Making Linked Data accessible for One Health Surveillance with the "One Health Linked Data Toolbox"
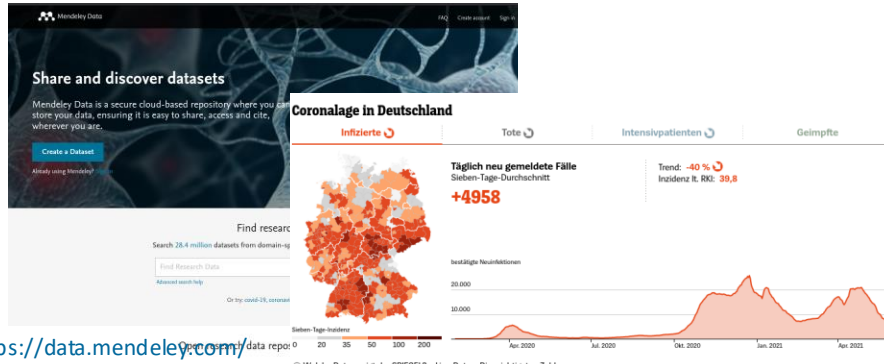
Presenter: Taras Günther

Matthias Filter, Fernanda Dórea
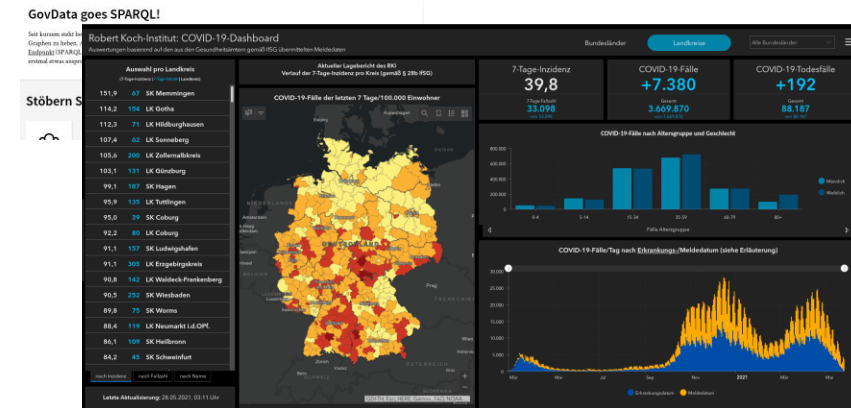
23.06.2021

