

# EHDS

HealthData@EU Pilot

## **European Health Data Space – Working with the future European metadata catalogue**

**Pascal Derycke (Sciensano)**

EUPHA conference 9 November 2023,  
9:00-10:00, Dublin

# Looking into the future

---

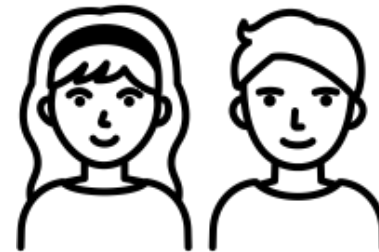
We are in 2028!

It's almost 3 years that the **European Health Data Space** has been launched providing the research community with a service-oriented architecture which comes with services supporting **data discoverability** and **access to health data**. A DXP

## A DXP fit for the secondary use of health data

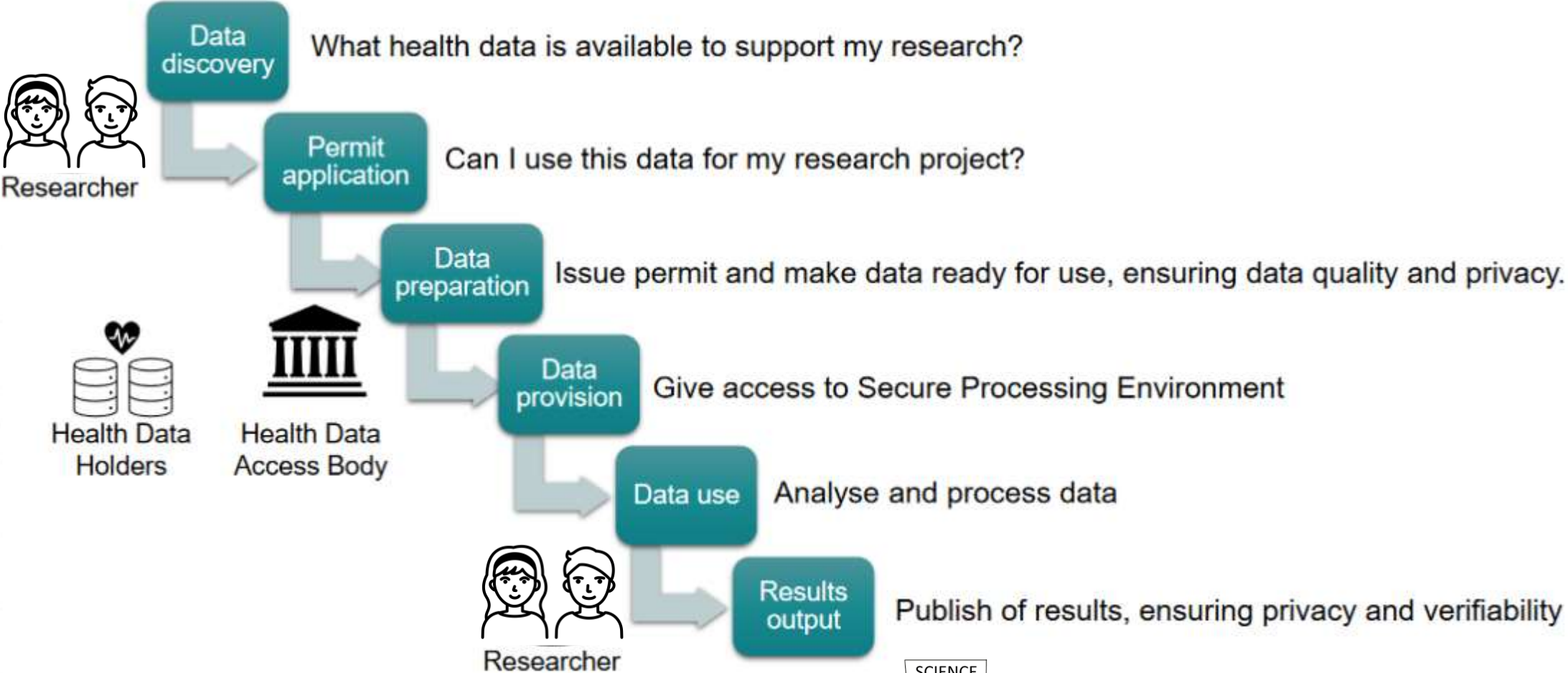
**Gaëlle** and **Diarmuid** are two newly graduated epidemiologists. They have just published their first scientific paper in which they establish a relationship between traffic pollution levels at certain locations, diseases and population types. For their research study, they have been processing data from 3 European data spaces:

- Environment (Green deal)
- Mobility
- Health



Concerning health data, they obtained access to sensitive population health data for all European capitals within 3 months by submitting one single application form at the **Healthdata@EU central platform**.

