



---

**Deliverable JIP1-3.3**  
**Revised OH**  
**Harmonisation**  
**Infrastructure Hub,**  
**including lessons learned**  
**from the OH pilots**  
**Workpackage 3**

Responsible Partner: 38-SVA

Contributing partners: 9-BfR; 10-FLI, 12-  
DTU; 13-SSI; 20-  
APHA; 21-PHE;  
31-RIVM; 37-  
FOHM



## GENERAL INFORMATION

European Joint Programme full title	Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards
European Joint Programme acronym	One Health EJP
Funding	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.
Grant Agreement	Grant agreement n° 773830
Start Date	01/01/2018
Duration	60 Months

## DOCUMENT MANAGEMENT

Deliverable	JIP1-3.3 - Revised OH Harmonisation Infrastructure Hub, including lessons learned from the OH pilots
WP and Task	JIP1 – WP3, Task T-3.4
Leader	Fernanda Dórea (SVA)
Other contributors	Taras Günther (BfR), Estibaliz Garrido (BfR), Mia Holmberg (SVA), Cecilia Jernberg (FoHM), Marika Hjertkvist (FoHM), Matthias Filter (BfR), Alessandro Foddai (DTU), Johanne Ellis-Iversen (DTU), Charlotte Cook (APHA), Joanna Lawes (APHA), Lesley Larkin (PHE); Ingrid Friesema (RIVM); Maria-Eleni Filippitz (sciensano); Mickael Cargnel (sciensano); Geraldine Boseret (sciensano); Karin Lagesen (NVI)
Due month of the deliverable	M42
Actual submission month	M42
Type <i>R: Document, report</i> <i>DEC: Websites, patent filings, videos, etc.</i> <i>OTHER</i>	R
Dissemination level <i>PU: Public</i> <i>CO: confidential, only for members of the consortium (including the Commission Services)</i>	PU



# TOWARDS AN INTEROPERABLE ONE HEALTH DATA INFRASTRUCTURE

---

WP3 in ORION was planned to tackle infrastructural resources related to data harmonization in One-Health surveillance (OHS). During the requirement analysis carried in year 1 (and reported in the deliverable JIP1-3.1 ([available here](#)), it became clear that to support collaborative data analysis across health sectors this WP should focus on solutions to document context and preserve meaning of surveillance data across health sectors – that is, solutions to promote **semantic interoperability**.

**Semantic interoperability** is concerned with ensuring the integrity and *meaning* of the data across systems<sup>1</sup>. This is particularly important in One-Health surveillance in order to allow *data reuse across sectors*, and even reuse of data for research and knowledge discovery<sup>2</sup>.

This WP therefore focused on two overarching goals:

- 1) Build a knowledge model for one-health surveillance that allows computers to understand and reason with current data terminologies in the same way that humans do, maximizing the benefit to cost ratio of the effort put into producing surveillance data;
- 2) Improve usability of data inside the institutions who own and/or use the data, as well as the potential for reuse by external stakeholders and for research and discovery.

This work was organized in three parallel working groups (*knowledge modeling, technical development and surveillance practice*).

The focus on semantic interoperability has allowed this WP to add value to the integratory activities already performed by EFSA and ECDC, and to develop tools for data interoperability that can be implemented in any scenario of data governance, that is, respecting current data sharing policies. None of the outcomes of this WP rely on, or require changes in current data sharing practices. Rather, they support and promote adoption of the FAIR principles of findability, accessibility, interoperability and reusability (<https://www.force11.org/group/fairgroup/fairprinciples>).

This report is a detailed account of the outcomes produced by this WP to contribute to data FAIRness in general, and data interoperability in particular. It is structured in the following way:

- **Introduction and background**. a brief overview on the underlying concepts of linked data, data interoperability and data FAIRness.
- **Section 1: OHS interoperability tools** – a description of data interoperability tools produced in this WP.
- **Section 2: Application of OHS interoperability tools to publish FAIR data** - a description of results from testing interoperability tools in practice. Examples of surveillance data production, reporting and sharing are described.
- **Section 3: opportunities to connect data in OHS**– reflections and recommendations on the opportunities created by connected OHS data.
- **Section 4: Lessons learned from ORION OH pilots** - a brief summary of the data interoperability issues and solutions created in selected national pilots.
- **Conclusions and lessons learned: considerations for a FAIR-ER OHS future**. We present reflections on the lessons learned from a general OHS perspective in the European Union.

