Preconference

WHAT IS METADATA? COMMON STANDARDS AND PROPERTIES

9 November | 13:30-17:00 CET

EUROPEAN PUBLIC HEALTH CONFERENCE

BERLIN | 9-12 NOVEMBER 2022

Pétronille BOGAERT
Sciensano

Truls KORSGAARD
Norwegian Directorate for e-Health

Hanna TOLONEN
Finnish Institute for Health and Welfare

Irini KESSISSOGLOU
Sciensano

Susanna-Assunta SANSONE
ELIXIR

Pascal DERYCKE
Sciensano

EUPHA
EUROPEAN PUBLIC HEALTH ASSOCIATION
PUBLIC HEALTH MONITORING AND REPORTING

PHIRI
Population Health Information Research Infrastructure

sciensano

NIPH
Norwegian Institute of Public Health

Direktoratet for e-helse

elixir

What is metadata?

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

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

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

Data

Metadata for humans

Metadata for machines

PHIRI
Population Health Information Research Infrastructure

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

Standards for dummies

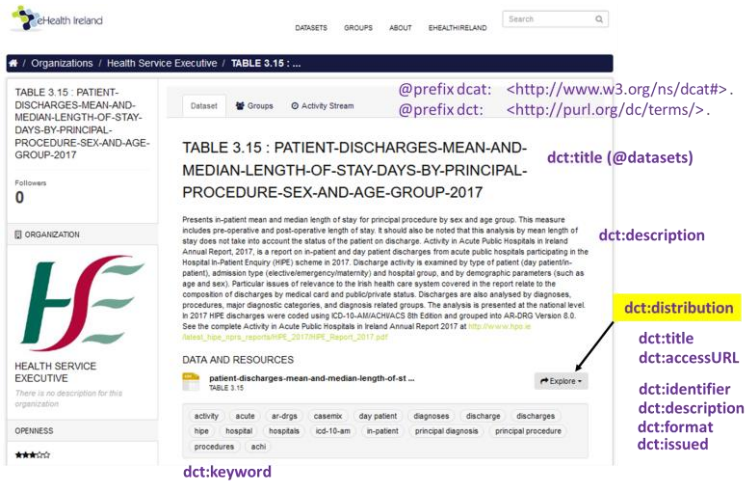
Standards used to structure metadata

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

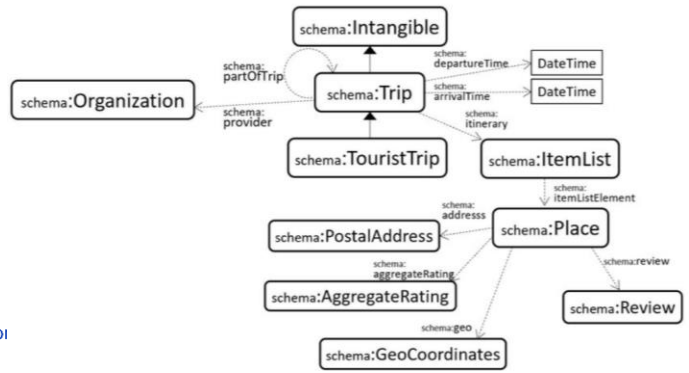
Descriptive metadata standards:

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

- **DCAT-AP**

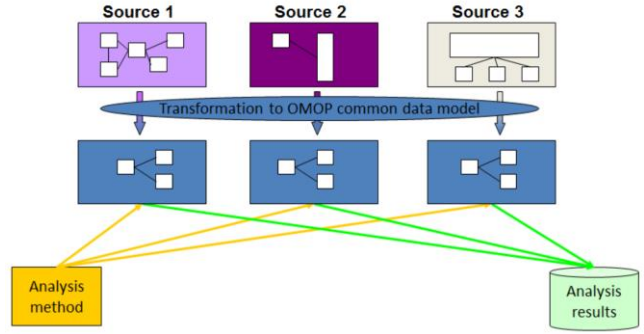


- **Schema.org**



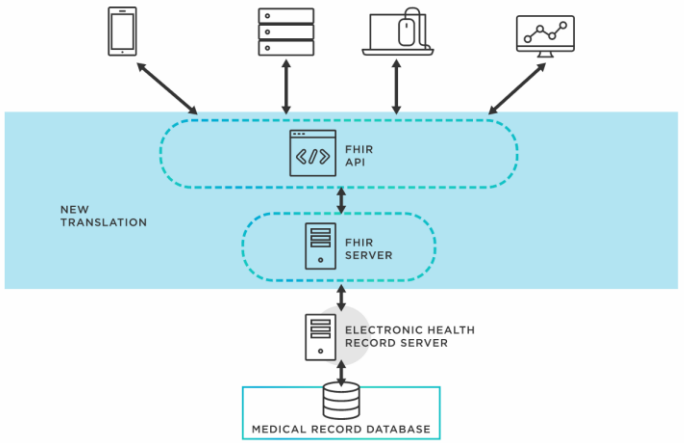
Structural metadata standards:

- **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.



PHIRI
Population Health Information Research Infrastructure

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

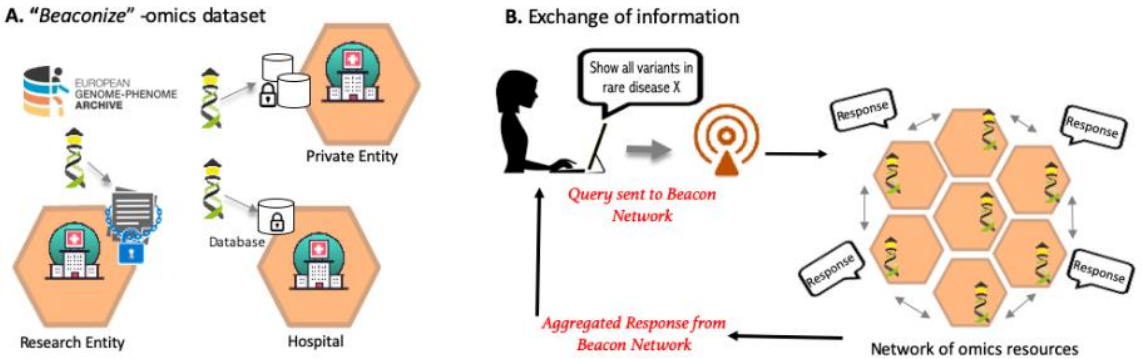
Standards used to structure images:

- **DICOM**

DICOM
Digital Imaging and Communications in Medicine

Standards used to make genomic data discoverable:

- **BEACON**



PHIRI
Population Health Information Research Infrastructure

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

Agenda

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

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**

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

Agenda

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

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

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

Agenda

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

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

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

PHIRI

Population Health Information Research Infrastructure

EUROPEAN PUBLIC HEALTH CONFERENCE

BERLIN | 9-12 NOVEMBER 2022

Conctact us: PHIRI.coordination@sciensano.be
Follow us on Twitter: @PHIRI4EU
@EU_HIS_unit
@PetronilleBo
@PHMRsection

sciensano

www.phiri.eu

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?

Direktoratet for e-helse

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

What is metadata?

Why is metadata important for researchers?

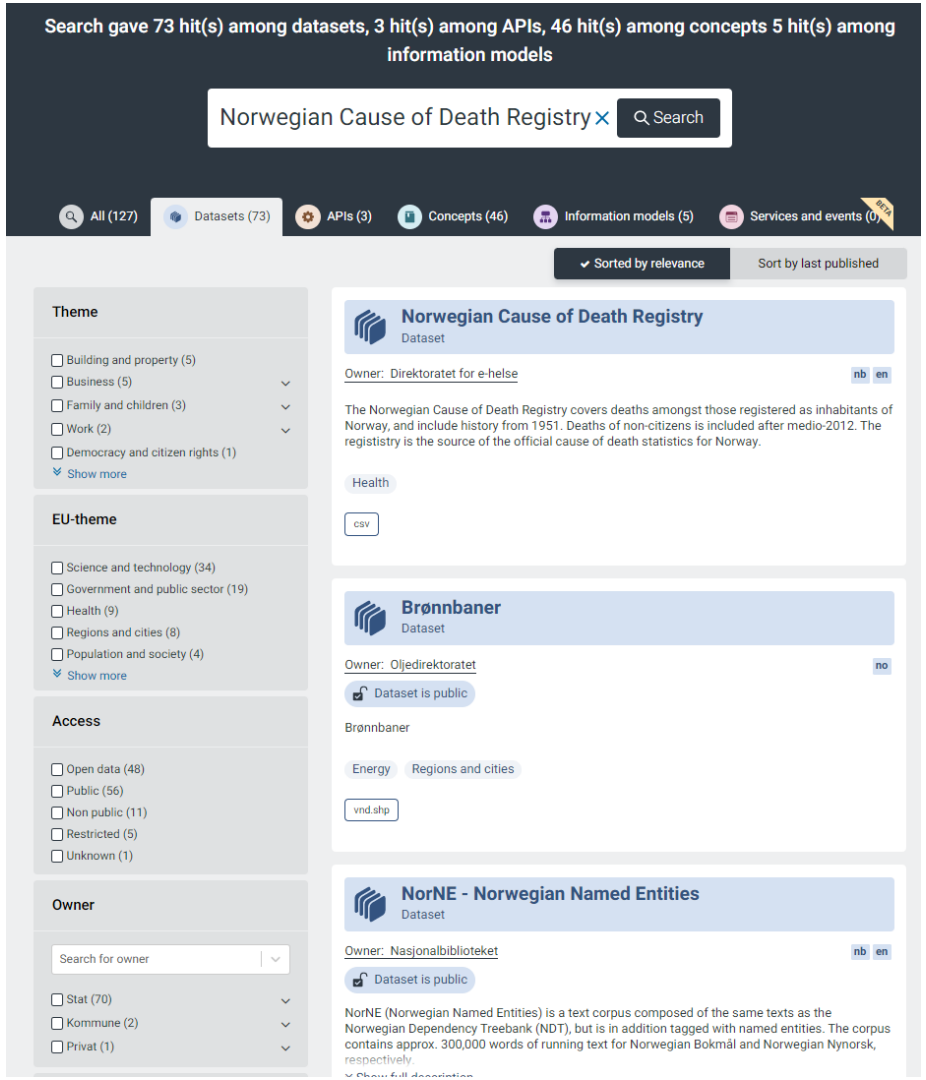
Why is common standards important?

Direktoratet for e-helse

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

What is metadata?

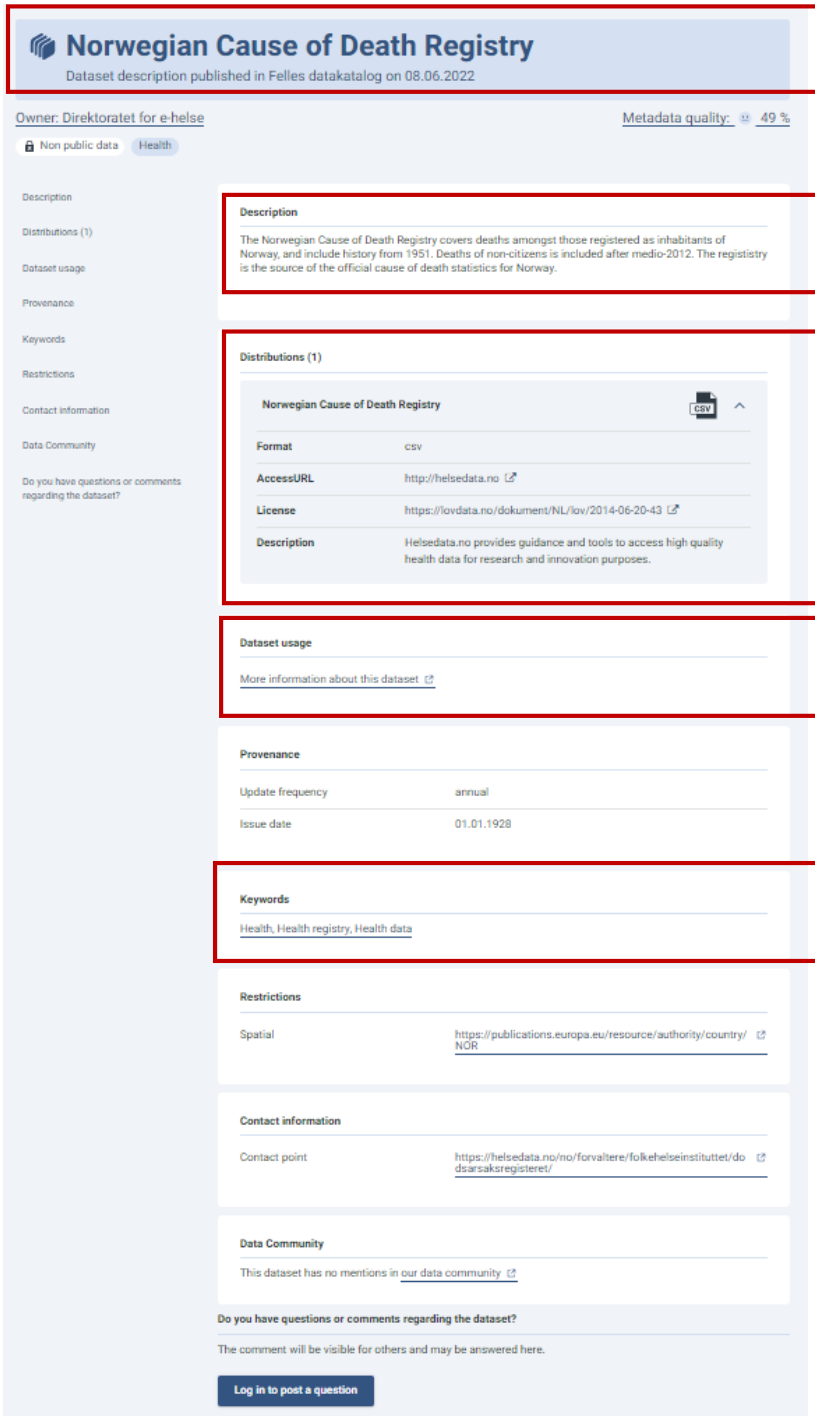
- **Metadata is information about for example a health registry. For example:**
  - **Title**
  - **Description**
  - **Contact point**
  - **Keywords**
  - **Coverage**
- **Metadata helps you to find the registry you are looking for**
  - **Filtering**
  - **Browsing**
- **And helps you to discover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



Direktoratet for e-helse

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 og 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 to much work…….**



DCAT-AP

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

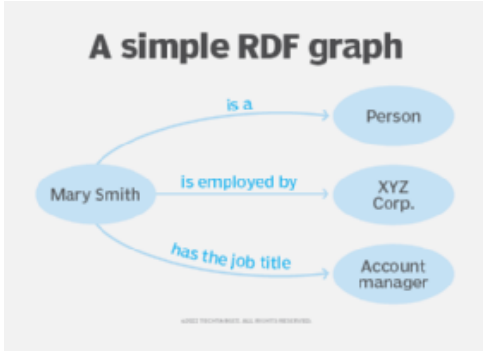
What is DCAT?

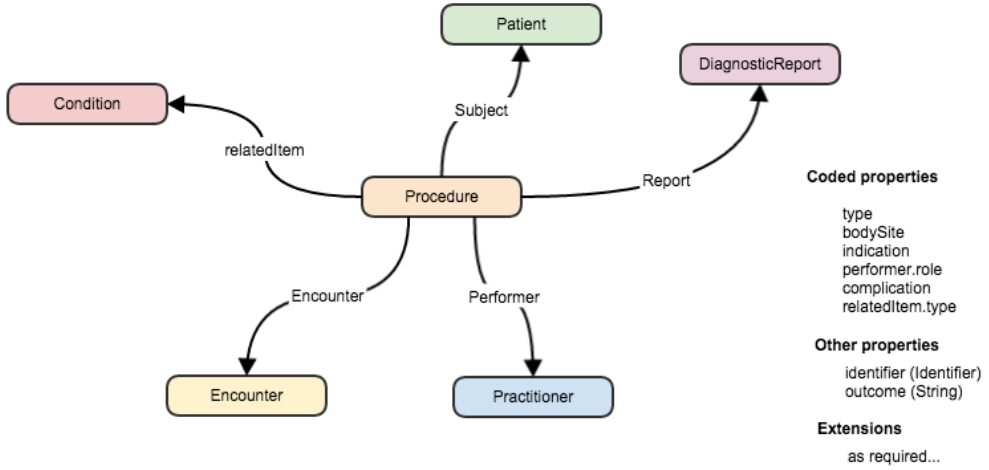
Data Catalog Vocabulary (DCAT) - Version 2

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.

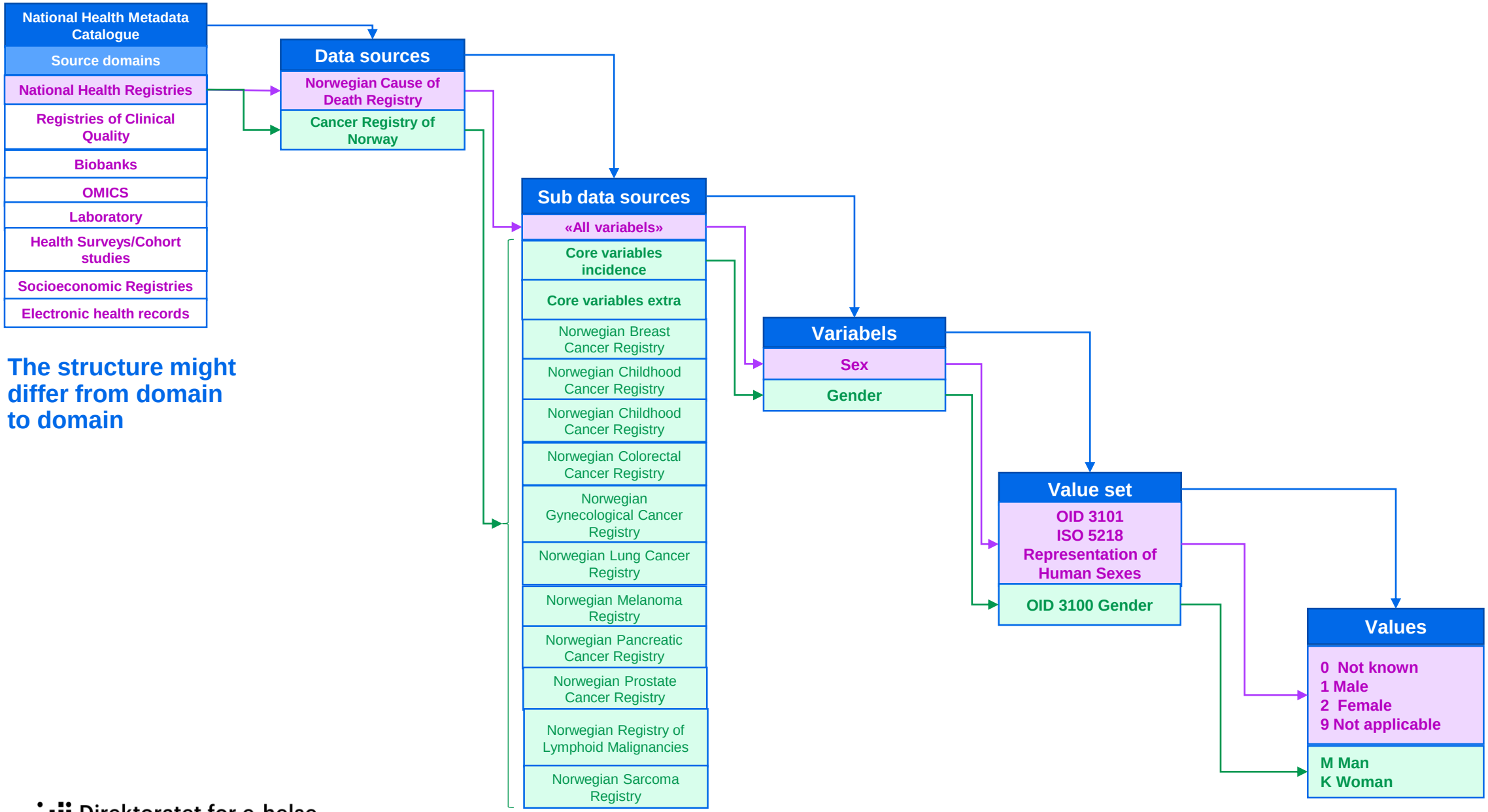
DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs. This can increase the discoverability of datasets and data services. It also makes it possible to have 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.





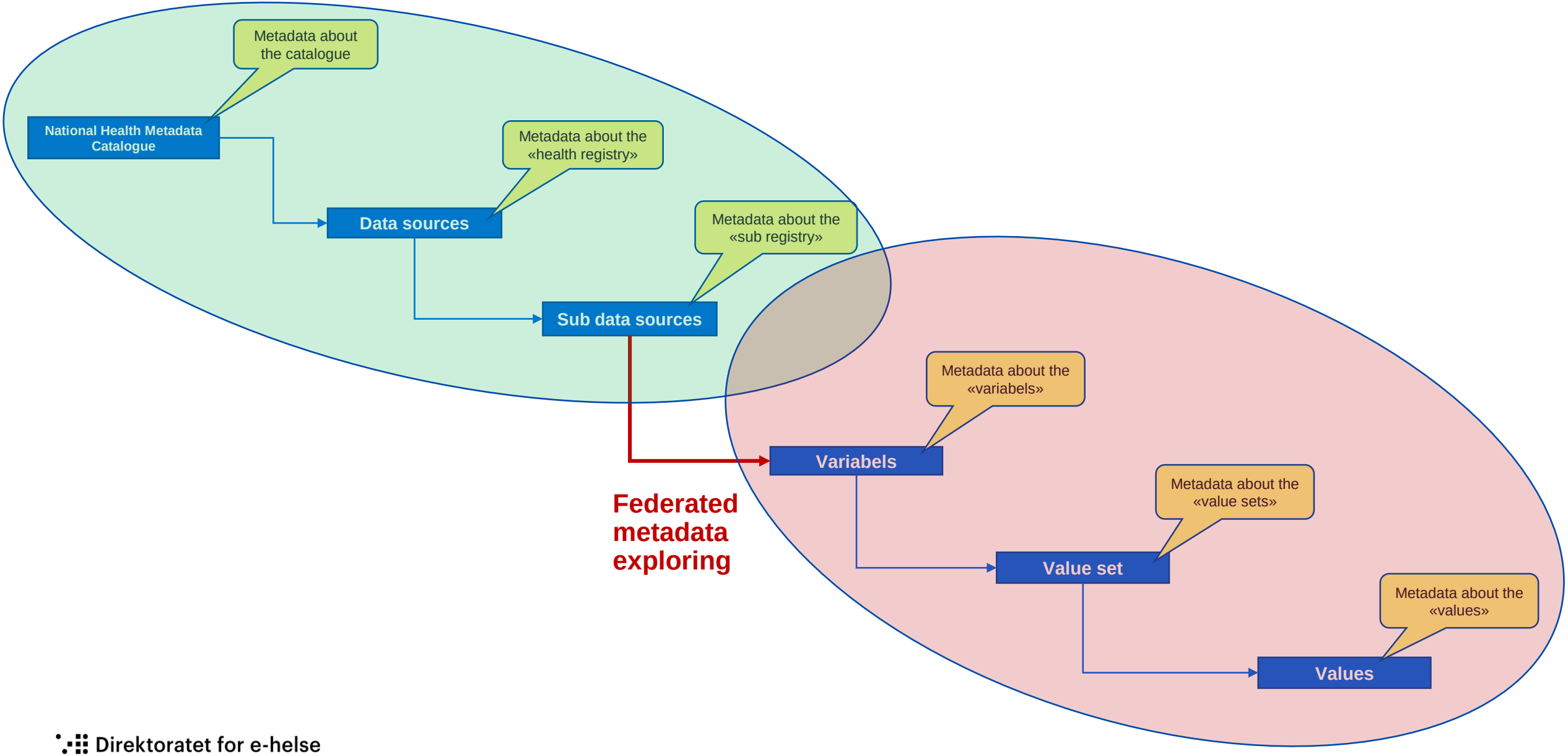
- RDF vocabulary
- Interopreability between catalogs
- Describe datasets and data services
- Vocabulary (controlled)
- Decentralized approach and federated research across catalogs
- Using the same query mechanism and structure
- This can increase the discoverability

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)



National Data Catalogue

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 | Distributions (1) | Dataset usage | Provenance | Keywords | Restrictions | Contact information | Data Community | Do you have questions or comments regarding the dataset?