One Health Surveillance Data - **Demands and challenges are rising!**

Fast and *interdisciplinary data exchange* and *integration* are key elements for One Health Surveillance
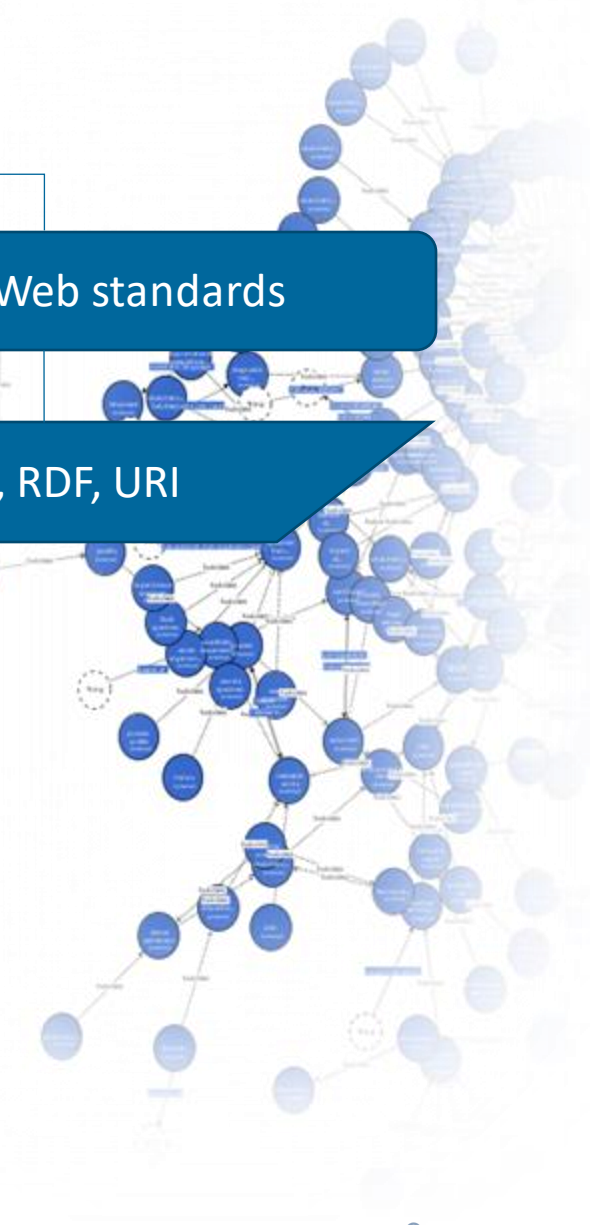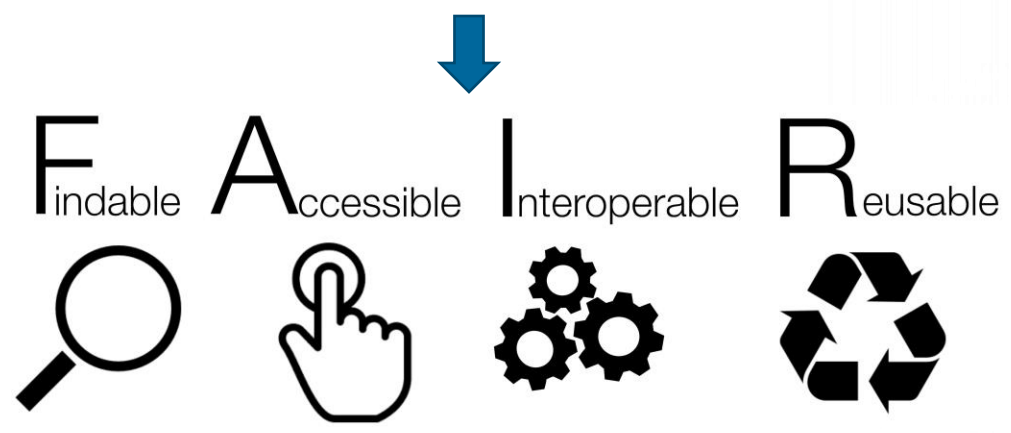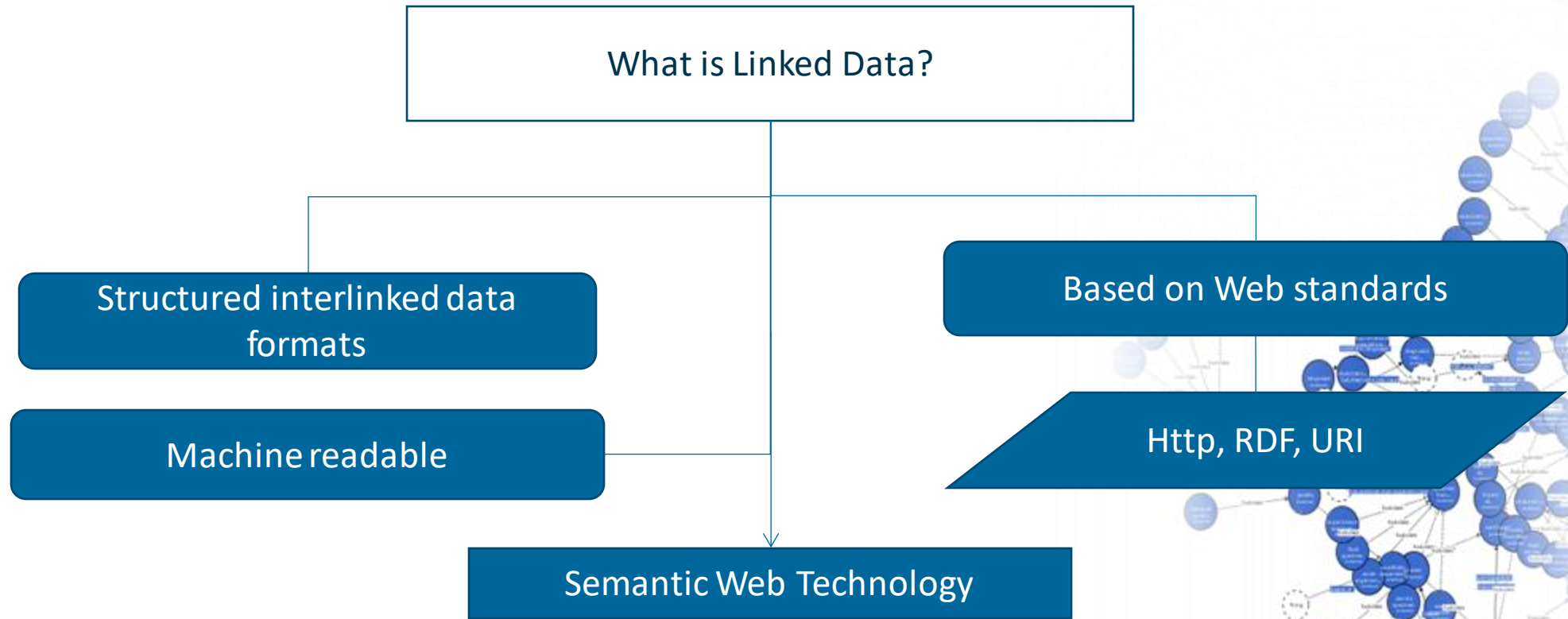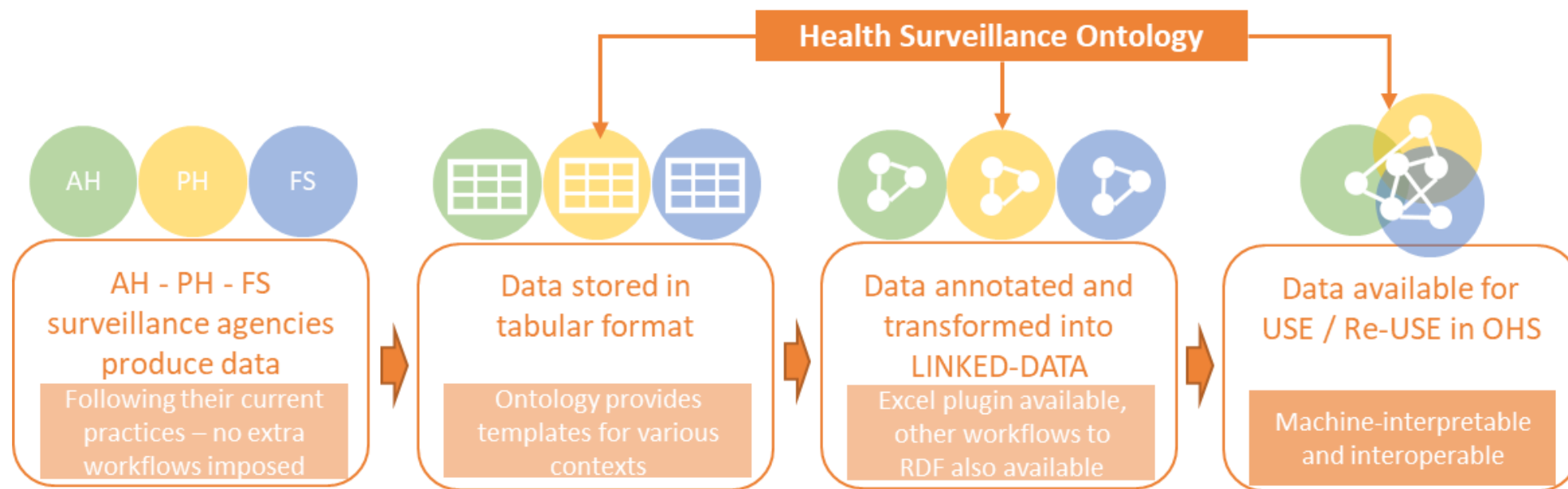
Source : https://data.mendeley.com/