---

<sup>1</sup> Definition of Interoperability. In: HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations. 2nd edition. 2010. p. 190

<sup>2</sup> Cardoso L, Marins F, Portela F, Santos M, Abelha A, Machado J. The next generation of interoperability agents in healthcare. Int J Environ Res Public Health. 2014



## INTRODUCTION AND BACKGROUND

---

### ***Connected data is smarter data –***

*Even in the face of intelligent applications, disconnected data result in dumb behaviour.*

*Dean Allemang and Jim Hendler<sup>3</sup>*

The construction of smarter applications does not depend on smarter data, but on making sure that “the right data can get to the right place, so that smart applications can do their work” (Allemang and Hendler, 2011)<sup>3</sup>. Surveillance is intrinsically an activity that depends on the connection of many disparate sources of data. In OHS, this complexity is amplified by the need for data across many different health sectors and knowledge domains. Connecting data from different disciplines, and most importantly, ensuring that the context of data is preserved is a great challenge. To add complexity still, knowledge is constantly evolving. If we are able to connect the right applications, we can develop an “ecosystem of solutions” for OHS, but the ecosystem needs to stay stable and synchronized along time.

The ***linked data model*** proposes that the development of applications that are able to connect to each other and produce consistent, future-proof results, relies on the ***separation of data from knowledge***. In this model, data should describe entities in the world, rather than store information (which is an interpretation of the data based on some previous knowledge). Information is generated by applications that query data based on a layer of knowledge. This “knowledge layer” models the connections between the entities represented in the data, and it is this layer that mediates the connections between applications. Any translations or assumptions needed to connect disparate data can be modelled in this layer, and as knowledge evolves, the model can be evolved, without losing the connection between the data sources. This can be better exemplified by an example.

Consider a diagnostic test which has a numerical result: for instance an ELISA reading. An absorbance value can only be interpreted if we know the cut-off value above which the result is to be interpreted as positive. If we test an animal serum sample with a given ELISA test, and record the result as “positive”, we are recording *information*. If we provide data to our applications only as “positive”, we won’t know if this value is compatible with other test protocols used in different laboratories and other countries, and changes in the protocol may mean that future results cannot be compared to old results. The context was not preserved (false conclusions can be taken when comparing these data to other sources), and the information is not future-proof. In this example, the actual entities in the process are the sample used, the diagnostic assay used, and the absorbance value in the result. If we record all of these data explicitly, then the translation of this information into a positive or negative result based on a cut-off value can be made by a human – but can also be delegated to applications. And if we are connecting various applications, they can “communicate” and translate values among them, by stating explicitly the diagnostic protocol and cut-off values used. “Applications” in this case can be automated systems (machines), people across different institutions, or it can even be ourselves looking at the data some time in the future.

Every time we record data based on an assumed knowledge, those data are only useful if the user interprets them with the same knowledge, and working under the same assumptions. *Within* health sectors, this can be sometimes enforced through data standardization. *Across domains*, as in the case of OHS, this is not possible or even desirable. When we reuse data collected within animal health, public health and food safety surveillance in order to generate OHS information, we want data to have been structured based on and preserving their original context of production. But reusing those data into a OHS context requires some knowledge about how these sectors connect. Some of this knowledge is

---

<sup>3</sup> Allemang, D., Hendler, J., 2011. *Semantic Web for the Working Ontologist: effective modeling in RDFS and OWL*. Morgan Kaufmann. 384p. Paperback ISBN: 9780123859655. eBook ISBN: 9780123859662



straightforward to humans, but requires proper knowledge modeling to empower smart data applications.

Allemang and Hendler (2011) state that modeling, explicitly, the knowledge needed to convert data into information enables:

- 1) Communication among people, as it makes the knowledge and assumptions used to process the data explicit;
- 2) Discovery of patterns in the data and predictions;
- 3) Mediation among multiple viewpoints; and
- 4) Development of a common collection of knowledge, which can be developed and evolved as a common effort.

