



# Making Linked Data accessible for One Health Surveillance with the “One Health Linked Data Toolbox”

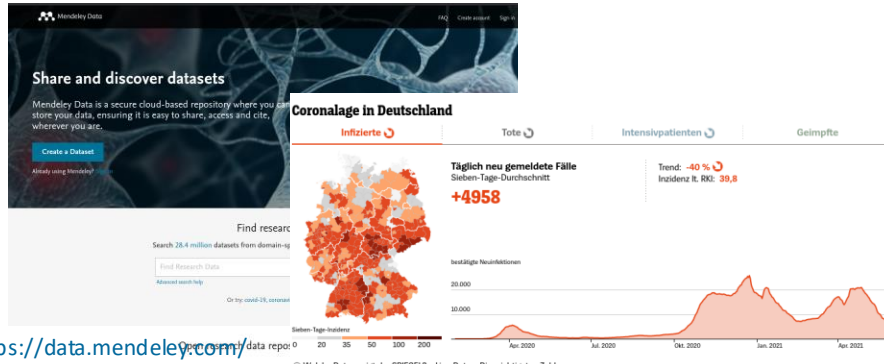
Presenter: Taras Günther

Matthias Filter, Fernanda Dórea

23.06.2021

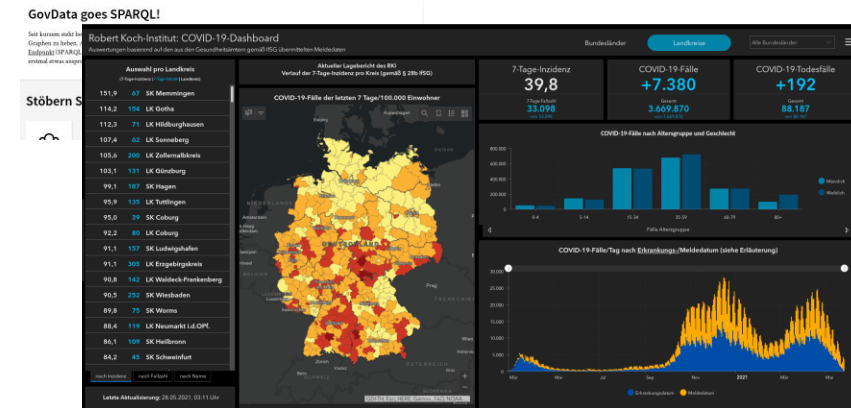
# One Health Surveillance Data - Demands and challenges are rising!

*Fast and interdisciplinary data exchange and integration are key elements for One Health Surveillance*