Source : https://www.spiegel.de/thema/coronavirus

Semantic Web Technologies such as **Linked Data** and **Ontologies**:
Tools to handle these challenges?

What is Linked Data?

Structured interlinked data formats

Based on Web standards

Machine readable

Http, RDF, URI

Semantic Web Technology

Findable   Accessible   Interoperable   Reusable

Semantic Interoperability with Linked Data in One Health

Source: One Health Surveillance Codex,2021

The Health Surveillance Ontology (HSO)

Interlinked Standard

Interlinked Schema

Semantic Logic

Available here:
https://github.com/nandadorea/HSO
https://bioportal.bioontology.org/ontologies/HSO
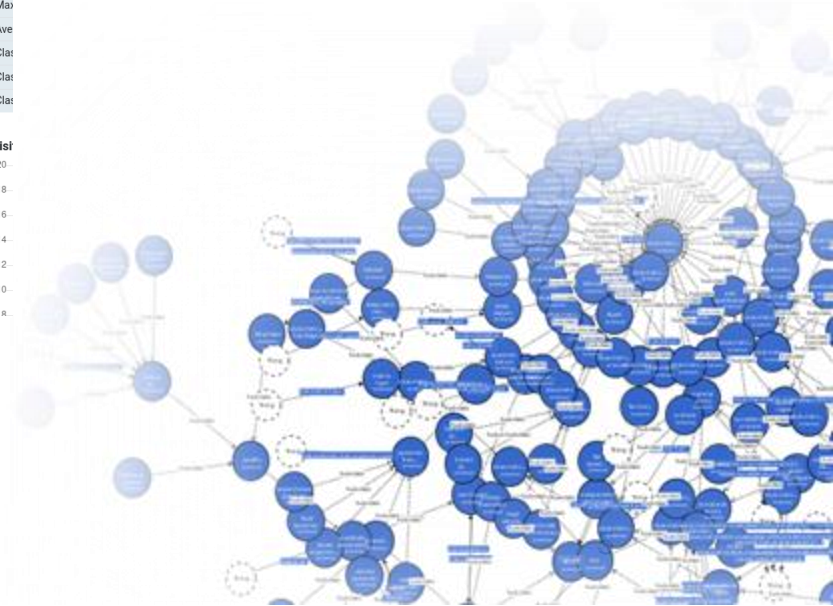
Health Surveillance Ontology toolbox

HSO Converter Tools

HSO Concept Search

SPARQL Endpoint   Application

EFSA ECDC Surveillance Data Connector

Use Case I : Data Harmonization and Standardization (Data management)

How?
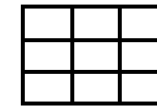
.xlsx → .rdf
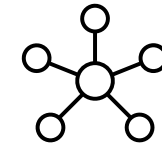
- Consistent use of URI's in the knowledge graph

- URI specific entities can be attached such as:
  - Relations / graphs / hierarchical structures
  - Different labels and languages

<http://purl.obolibrary.org/obo/HSO_0000368>

Hospitalised cases (HSO)

Hospitalized cases (ECDC)

Benefits:
• Standardized data structures attached to corresponding metadata
• Automatic parsing of existing data bases and files