Points 1) and 3) are particularly important across domains – contexts are intrinsically different between the health sectors involved, and this impacts both how they record and how they use their data. Knowledge models allow explicit exploration of the differences, while creating a representation for the commonalities or the possible links and translations that can be applied to use data across sectors.

### **Semantic modeling – communicate, explain/predict, mediate**

The separation of data and knowledge under the linked data model relies on three main necessary pillars:

- 1) Creating a knowledge model for the desired application
- 2) Linking data to the knowledge model
- 3) Implementing applications which use the knowledge model to query the data and create information.

There are a number of languages available for creating a knowledge model, with different levels of expressivity:

- **RDF – the Resource Description Framework**, is the basic framework used for the construction of interlinked applications in the “Semantic Web”. RDF provides a mechanism for allowing anyone to make a basic statement about anything, and for layering statements into a single model. It has been a recommended language of the W3C (the World Wide Web Consortium, an international community that develops open standards to ensure the long-term growth of the Web) since 1999.
- Since 2004 the W3C recommends the use of **RDFS, the RDF Schema language**, which adds to RDF by allowing expressivity of the basic notions of commonality and variability familiar from object languages, namely classes, subclasses and properties.
- **OWL - Web Ontology Language**, adds *logic* to semantic modeling, allowing expression of detailed constraints between classes, entities and properties.

RDF allows us to declare concepts – or classes – and the properties connecting them. We can for instance declare the concept, or *class* of a “person”, and model one of the *relationships* between persons to be “is biological father of”. RDF can be used to declare that individual people are instances of the class “person” and to connect two people by this relationship. Figure 1 shows how we would normally see this in an Excel spreadsheet (A) and how this would be declared in RDF (B). If we also want to use logic to model constraints about this relationship, for instance that a person can only have one biological father, we can model this logic in OWL (Figure 1-C). The logic modelled in OWL can be used to check data for error, infer relationships between instances based on logic (if *A is father of B*, then *B has father A*), or discover knowledge based on the logic, for instance to discover who are the grandparents of a person, using the logic that grandparents are parents of their parents.

Resources described using RDF can have “human friendly” labels such as shown in Figure 1, but they are made unique by a **Uniform Resource Identifier (URI)**. In the example shown, all concepts, relationships and individual persons listed would have their own URI. The concept “person” will have a unique URI, and the specific instances of a person - “Ada Lovelace”, “Lord Byron” – will also have their



own unique URIs. All things “said” about “Ada Lovelace” anywhere in the web can be mapped to that unique person using this URI. Similarly, any model can be built to say additional things about a “person”. As long as the properties in these new models point to the same URI used for “person” in the model shown in Figure 1, these models will complement each other. Reasoners can be used to check if the logic between these models is consistent, or discover knowledge in data based on these models combined logic.

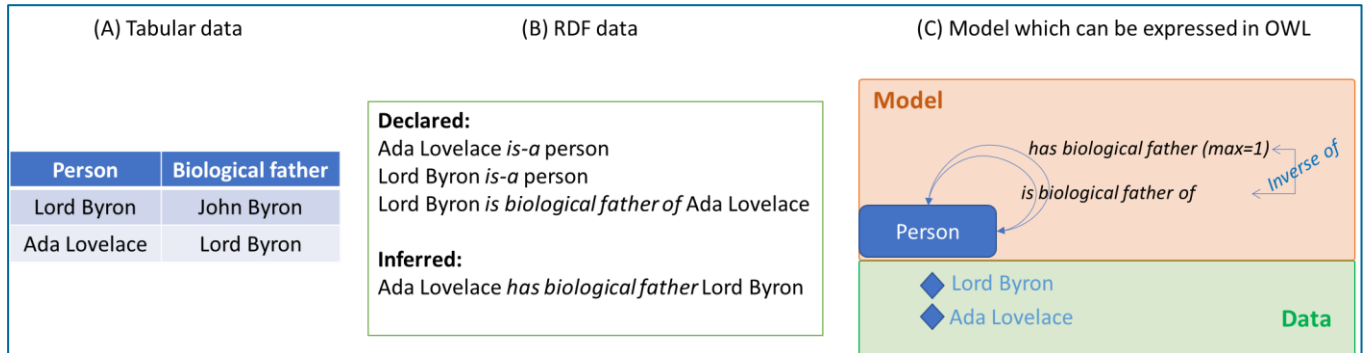


Figure 1. Example data in tabular format (A), RDF format (B), and a simple model of how the concepts in the data are connected, which can be modelled in OWL (C).