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 regististry is the source of the official cause of death statistics for Norway.

Distributions (1)

Norwegian Cause of Death Registry (csv)

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/dodsarsaksregisteret/

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

National Health Metadata Catalogue

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.

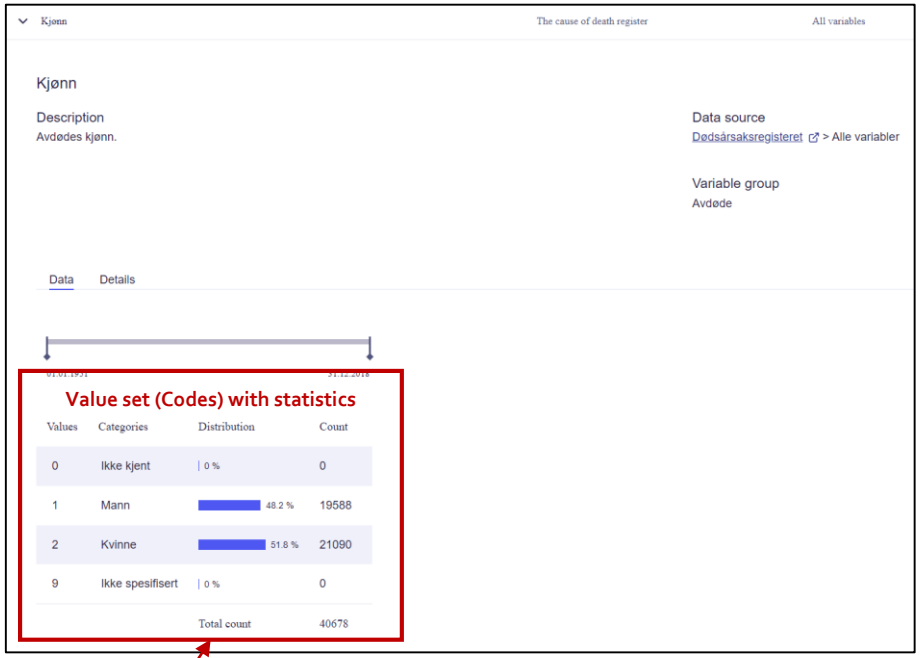
+ Create new application

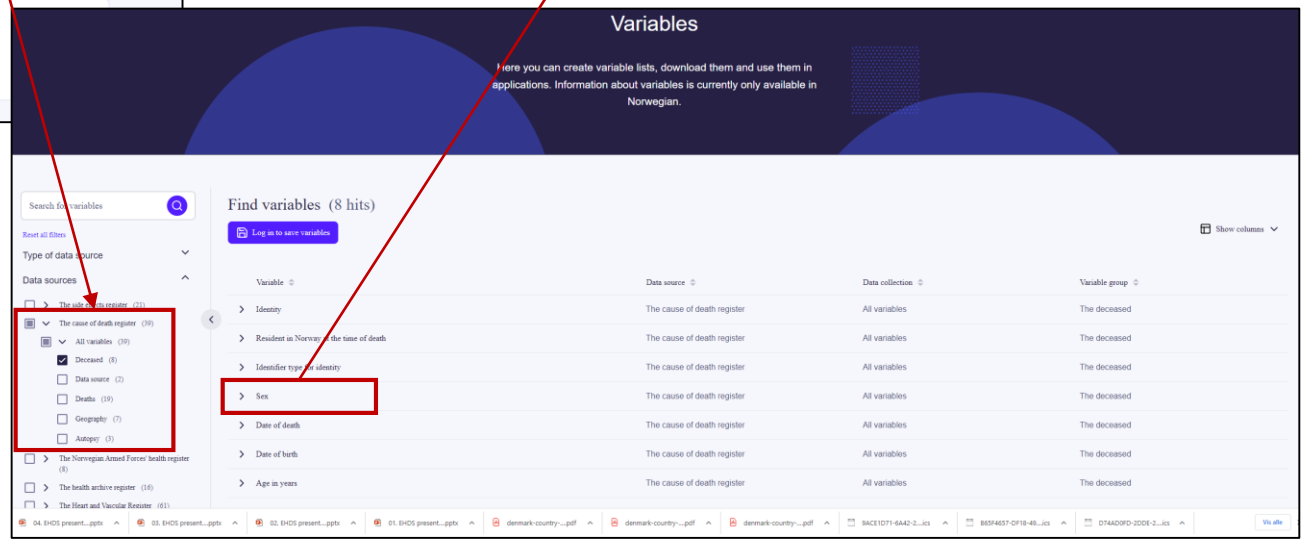
Contact

datatilgang@fhi.no

LINKS

Data source homepage





What is metadata?

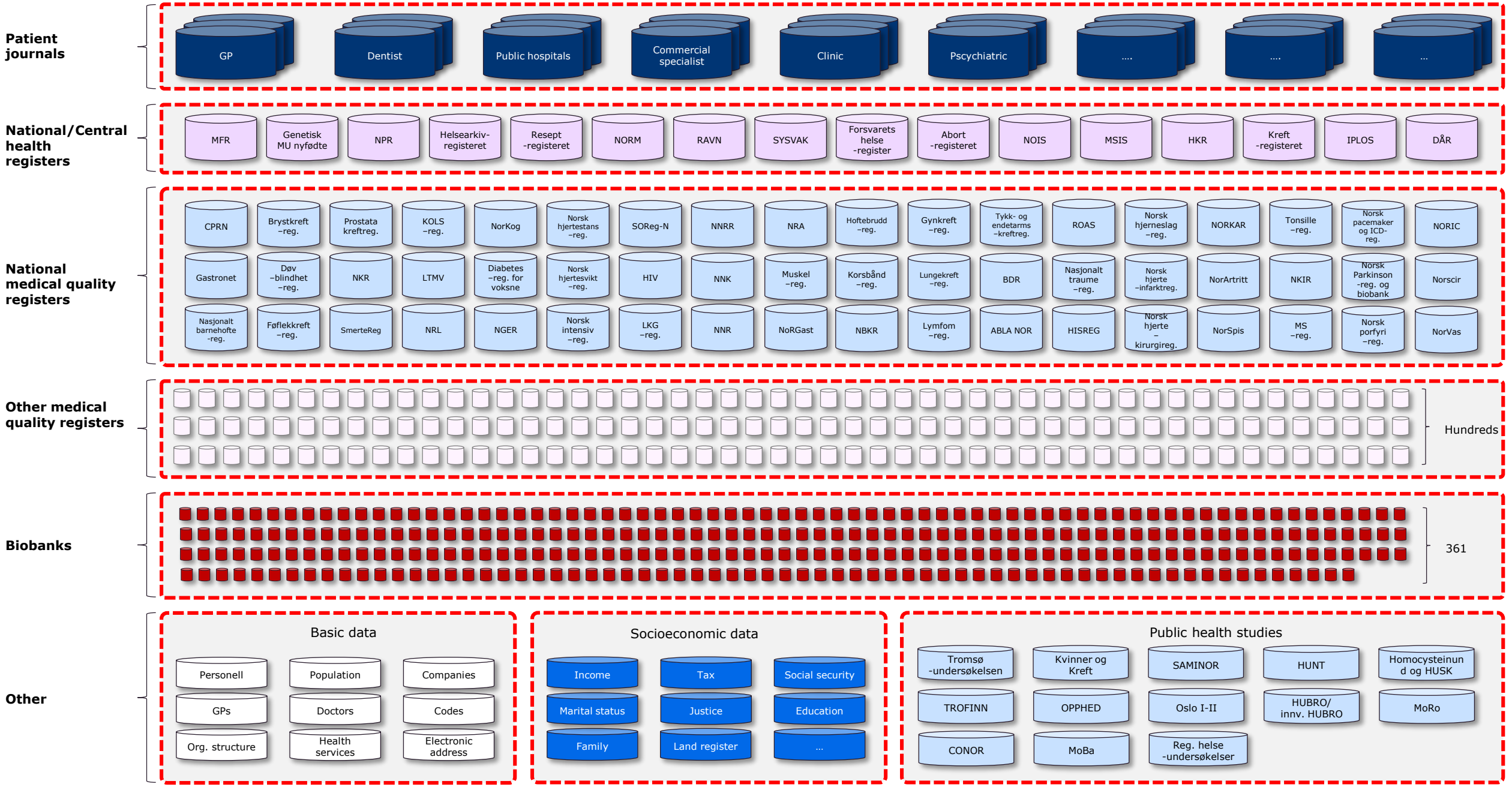
Why is metadata important for researchers?

Why is common standards important?

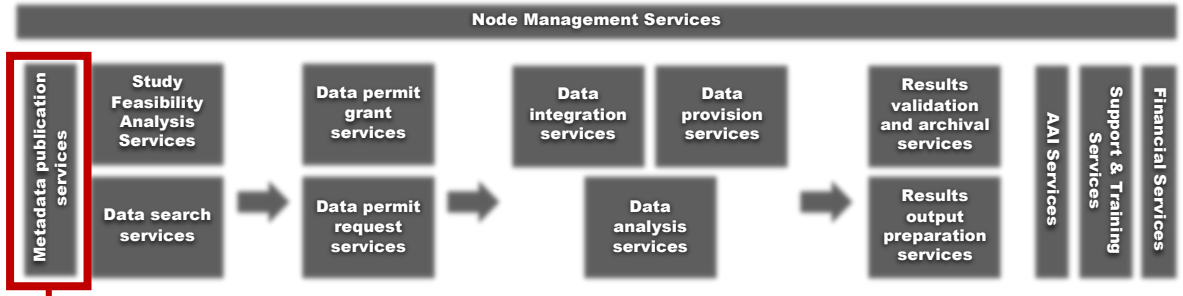
Direktoratet for e-helse

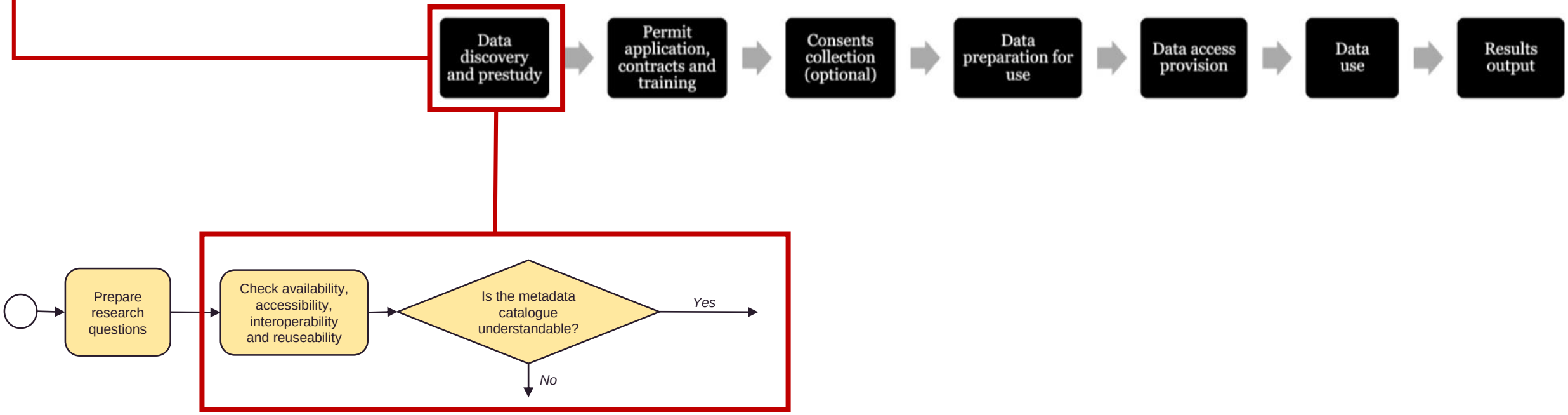
Truls Korsgaard
Senior advisor
The Norwegian Directorate of eHealth
truls.Korsgaard@ehelse.no

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

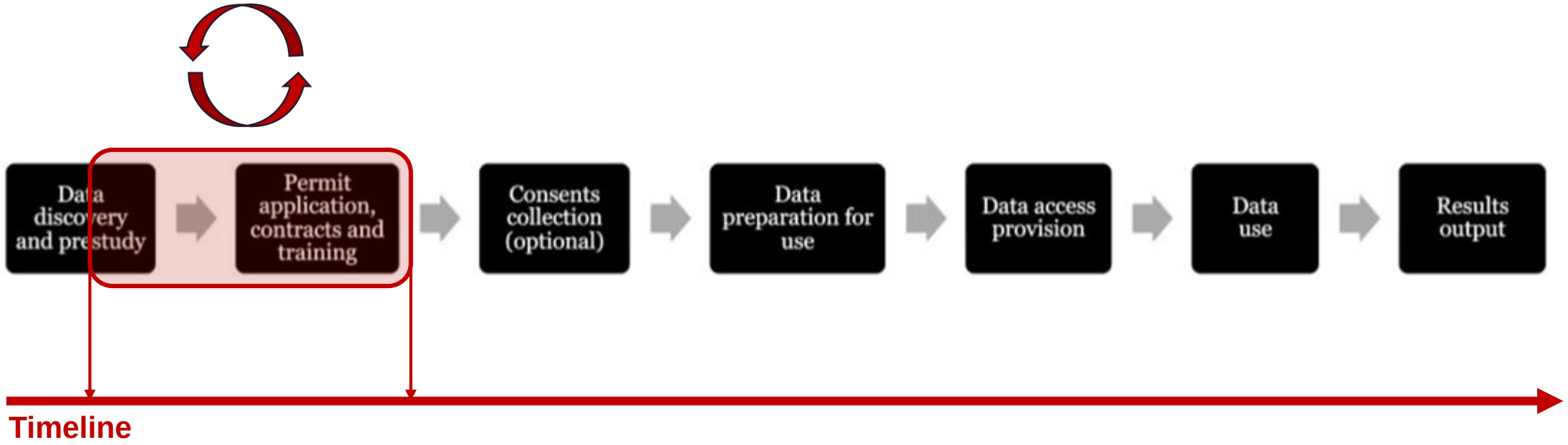


The first step in a users journey

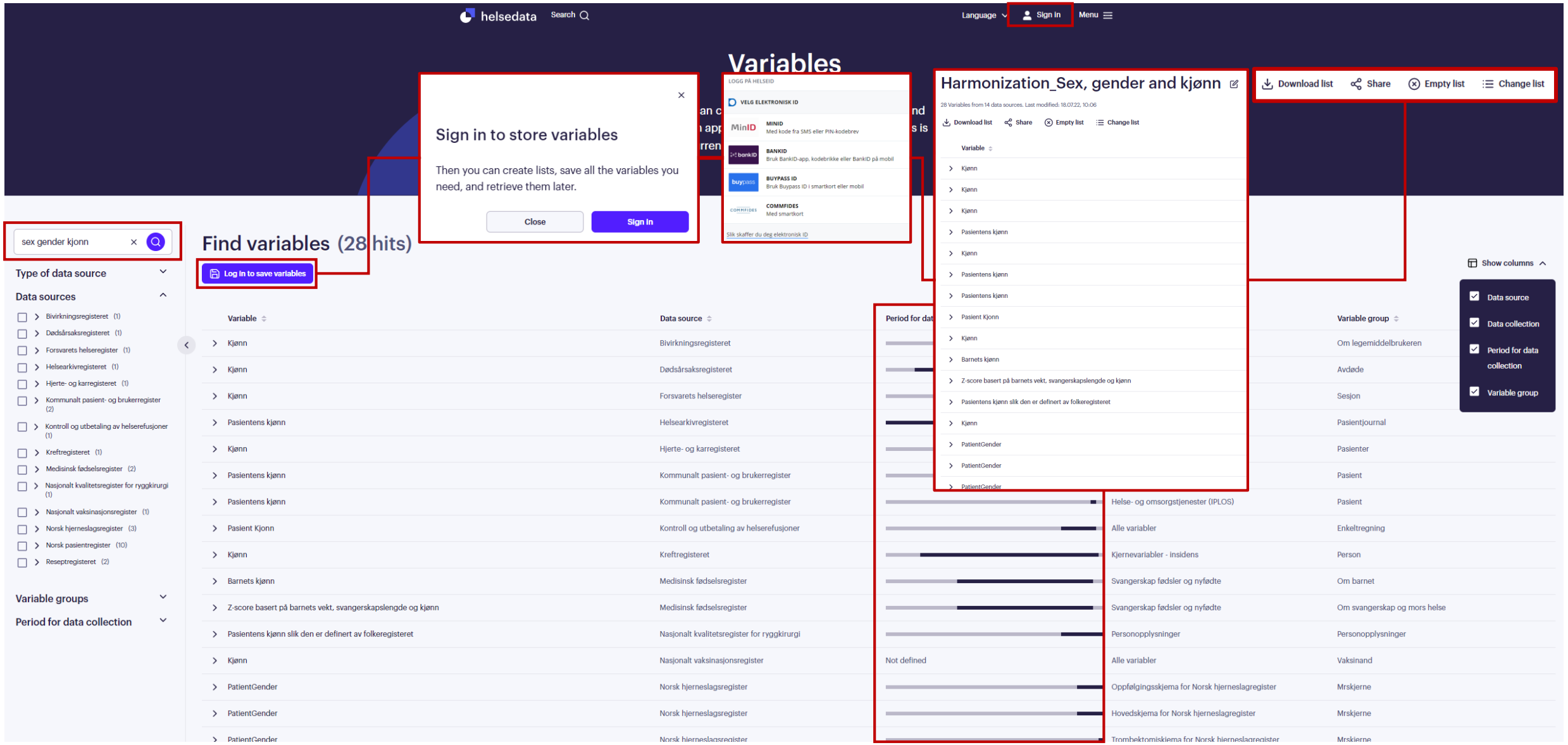




Reducing time to find and choose the right variabels



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



Direktoratet for e-helse

What is metadata?

Why is metadata important for researchers?

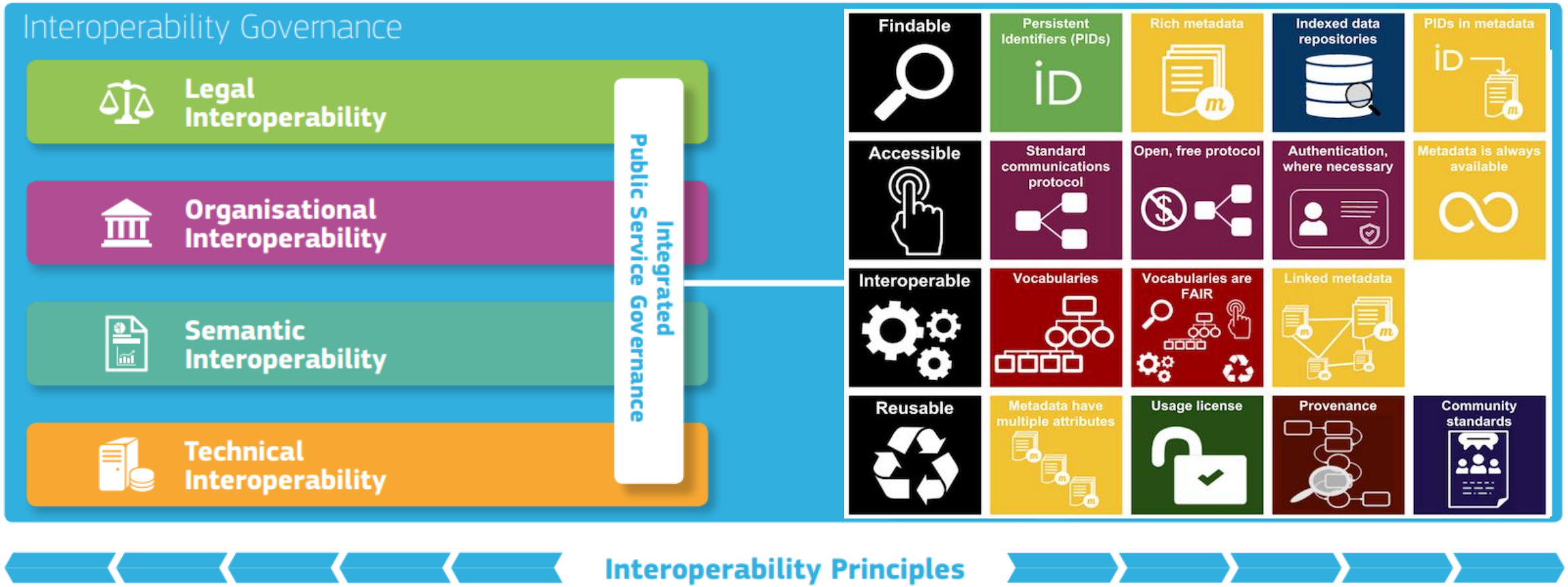
Why is common standards important?

