



PHIRI

Population Health Information
Research Infrastructure

Benefits of common standards for the secondary use of health data

Nienke Schutte & Miriam Saso



EU Health Information System Unit - Sciensano

Aim: to facilitate the secondary use of health information for research and evidence-informed health policies



Nienke Schutte



Miriam Saso



Shona Cosgrove



Charles-Andrew Vande Catsyne



Pascal Derycke



Barthélémy Moreau de Lizoreux



Iris Van Dam



Sasha Milbeck



Primary and secondary use of health data

Primary Use of Health Data

- Processing of personal electronic health data for the provision of health services
- Examples include diagnosis, treatment, and ongoing patient care

Secondary Use of Health Data

- Processing of electronic health data for other than the original reason for which the data was collected
- Key for research, public health monitoring, and system improvements
- Analytical use to derive insights beyond individual patient care



PHIRI

Population Health Information
Research Infrastructure

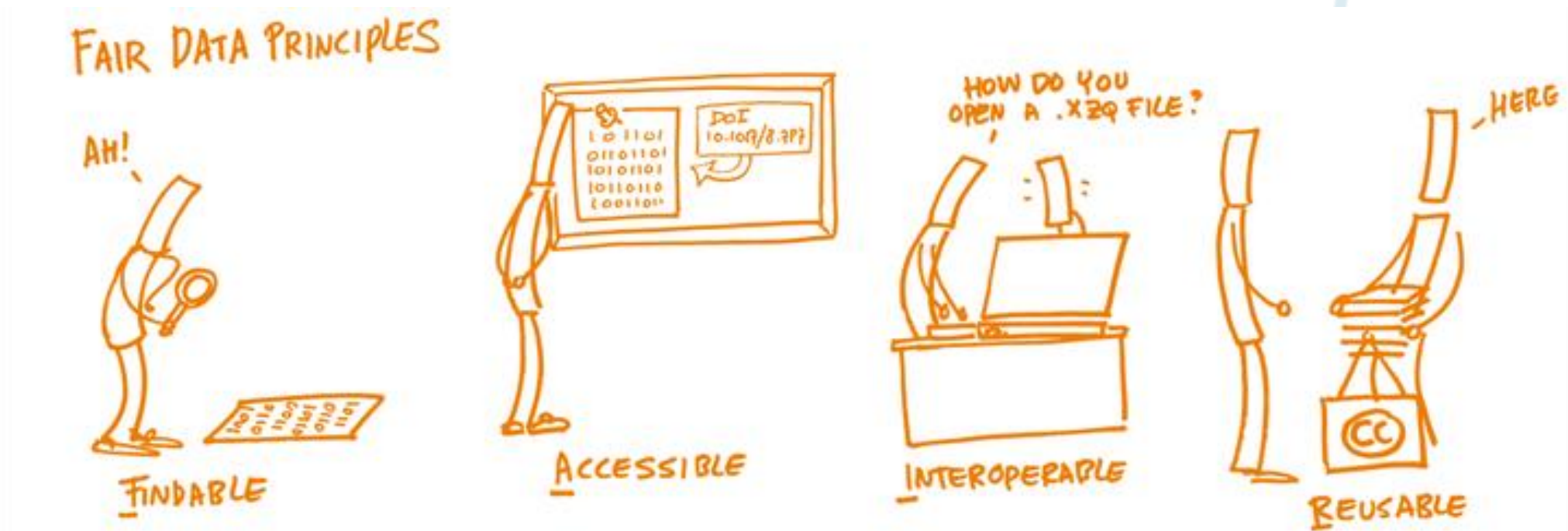
FAIR Data Principles

"I'm equitable and courteous, is my data too?"



FAIR Data Principles (1)

Aim: To enable and enhance the reuse of data by both humans and machines



FAIR Data Principles (2)

Findable: Your data can be discovered by others

- Reach metadata & persistent identifier (e.g. DOI)