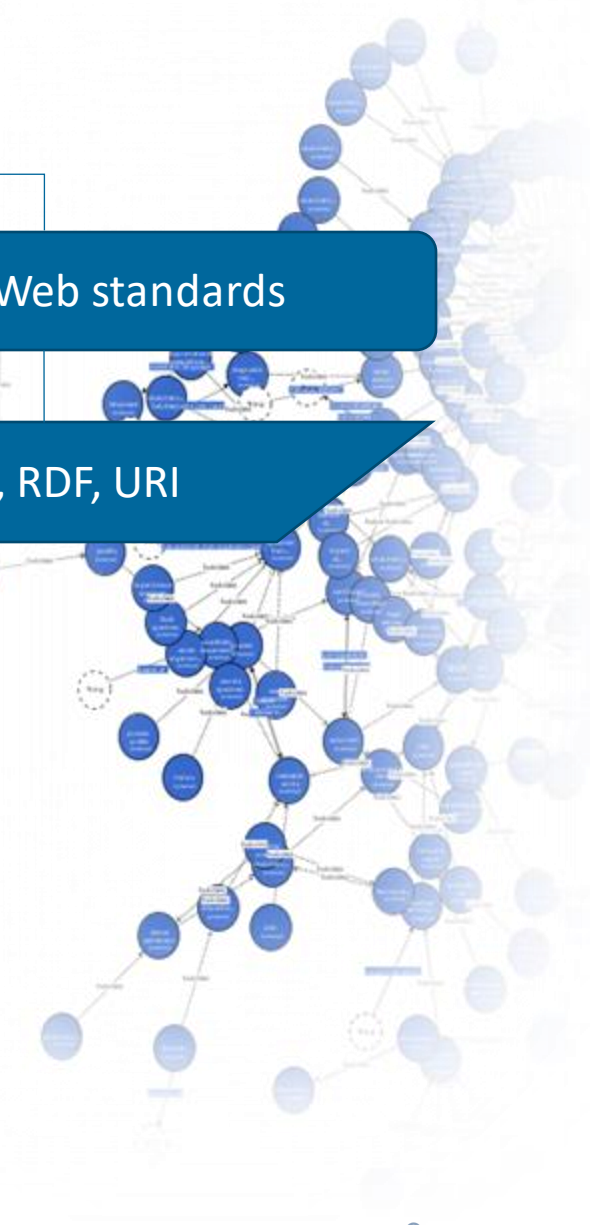
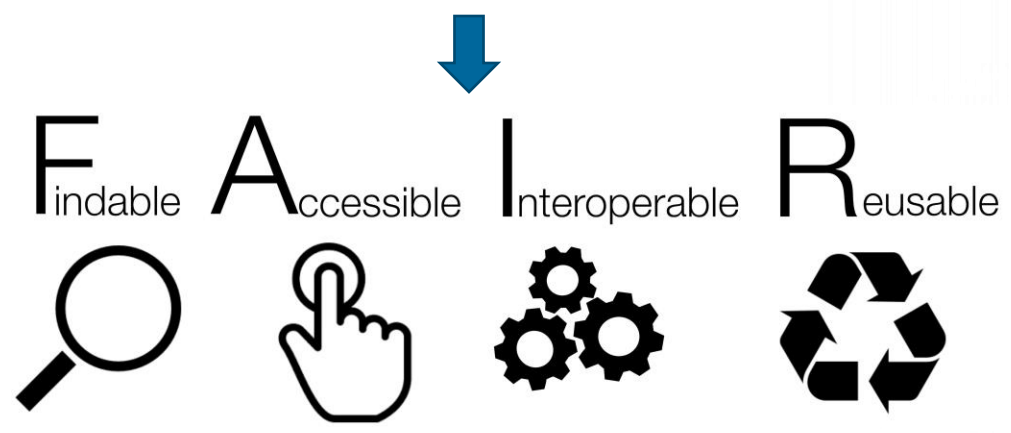
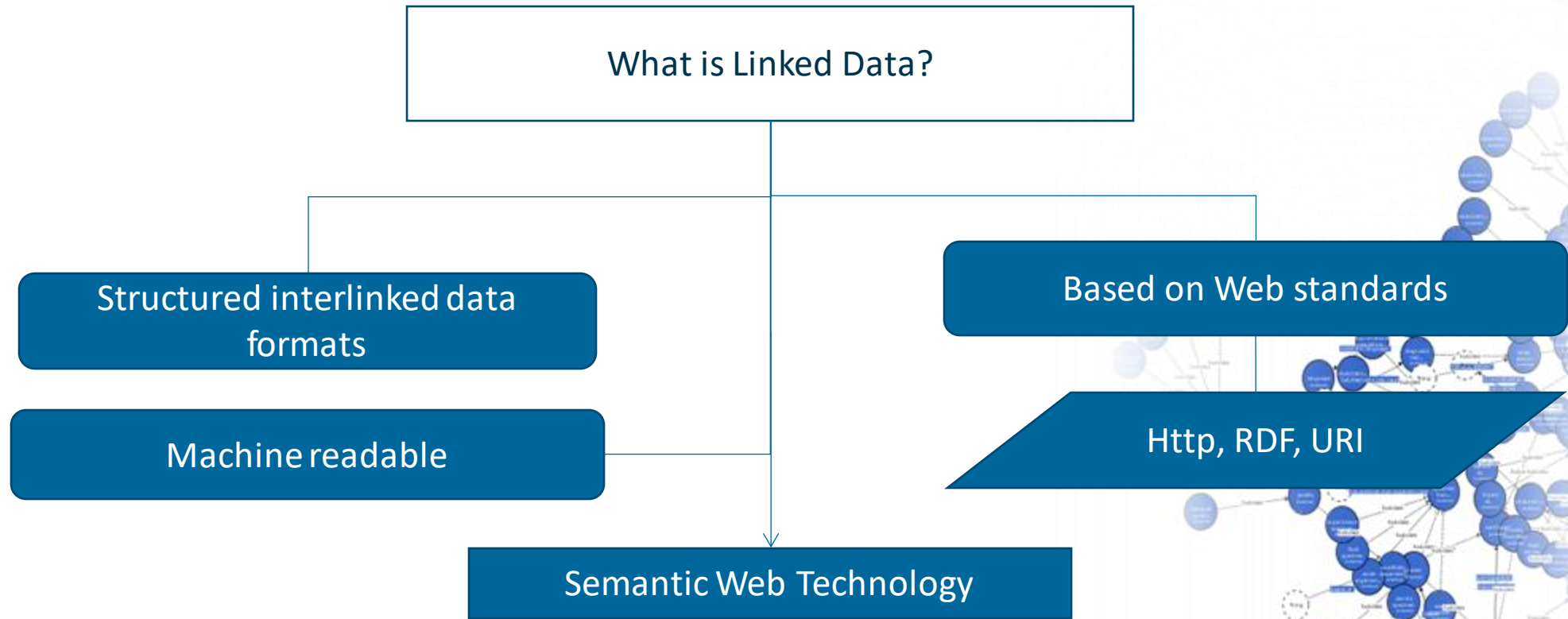


Source : <https://data.mendeley.com/>

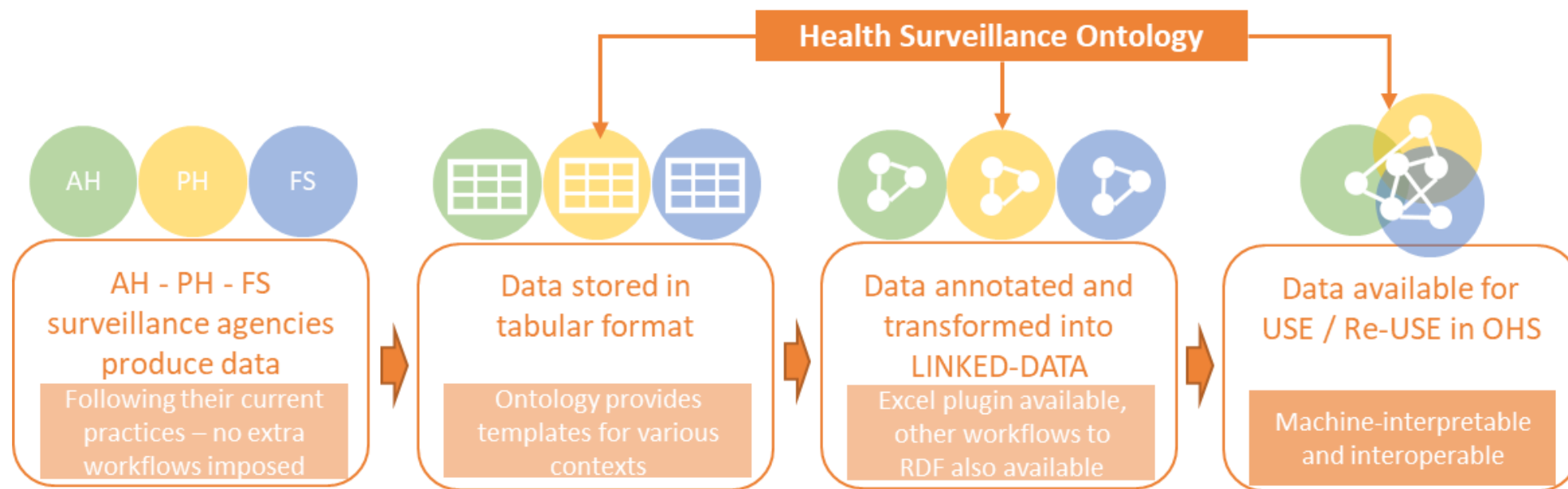
Source : <https://www.spiegel.de/thema/coronavirus>