Use Case I: XLSX to RDF Converter

HSO Converter Tools

HSO Concept Search

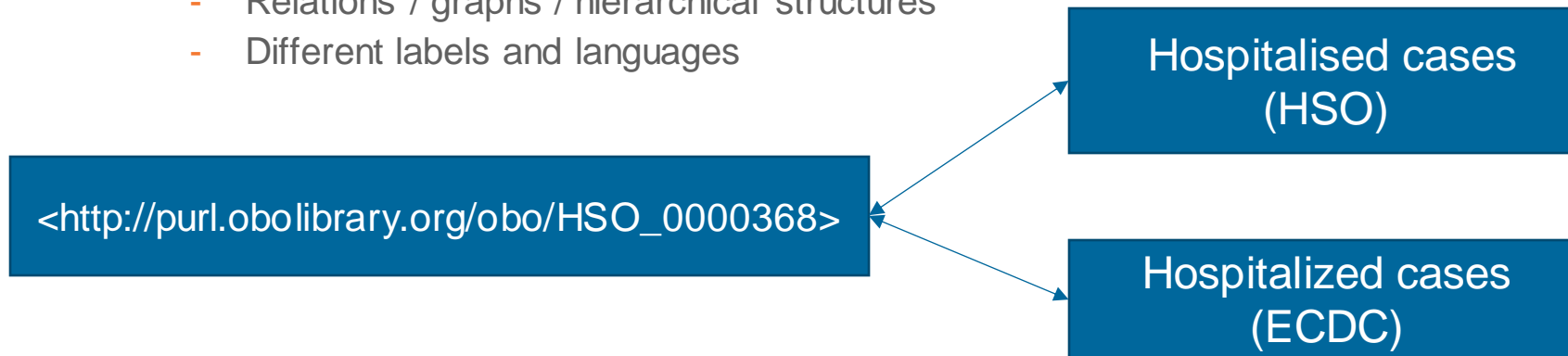EFSA ECDC Surveillance Data Connector

SPARQL Endpoint   Application

European Food Safety Authority
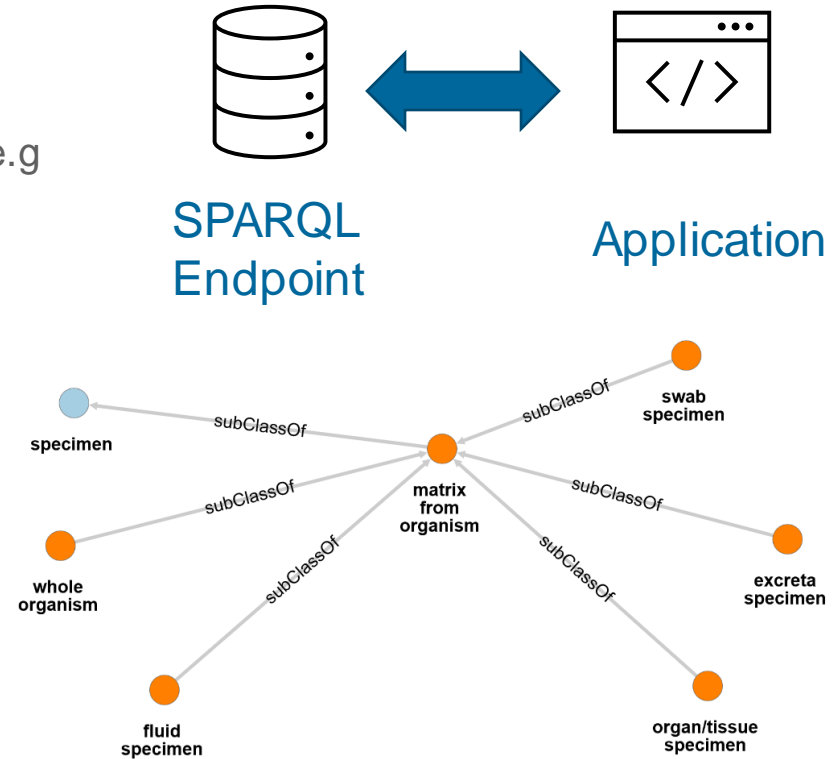
ecdc
EUROPEAN CENTRE FOR
DISEASE PREVENTION
AND CONTROL

Use Case II: Tools and Applications

How?

- Databases can be graph databases(e.g Neo4j) or triplestore databases (e.g. Apache Jena Fuseki)
- Data is stored as graph or as triplet (subject:predicate:object)

SPARQL Endpoint

Application

Benefits:
- Reasoning of data, Semantic search queries get possible e.g. „Suveillance Activity in Germany from 2001 to 2020 of EFSA and ECDC"
- Automatic interpretation and translation of repeated data processes (no humans needed)
- Automatic context based analyses with data sets
- Easier linking to external sources

Use Case II: Context based dashboard

Service that ceates based on the data a specific dashboard:

First picture show: Campylobacter Surveillance in Sweden over 9 Years
→ The service knows that it has to create a line chart