Accessible: Your data can be made accessible by others

- Protocol to access the data

Interoperable: Your data can be integrated with other data

- Implementation of recognized standards

Reusable: Your data can be reused by others

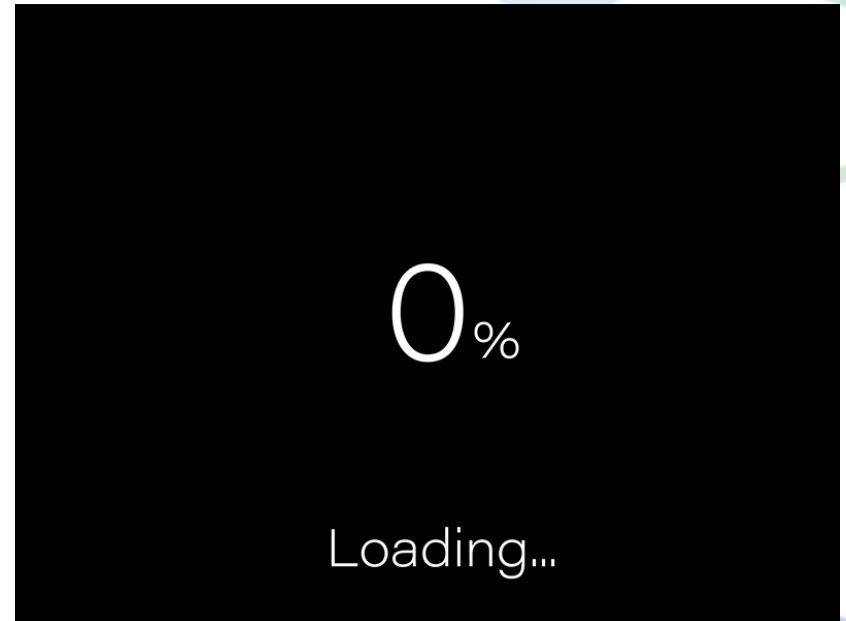
- Context of the data & License

FAIR or not FAIR?

FAIR is a *spectrum*

Add your data in domain specific repositories to start your


FAIR data journey!

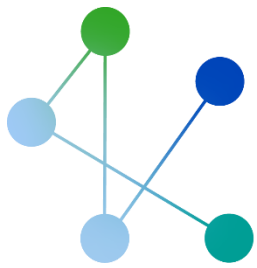




Interoperability in secondary data use

Two examples

1. Create and expose metadata in a FAIR manner
 2. Creating common data models to address pan-European research questions
- 



PHIRI

Population Health Information
Research Infrastructure

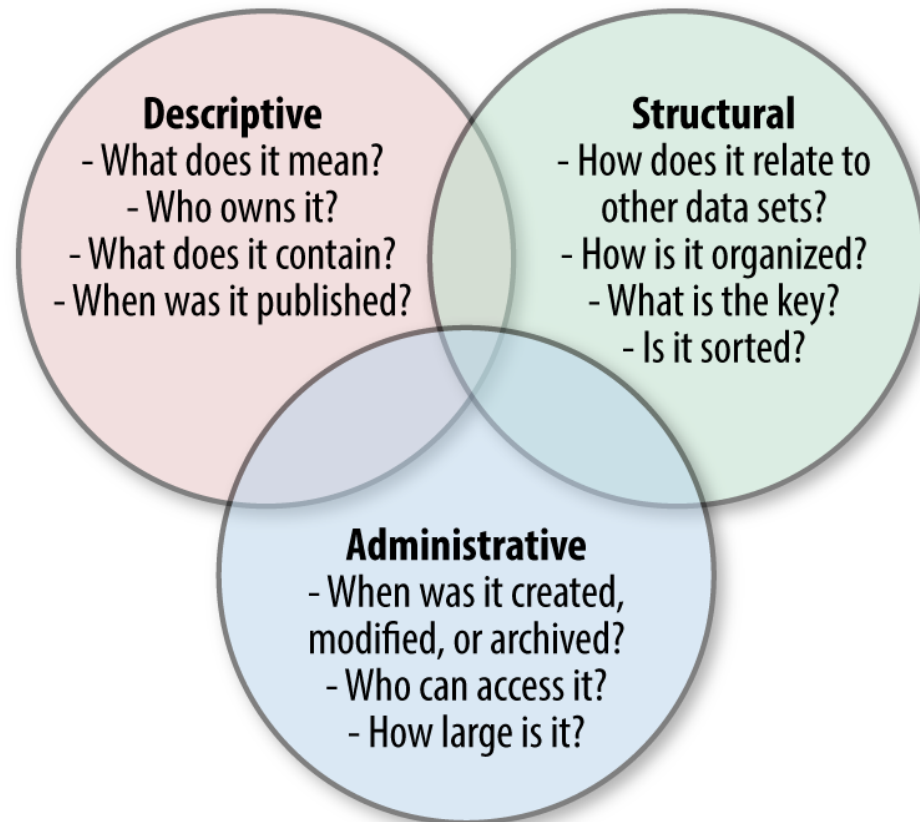
Metadata

"I never met a date I did not like"



What is metadata?