# Overview of a generic data access approval process



# Creating a metadata record

---

The scientific paper of Gaëlle and Diarmuid comes along with a **health research dataset**.



For the purpose of making their work findable (FAIR), they have created a metadata record in the European Health metadata catalogue.

The DCAT Application profiles\* within EU data spaces provide all necessary templates to create a rich metadata record for their dataset.

\* **DCAT-AP** is a specification based on W3C's Data Catalogue vocabulary ([DCAT](#)) for describing public sector datasets in Europe

# Describing a dataset

---

Describing a dataset consists in responding to the following questions:

**What?**

**How?**

**Why?**

**Who?**

**When?**

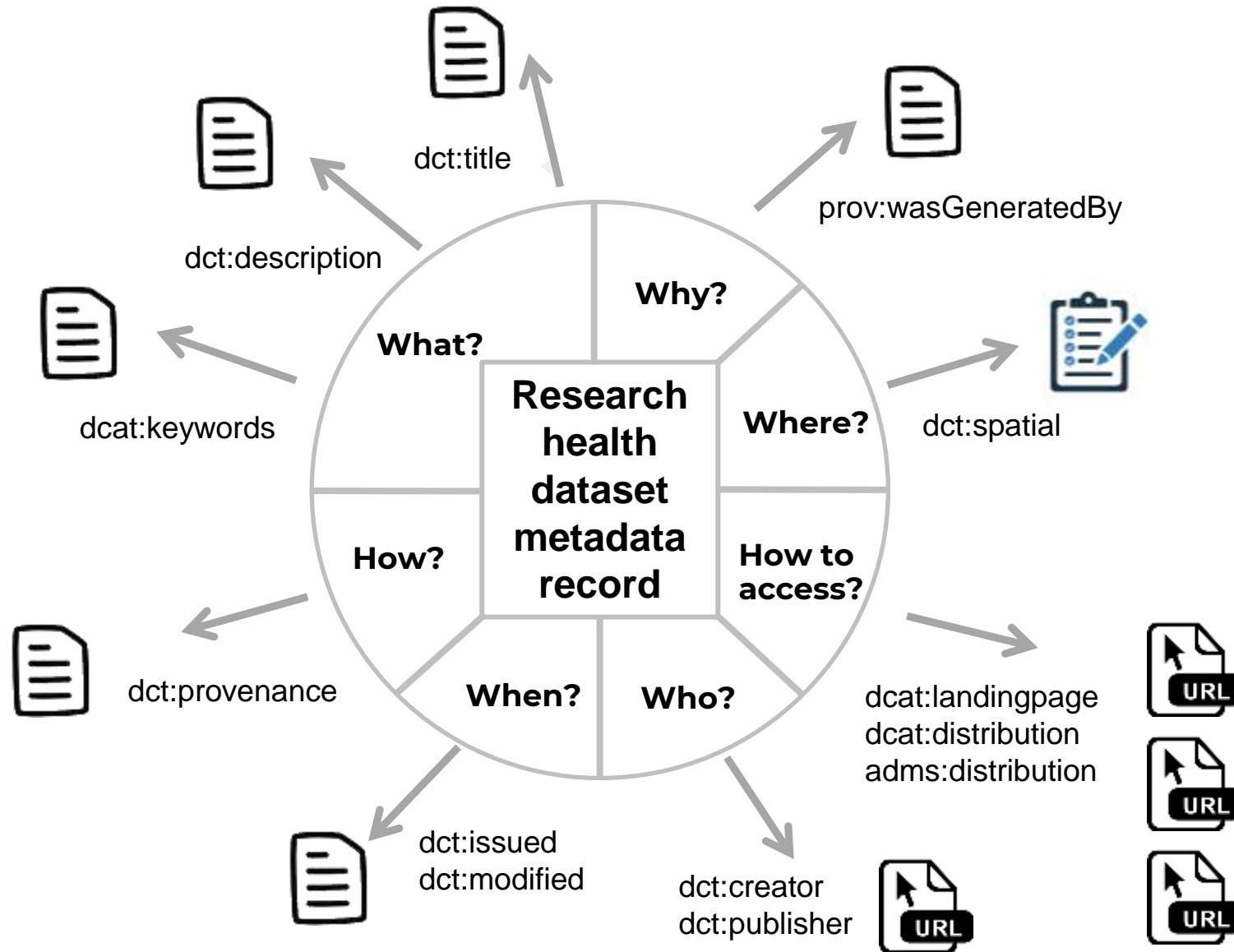
**Where?**

**And ... How to access?**

Each Application Profile is designed to address the specificity of a particular data type and/or data space:

- GeoDCAT for spatial datasets
- StatDCAT for statistical open datasets
- HealthDCAT for Health datasets

# DCAT metadata model for describing datasets



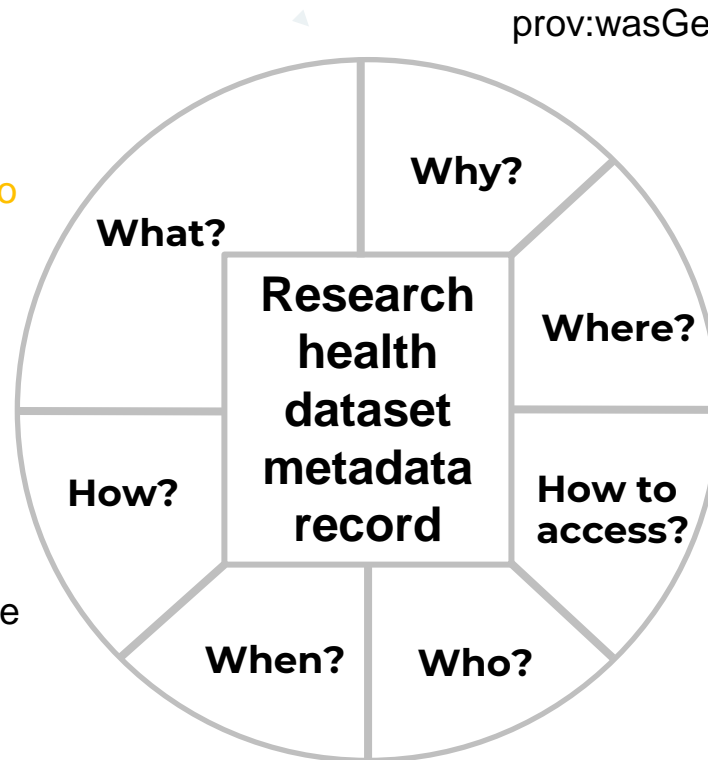
# DCAT metadata model for describing datasets

  
Text@  
language

  
Code list

  
www

dct:title  
dct:description  
dcat:keywords  
foaf:page  
dct:language  
dcat:theme  
dct:conformsTo  
dct:type



dct:spatial  
dcat:spatialResolutionInMeters

dct:hasVersion  
dct:isReferencedBy  
dct:isVersionOf  
prov:qualifiedAttribution  
dcat:qualifiedRelation  
dct:relation  
dct:source


dct:provenance

dcat:landingpage  
dcat:distribution  
adms:distribution  
dct:accessRights

dct:issued  
dct:modified  
dct:temporal  
dct:accrualPeriodicity  
dcat:temporalResolution