Second picture show: Salmonella in egg surveillance EFSA data just in 2020

→ Service knows that only one year was sampled so it creates a Barchart
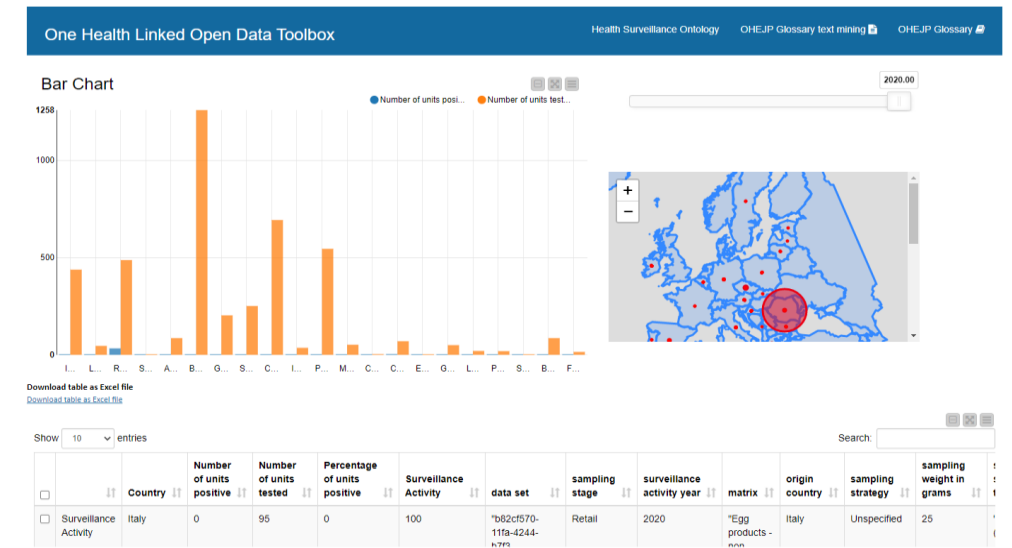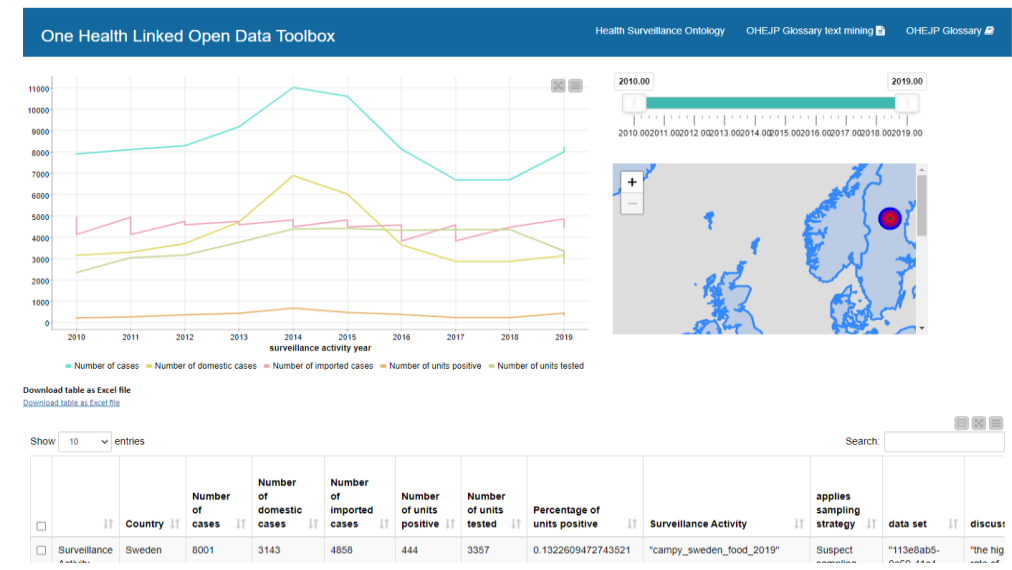
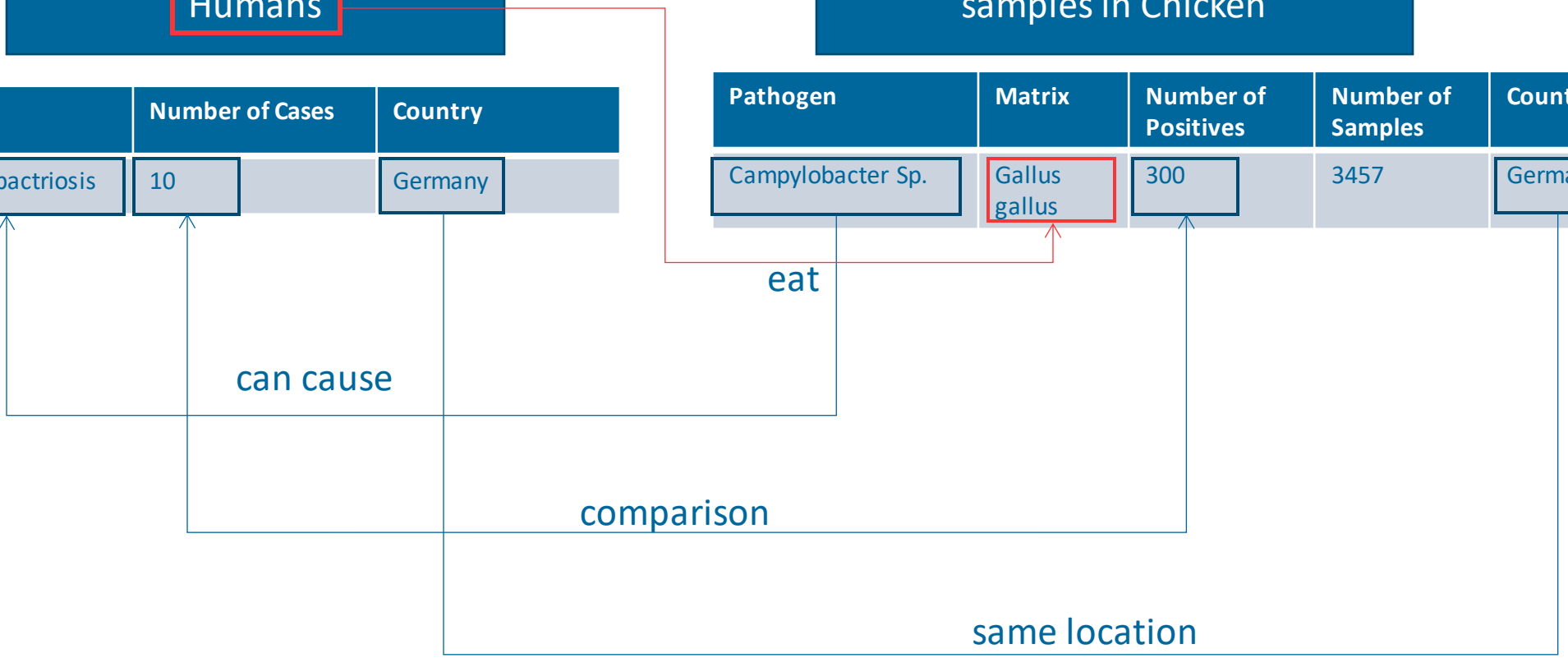**Data set 1 Surv. Data on Campylobacteriosis Cases in Humans**