Direktoratet for e-helse

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

FAIR operationalizes the European Interoperability Framework

A common language between, arhitects, technicans and researchers



https://ec.europa.eu/isa2/sites/isa/files/eif_leaflet_final.pdf

DCAT operationalize the FAIR criterias

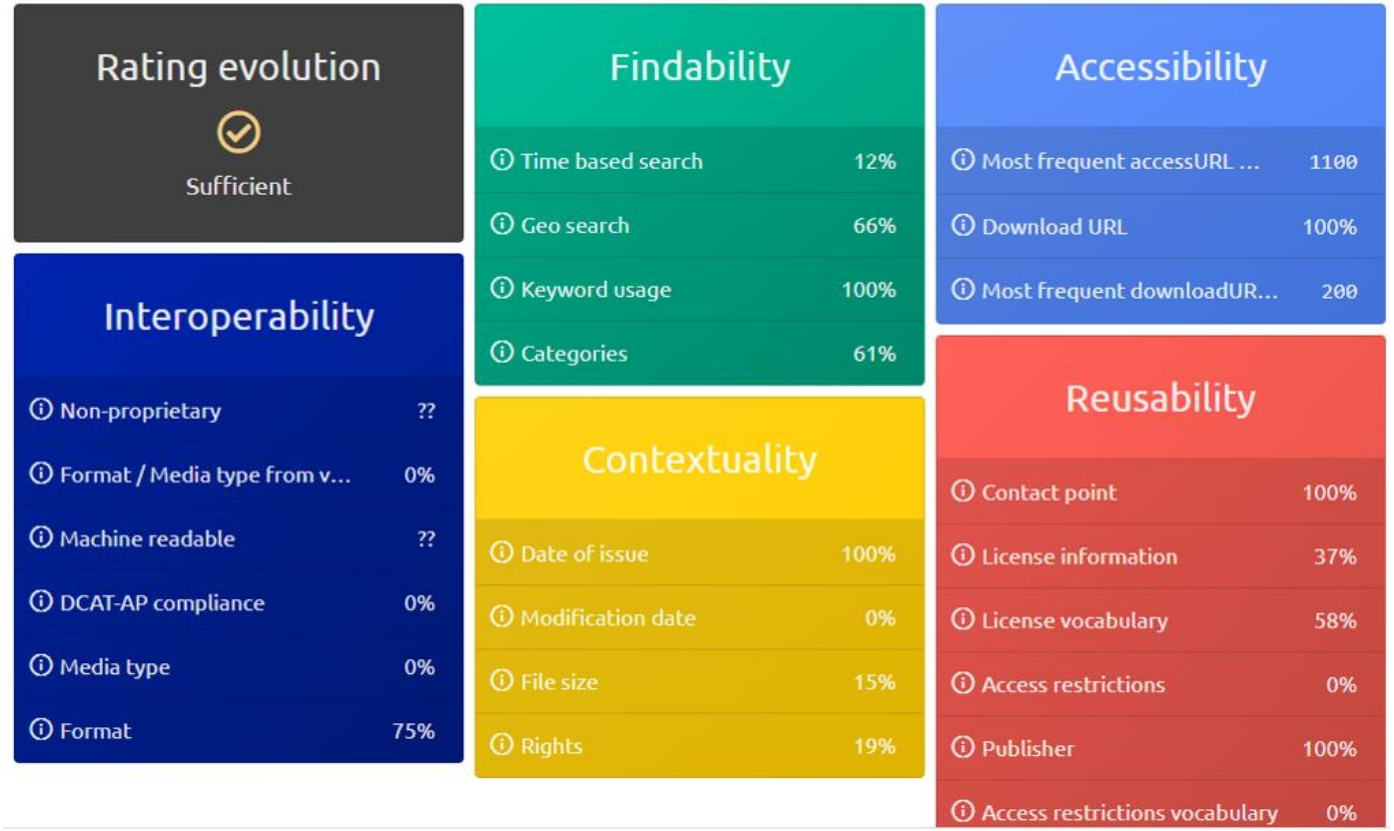
…and makes it possible for machines to share and validate the quality of metadata…. among others

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.

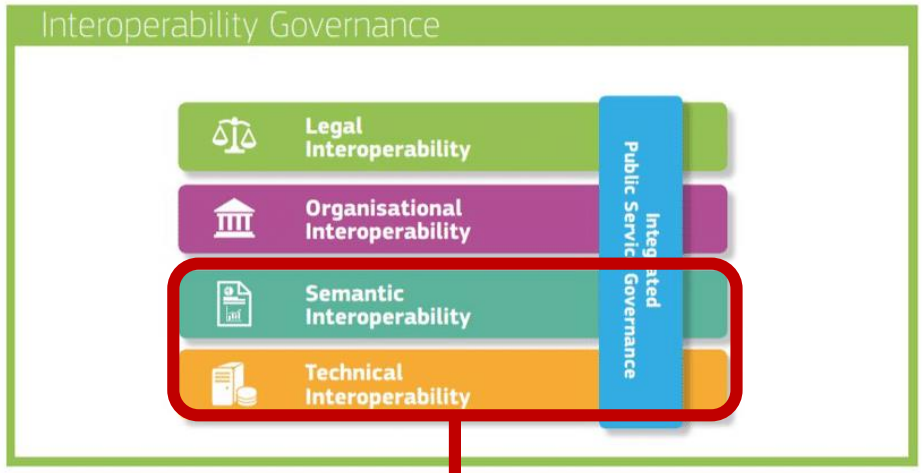
TO BE INTEROPERABLE:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

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.

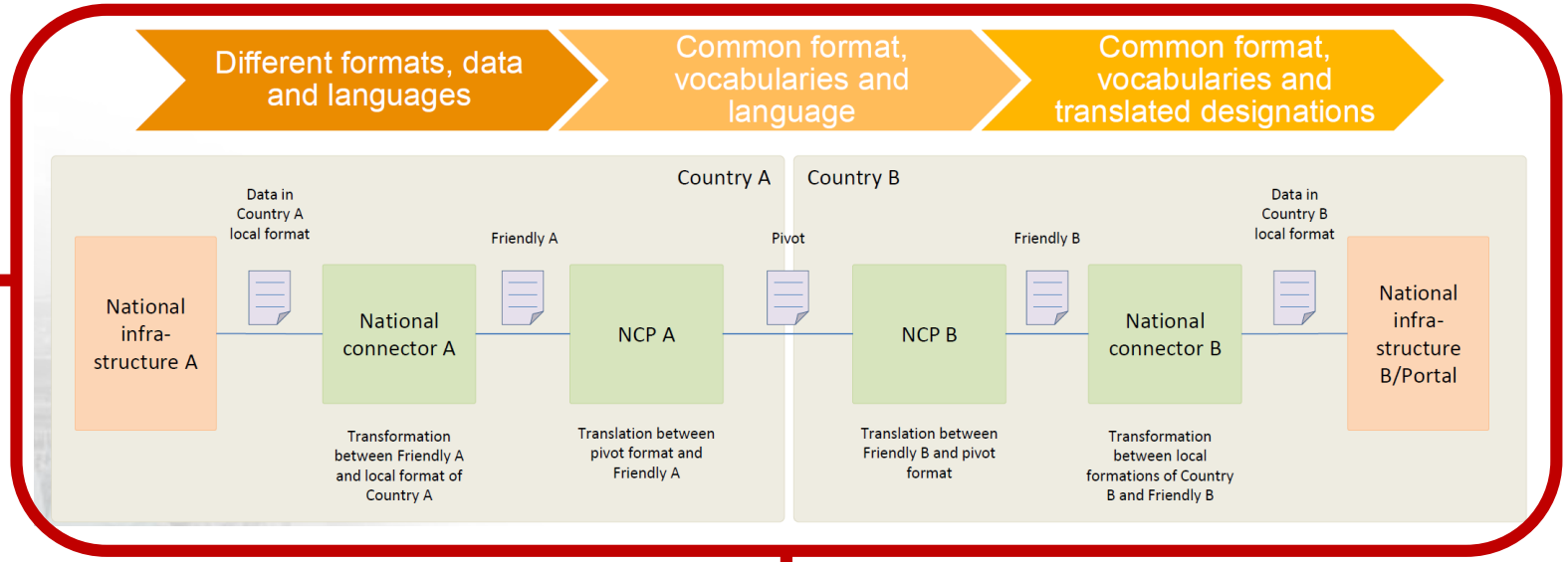


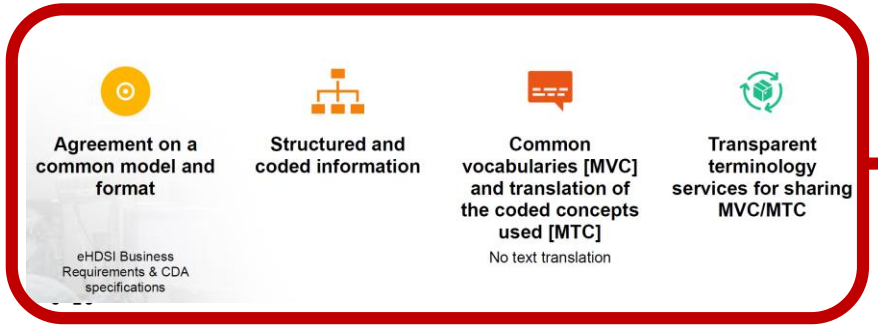
Common framework for interoperability and standards

European Interoperability Framework (EIF)



NB! Data Quality





Standardization is prerequisites for an efficient federated ecosystem of metadata catalogues

EHDS
Central Services Portal

FAIR DATA POINT WEB CLIENT

FAIR DATA POINT

A P I

European Data Catalogue

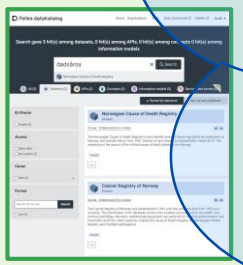
A P I

Domain based metadata catalogues

A P I

Machine-actionable
FAIR Metadata

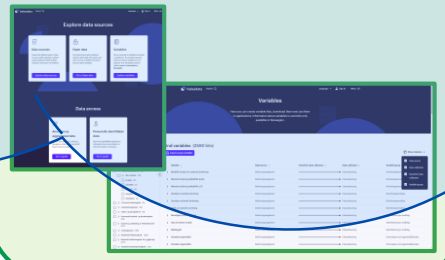
National Data Catalogue



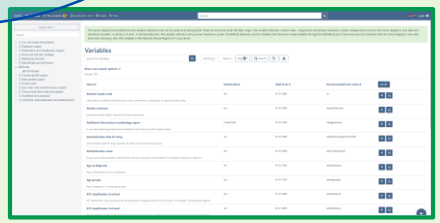
A P I

A P I

National Health Metadata Catalogue



Data holders metadata catalogues



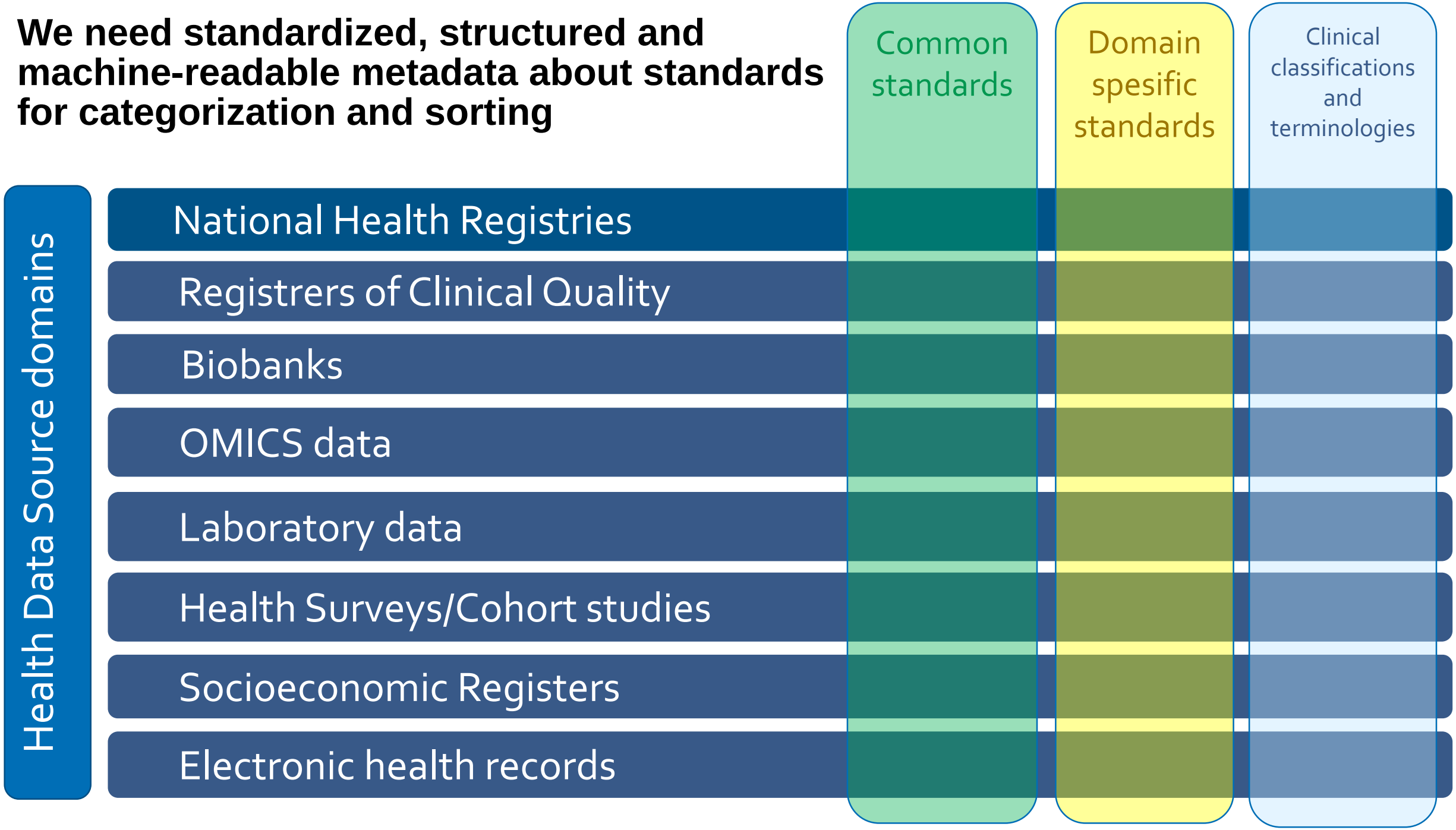
A P I

Direktoratet for e-helse

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-priciples ISO 11179 (Metadata) GSIM (Metadata) HL7 FHIR (Interoperability)
”High level” metadata standards	Attributes/properties to describe the dataset and more	Title, Descrition, 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 diseeases) 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



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 promotoing using Beacon.	Variant-Disease associations collected from curated resources and the litterature. 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 substanses)		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 applicatoins handling medical imaging and was established in 1992 . All hospitals in Norway use DICOM for medical image communicaton.		

Overview clinical classifications and terminologies per 02.09.2022

Domain	Sub domain	International	Finland	Norway	Denmark	Sweden	Iceland	NC preferred	TEHDAS preferred	1+MG	OMOP	National experts: xpert (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/UX , 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 regsitries				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									GATS						
	Lifteime smoking status									Lifteime smoking status						
	Pack-yers									Pack-yers						
Physical aktivity	IPAQ									IPAQ						
Obesity	BMI									BMI						
	Waist circumference									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 include[4]:

- 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]

Preparing for federated analyses

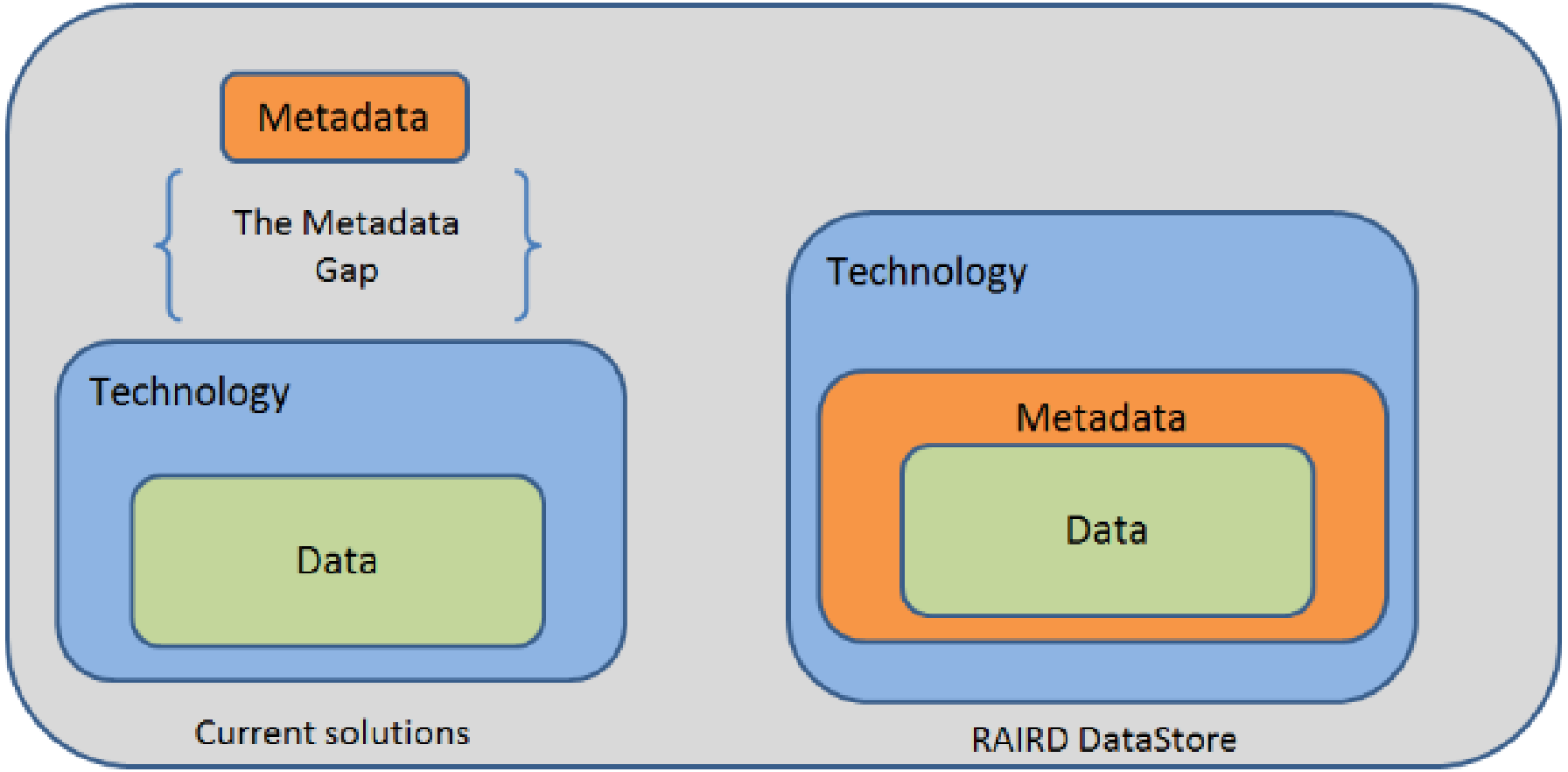


Figure 5 - The Metadata Gap and the RAIRD DataStore

Mapping key variables from national key registers to a common information model

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

Table 1 Key Variables in the Nordic Total Population 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

Table 2 Selected Key Variables in the Nordic Birth Registries

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]

Table 3 Selected Key Variables in the Nordic Patient Registries

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

Table 7 Selected Key Variables in the Nordic Causes of Death Registries

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[a] 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

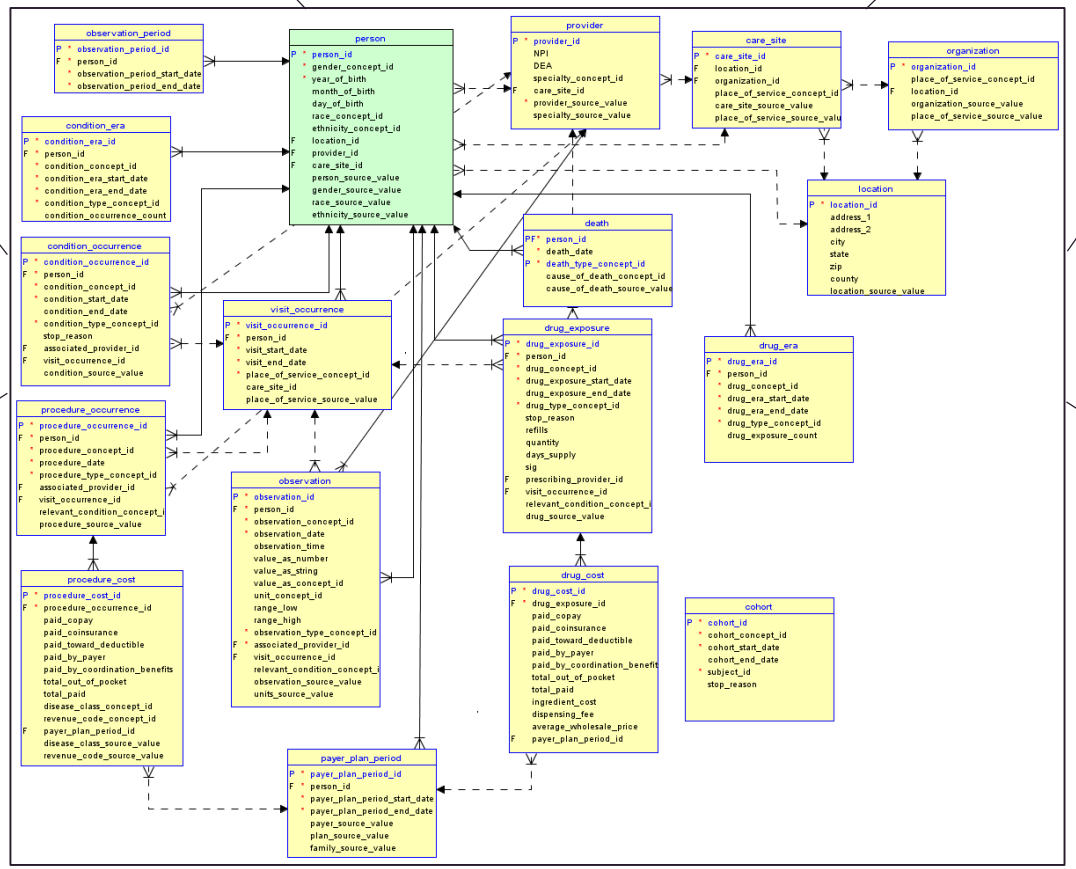


Table 6 Selected Key Variables in the Nordic Prescription Registries

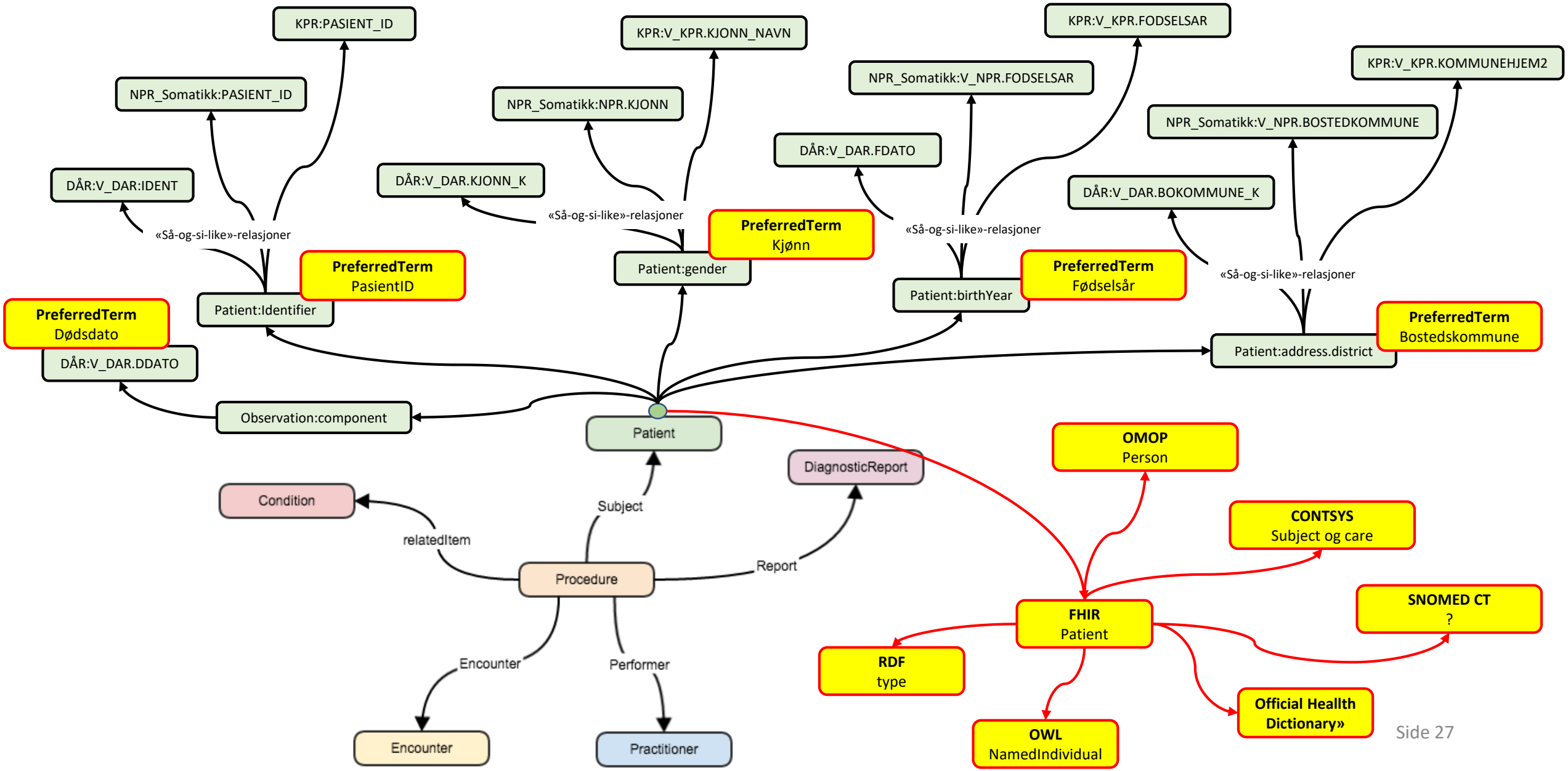
Information Type	Variables
The patient	Personal identity number (or pseudonymized number) Sex Date of birth
The prescriber	Prescriber type (eg, GP, hospital physician, or private physician)
The pharmacy	Identifier Location of the pharmacy[a]
The drug	Date of dispensing Nordic article number (unique identifier) 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