dct:contactPoint  
dct:creator  
dct:publisher

  
Organisation  
& people

  
Dataset

  
Resource

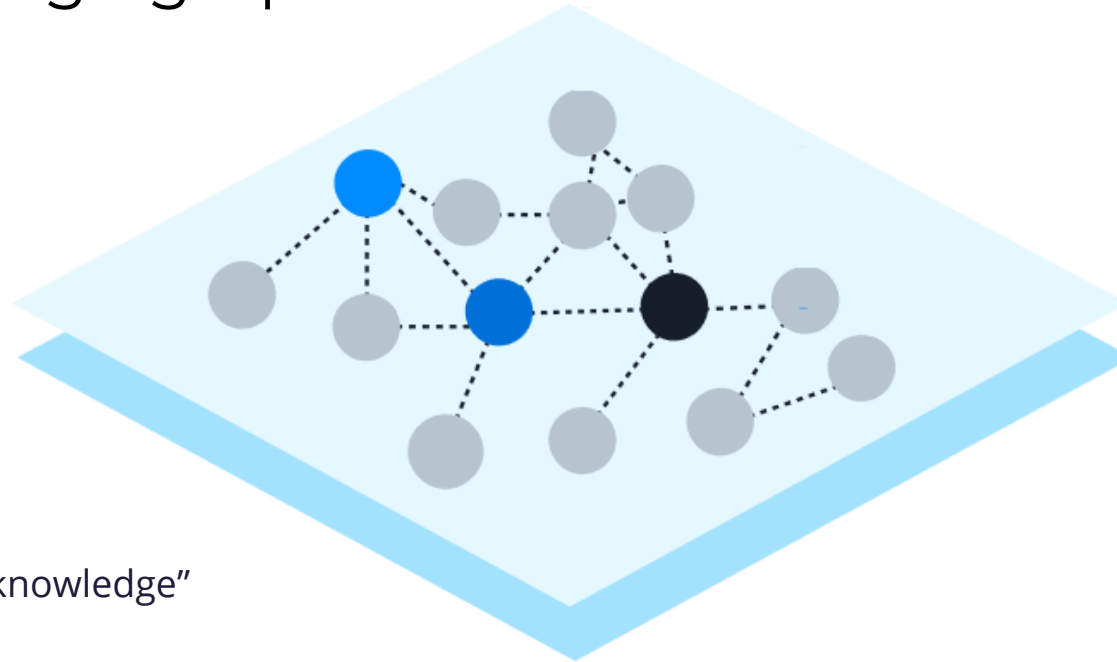




# Data Catalogue Vocabulary (DCAT): a RDF vocabulary

DCAT is a representation of metadata using Semantic Web Technologies

DCAT is a knowledge graph



## Semantic web

"enabling machines to create new knowledge"

## Web of data or linked data

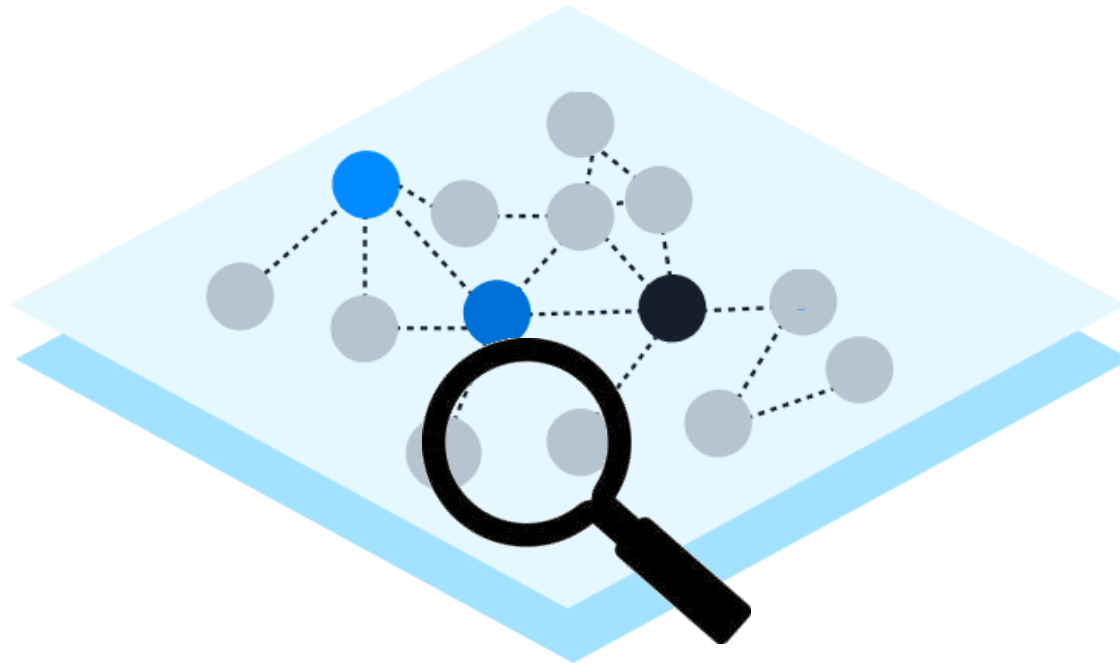
"enabling machines to act on our behalf to find and aggregate information in more flexible, intuitive and intelligent way"

**RDF** stands for Resource Description Framework. It is a standard for describing resources on the web, using a variety of data formats. It provides a way to structure data, and link it to other data, effectively allowing for the data to be mixed, exposed, and shared across different applications. RDF enables the creation of a web of data that can be easily interconnected and has a wide range of applications, particularly in representing information about resources in a machine-readable way.