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



# Semantic Interoperability with Linked Data in One Health



Source: One Health Surveillance Codex, 2021



# The Health Surveillance Ontology (HSO)

The screenshot shows the GitHub repository for the Health Surveillance Ontology (HSO) and its corresponding BioPortal entry. The GitHub interface includes a navigation bar with 'Projects', 'Wiki', 'Security', and 'Insights'. Below this, there are buttons for 'Go to file', 'Add file', and 'Code'. A status bar indicates 'This branch is 2 commits ahead, 9 commits behind SVA-SE:master.' The repository files list includes 'data\_annotation\_examples/campy\_s...', 'docs', 'src/ontology', '.gitignore', 'README.md', 'hso-full.owl', and 'hso.owl'. The BioPortal entry for 'Health Surveillance Ontology' is displayed, showing its last upload date as June 1, 2020. The 'Details' section provides information about the ontology's acronym (HSO), visibility (Public), description, status (Alpha), format (OWL), and contact (Fernanda Dórea, fernanda.dorea@sva.se). A 'Submissions' table lists various versions of the ontology, including 3.1.0 (Released 06/01/2020) and several 2.1.0 versions (Released 03/13/2020 and 09/03/2019). The table also includes columns for 'Released', 'Uploaded', and 'Downloads'.

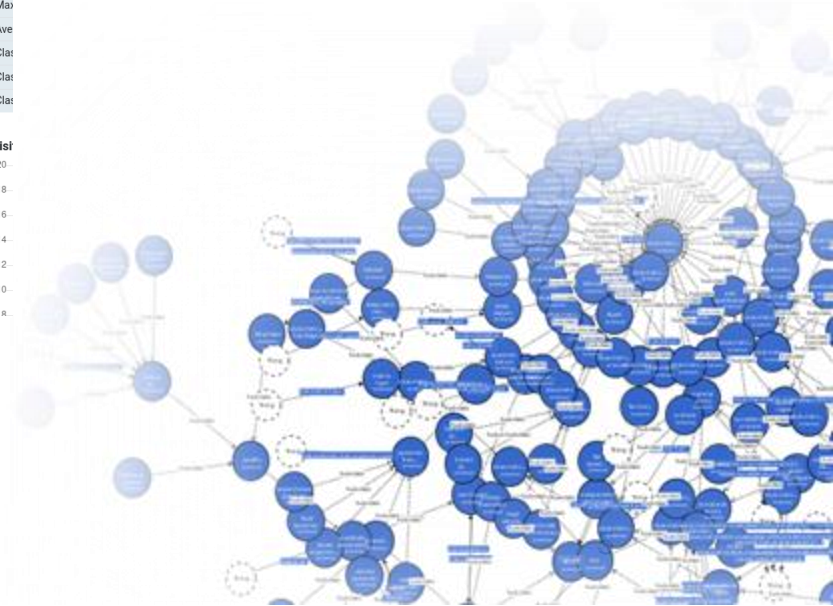
Version	Released	Uploaded	Downloads
3.1.0 (Parsed, Indexed, Metrics, Annotator)	06/01/2020	06/01/2020	<a href="#">OWL</a>   <a href="#">CSV</a>   <a href="#">RDF/XML</a>   <a href="#">Diff</a>
2.1.0 (Archived)	03/13/2020	03/16/2020	<a href="#">OWL</a>   <a href="#">Diff</a>
2.1.0 (Archived)	03/13/2020	03/16/2020	<a href="#">OWL</a>
2.1.0 (Archived)	09/03/2019	03/16/2020	<a href="#">OWL</a>
unknown (Archived)	09/03/2019	11/04/2019	<a href="#">OWL</a>

Interlinked Standard

Interlinked Schema

Semantic Logic

Available here:  
<https://github.com/nandadorea/HSO>  
<https://bioportal.bioontology.org/ontologies/HSO>