Table 5 Selected Key Variables in the Nordic Cancer Registries

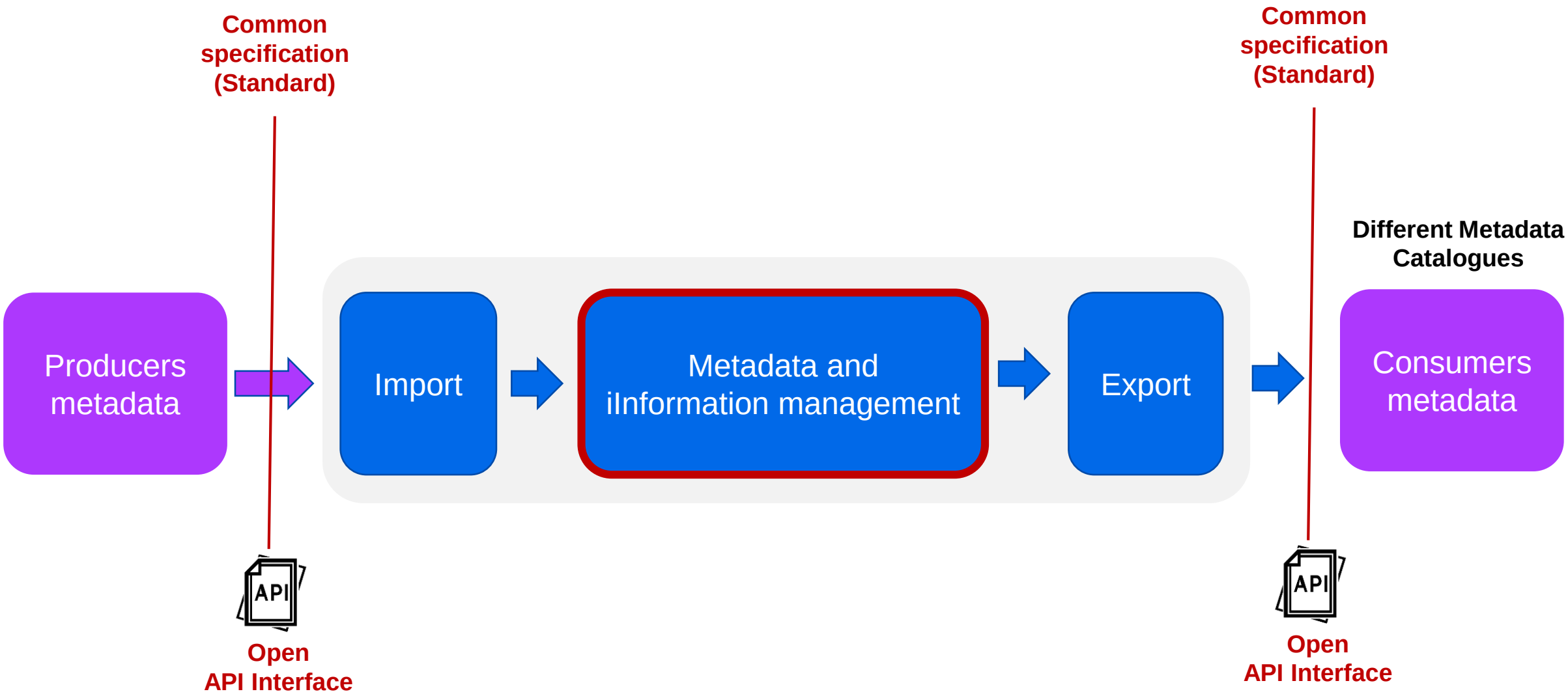
Information Type	Variables
The patient	Personal identity number Date of birth Sex Place of residence (unit) Vital status Date of death
Tumor characteristics	Date of diagnosis Topography (primary site) Morphology/histology Tumor stage or grade[a] Method of confirmation Behavior (malignant, premalignant,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

Direktoratet for e-helse

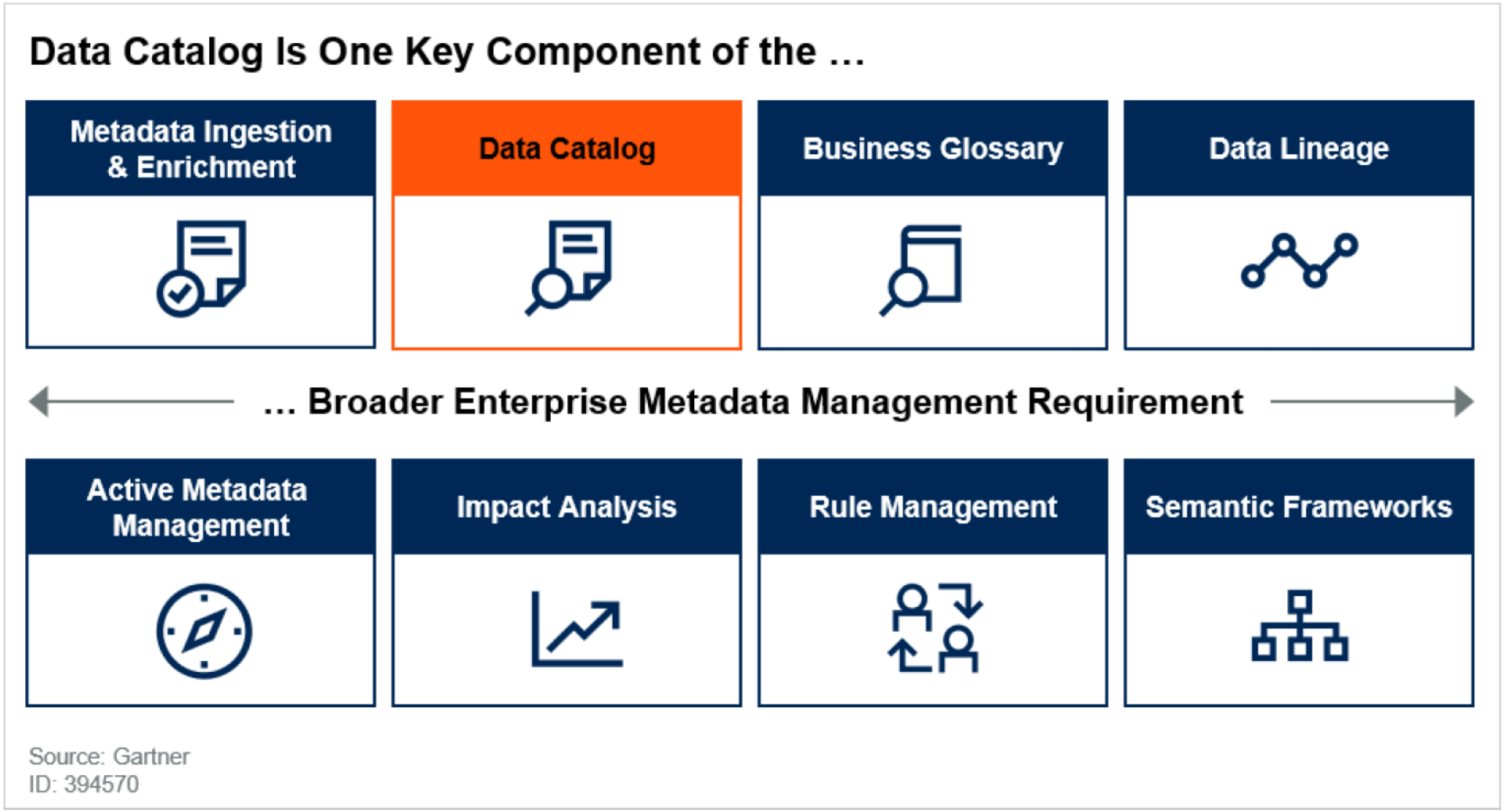
Understanding the importance of information management



The Metadata Management Process



Broader Metadata Management is required!



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

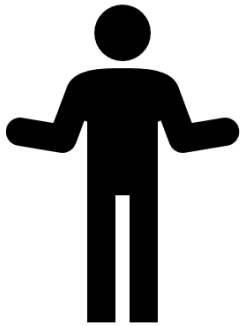
Hanna Tolonen, Finnish Institute for Health and Welfare (THL), Finland

Petronille Bogaert, Sciensano, Belgium

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

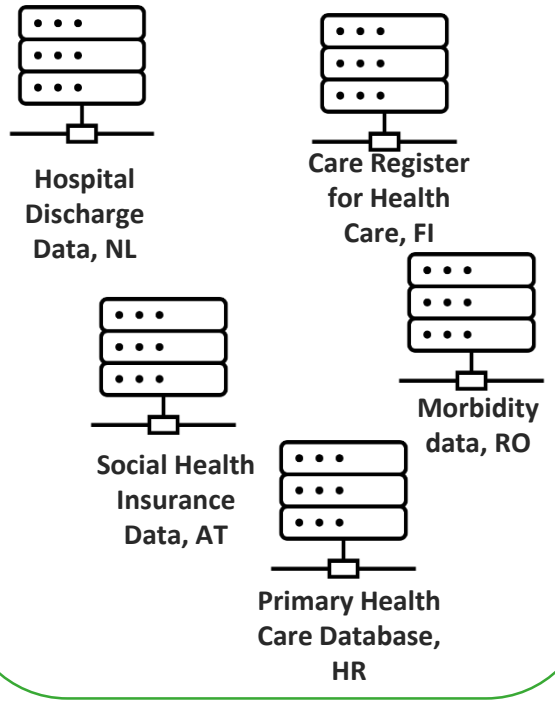
Why metadata?

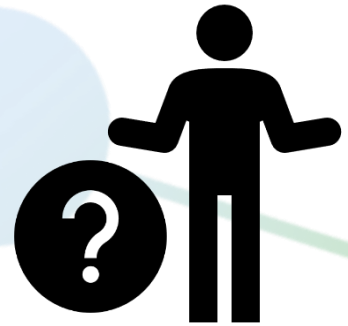
For my research, I would need data about prevalence of NCDs in EU MSs. Where to find it?



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

Data without metadata



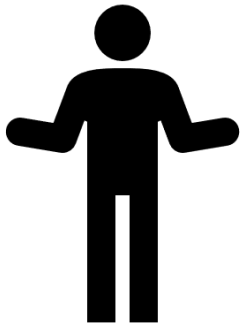


- What type of data are these?
- What age groups they covers?
- What are their geographical coverage?
- What type of variables they have?
- How do I get access to these data sources?

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

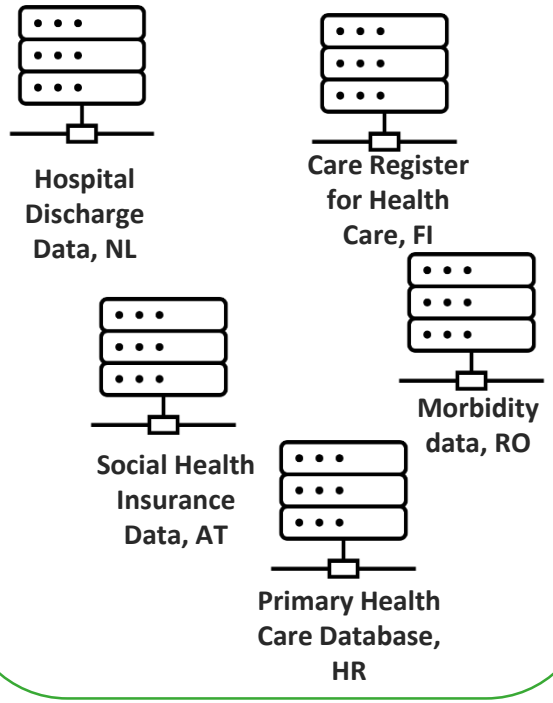
Why metadata?

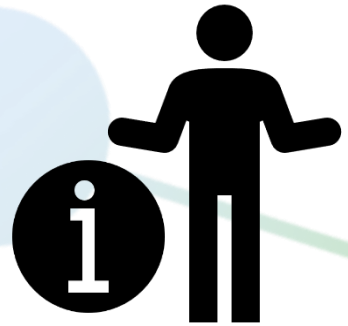
For my research, I would need data about prevalence of NCDs in EU MSs. Where to find it?



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

Data with metadata

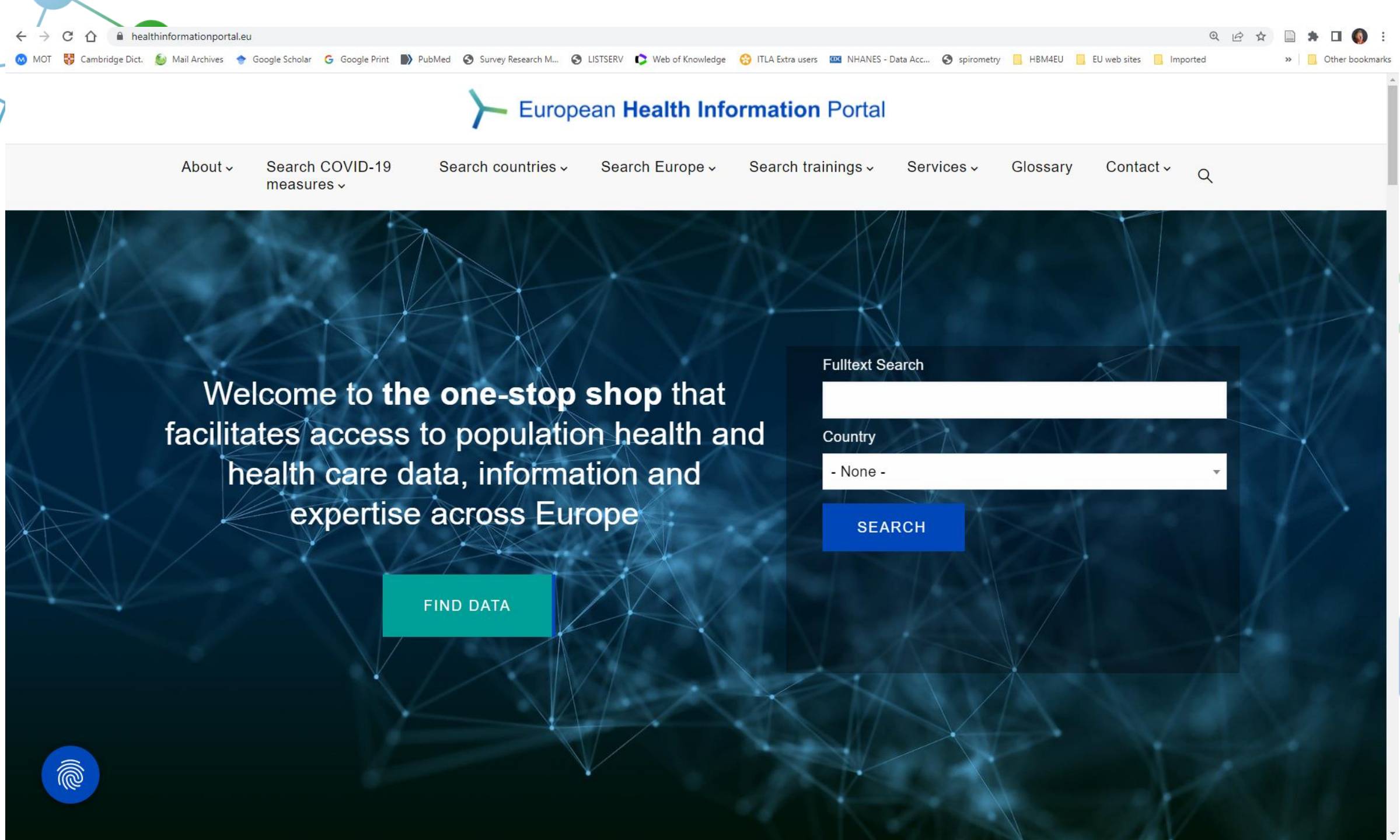




In structured format

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

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


healthinformationportal.eu
MOT
Cambridge Dict.
Mail Archives
Google Scholar
Google Print
PubMed
Survey Research M...
LISTSERV
Web of Knowledge
ITLA Extra users
NHANES - Data Acc...
spirometry
HBM4EU
EU web sites
Imported
Other bookmarks
European Health Information Portal
About
Search COVID-19 measures
Search countries
Search Europe
Search trainings
Services
Glossary
Contact
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

Type of information Registry data

URL of the data source

Link to Care Register for Health Care

Description

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

Governance and legal framework

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)

Free keywords

ICD-10; disease; Hospital admission

Topics

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

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

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

CARE REGISTER FOR HEALTH CARE

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)

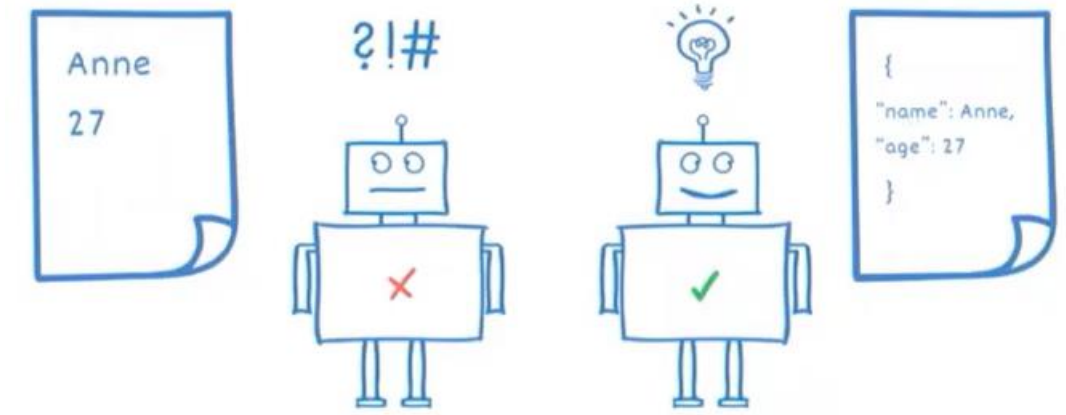
Linkage possible	Yes

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

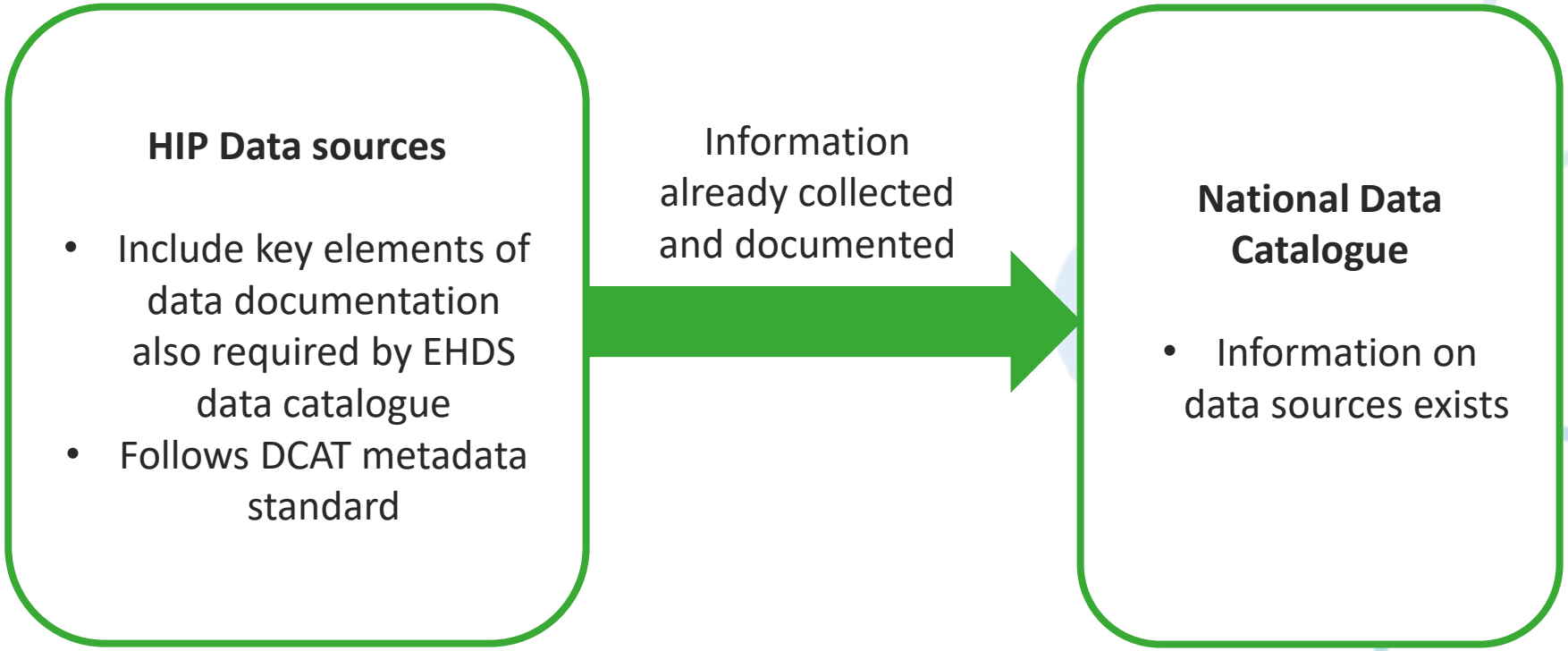
DCAT is a vocabulary for publishing data catalogs on the web