# Health data exploration

---

Back in 2025, when Gaëlle and Diarmuid started their research project, they searched for relevant health datasets in the **European Health Data Space** metadata catalogue looking for specific diseases, population types, locations, time periods, ...



# Challenges encountered when exploring sensitive datasets

---

For health data, they are facing several challenges: They will only be able to evaluate the fitness for their purpose if the dataset is properly described.

- what are the variables of the dataset?
- can it be linked to another dataset – technically and legally?
- what is the coding system in use?
- will the reuse be authorised for the research project I want to conduct? (Is my project in line with its scope?)
- ...



# How to address these challenges?

---

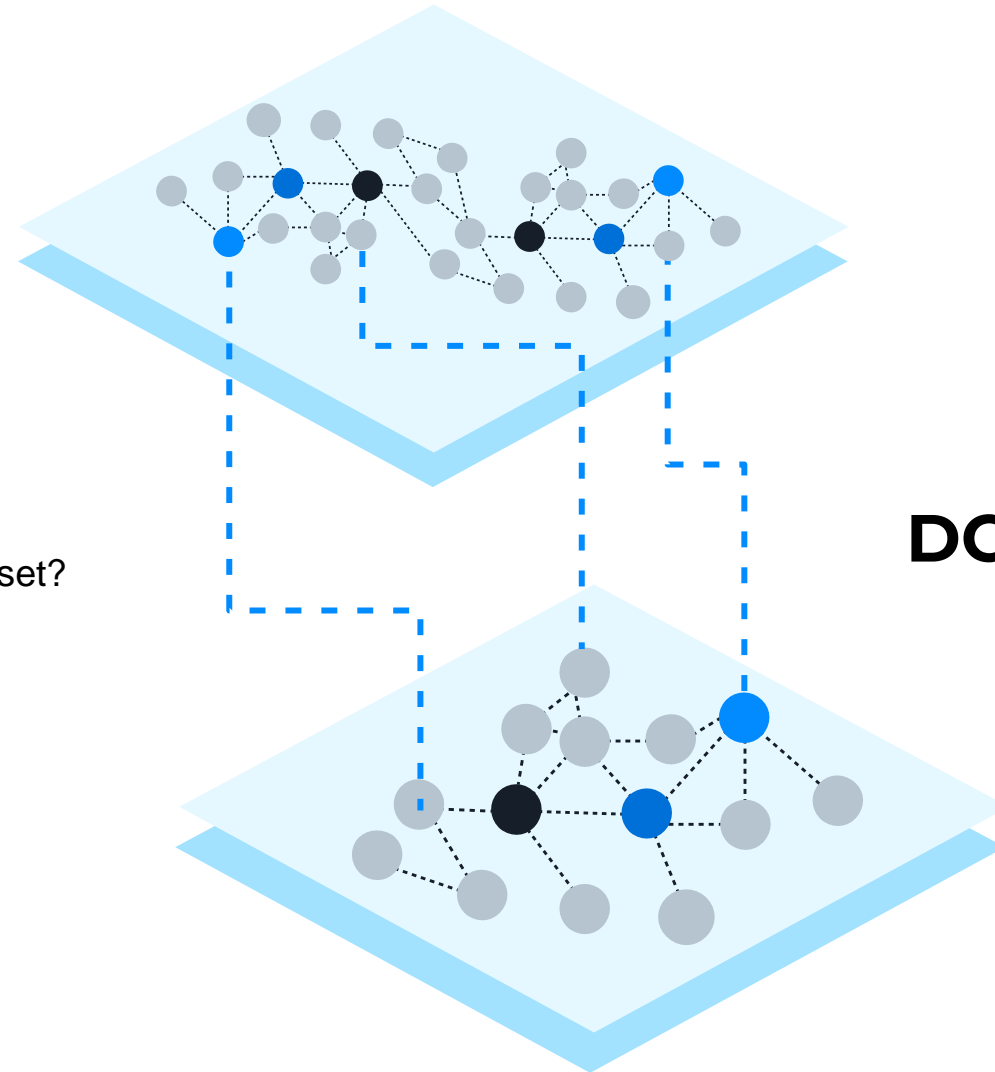


# HealthDCAT-AP in interoperable European data spaces

**HealthDCAT-AP will provide answers to the following questions:**

- Who is the Health Data Access Body?
- What is the population type in the dataset?
- What is the population size?
- How are individuals identified?
- What is the medical coding system in use in the dataset?
- What is the dataset quality?