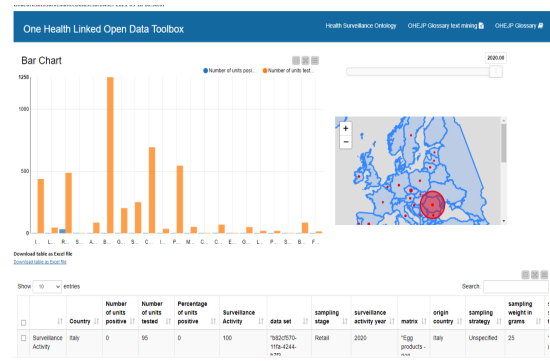
# Health Surveillance Ontology toolbox

One Health Linked Open Data Toolbox

```

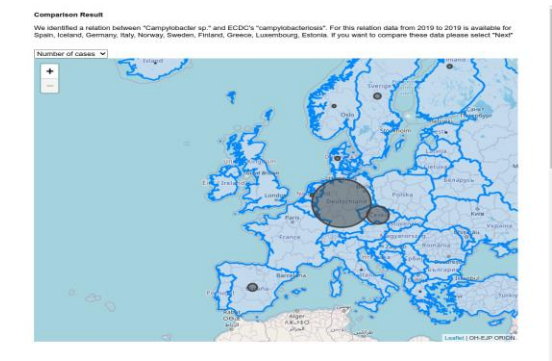
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:skos="http://www.w3.org/2004/02/skos#" xmlns:foaf="http://xmlns.foaf.org/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:geo="http://www.opengis.net/ont/geosparql#" xmlns:time="http://www.w3.org/2006/time#" xmlns:prov="http://www.w3.org/ns/prov#" xmlns:org="http://www.w3.org/ns/org#" xmlns:location="http://www.w3.org/ns/location#" xmlns:unit="http://www.w3.org/ns/unit#" xmlns:quantity="http://www.w3.org/ns/quantity#" xmlns:quantityunit="http://www.w3.org/ns/quantityunit#" xmlns:quantityunitquantity="http://www.w3.org/ns/quantityunitquantity#" xmlns:quantityunitquantityunit="http://www.w3.org/ns/quantityunitquantityunit#" xmlns:quantityunitquantityunitquantity="http://www.w3.org/ns/quantityunitquantityunitquantity#" xmlns:quantityunitquantityunitquantityunit="http://www.w3.org/ns/quantityunitquantityunitquantityunit#" xmlns:quantityunitquantityunitquantityunitquantity="http://www.w3.org/ns/quantityunitquantityunitquantityunitquantity#" xmlns:quantityunitquantityunitquantityunitquantityunit="http://www.w3.org/ns/quantityunitquantityunitquantityunitquantityunit#" xmlns:quantityunitquantityunitquantityunitquantityunitquantity="http://www.w3.org/ns/quantityunitquantityunitquantityunitquantityunitquantity#" xmlns:quantityunitquantityunitquantityunitquantityunitquantityunit="http://www.w3.org/ns/quantityunitquantityunitquantityunitquantityunitquantityunit#" xmlns:quantityunitquantityunitquantityunitquantityunitquantityunitquantity="http://www.w3.org/ns/quantityunitquantityunitquantityunitquantityunitquantityunitquantity#" xmlns:quantityunitquantityunitquantityunitquantityunitquantityunitquantityunit="http://www.w3.org/ns/quantityunitquantityunitquantityunitquantityunitquantityunitquantityunitquantity" />
  
```

HSO Converter Tools



HSO Concept Search

SPARQL Endpoint ↔ Application



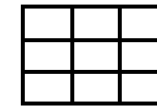
EFSA ECDC Surveillance Data Connector

efsa ↔ ecdc

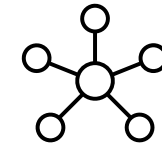
European Food Safety Authority ↔ European Centre for Disease Prevention and Control



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



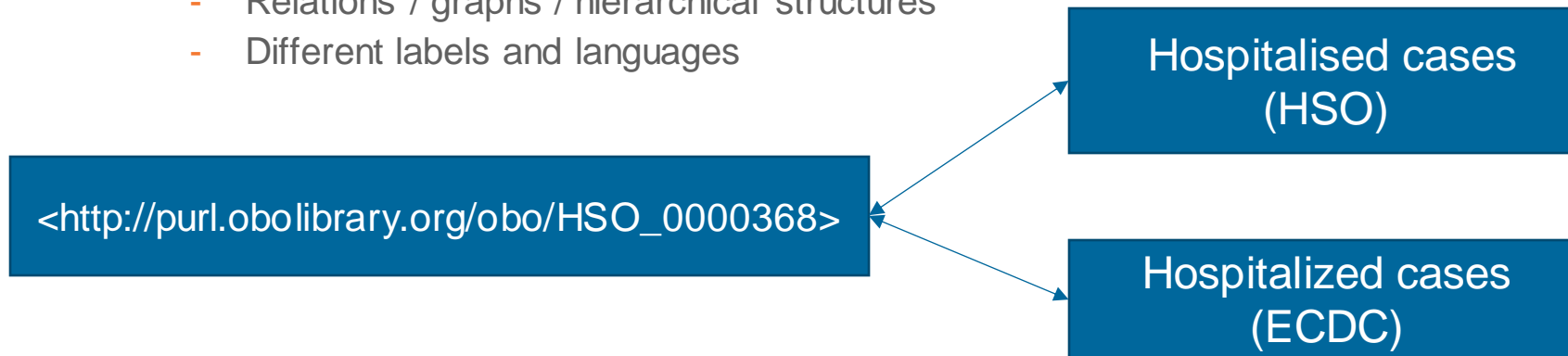
.xlsx



.rdf

How?

- 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



Benefits:

- Standardized data structures attached to corresponding metadata
- Automatic parsing of existing data bases and files



# Use Case I: XLSX to RDF Converter

## One Health Linked Open Data Toolbox

Health Surveillance Ontology

OHEJP Glossary text mining

OHEJP Glossary

### Excel to RDF

This service can convert Excel to HSO-RDF. This can be used to link the data with the Health surveillance ontology. Please upload your Excel file and choose which spreadsheet you want to convert. All entities need to be labeled according to HSO. The service will check if the provided entities are HSO compliant.

Please upload your file here

Select File

<no file selected>

Please select which sheet you want to choose from your file

Read first Excel sheet

Sample data for testing purposes.