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

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

National data catalogue and HIP data sources

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



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

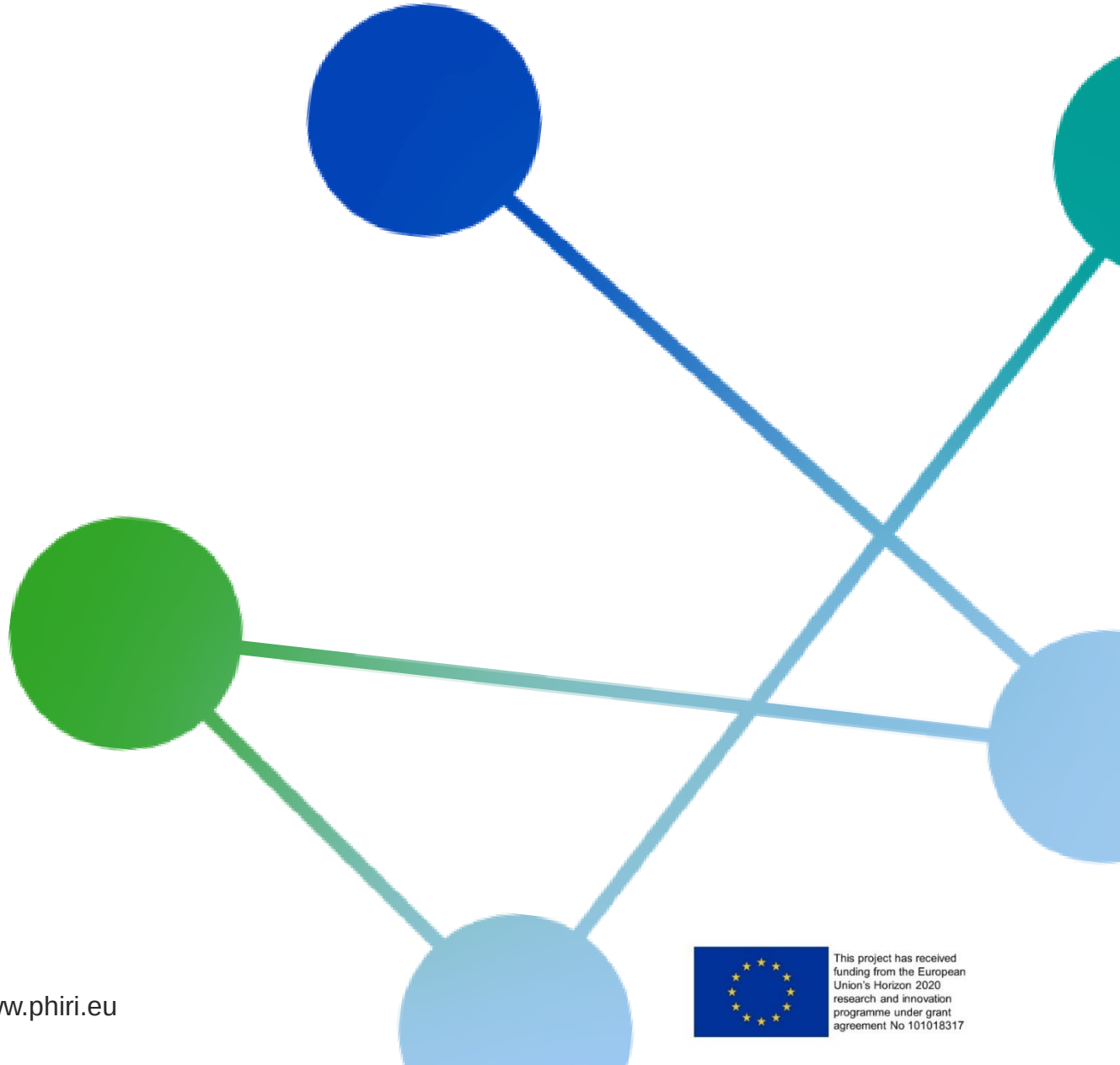
PHIRI

Population Health Information Research Infrastructure

Thank you for your attention!

@PHIRI4EU

/company/phiri



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

sciensano

healthy all life long

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

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

A major consortium…

National platforms
(France, Finland, Germany, Belgium, Denmark, Hungary, Croatia, Norway)

European agencies
(EMA, ECDC)

International research infrastructures and associations
(BBMRI, Elixir, eBrains, EUPHA)

… 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

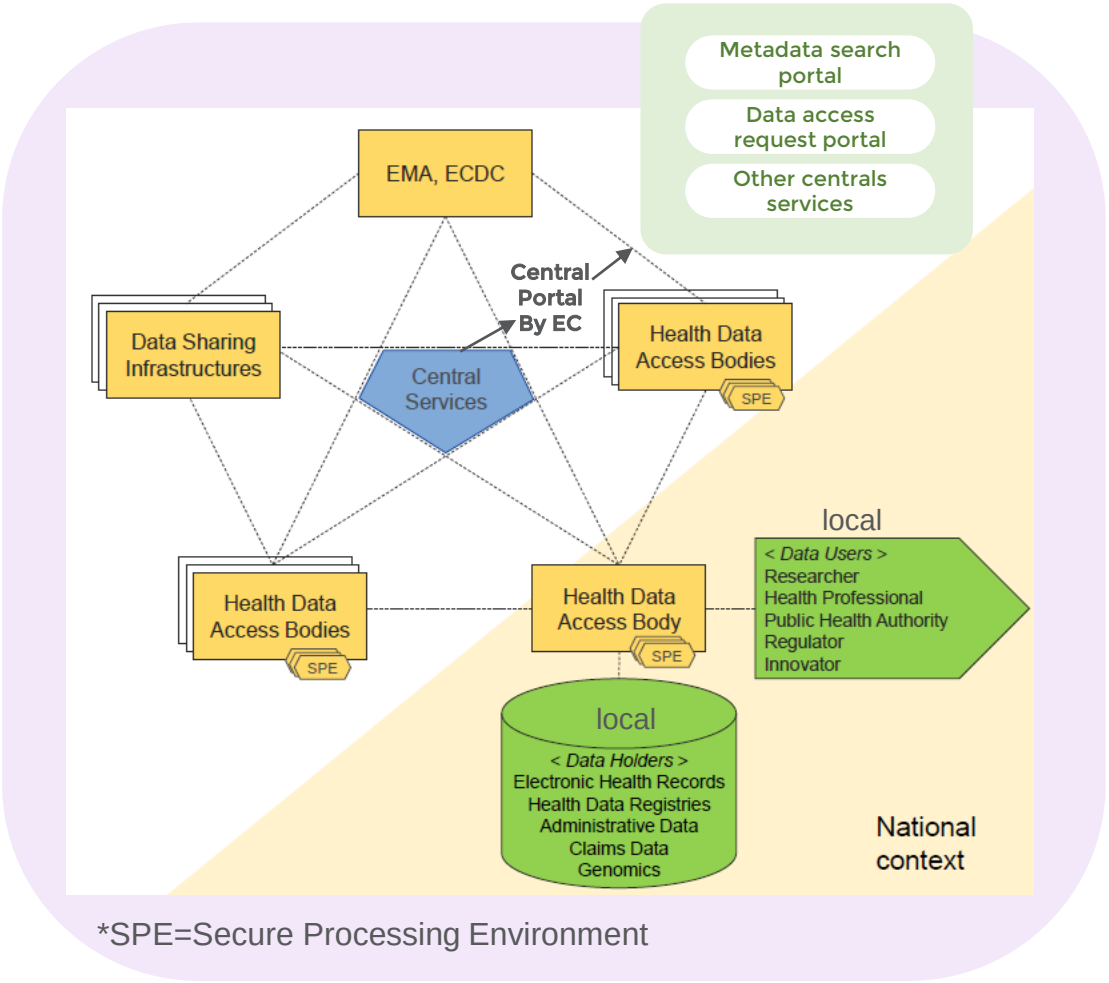
A consortium of 16 major partners

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

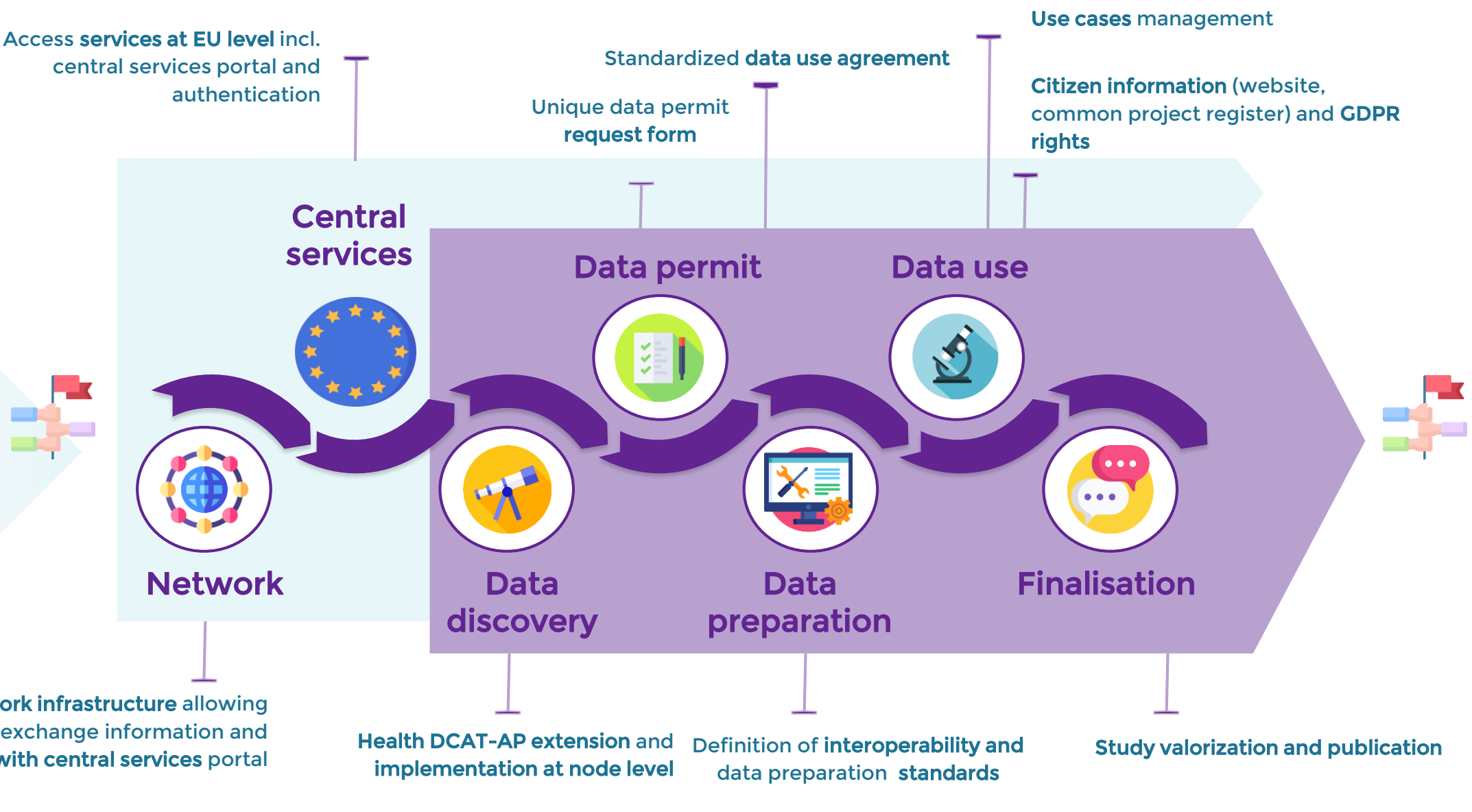
Within this networks of nodes researchers will be able to:

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

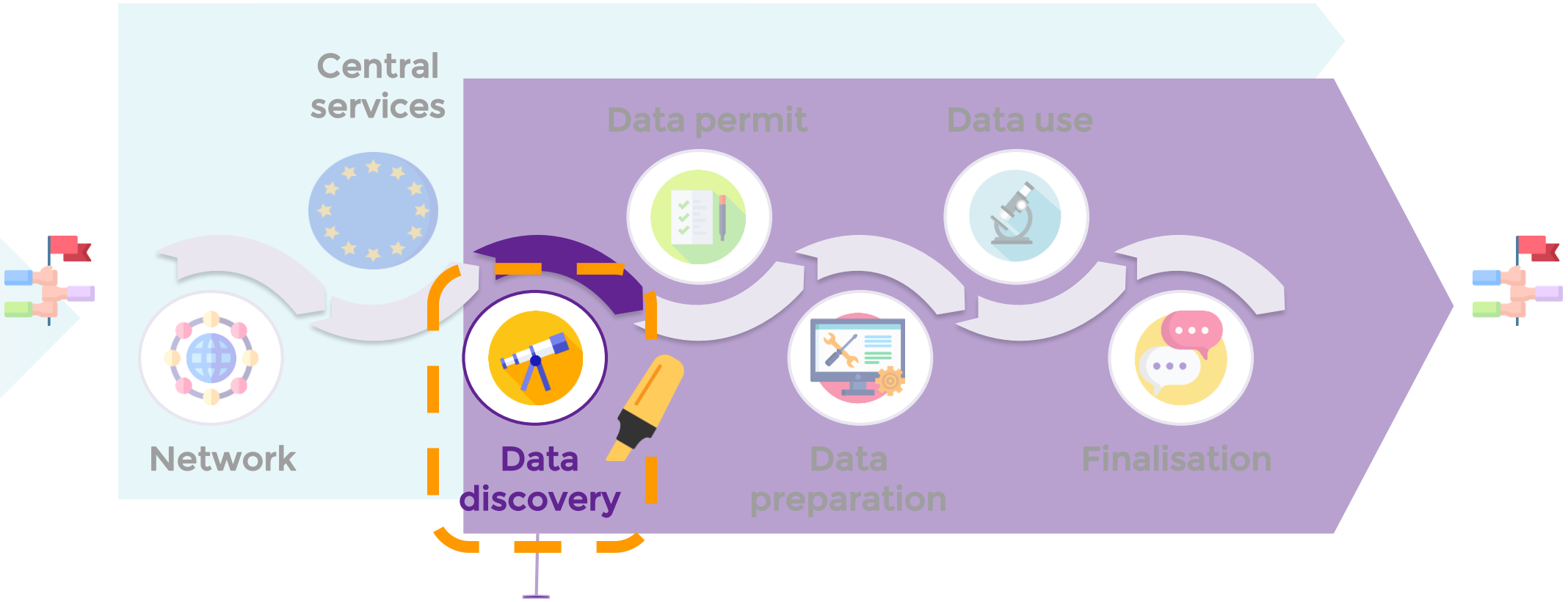


*SPE=Secure Processing Environment


Access **services at EU level** incl. central services portal and authentication
Standardized **data use agreement**
Unique data permit **request form**
Use cases management
Citizen information (website, common project register) and **GDPR rights**
Central services
Data permit
Data use
Network
Data discovery
Data preparation
Finalisation
Network infrastructure allowing nodes to exchange information and **link with central services** portal
Health DCAT-AP extension and **implementation at node level**
Definition of **interoperability and** data preparation **standards**
Study valorization and publication

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	**Health Data Lab** *(co-leader)* **Health Data Hub** *(co-leader)*
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	**Sciensano** *(co-leader)* **Norwegian Directorate of eHealth** *(co-leader)*
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**	**BBMRI** *(leader)* **FinData** *(Task leader : data application)*
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.	**BBMRI** *(leader)*
WP9	**USE CASES MANAGEMENT** Ensure project management for implementing standards on chosen data use cases, provide se cases management (completion, opening of HDW…), and produces recommendations on use cases management procedures and governance.	**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

sciensano

healthy all life long

Contact

EU health information system unit
SD Epidemiology and Public Health

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

.be

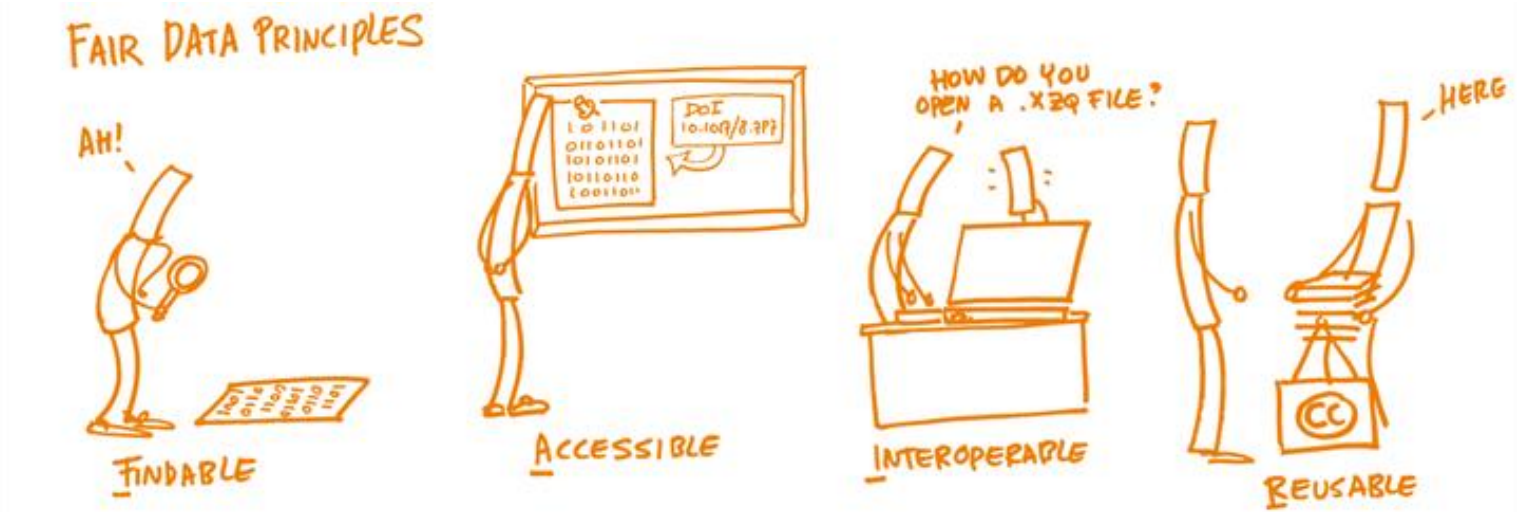
HEALTHYCLOUD
Health Research & Innovation Cloud

Assessing the FAIRness level of health-related data collections

HEALTHYCLOUD
Health Research & Innovation Cloud

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:



sciensano

HEALTHYCLOUD
Health Research & Innovation Cloud

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

sciensano

HEALTHYCLOUD
Health Research & Innovation Cloud

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…**

sciensano

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
Health Research & Innovation Cloud

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%

sciensano

HEALTHYCLOUD
Health Research & Innovation Cloud

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)

sciensano

HEALTHYCLOUD
Health Research & Innovation Cloud

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)

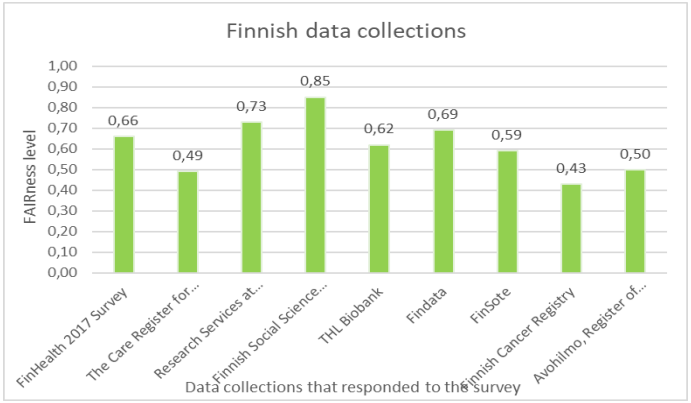
Is it possible for third party users to access the data and re-use it for more than one purpose/project?

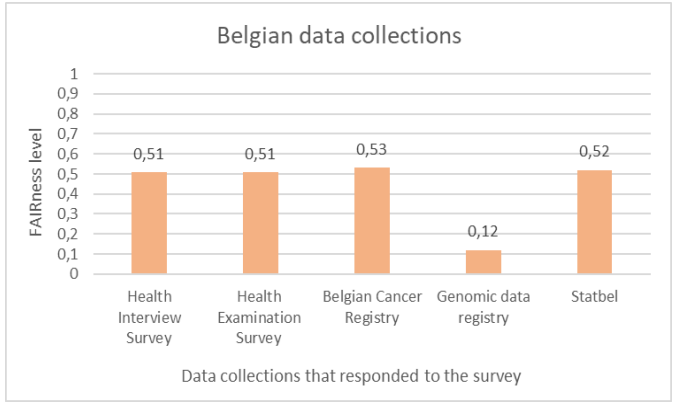
I don't know (0)

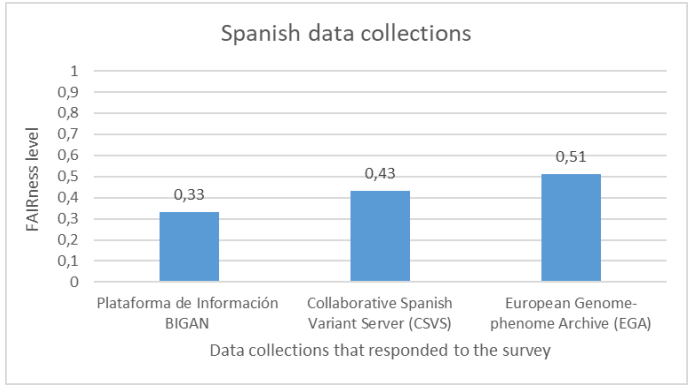
sciensano

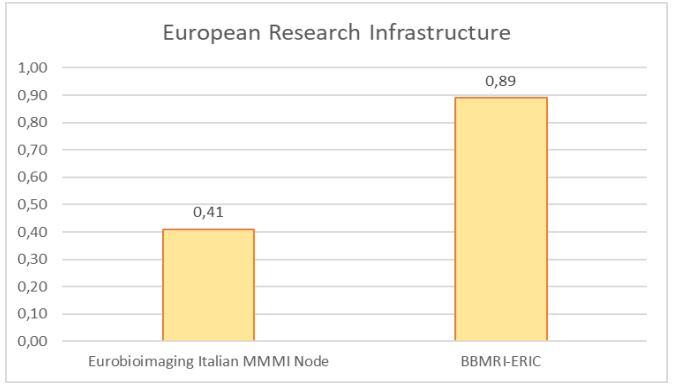
HEALTHYCLOUD
Health Research & Innovation Cloud

Results







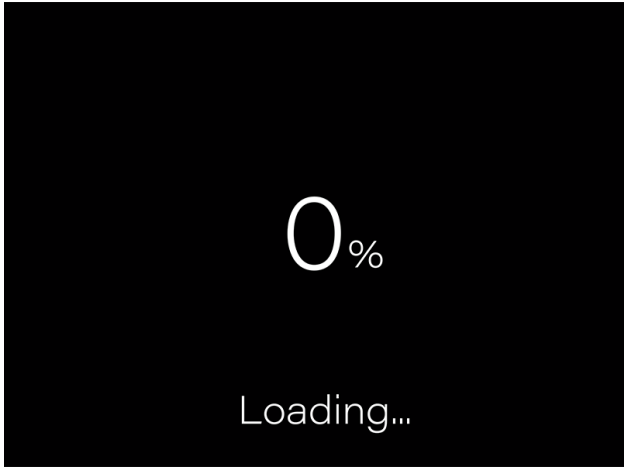


sciensano

HEALTHYCLOUD
Health Research & Innovation Cloud

FAIR is a *spectrum*

Add your data in domain specific repositories to start your **FAIR data journey!**



sciensano

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 № 965345.