...



**GeoDCAT-AP**  
**StatDCAT-AP**  
...

**DCAT-AP**

**European Health Data Space**

**But also ...**

---

**expanding knowledge**

# European Health Data Space: a virtuous circle platform

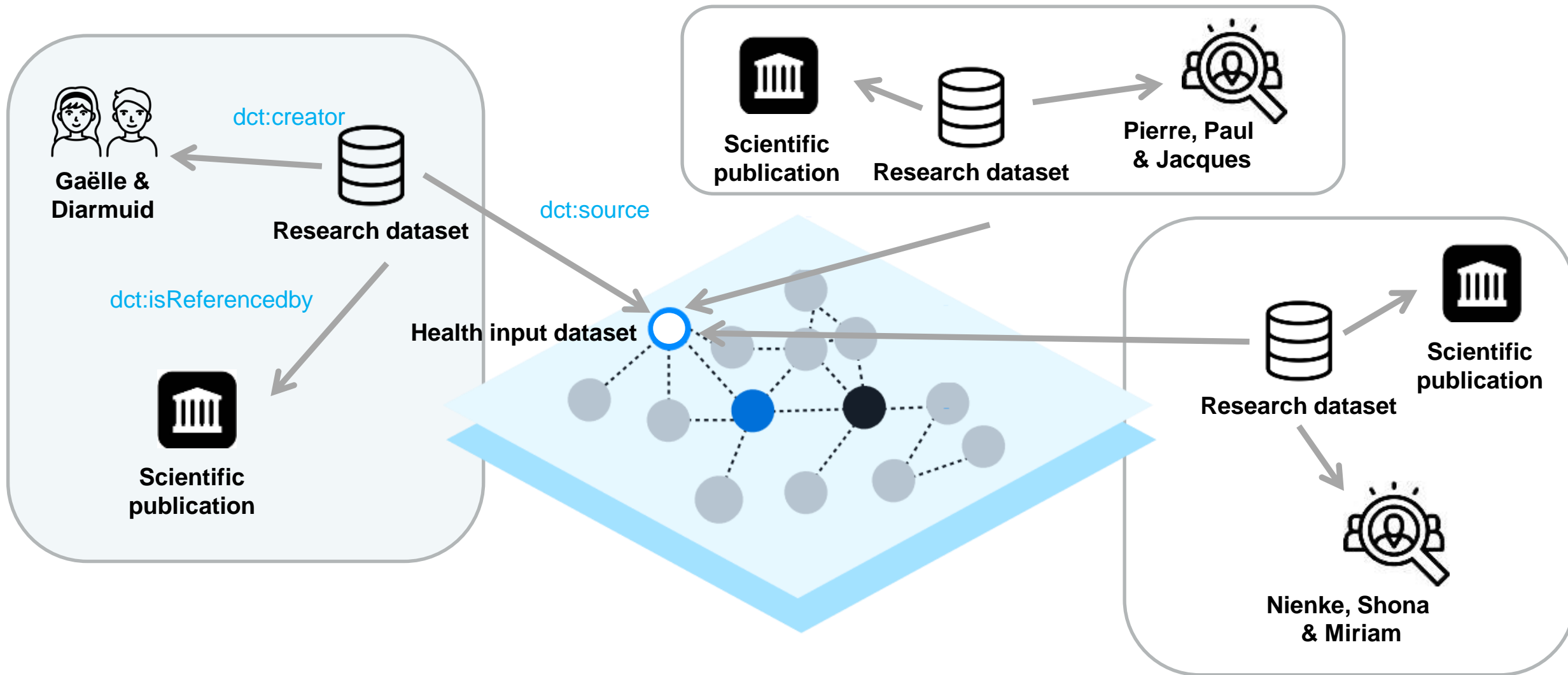
---

Through their work, Gaëlle and Diarmuid have extended the DCAT knowledge graph by creating and connecting new dots.