In the simple ELISA example given before, if we store the cut-off information in the knowledge model, we could retrieve positive *results* with the query: "all observations which absorbance reading was equal or greater than a given cut-off value". This is in contrast to declaring in the data itself, that an observation is positive or negative. With the linked data concept an end user is able to choose the constraints applied to the data at the time of retrieval.

This has an obvious benefit for the mediation between contexts of data creation and data usage.

In section 3, we discuss the view of OHS as a knowledge model that connects data from different domains, without imposing any changes to the data structure within individual health sectors.



## SECTION 1: OHS INTEROPERABILITY TOOLS

As stated in the Introduction, the promotion of semantic interoperability through the use of knowledge models requires three main steps:

- 1) the development of a knowledge model covering the knowledge areas among which interoperability is to be achieved
- 2) annotation of existing data using this knowledge model
- 3) tools to consume these semantically annotated data

Figure 2 depicts this process in a One-Health surveillance (OHS) scenario, where the knowledge model is provided by a dedicated **ontology**.

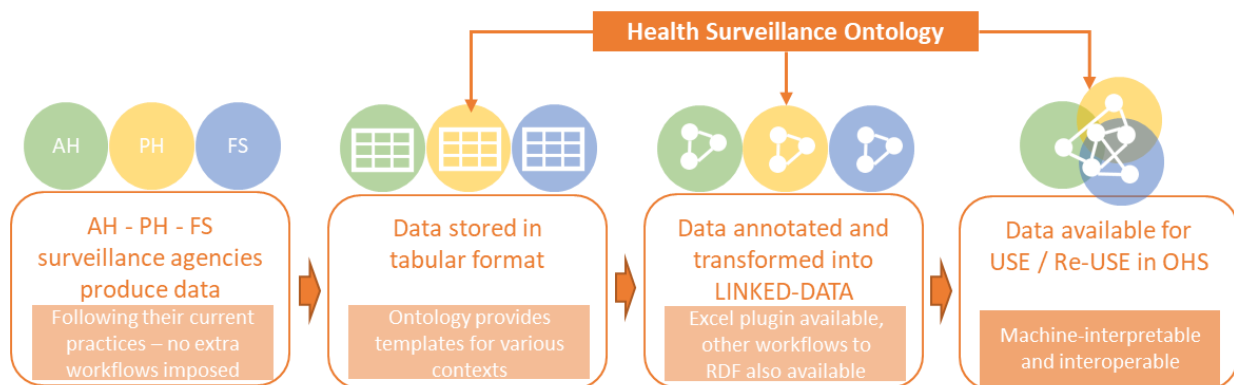


Figure 2. Vision of semantic interoperability in One Health supported by the Health Surveillance Ontology.

In the pursuit of supporting this vision, ORION has developed the **Health Surveillance Ontology** and provided proof-of concept workflows for the other steps of the semantic interoperability continuum. These are detailed in this section.

### Health Surveillance Ontology (HSO)

*“An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them”<sup>4</sup>.*

In order to support the need for a human- and machine-readable knowledge model in One Health surveillance, ORION has developed a Health Surveillance Ontology reusing knowledge from existing ontologies, as well as reusing terminologies already commonly used in practice, such as those adopted by EFSA and ECDC. Identification of concepts and their specialization was informed by data examples from the various “OH pilots” carried out in ORION. Figure 3 depicts some of the main concepts in HSO, and how they can be used to annotate specific surveillance results. The goal is for surveillance methodology to always be associated with its results, and for numbers (such as number of positive samples) to be fully annotated with the context in which they were produced (such as surveillance objective, sampling design, etc). Figure 3 also shows some of the EFSA catalogues that were reused and made fully compatible with HSO.