What is metadata? Common standards and properties. EHP Workshop, November 9, 2022
Slides: https://www.slideshare.net/SusannaSansone

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

DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF OXFORD

Group: datareadiness.eng.ox.ac.uk

Founding

Academic Editor

scientific data
SPRINGER NATURE

nature.com/sdata

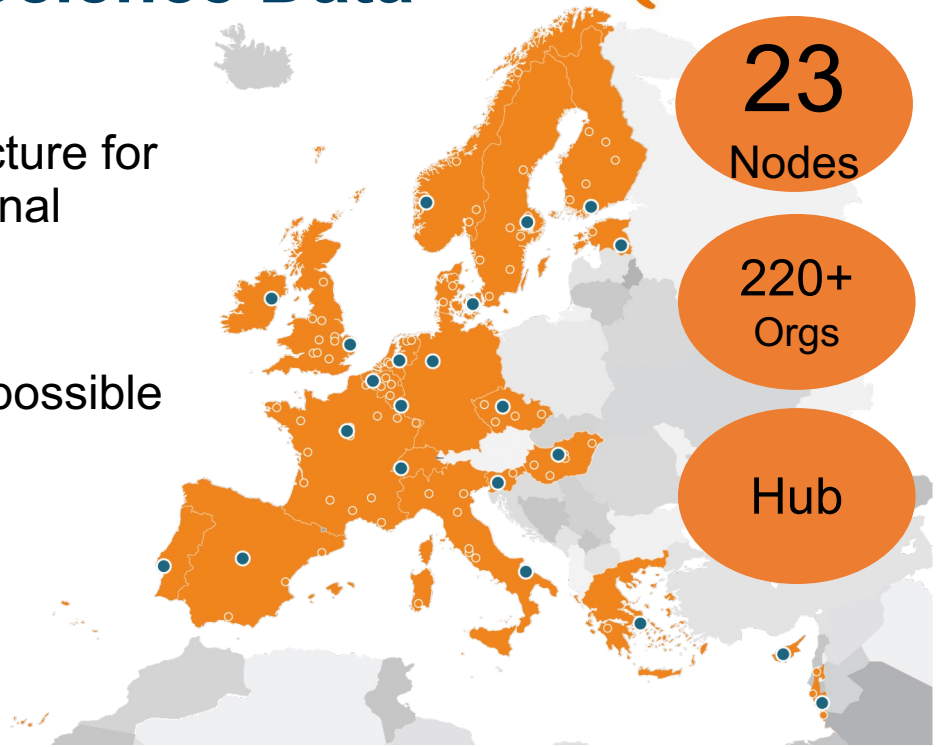
ELIXIR European Research Infrastructure for Life Science Data

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)

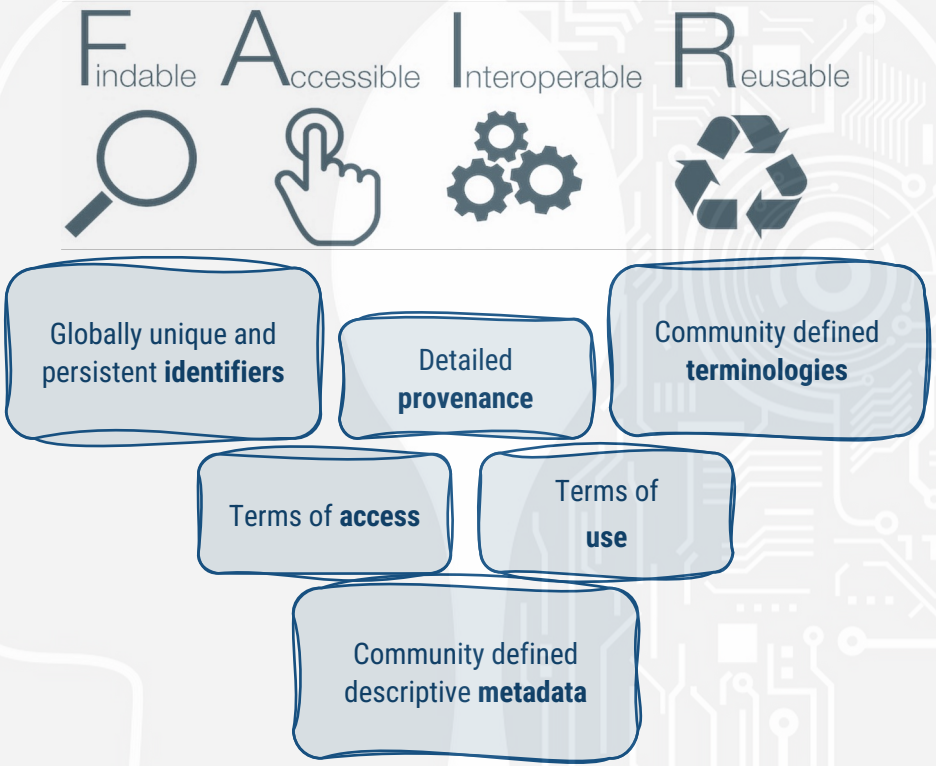
Chris Evelo
(ELIXIR Netherlands)

Tony Burdett
(EMBL-EBI)

Peter Maccallum
(*ad interim* Platform Coordinator, ELIXIR Hub)

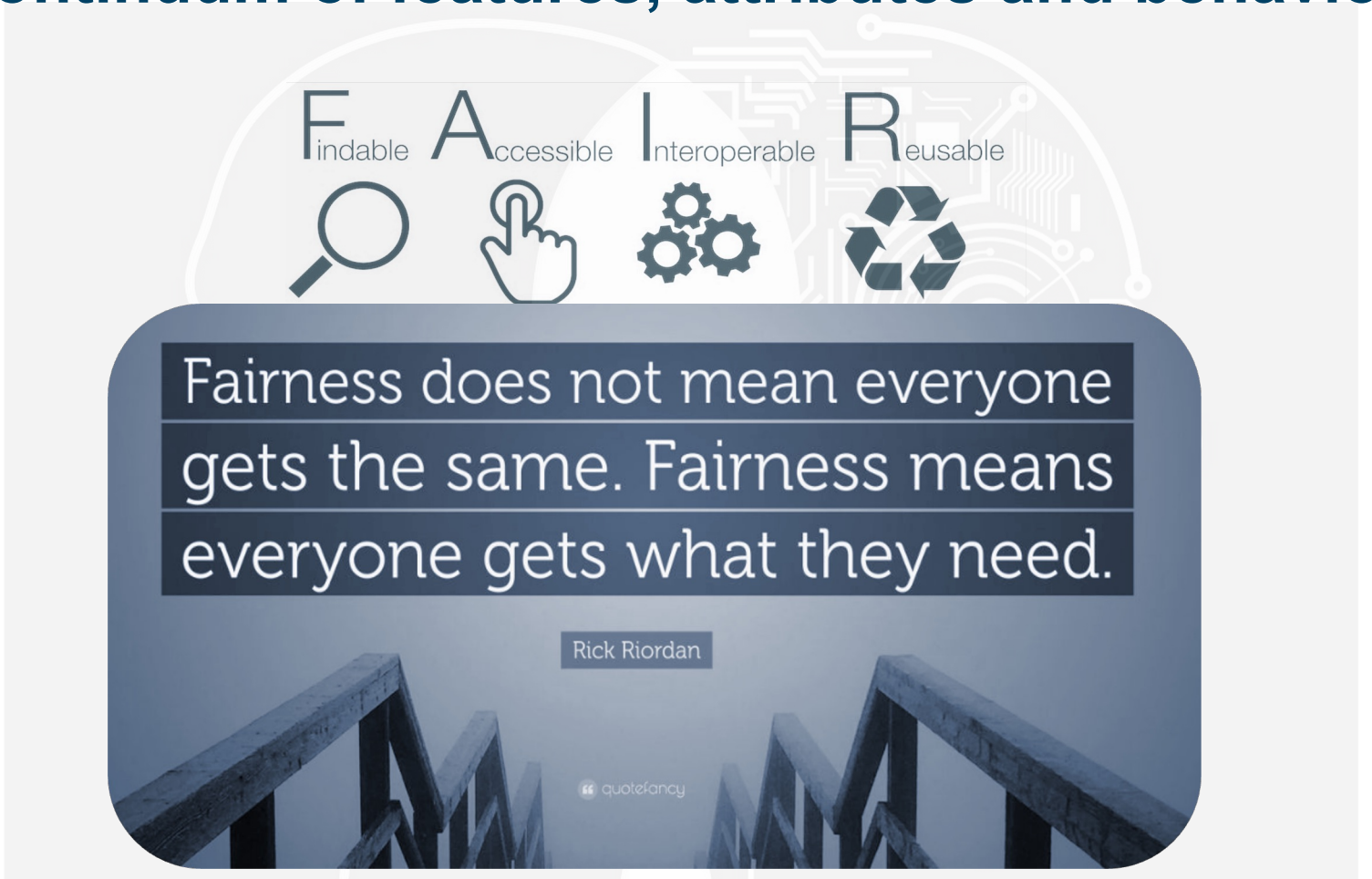
Clare Garrard
(Platform Officer, ELIXIR Hub)

Metadata make data count

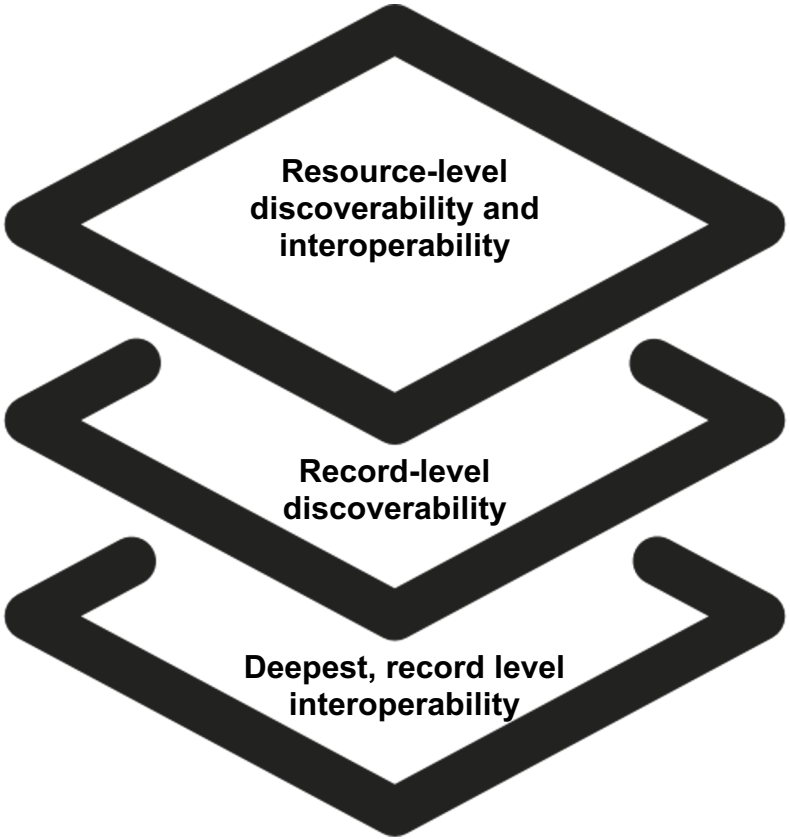


DOI: 10.1038/sdata.2016.18

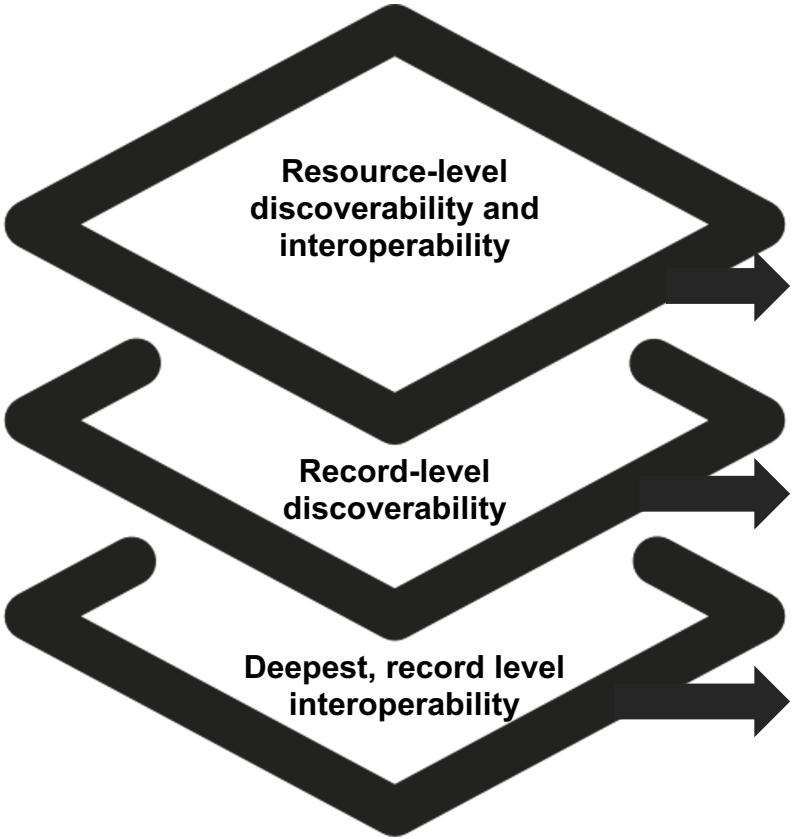
A continuum of features, attributes and behaviours



Metadata standards for different purposes



Metadata standards for different purposes

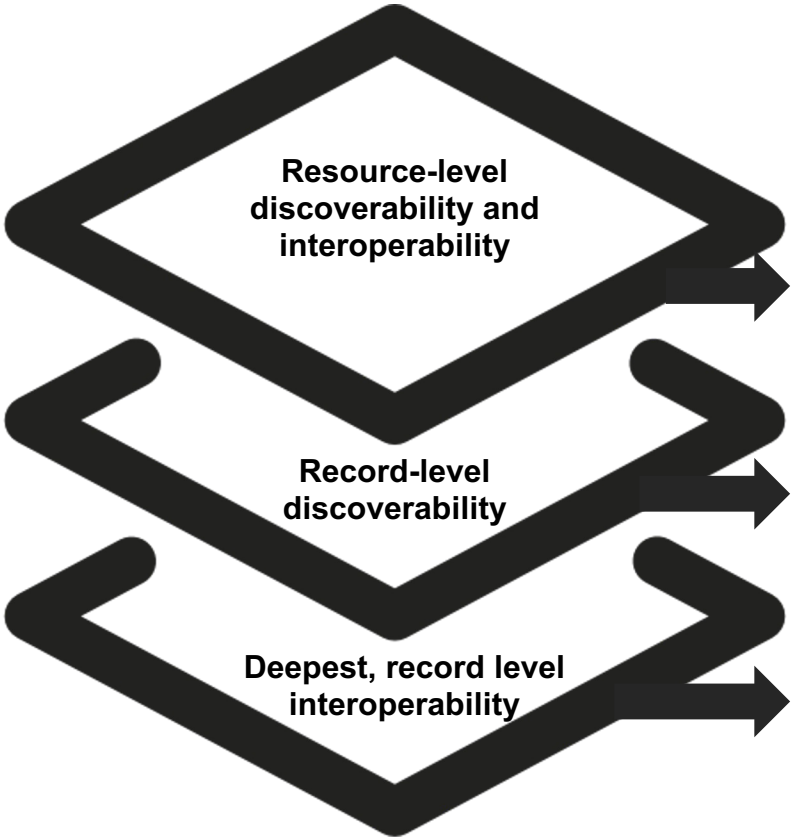


A database
or among databases

Datasets in a database

Datasets and data reuse

Metadata standards for different purposes



A database or among databases

Datasets in a database

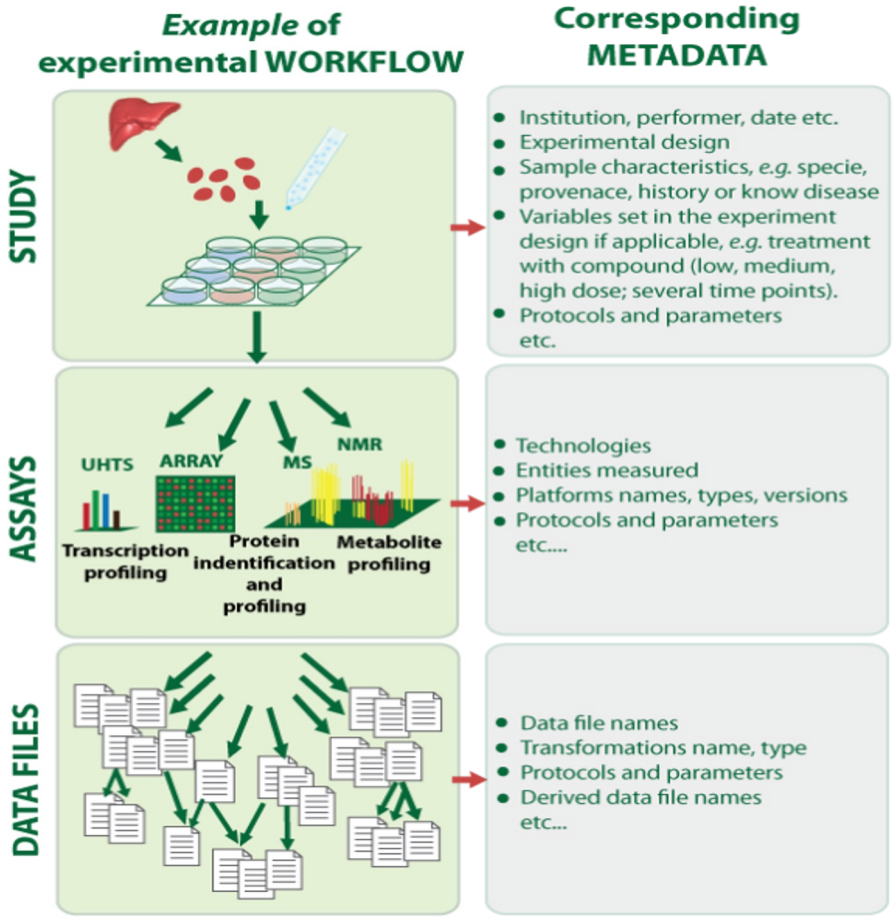
Datasets and data reuse

W3C Data Catalog Vocabulary (DCAT)

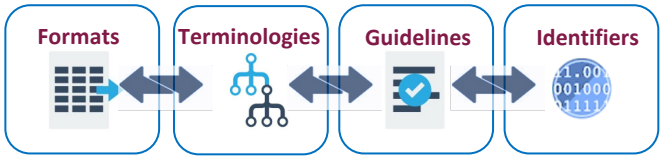
Schema.org

DublinCore

Standards to report metadata at dataset level



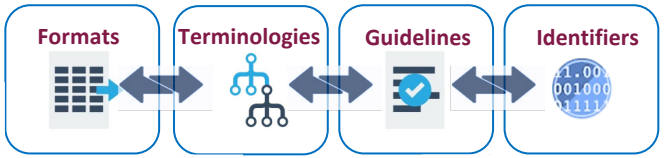
Standards to report metadata at dataset level



Source:
FAIRsharing.org

Standards to report metadata at dataset level

Conceptual model, conceptual schema, exchange formats to represent, contain and move information



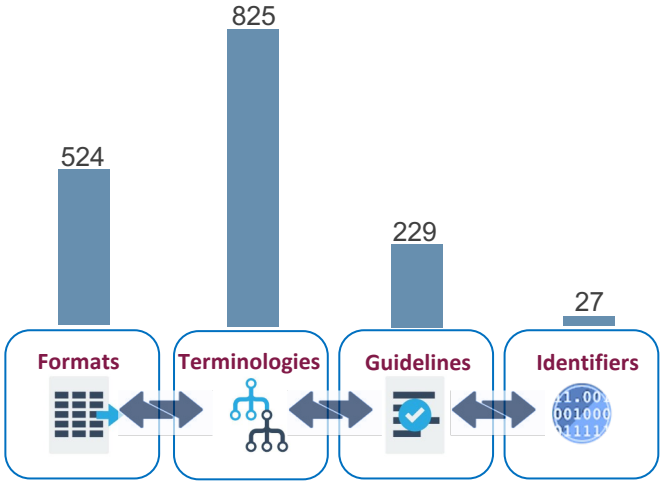
Unambiguous, persistent and context-independent schema to identify data and metadata elements

Controlled vocabularies, thesauri, ontologies to disambiguate terms and enable semantic relationships

Minimum information reporting requirements, or checklists to report the same core, essential information

Source:
FAIRsharing.org

Natural, engineering, humanities & social sciences



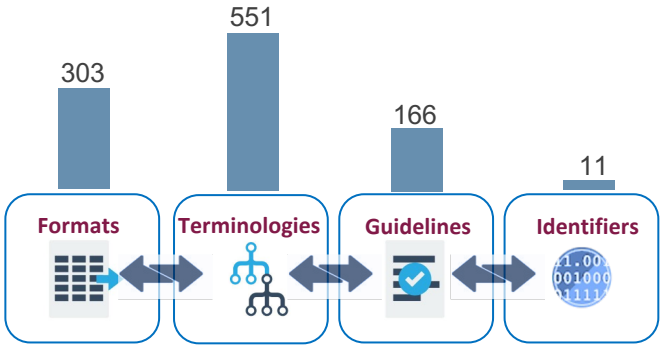
More than 1600 data and metadata standards

Source:
FAIRsharing.org

Life and biomedical sciences

Standard organizations, e.g.:

Grass-roots groups, e.g.:



More than 1000 data and metadata standards

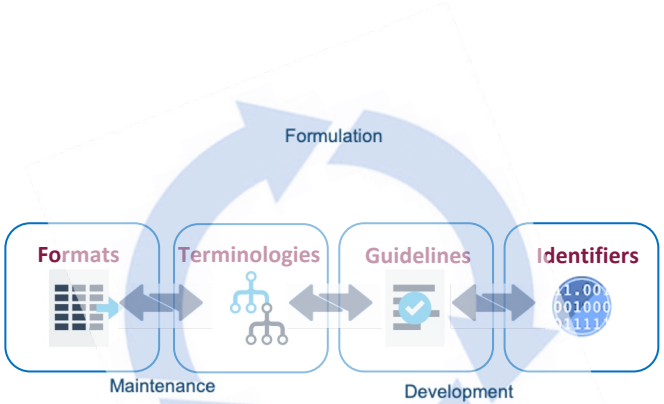
SRAxml VO AAO MIAME EC number
SDTM FASTA OBI CHEBI MIRIAM DOI
URL LSID PURL
ISA CML DICOM PATO ENVO MIX MIQAS
MOD TEDDY MIGEN REMARK
ORCID Handle IVOA ID
CDASH SBRML MIAPE CONSORT
XAO BTO RRID InChI
MITAB OMOP SEDML ARRIVE MIASE MISFISHIE
DO PRO IDO