Download Link

Back

Upload



## One Health Linked Open Data Toolbox

Health Surveillance Ontology

OHEJP Glossary text mining

OHEJP Glossary

Download file

Download Size

HSO RDF - Preview

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE rdf:RDF [
  <ENTITY base "http://www.
  <rdf:RDF xmlns="http://www.w3.org/2002/07/owl#"
    xmlns:base="http://www.w3.org/2002/07/owl#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:hs0="https://w3id.org/hs0#"
    xmlns:obo="http://purl.obolibrary.org/obo/"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xml="http://www.w3.org/XML/1998/namespace"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:core="http://purl.obolibrary.org/obo/core#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:obo1="http://data.bioontology.org/metadata#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#"
    xmlns:terms="http://purl.org/dc/terms/"
    xmlns:metadata="http://data.bioontology.org/metadata#"
    xmlns:oboInOwl="http://www.geneontology.org/formats/oboInOwl#"
    xmlns:ncbitaxon="http://purl.obolibrary.org/obo/ncbitaxon#"
  ]>
  <obo:HSO_0000001 rdf:about="compy_sweden_slaughterhouse_2019"
    <obo:HSO_0000266 rdf:resource="<obo:HSO_0000271"/>
    <obo:NCIT_C25464 rdf:resource="<obo:GAZ_00002729"/>
    <obo:HSO_0000301 rdf:resource="<obo:NCBITaxon_205"/>
    <obo:HSO_0000302 rdf:resource="<obo:NCBITaxon_9031"/>
    <obo:HSO_0000267 rdf:resource="<obo:HSO_0000277"/>
    <obo:HSO_0000299 rdf:resource="<obo:NCBITaxon_9031"/>
    <obo:HSO_0000300 rdf:resource="<obo:HSO_000020"/>
    <obo:HSO_0000244 rdf:resource="<obo:NCBITaxon_9031"/>
    <obo:HSO_0000243 rdf:resource="<obo:UBERON_0001153"/>
    <obo:HSO_0000213 rdf:datatype="<xsd:integer">2019</obo:
    <obo:HSO_0000260 rdf:datatype="<xsd:integer">1991</obo:
    <obo:HSO_0000295 rdf:datatype="<xsd:string">Monitorin
    <obo:HSO_0000304 rdf:datatype="<xsd:string">Census</o
    <obo:HSO_0000305 rdf:datatype="<xsd:string">slaughter
    <obo:HSO_0000306 rdf:datatype="<xsd:string">off444-1
    <obo:HSO_0000307 rdf:datatype="<xsd:string">slaught
    <obo:HSO_0000600 rdf:datatype="<xsd:string">animal
    <obo:HSO_0000245 rdf:datatype="<xsd:string">organ/tis
    <obo:HSO_0000368 rdf:datatype="<xsd:string">employes
    <obo:HSO_0000310 rdf:datatype="<xsd:string">Detection</obo:
    <obo:HSO_000074 rdf:datatype="<xsd:integer">4363</obo:HSO_000074>
```



Download



oolbox

HSO  
nd in Ontology

sampled from anatomical entity  
specimen from organism type