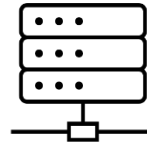
Metadata is "data that provides information about other data"



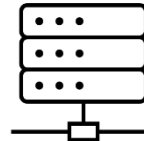
Why metadata?

For my research I need data on prevalence of NCD's in European countries. Where to find it?

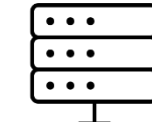
Data **without** metadata



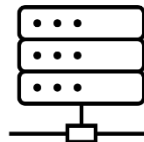
Hospital Discharge Data, NL



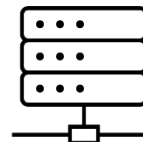
Social Health Insurance Data, AT



Care Register for Health Care, FI

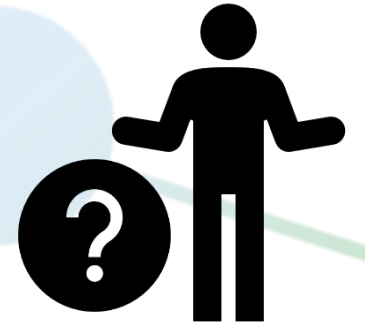


Primary Health Care Database, HR



Morbidity data, RO

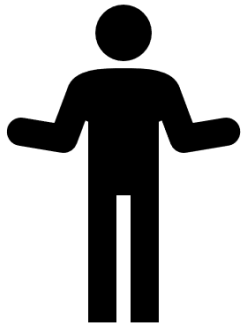
Through my networks and internet search I was able to identify the following potential data sources



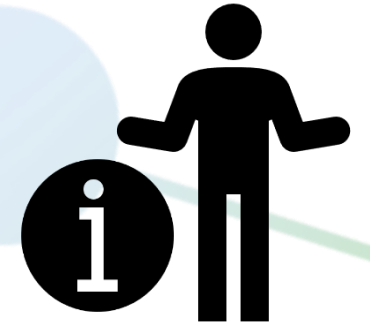
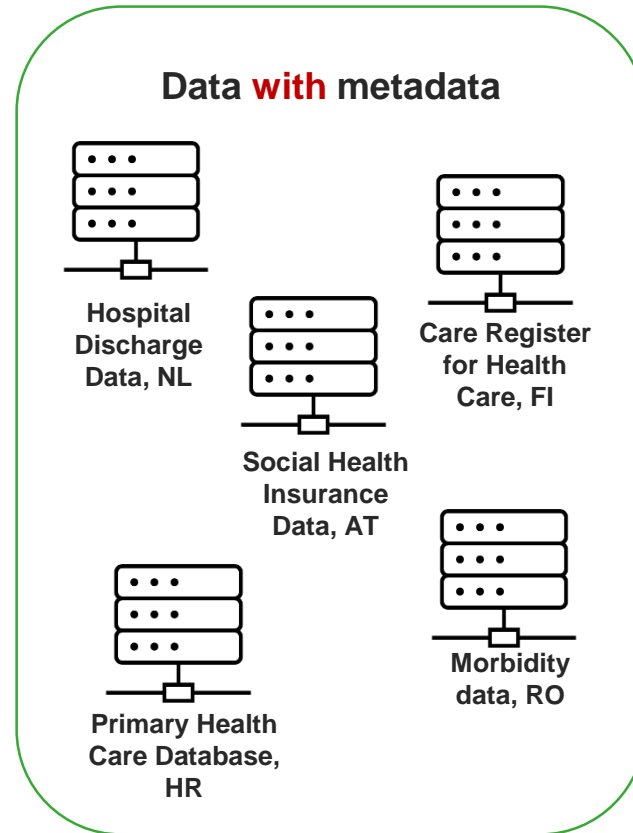
- What type of data are these?
- What age groups do they cover?
- What information do the variables cover?
- Who owns these data?
- How do I get access to these data sources?

Why metadata?

For my research I need data on prevalence of NCD's in European countries. Where to find it?