<sup>4</sup> Natalya F. Noy and Deborah L. McGuinness. 2001. Ontology Development 101: A Guide to Creating Your First Ontology. Available at [http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf)



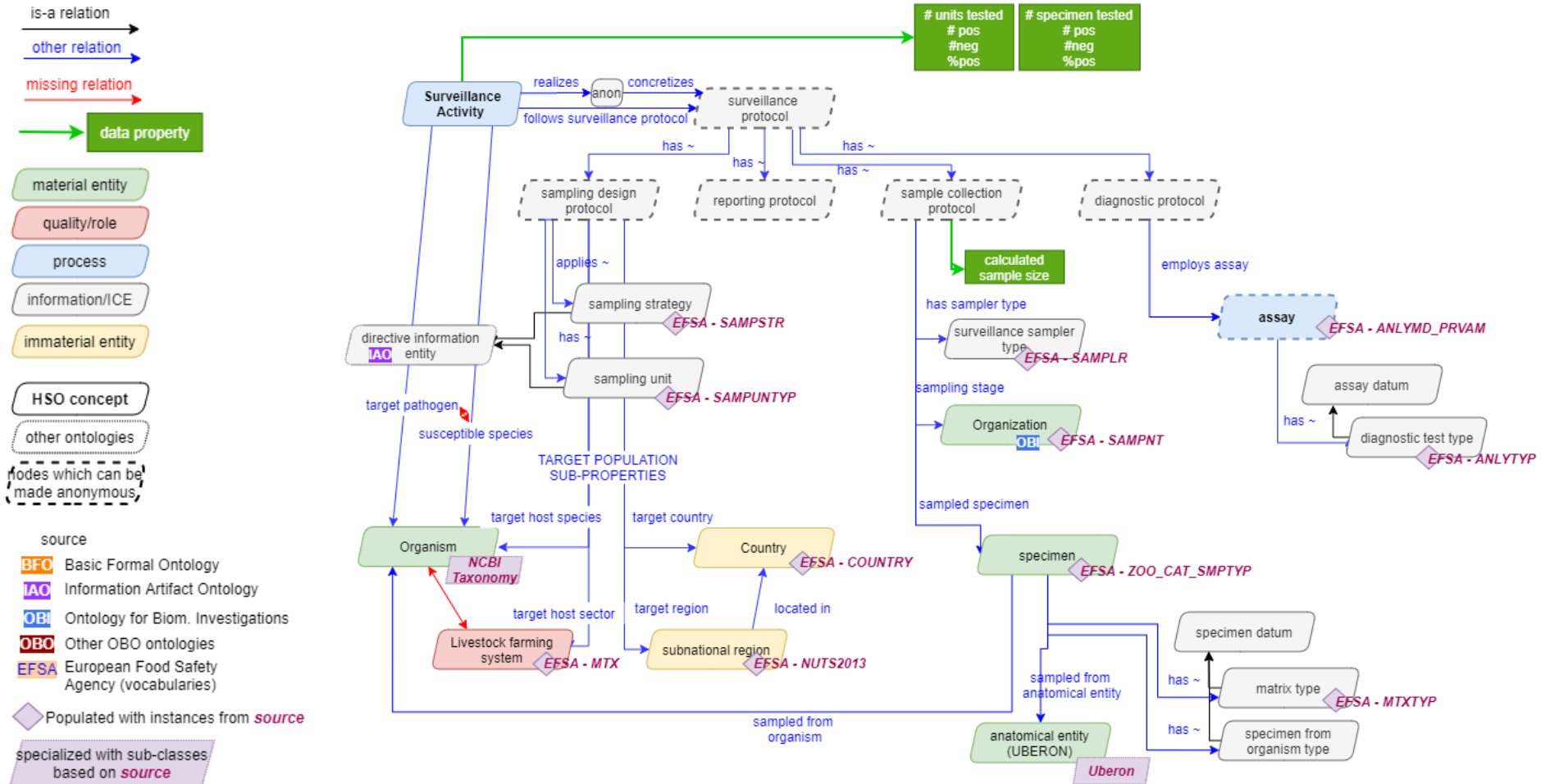


Figure 3: Some of the key classes in the Health Surveillance Ontology, and a schematic view of how they can be used to annotate surveillance results (strong-green rectangular boxes).





Full and up to date documentation of the ontology can always be found in HSO's globally unique and eternally persistent identifier: <https://w3id.org/hso>

This link is a permanent address provided by the W3C Permanent Identifier Community Group<sup>5</sup>. It ensures that the ontology will always have the same address, even if its hosting location is changed (for instance if in the future the ontology is maintained by other groups or projects).

