facilitates federated dataset search across sites. (www.w3.org/TR/vocab-dcat)



www.phiri.eu

Interoperability framework for exchanging information about data on the Web



Interoperability in practice:

The screenshot shows the data.europa.eu website. The header includes the logo and navigation links like 'Data', 'Studies', 'data.europa academy', 'News', and 'Contact'. A search bar is present. The main content area displays the dataset 'TABLE 3.15: patient-Discharges-mean-and-Median-length-of-stay-days-principal-procedure-sex-and-age-group-2017' by 'GovData'. It includes a machine translation notice and a detailed description of the dataset, which is based on the 2017 HIPE Discharges report from acute public hospitals in Ireland.

<https://data.europa.eu/data/datasets/b1a505b6-bd08-4063-90a1-1816eba20916?locale=en>

The screenshot shows the DATA.GOV.IE website. The header includes the logo and navigation links like 'Home', 'Datasets', 'Publishers', 'Suggest Data', 'Showcases', 'Contact', 'About', 'More', and 'EN/GA'. The main content area displays the dataset 'TABLE 3.15 : patient-discharges-mean-and-median-length-of-stay-days-by-principal-procedure-sex-and-age-group-2017' published by 'Health Service Executive'. It includes a 'Like' button, a 'Theme: Health' tag, and a detailed description of the dataset, which is based on the 2017 HIPE Discharges report from acute public hospitals in Ireland.

https://data.gov.ie/dataset/patient-discharges-mean-and-median-length-of-stay-days-by-principal-procedure-sex-and-age-2017?package_type=dataset

The metadata is exchanged in a structured format via API endpoints

- Structured formats: turtle, rdf/xml, json-ld data
- Metadata schema: DCAT vocabulary
- The API endpoint of a metadata catalogue (or any other information system) is a digital location (URL) from where the server receives requests and sends out responses in order to share its data with third-party apps following a set of rules (e.g.: RESTful API).

<https://data.europa.eu/api/hub/repo/datasets/b1a505b6-bd08-4063-90a1-1816eba20916.ttl?useNormalizedId=true&locale=en>

Practical exercise:

- Fill the form: <http://search.healthdataportal.eu/form/>
 - Part1: Fill few DCAT properties of the class Datasets for your dataset
 - Part2: Provide some suggestions for a health DCAT extension
 - Part3: Select some extra properties of the HealthInformationPortal.eu metadata model.
- Search for your data: <http://search.healthdataportal.eu>
- Download your metadata record as turtle, rdf/xml, json-ld data



Q&A

- Was this exercise useful?
- If we keep the form active, would you be interested to create metadata records for your health data? And have them published in the [HealthInformationPortal.eu](https://healthinformationportal.eu)?



PHIRI

Population Health Information
Research Infrastructure



BERLIN | 9-12 NOVEMBER 2022

Thank you

<https://HealthInformationPortal.eu>

Pascal Derycke: Pascal.Derycke@sciensano.be