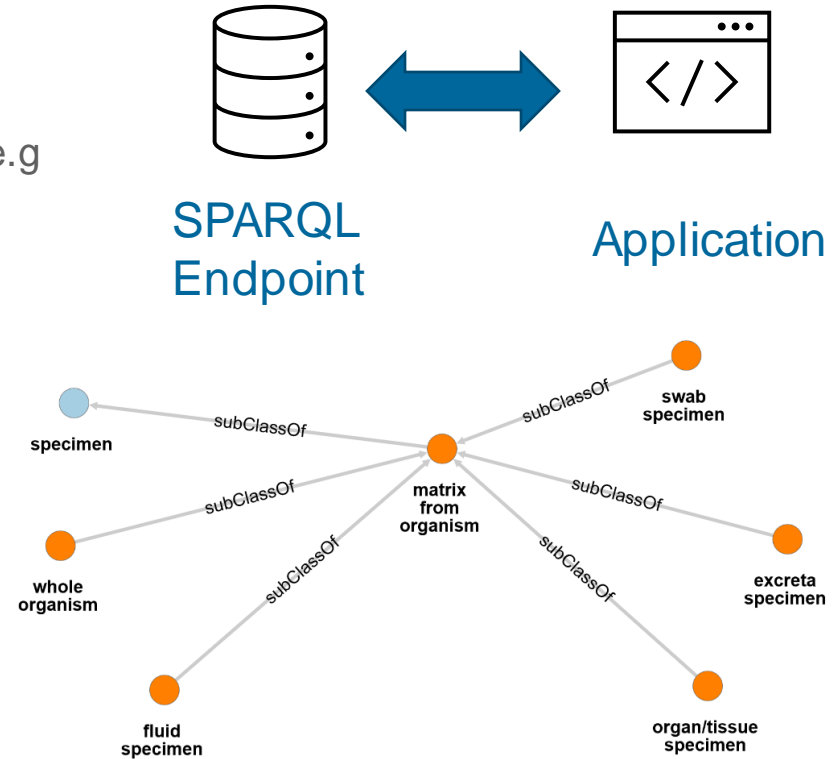




# Use Case II: Tools and Applications

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)



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

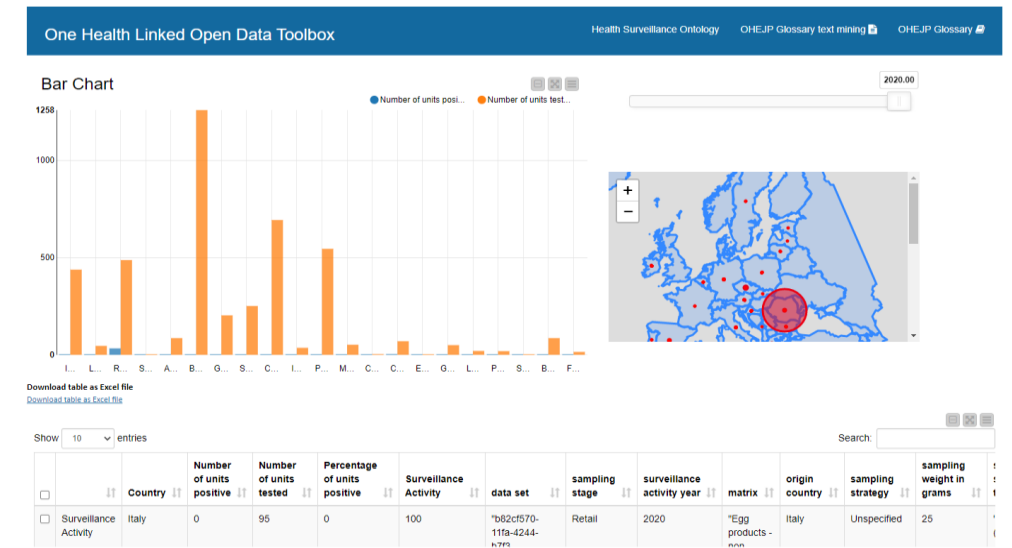
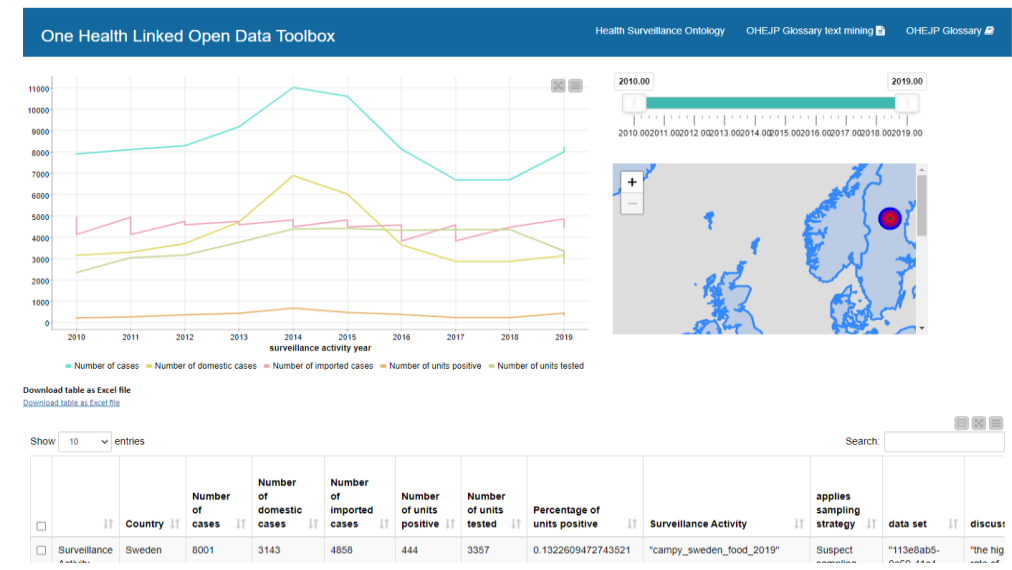


# Use Case II: Context based dashboard

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



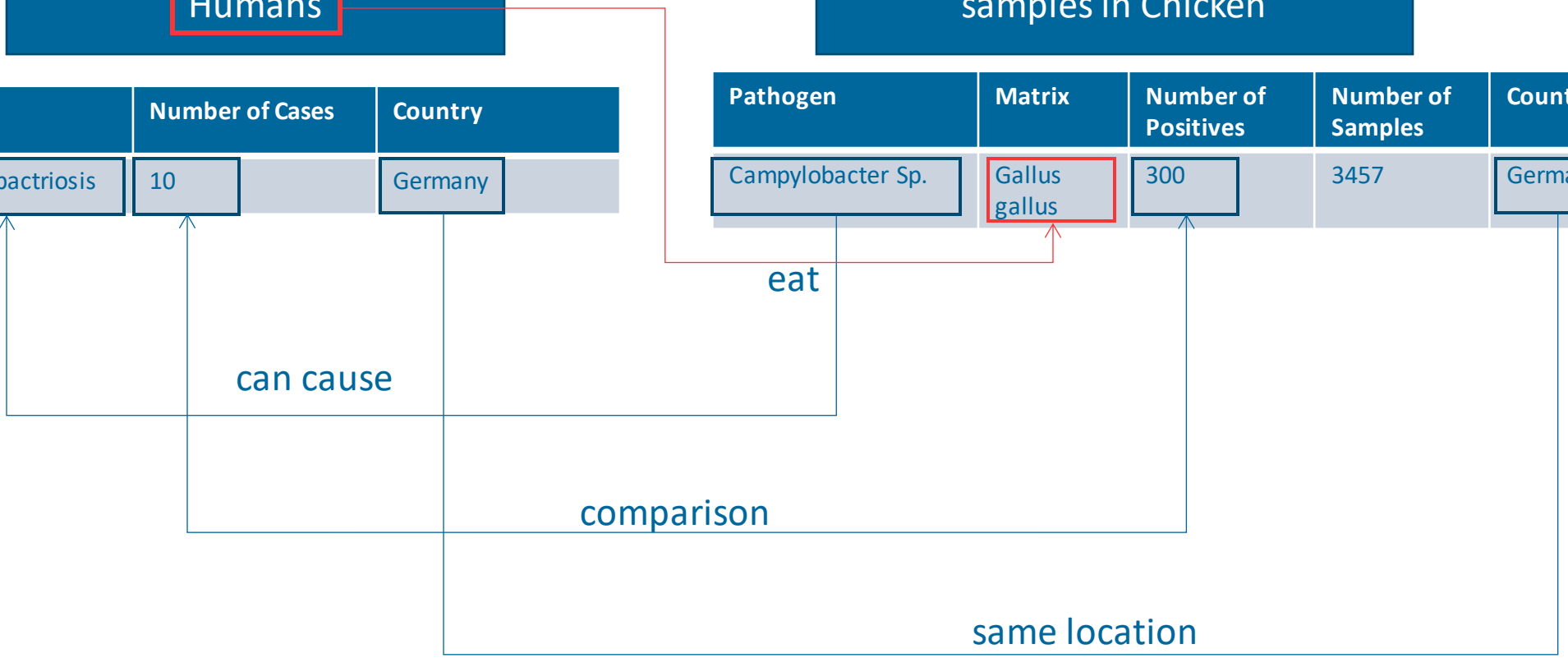


Data set 1 Surv. Data on  
Campylobacteriosis Cases in  
**Humans**

Disease	Number of Cases	Country
campylobacteriosis	10	Germany

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

Pathogen	Matrix	Number of Positives	Number of Samples	Country
Campylobacter Sp.	<b>Gallus gallus</b>	300	3457	Germany



EFSA ECDC Surveillance Data Connector



# Workflow of the application

**Choose data set**

- [http://www.w3.org/2002/07/Campylobacter in RTE milk and milk products, EU](http://www.w3.org/2002/07/Campylobacter%20in%20RTE%20milk%20and%20milk%20products%20EU)
- [http://www.w3.org/2002/07/Campylobacter in raw milk, EU](http://www.w3.org/2002/07/Campylobacter%20in%20raw%20milk%20EU)
- [http://www.w3.org/2002/07/Campylobacter in pigs, EU](http://www.w3.org/2002/07/Campylobacter%20in%20pigs%20EU)
- [http://www.w3.org/2002/07/Campylobacter in RTE fish and fishery products, EU](http://www.w3.org/2002/07/Campylobacter%20in%20RTE%20fish%20and%20fishery%20products%20EU)
- [http://www.w3.org/2002/07/Campylobacter in fresh broiler meat, EU](http://www.w3.org/2002/07/Campylobacter%20in%20fresh%20broiler%20meat%20EU)
- [http://www.w3.org/2002/07/Campylobacter in cattle, EU](http://www.w3.org/2002/07/Campylobacter%20in%20cattle%20EU)
- [http://www.w3.org/2002/07/Campylobacter in RTE other processed food products and prepared dishes, EU](http://www.w3.org/2002/07/Campylobacter%20in%20other%20processed%20food%20products%20and%20prepared%20dishes%20EU)
- [http://www.w3.org/2002/07/Campylobacter in fresh turkey meat, EU](http://www.w3.org/2002/07/Campylobacter%20in%20fresh%20turkey%20meat%20EU)
- [http://www.w3.org/2002/07/Campylobacter in RTE spices and herbs, EU](http://www.w3.org/2002/07/Campylobacter%20in%20RTE%20spices%20and%20herbs%20EU)
- [http://www.w3.org/2002/07/Campylobacter in RTE sprouted seeds, EU](http://www.w3.org/2002/07/Campylobacter%20in%20RTE%20sprouted%20seeds%20EU)
- [http://www.w3.org/2002/07/Campylobacter in cats and dogs, EU](http://www.w3.org/2002/07/Campylobacter%20in%20cats%20and%20dogs%20EU)