This link is also subjected to content negotiation: humans accessing this link via browser will be referred to a page listing all ontology documentation and additional resources, such training materials. Software agents pointed to the same address will find the machine-readable codes for the knowledge model (written using the Web Ontology Language - OWL).

The ontology is also browsable on Bioportal: <https://bioportal.bioontology.org/ontologies/HSO>

The development of the ontology was described in deliverable JIP1-3.2 ([available here](#)). Since then, HSO has been accepted as a member of the Open Biological and Biomedical Ontology (OBO) Foundry (<http://www.obofoundry.org/>).

*The Open Biological and Biomedical Ontologies Foundry is a group of people dedicated to build and maintain ontologies related to the life sciences. The OBO Foundry establishes a set of principles for ontology development for creating a suite of interoperable reference ontologies in the biomedical domain<sup>6</sup>.*

HSO's persistent uniform resource locator (PURL) under OBO is:

<http://purl.obolibrary.org/obo/hso.owl>

Acceptance into OBO attests that HSO fulfils the community's principles<sup>7</sup>. We highlight the compliance to openness, unique identifiers for concepts, and interoperability (while orthogonal, that is, avoiding duplication) to other ontologies in the foundry (Figure 4).

More information about ontologies and their use and other supporting materials are available at <http://datadrivensurveillance.org/ontology/>.

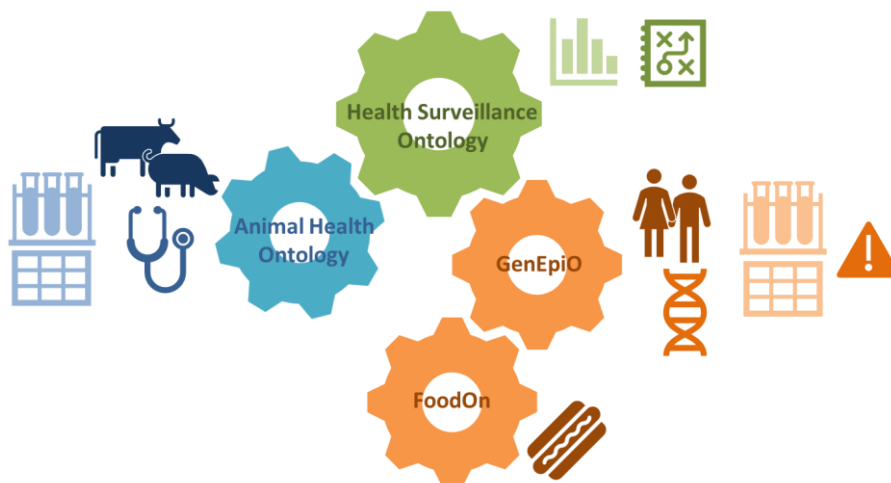


Figure 4: HSO is interoperable with other ontologies in the Open Biological and Biomedical Ontology (OBO) Foundry, such as the Genetic Epidemiology Ontology and the Food Ontology.

<sup>5</sup> <https://www.w3.org/community/perma-id/>

<sup>6</sup> [https://en.wikipedia.org/wiki/OBO\\_Foundry](https://en.wikipedia.org/wiki/OBO_Foundry)

<sup>7</sup> <http://www.obofoundry.org/principles/fp-000-summary.html>



## Tools to annotate data using HSO

As HSO is, on itself, FAIR, it provides the required data annotation model for any data source to attend the FAIR principle of interoperability I2 (“To be interoperable: I2 (meta)data use vocabularies that follow FAIR principles”).

The data annotation process is highly dependent on the data management tools used at each institution. In ORION we have identified that epidemiologists most frequently manipulate and exchange data in flat formats, such as “.xls”, “.xlsx” or “.csv” formats. For that reason, we have developed tools for semantic annotation of data in Excel, and subsequent exportation of the data in Resource Description Framework (RDF) format (see Introduction for information about RDF).

### 1.1. The ExcelRDF plug-in

The Excel plug-in is free and open source, and it was developed in conjunction with the RealEstateCore project in Sweden. Codes for developers, as well as a guide to install the plug-in for users are available at <https://github.com/RealEstateCore/ExcelRDF>. ExcelRDF is a Visual Studio Tools for Office (VSTO) plugin.