Through the Health Information Portal I was able to identify the following potential data sources



In structured format

- Data on hospitalizations
- Covers all 18+
- National coverage
- Basic background variables on sex, age and diagnostic information
- Contact details for access requests

Schema.org & DCAT metadata standards

Why Schema.org

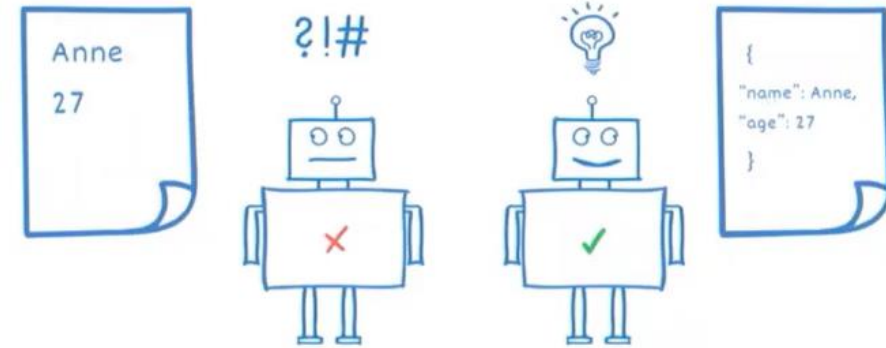
Schema.org is a metadata standard for the indexation of web pages

- 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





The Health Information Portal

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

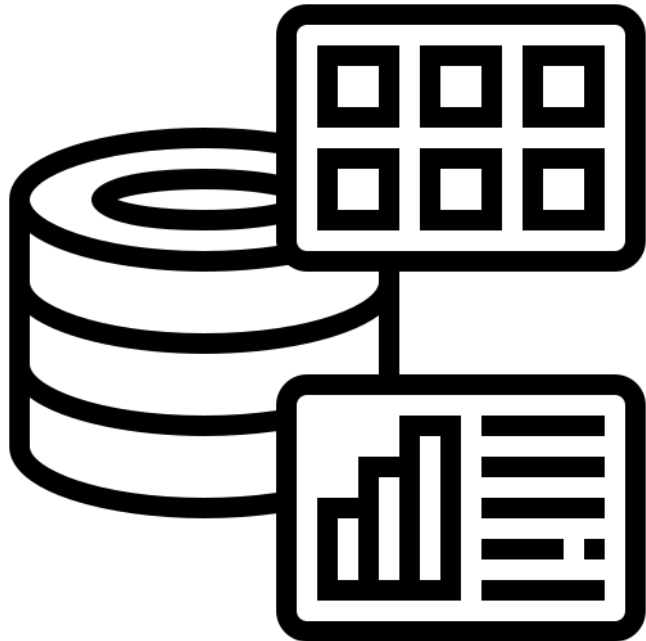
You can use boolean operators like AND, OR and NOT (in capital letters) to narrow your search.

Country

- None -

SEARCH

What can you discover in the metadata catalogue?



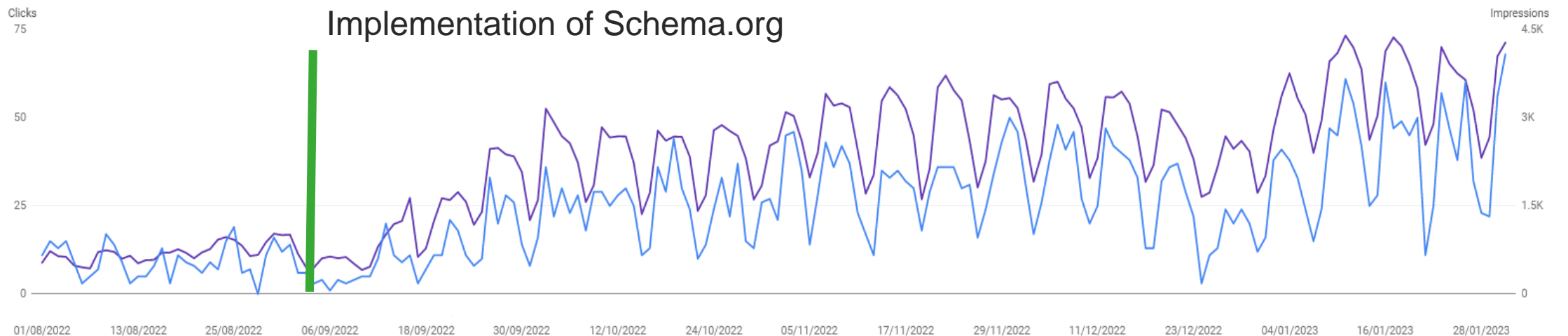
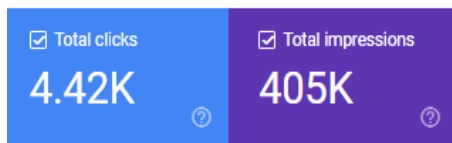
The Health Information Portal provides structured information on

- Population registries such as cancer or mortality registries
- Health survey data
- Surveillance data for infectious diseases
- Hospitalizations data (in- and out-patient)
- Healthcare resources, costs and activities
- Etc.