- [http://www.w3.org/2002/07/Campylobacter in RTE meat and meat products, EU](http://www.w3.org/2002/07/Campylobacter%20in%20RTE%20meat%20and%20meat%20products%20EU)
- [http://www.w3.org/2002/07/Campylobacter in other foods, EU](http://www.w3.org/2002/07/Campylobacter%20in%20other%20foods%20EU)
- [http://www.w3.org/2002/07/Campylobacter in fresh meat from birds other than broilers and turkeys, EU](http://www.w3.org/2002/07/Campylobacter%20in%20fresh%20meat%20from%20birds%20other%20than%20broilers%20and%20turkeys%20EU)
- [http://www.w3.org/2002/07/Campylobacter in other animals, EU](http://www.w3.org/2002/07/Campylobacter%20in%20other%20animals%20EU)
- [http://www.w3.org/2002/07/Campylobacter in other processed food products and prepared dishes, EU](http://www.w3.org/2002/07/Campylobacter%20in%20other%20processed%20food%20products%20and%20prepared%20dishes%20EU)
- [http://www.w3.org/2002/07/Campylobacter in fruits, vegetables and juices, EU](http://www.w3.org/2002/07/Campylobacter%20in%20fruits%20vegetables%20and%20juices%20EU)
- [http://www.w3.org/2002/07/Campylobacter in broilers, EU](http://www.w3.org/2002/07/Campylobacter%20in%20broilers%20EU)

Cancel Next

Select data set

**Comparison Result**

We identified a relation between "Campylobacter sp." and "campylobacteriosis" in ECDC's data base. For this relation data from 2019 to 2019 is available for Spain. If you want to compare these data please select "Next".

Select the Surveillance Countries

Search:

<input type="checkbox"/>	Austria
<input type="checkbox"/>	Germany
<input type="checkbox"/>	Slovakia
<input type="checkbox"/>	Spain

Showing 1 to 4 of 4 entries

Number of cases:

Back Next

Select Country of interest

Show 10 entries

RowID	Country	Hospitalized cases	Notification rate in domestic cases	Notification rate	Number of cases	Number of deaths	Bacterial infectious disease	data set	surveillance activity year
Row_1	Slovakia	1330							2019
Row_2									

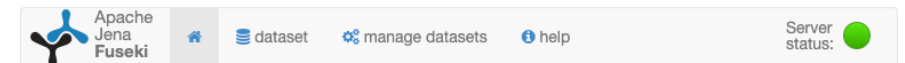
Back Cancel

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...



## Apache Jena Fuseki

Version 3.17.0. Uptime: 42m 34s

### Datasets on this server

There are no datasets on this server yet. [Add one.](#)

Use the following pages to perform actions or tasks on this server:

- [Dataset](#) Run queries and modify datasets hosted by this server.
- [Manage datasets](#) Administer the datasets on this server, including adding datasets, uploading data and performing backups.
- [Help](#) Summary of commands and links to online documentation.