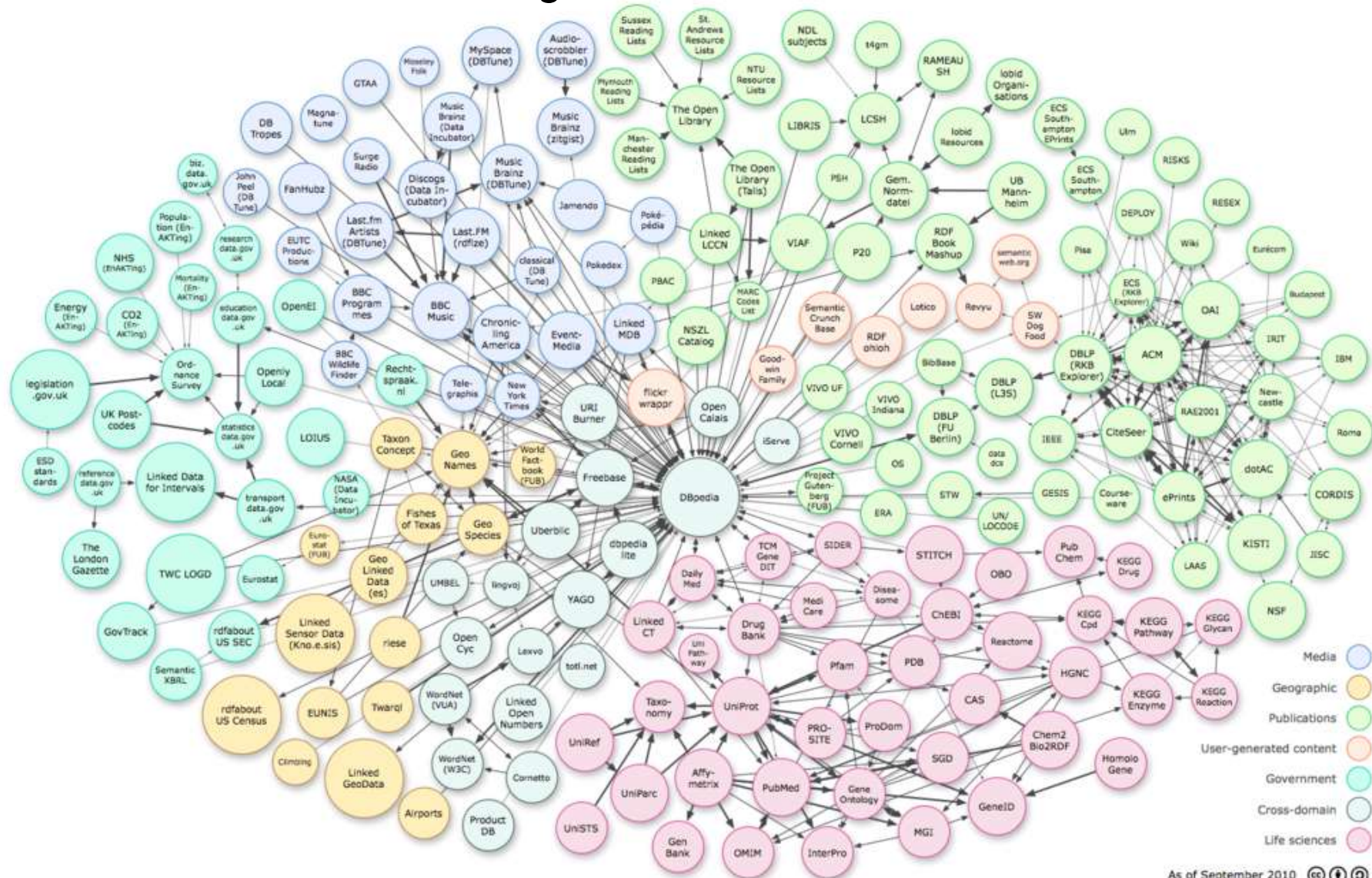
Although we don't know yet how the health dataset created by their research project will be used, we do now know for which type of research the sensitive health data that they have been processing can be used... thanks to the newly created dots connections



# Connecting the dots



# The European Health Data Space: a knowledge hub designed for the secondary use of data





Q&A session

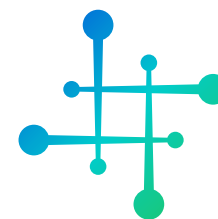


## Dr. Pascal Derycke, PhD

Researcher

**EU health information system unit  
Epidemiology and public health**

Pascal.derycke@sciensano.be



# EHDS

HealthData@EU Pilot

Follow us on social media!



@ehds2pilot



<https://www.ehds2pilot.eu/>



[ehds2pilot@health-data-hub.fr](mailto:ehds2pilot@health-data-hub.fr)



**Co-funded by  
the European Union**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.