Effects of Schema.org

The Health Information Portal: Google search analytics:

- Clicks: number of times a user clicked on a link provided by Google to the HIP.
- Impressions: how often someone saw a link of the HIP on Google.



Why create metadata?

- Valorization of your work
 - Your data can be **used** by others (instead of collecting dust)
 - Your data will be **findable** to search engines (Google, Yahoo...)
 - Your data will **increase your exposure** as a researcher
 - Included in interesting collaborations and publications
 - You will easily **find other datasets** with the same features (increase your sample size)
- Only-once principle
 - Metadata only needs to be entered once, **because the use of standards**
- Link with existing strategies/regulations
 - Local data management strategies
 - European Health Data Space
- **Stimulate cross-institutional and cross-border research**



PHIRI


Population Health Information
Research Infrastructure

Common data models

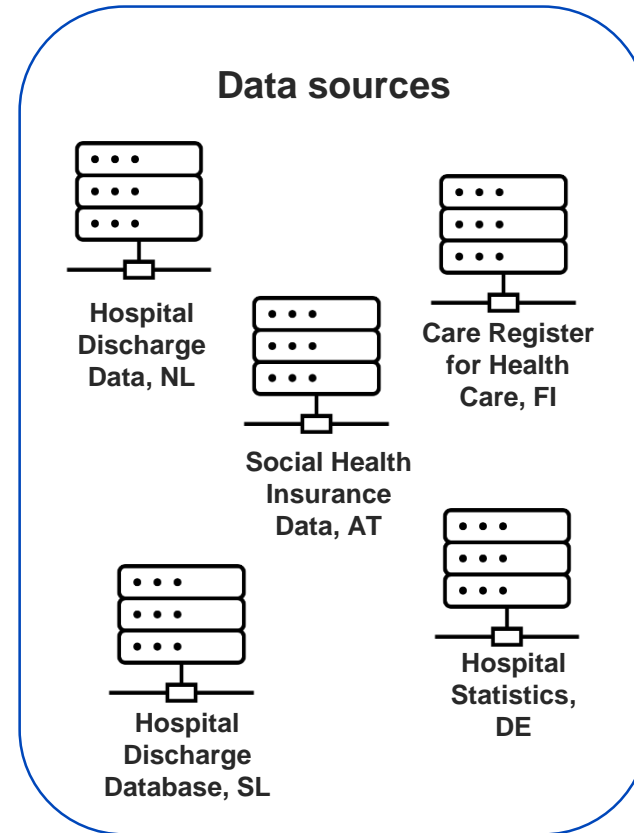
Common grounds make data flourish



Data from difference sources



For my research I need data from disparate data sources



But:

- Depending on their primary needs, none of the databases capture variables equally well
- Extracting data for analysis purposes requires strict data use agreements and complex access control

Data need to be harmonized into a common standard

The Common Data Model


A common framework that defines how data is represented in a systematic, consistent, interoperable manner

- to provide data organized in a way optimal **for analysis**
- using **standardized vocabularies** containing all necessary and appropriate corresponding standard healthcare concepts
- being '**technology neutral**': it does not require a specific technology, can be realized in any relational database