```
##{r}
endpoint <- ("http://localhost:3030/HSO")

##{r}
# Surveillance data of campylobacteriosis ECDC

##{r cars}
Query = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX obo: <http://purl.obolibrary.org/obo/#>

SELECT ?subject ?object
WHERE {
  ?subject <http://purl.obolibrary.org/obo/D0ID_104> ?object
}
"

datasets <- SPARQL::SPARQL(endpoint,Query)

# Query provide all years, number of cases and the corresponding count

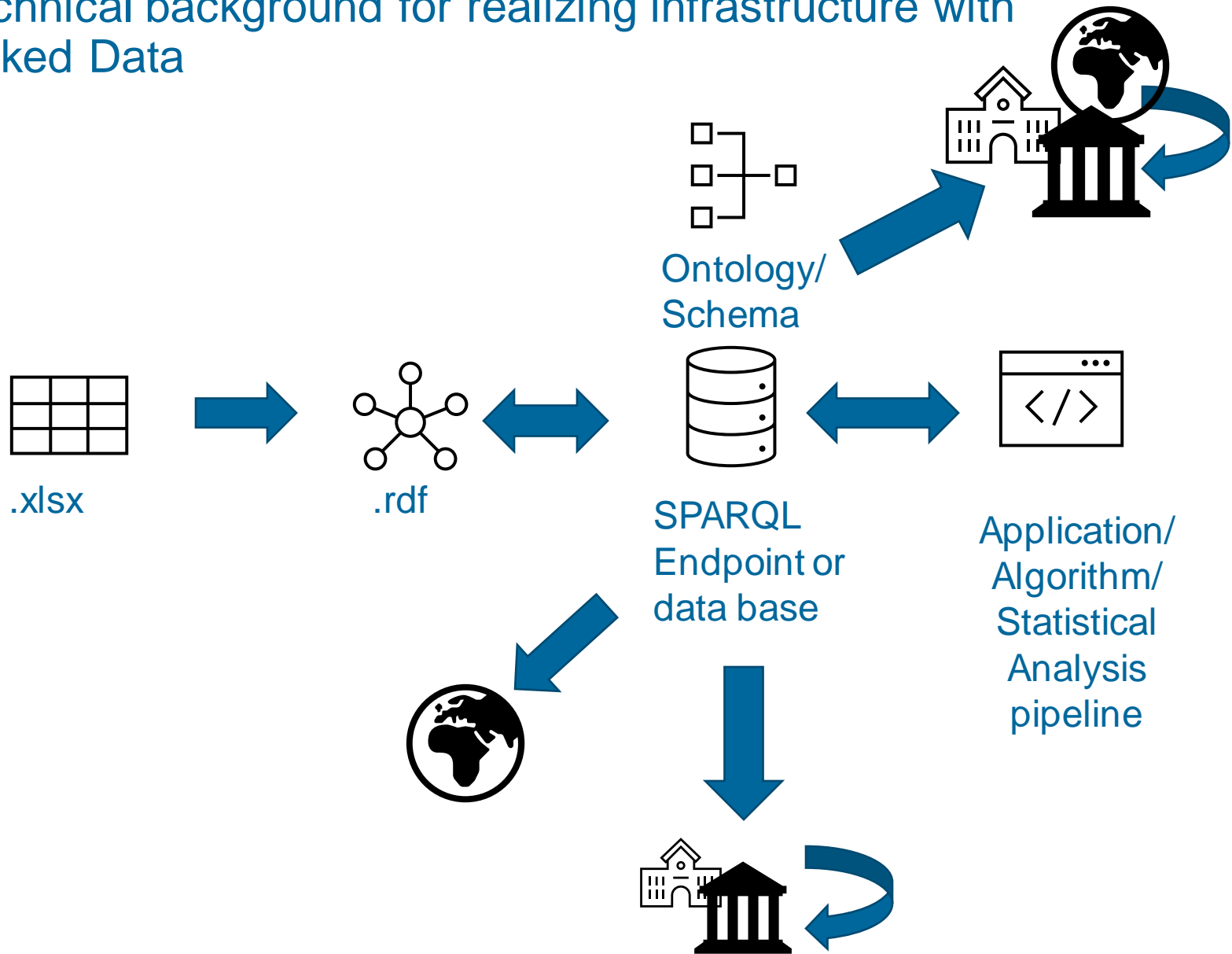
query2 = paste("PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX obo: <http://purl.obolibrary.org/obo/#>

SELECT ?year ?number_of_cases ?surveillance_country
WHERE {", datasets$results$subject, " <http://purl.obolibrary.org/obo/HSO_0000213> ?year.
", datasets$results$subject, " <http://purl.obolibrary.org/obo/HSO_0000321> ?number_of_cases.
", datasets$results$subject, " <http://purl.obolibrary.org/obo/NCIT_C25464> ?Country.
?Country rdfs:label ?surveillance_country.
}", sep=" ") %>% as_tibble()
```

```
##{r}
ggplot(data=Germany_campylobacteriosis, aes(x=year, y=
geom_line()+
geom_point()+
labs(title="Campylobacteriosis Cases in Germany",x
```

Year	Number of Cases
2010	66000
2011	64500
2012	62500
2013	65000
2014	71000
2015	63000
2016	71000
2017	68000

# Technical background for realizing infrastructure with Linked Data



→ logic can also be build on top of existing systems

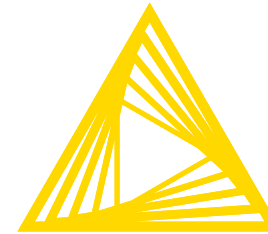


# Tools for Linked Data



Some tools to work with Linked Data & Ontologies

Support for Linked Data is everywhere...



Open for Innovation

# KNIME



....



# 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