Source:
FAIRsharing.org

Understanding their life cycle and landscape

Standard organizations, e.g.:

Grass-roots groups, e.g.:



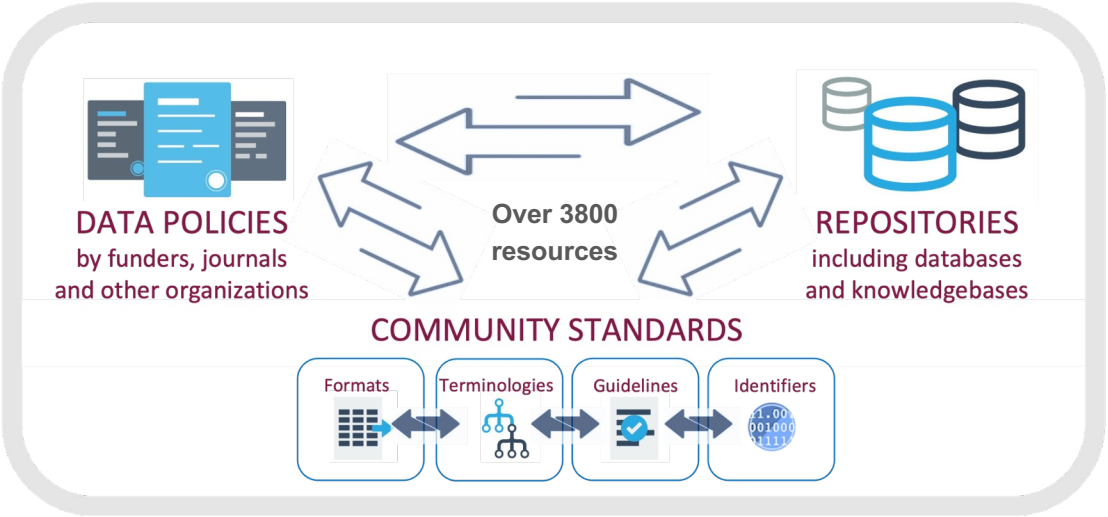
- **Industry-level standards**
- Mostly regulators-driven
- Participation is often regulated
- Standards are sold or licenced
- Formal development process, often less flexible, could be lengthy
- Charges apply to advanced training or programmatic access

- **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

Informative and educational resource

FAIRsharing.org



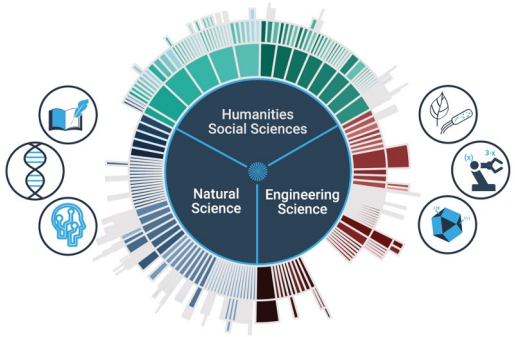
Guides ***consumers*** to *discover*, *select* and *use* these resources with confidence

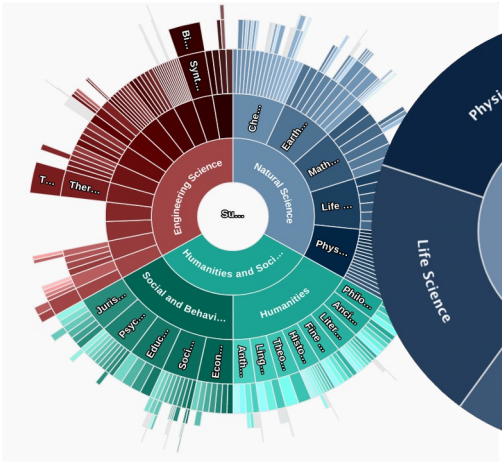
Helps ***producers*** to make their resources more *visible*, more widely *adopted* and *cited*

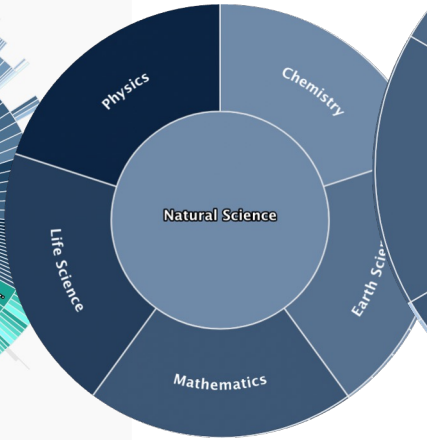
Browse by subject

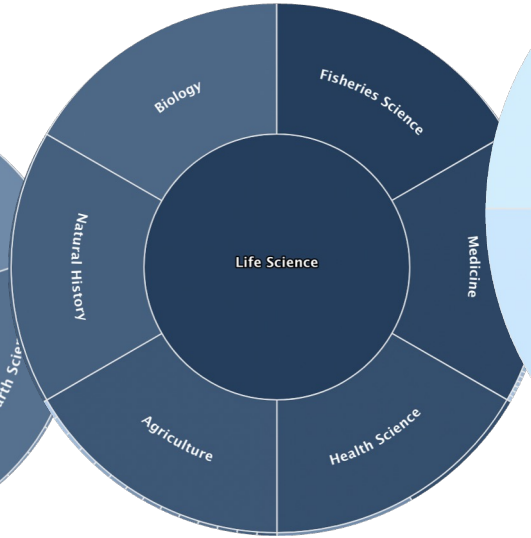
- Natural Science 8284
  - Chemistry 372
  - Earth Science 1242
  - Mathematics 49
  - Life Science 6831
  - Physics 251
- Humanities and Social Science 679
- Subject Agnostic 274
- Engineering Science 1381

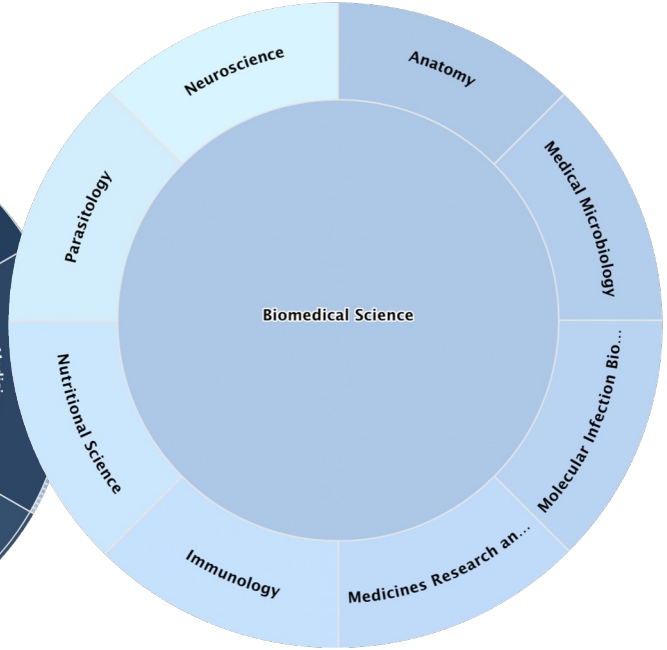
FAIRsharing.org



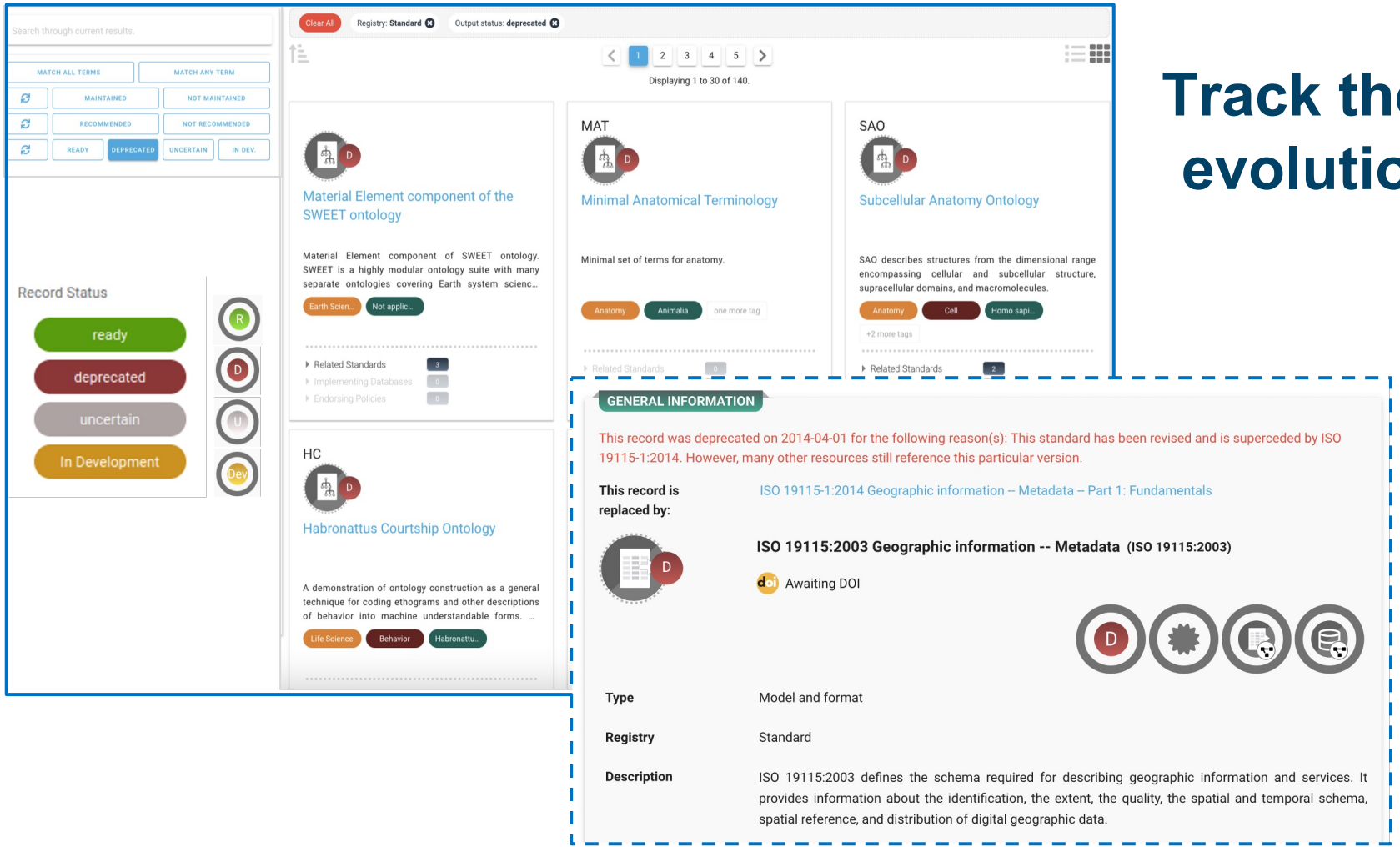




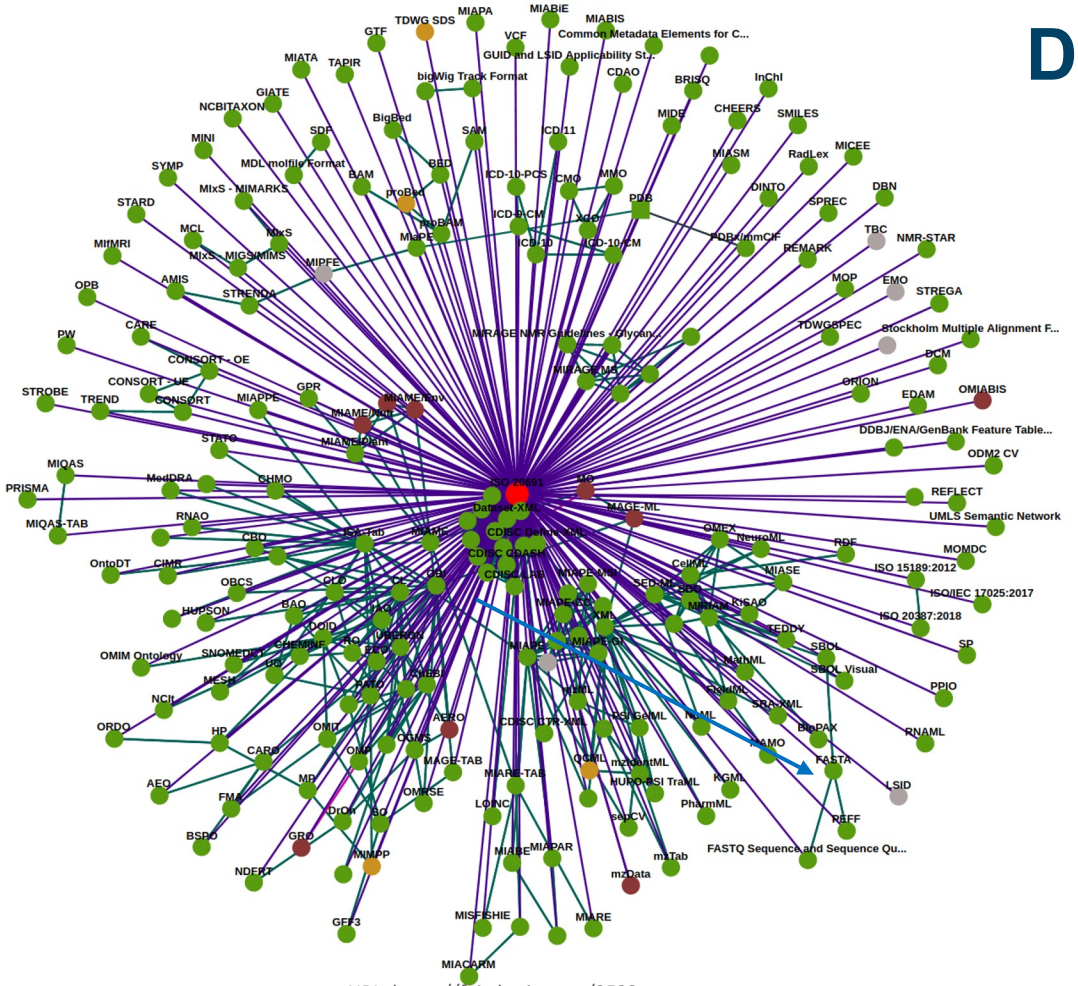




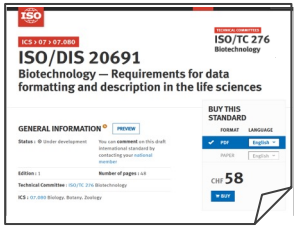
Track their evolution



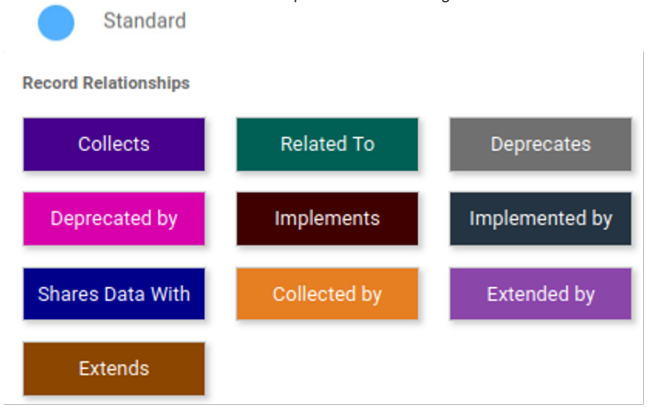
Displaying relations among metadata standards



URL: https://fairsharing.org/3533



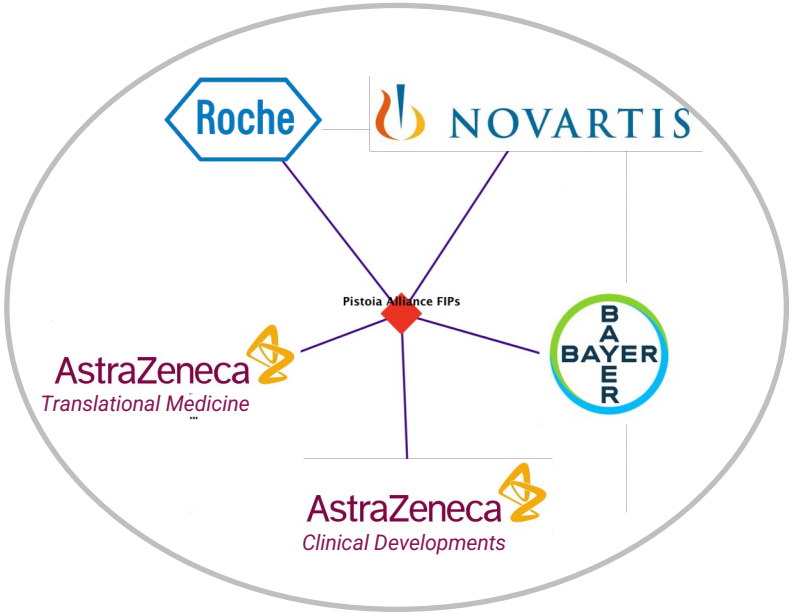
URL: https://committee.iso.org/standard/68848.html



Building and comparing "FAIR profiles"

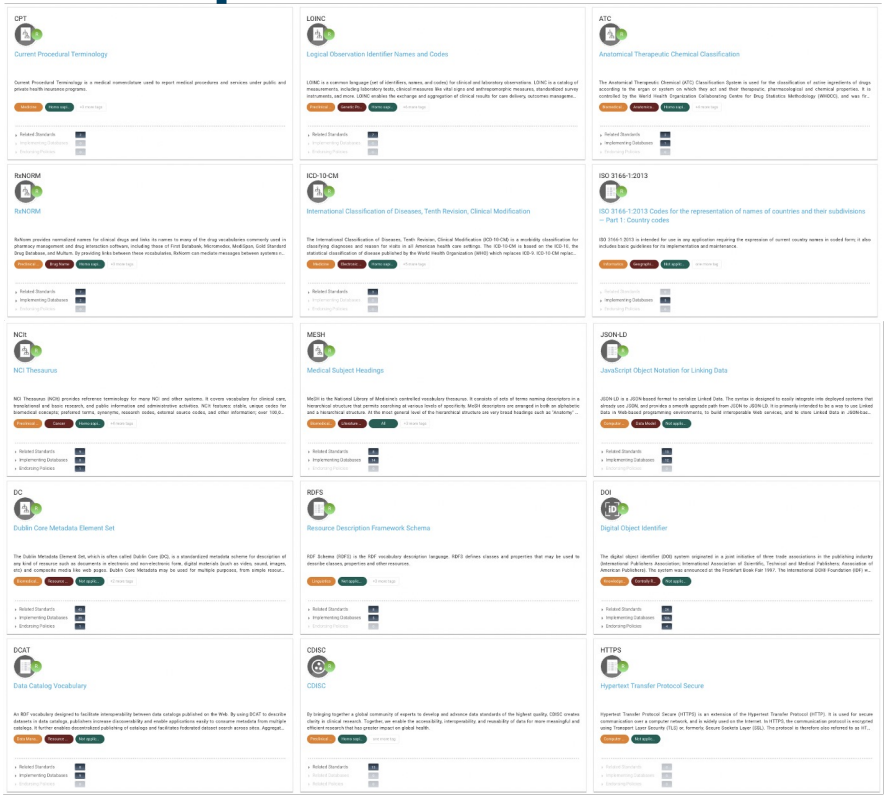
Pistoia Alliance

A collaboration with their FAIR Implementation WG



URL: https://fairsharing.org/3519

(work in progress!)



Disclaimer: These profiles speak for a limited community and do not represent any company standards

Pistoia Alliance

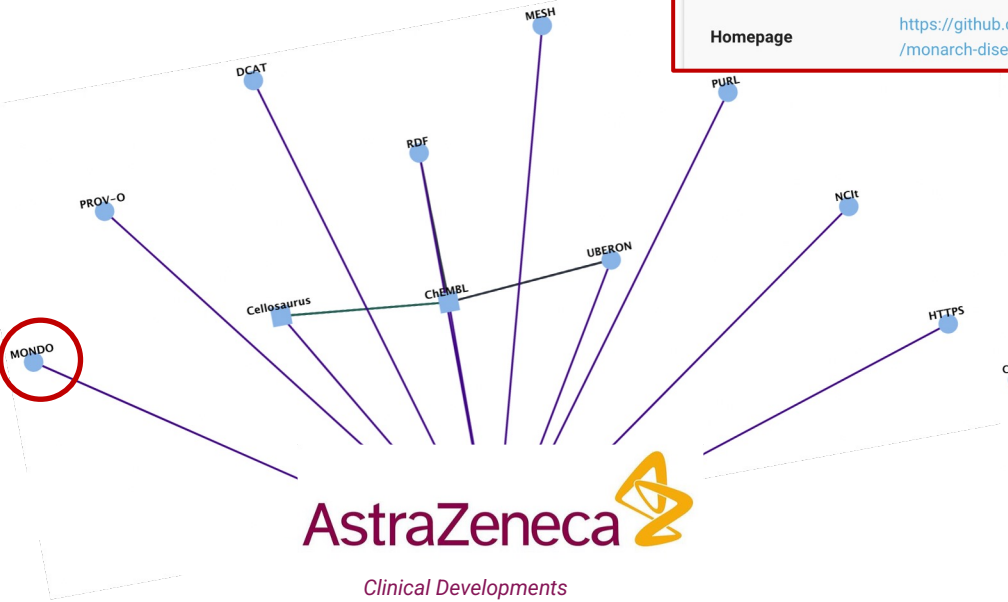
GENERAL INFORMATION

Monarch Disease Ontology (MONDO)

10.25504/FAIRsharing.b2979t

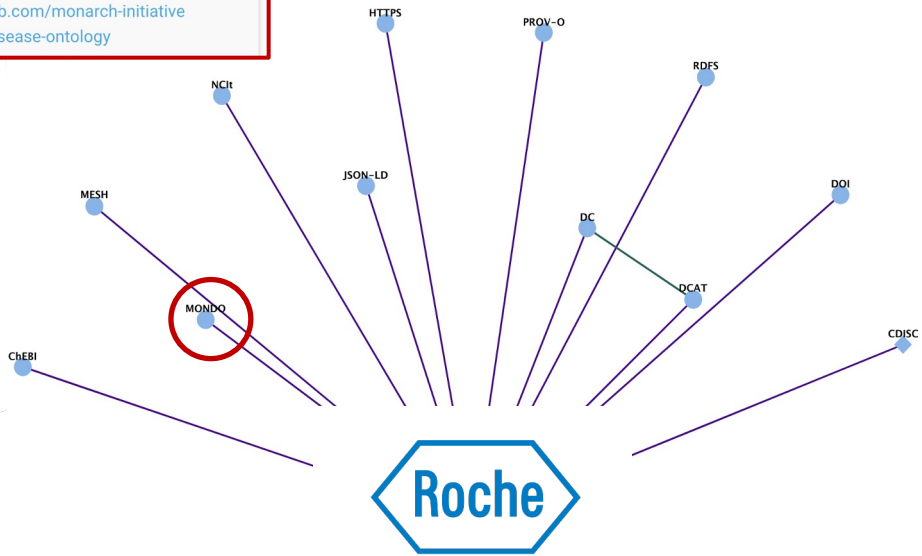
Type	Terminology artefact
Registry	Standard
Description	MonDO (Monarch Disease Ontology) is a semi-automatically constructed ontology that merges in multiple disease resources to yield a coherent merged ontology.
Homepage	https://github.com/monarch-initiative/monarch-disease-ontology