Data set 2: Surv. Data Campylobacter Sp. positive samples in Chicken

Disease	Number of Cases	Country
campylobactriosis	10	Germany

Pathogen	Matrix	Number of Positives	Number of Samples	Country
Campylobacter Sp.	Gallus gallus	300	3457	Germany

eat

can cause

comparison

same location

EFSA ECDC Surveillance Data Connector

Workflow of the application

Select data set

Select Country of interest
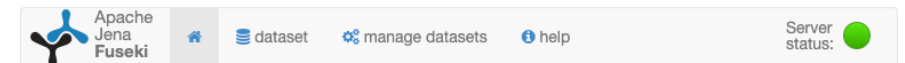
Compare

Use case III: Analysis pipelines in R with Linked Data (Data Science perspective)

- RDF is widely supported by different programming languages
  - R, Python, Java, KNIME....
- Open Source libraries available
- Automatic document generation…

Technical background for realizing infrastructure with Linked Data

.xlsx

.rdf

Ontology/
Schema

SPARQL
Endpoint or
data base

Application/
Algorithm/
Statistical
Analysis
pipeline
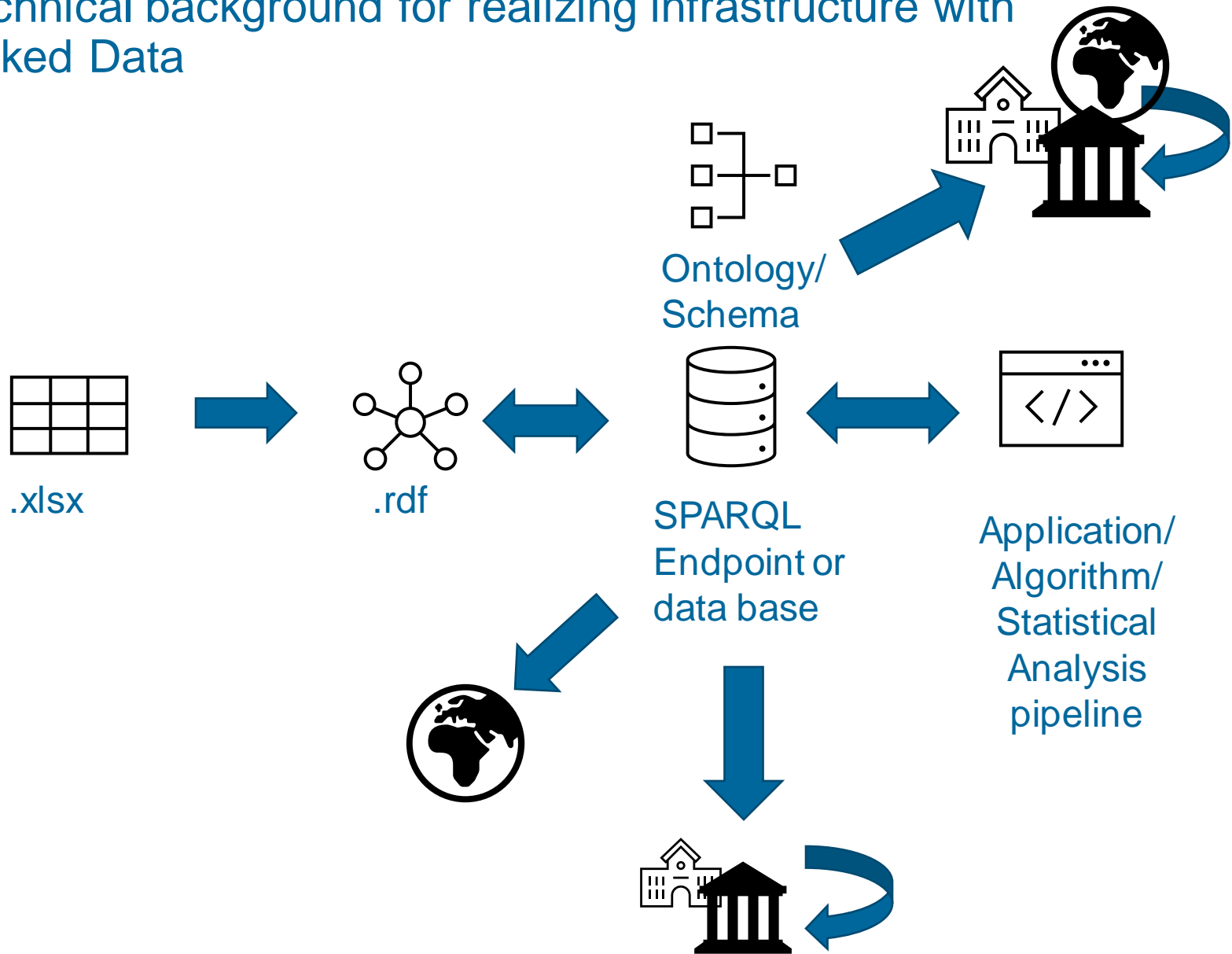
→ logic can also be build on top of existing systems

Tools for Linked Data

Support for Linked Data is everywhere...

Open for Innovation

Some tools to work with Linked Data & Ontologies

....
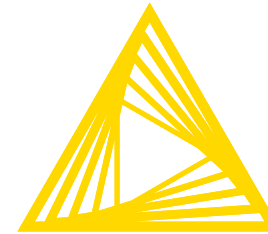
Thank you for your attention!

Your comments & feedback is
highly appreciated:
taras.guenther@bfr.bund.de

Special thanks to:
- Fernanda Dórea
- Matthias Filter
- Estibaliz Lopez-de-Abechuco-Garrido
- Nazareno Scaccia

@OneHealthEJP
/company/h2020-One-Health-EJP
OneHealthEJP.eu