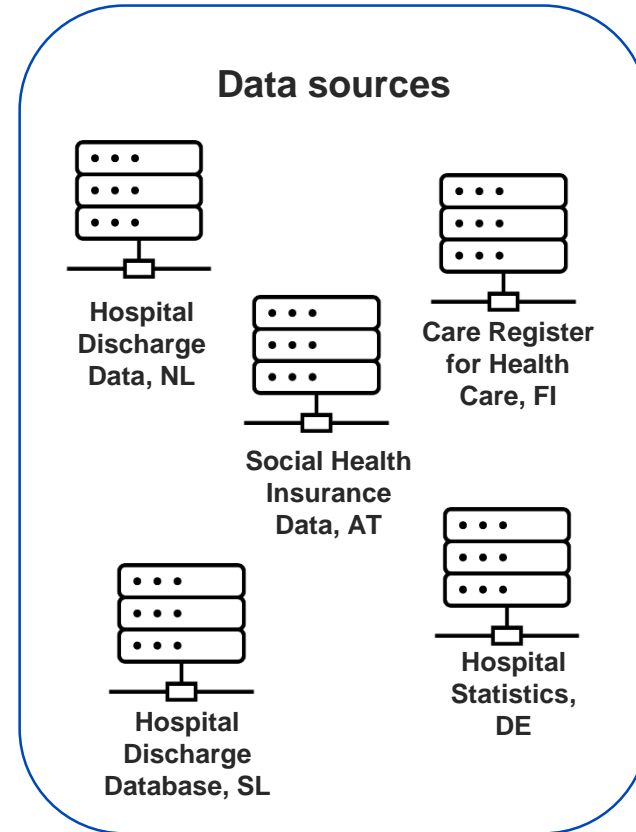
The Common Data Model

Model entity	Variables	Variables information						
entity	variable label	variable description (concept)	encoding	variable format	variable type	units	requirement_level	variable validation rules
person	person_id	Pseudoid of the person included in the cohort	UTF-8	string / character	Categorical variable		Required	
person	sex_cd	Sex of the person included in the cohort (at the time of entering the cohort)	ISO/IEC 5218	integer	Categorical variable		Recommended	Options: 0 (not known), 1 (male), 2 (female), 9 (not applicable)
person	age_nm	Age of the person included in the cohort (at the time of entering the cohort)		integer	Numerical variable	years	Required	age_nm >= 18 age_nm <= 115
person	age_cd	Age group of the person included in the cohort		string / character	Categorical variable		Required	Options (>= x & <y): '18-25', '25-35', '35-45', '45-55', '55-65', '65-75', '75-85'

Data from difference sources



For my research I need data from disparate data sources



But:

- Depending on their primary needs, none of the databases capture variables equally well
- Extracting data for analysis purposes requires strict data use agreements and complex access control