Snapshot of the semantic and syntactic standards used



AstraZeneca

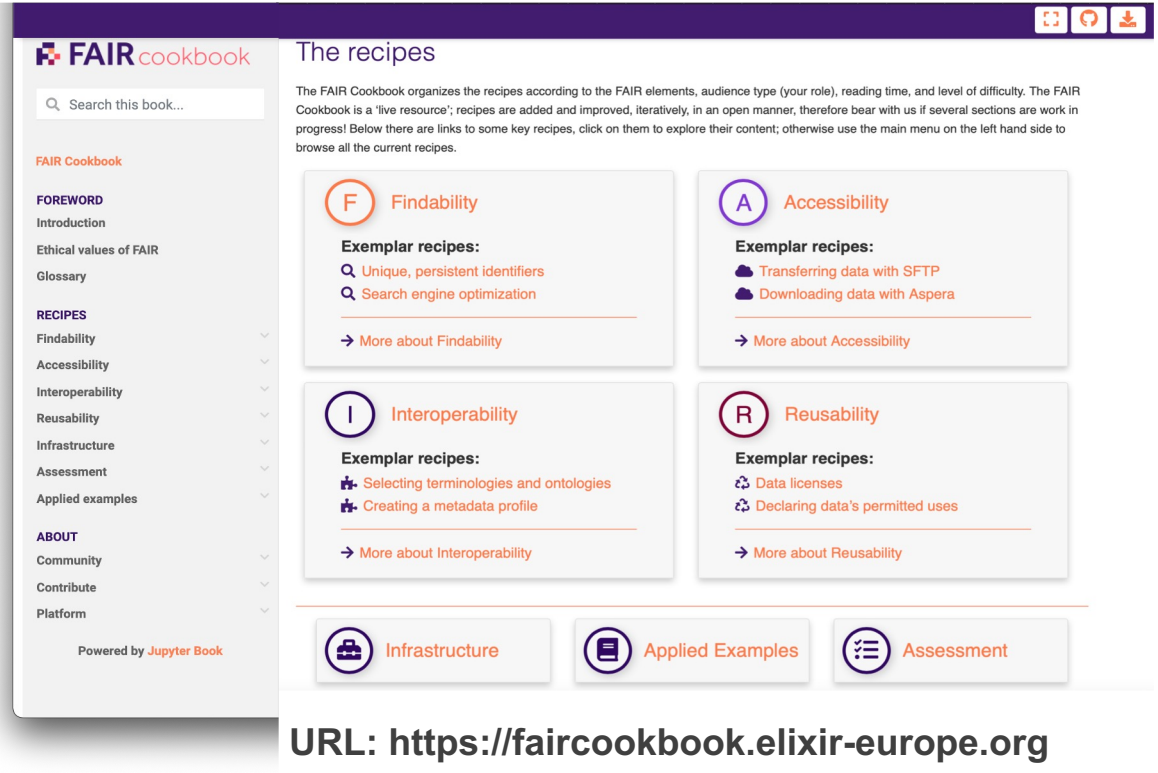
Clinical Developments



Roche

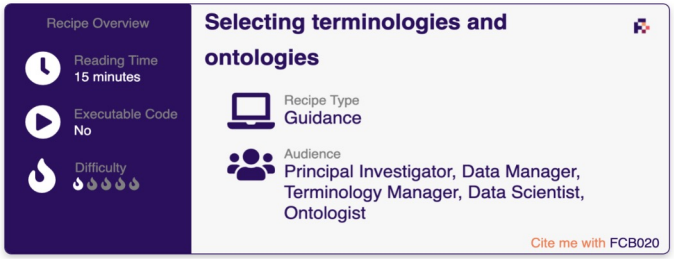
Disclaimer: These profiles speak for a limited community and do not represent any company standards

FAIR cookbook: from knowledge to recipes



URL: https://faircookbook.elixir-europe.org

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

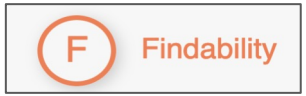


Authored by almost 100 data professionals from industry and academia, including:

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

Define what your needs are

Goal: **improving visibility of content**



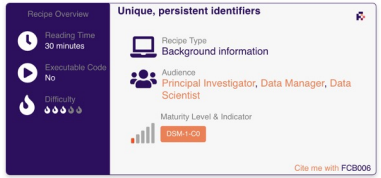
Goal: **semantic integration of datasets from multiple sources**

Goal: **security compliance and with regulators**

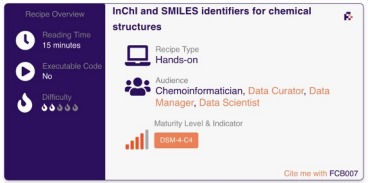
Define what your needs are

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

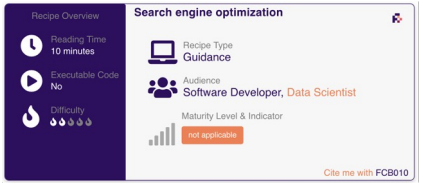
F Findability



https://w3id.org/faircookbook/FCB006



https://w3id.org/faircookbook/FCB007

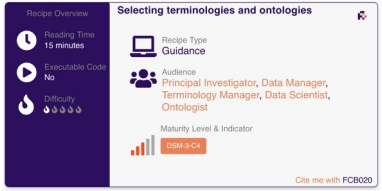


https://w3id.org/faircookbook/FCB010

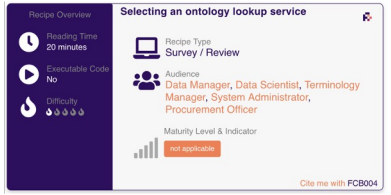
Goal: **semantic integration of datasets from multiple sources**, e.g.:

I Interoperability

Infrastructure



https://w3id.org/faircookbook/FCB020

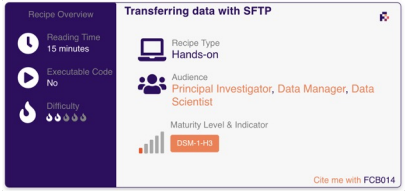


https://w3id.org/faircookbook/FCB004

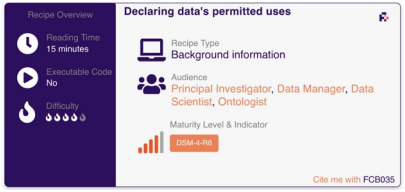
Goal: **security compliance and with regulators**, e.g.:

A Accessibility

R Reusability

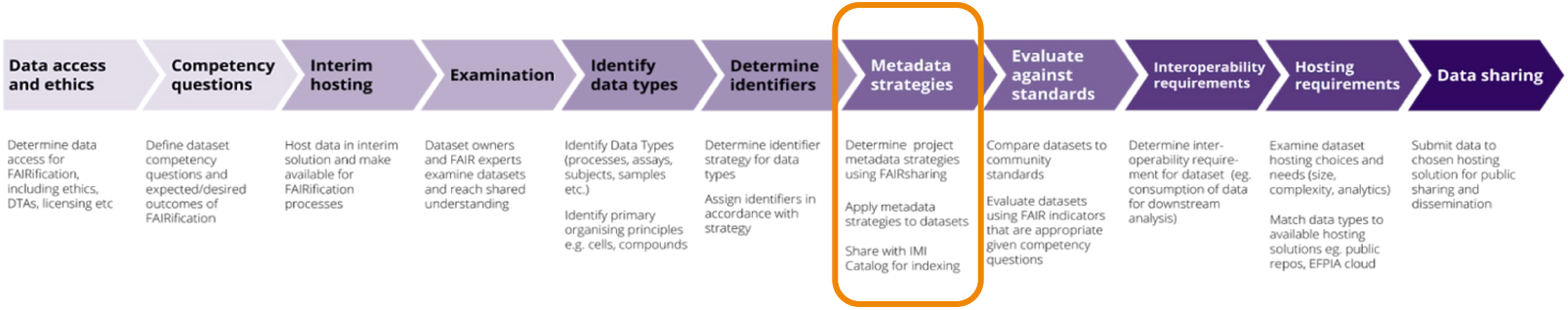


https://w3id.org/faircookbook/FCB014



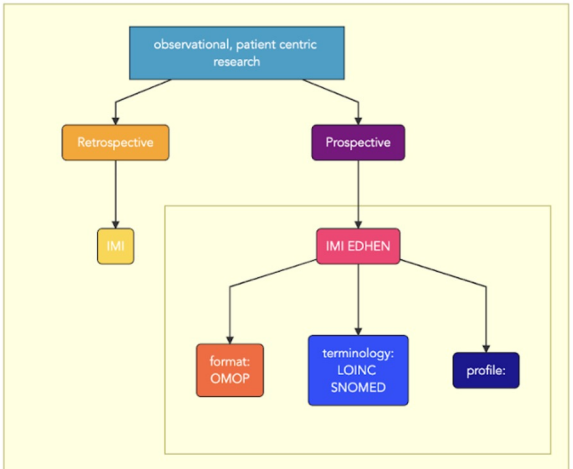
https://w3id.org/faircookbook/FCB035

FAIRification paths: one size does not fit all

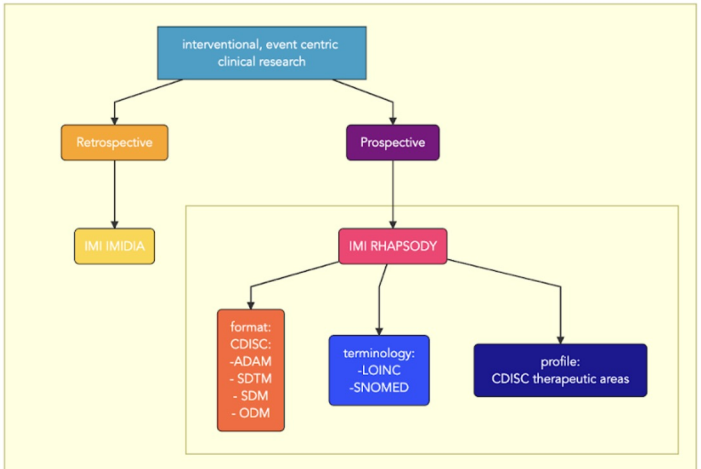


Different contexts mandate different metadata strategies

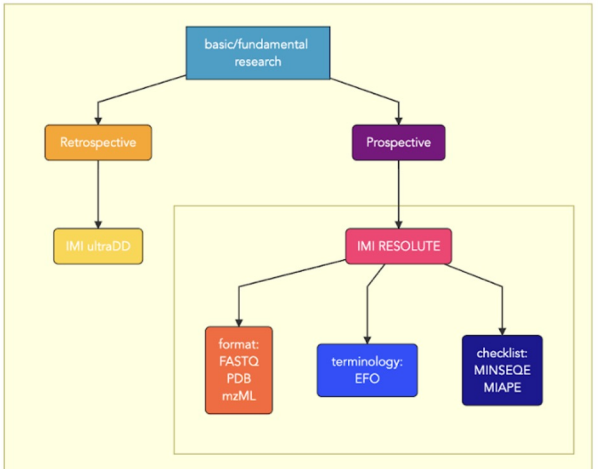
Clinical (observation based)



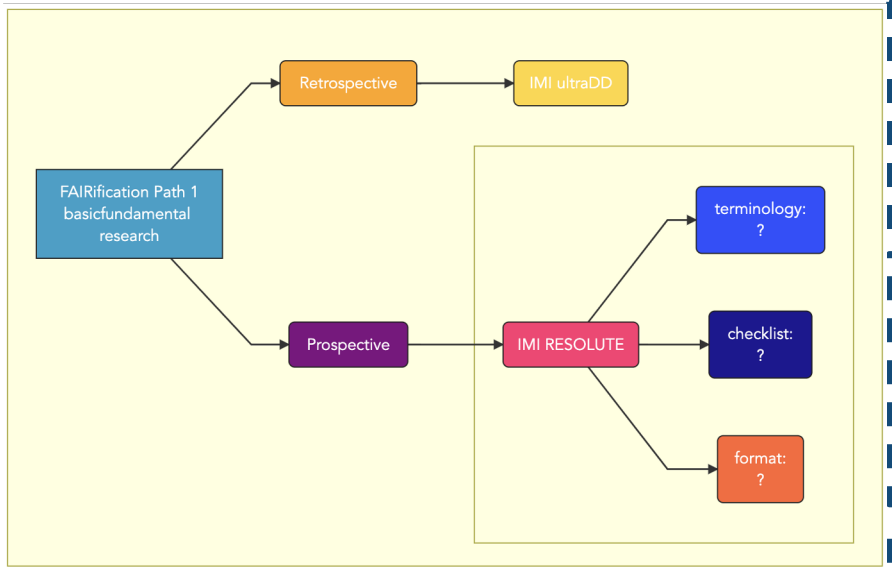
Clinical trial (event based) data



Molecular data



Selecting a ‘standard stacks’ for the FAIRification



Terminologies

FAIRsharing.org

FAIRsharing.org
EFO
Experimental Factor Ontology
TERMINOLOGY ARTIFACT
standards > terminology artifact > doi:10.25504/fairsharing.1gr4tz

Guidelines

FAIRsharing.org
MINSEQE
Minimal Information about a high throughput SEQuencing Experiment
REPORTING GUIDELINE
standards > reporting guideline > doi:10.25504/fairsharing.a55z32

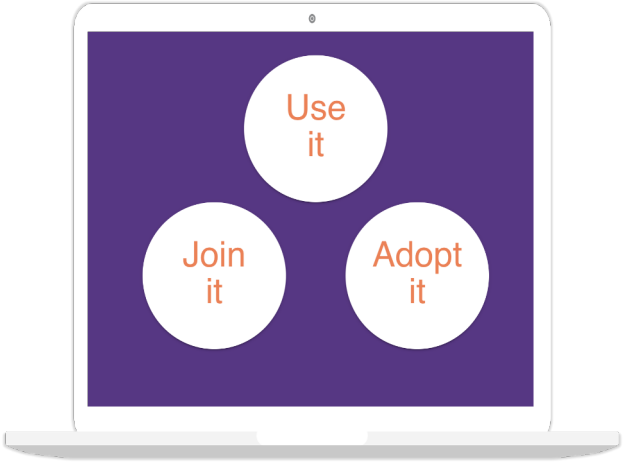
FAIRsharing.org
MIAPE
Minimum Information About a Proteomics Experiment
REPORTING GUIDELINE
standards > reporting guideline > doi:10.25504/fairsharing.8vv5fc

Formats

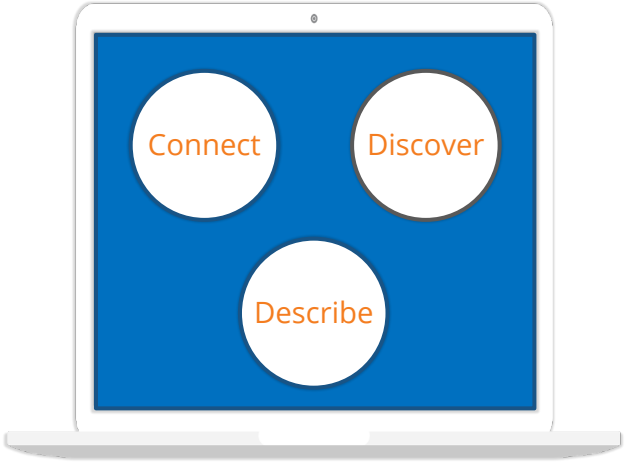
FAIRsharing.org
FASTQ Sequence and Sequence Quality Format
MODEL/FORMAT
standards > model/format > doi:10.25504/fairsharing.r2ts5t

FAIRsharing.org
PDB
Protein Data Bank Format
MODEL/FORMAT
standards > model/format > doi:10.25504/fairsharing.9y4cqw

FAIRsharing.org
mzML
mz Markup Language
MODEL/FORMAT
standards > model/format > doi:10.25504/fairsharing.26dmba



elixir



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

fairsharing.org
contact@fairsharing.org

PHIRI
Population Health Information Research Infrastructure

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

What is metadata?

Practical exercise: Create your own metadata documentation using DCAT standard

Pascal Derycke

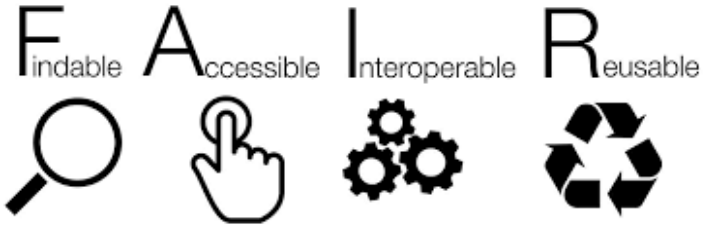
Pre-conference 09-11-2022

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

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.

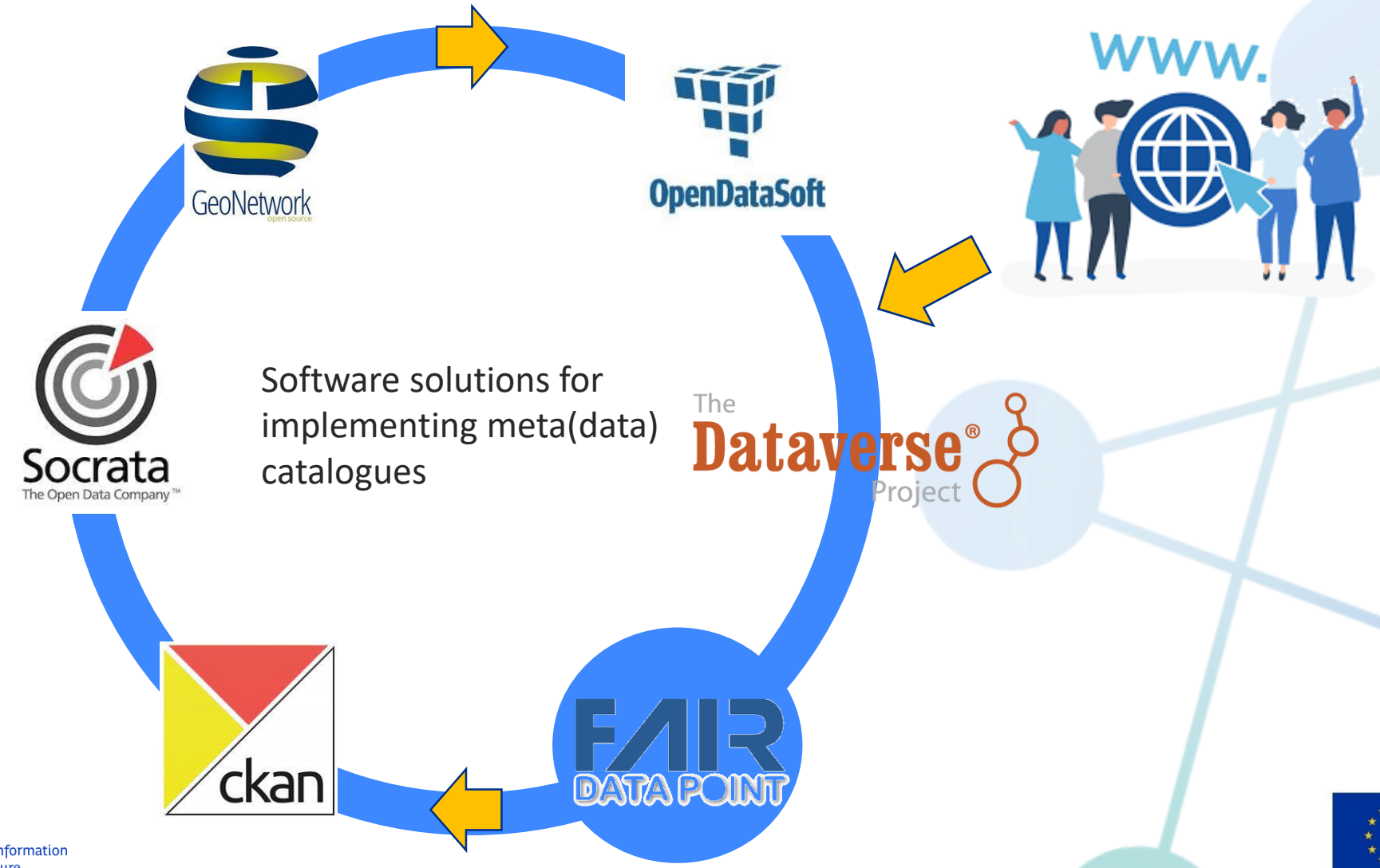
DCAT is designed to facilitate interoperability between data catalogs published on the Web.

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)



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

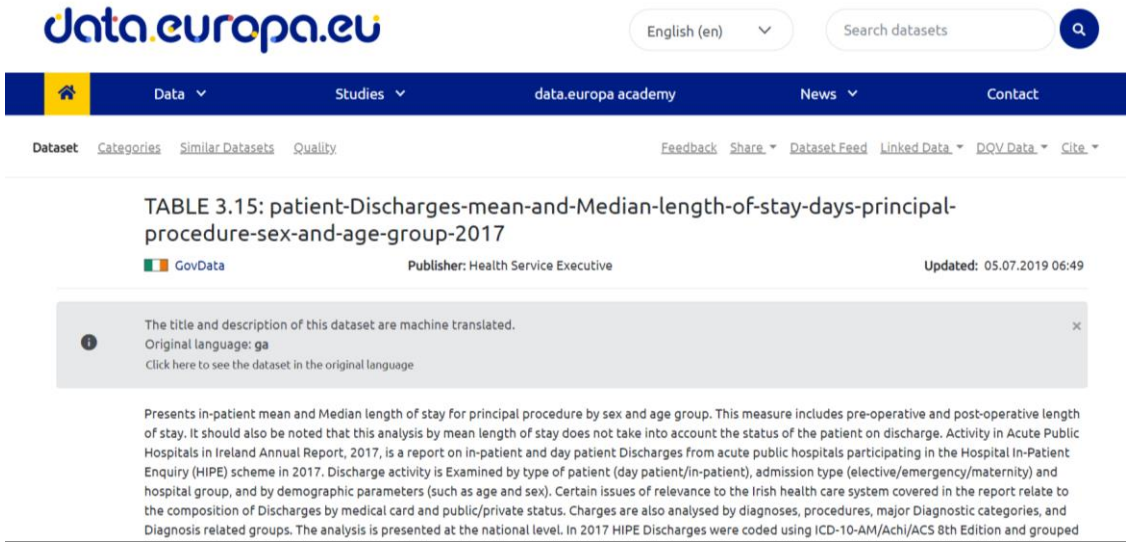
Interoperability framework for exchanging information about data on the Web



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

Interoperability in practice:

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022



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

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

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

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

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

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

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

Q&A

EUROPEAN PUBLIC HEALTH CONFERENCE

BERLIN | 9-12 NOVEMBER 2022

- 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?

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

PHIRI
Population Health Information Research Infrastructure

EUROPEAN PUBLIC HEALTH CONFERENCE
BERLIN | 9-12 NOVEMBER 2022

Thank you

https://HealthInformationPortal.eu

Pascal Derycke: Pascal.Derycke@sciensano.be

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