Data need to be harmonized into a common standard

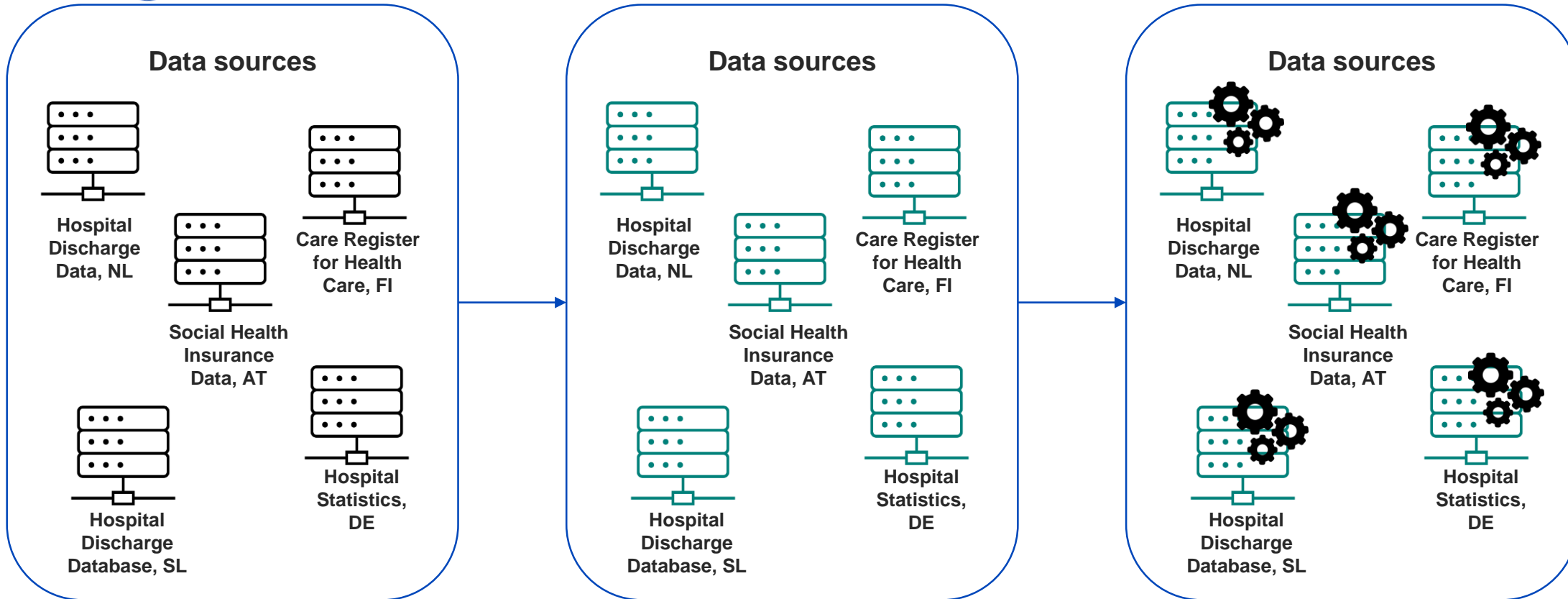
Applying the Common Data Model



Common Data Model



Analytical script



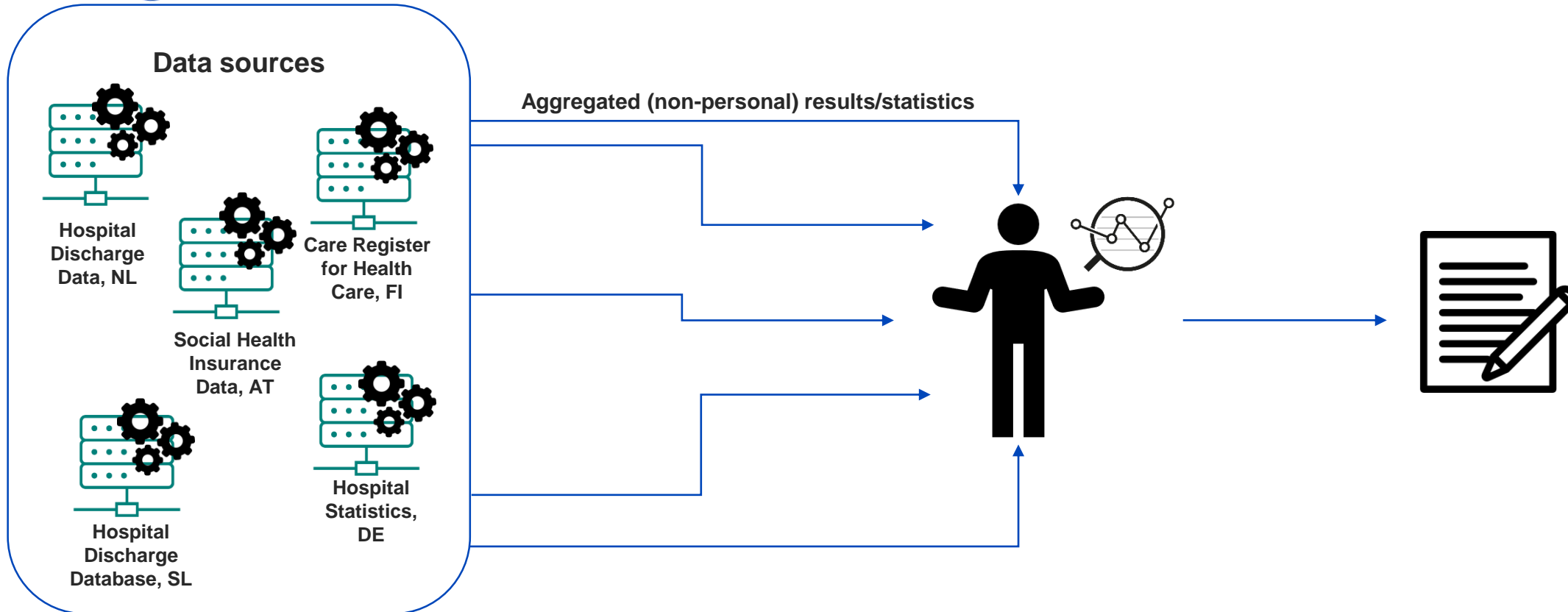
Applying the Common Data Model



Common Data Model



Analytical script



Common Data Models in PHIRI



Direct and indirect determinants of COVID-19 infection and outcomes in vulnerable population groups with reference to inequalities



COVID-19 related delayed care in breast cancer patients



The impact of COVID-19 on perinatal health and perinatal health inequalities



COVID-19 related changes in population mental health

4 Real life research use cases measuring the impact of COVID-19 on population health

 European Health Information Portal

About ▾ Search COVID-19 measures ▾ Search countries ▾ Search Europe ▾ Search trainings ▾ Services ▾

FIND DATA

FIND PUBLICATION

FIND EUROPEAN INITIATIVE

FEDERATED DEMONSTRATORS ▾

PHIRI DEMONSTRATORS

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

Country

- None -

SEARCH

FIND DATA

www.phiri.eu

Take home messages: Recap

01

Primary use and secondary use of health data

02

FAIR principles

03

Metadata interoperability

04

Common Data Models

05


Interoperability as key to pan-European research

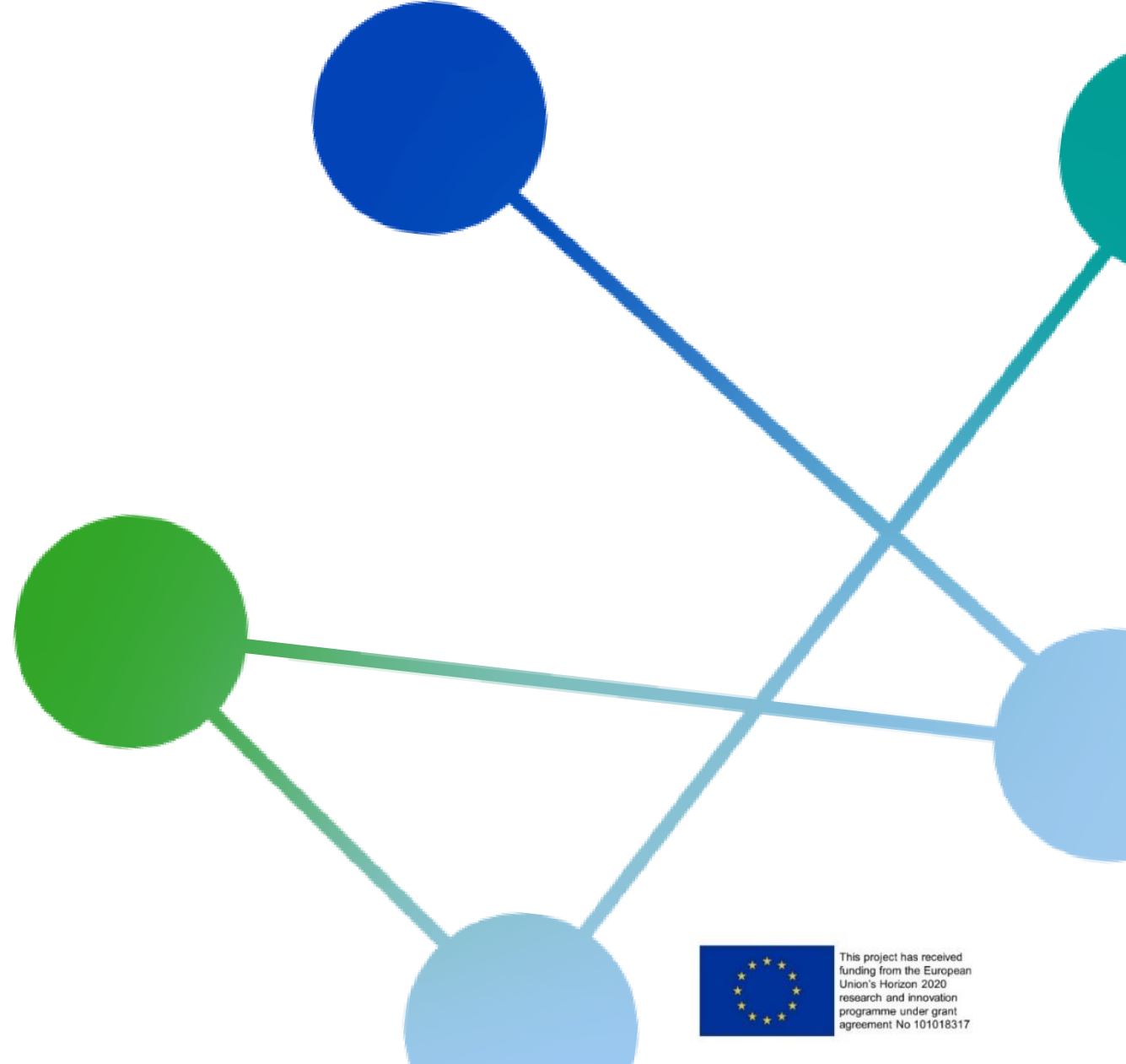


PHIRI

Population Health Information
Research Infrastructure

Thank you for your attention and FAIRwell

Nienke Schutte & Miriam Saso
Phiri.coordination@sciensano.be
Follow us on Twitter:  @PHIRI4EU



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