The plugin reproduces, semantically, the way people understand tabular data (Figure 5). It allows a user to associate each column of a tabular dataset to a specific concept in the ontology, and all content of that column is then understood as subclasses or instances of the class assigned to the column.

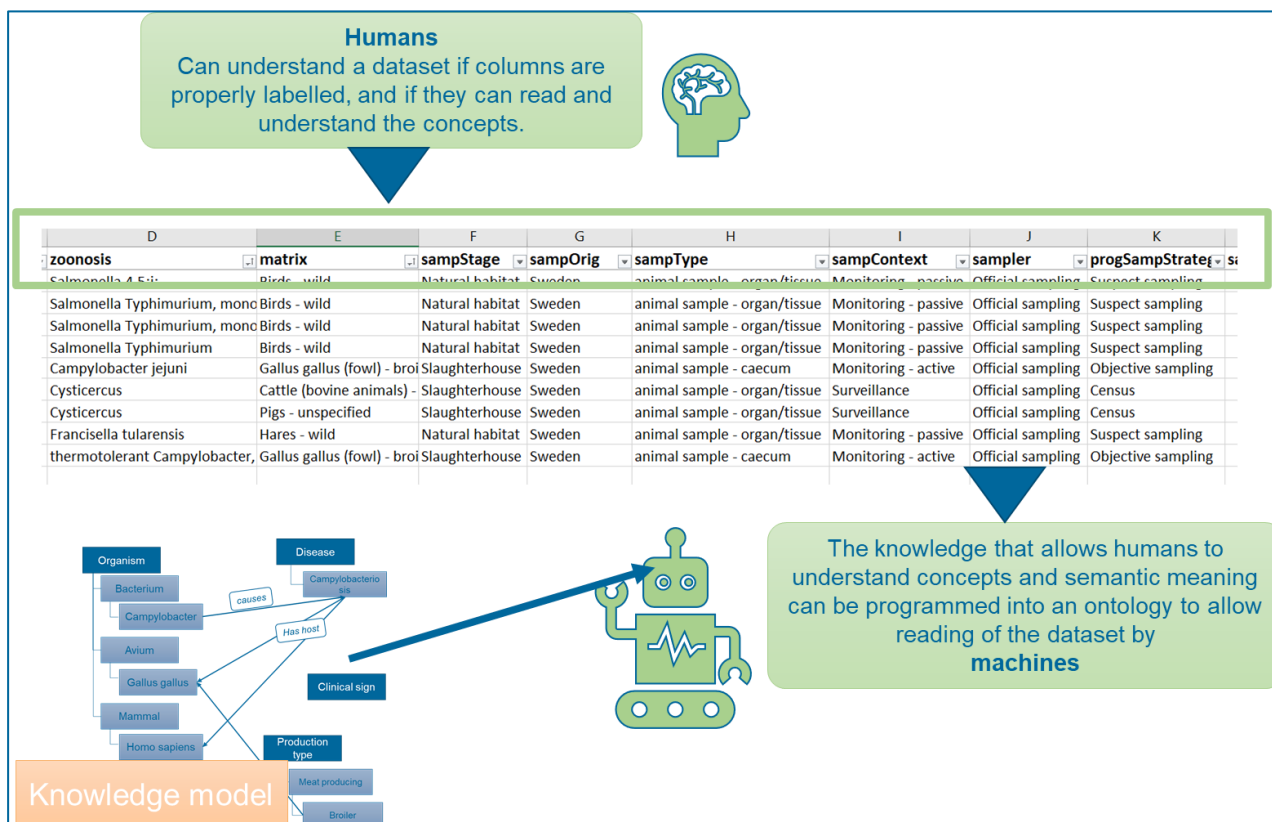


Figure 5. Processing of tabular data by humans and by machines, and the role of a knowledge model.

Once installed, the plugin appears in the “DATA” menu of Microsoft Excel, as shown in Figure 6. Users can use “Load ontology” to capture concepts directly from the ontology, and populate a blank Excel spreadsheet with columns which represent concepts form the ontology. As the ontology is quite extensive, we have created an empty template which users can download directly from the ontology homepage (<http://datadrivensurveillance.org/health-surveillance-ontology-hso/>).



This meeting is part of the European Joint Programme One Health EJP. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

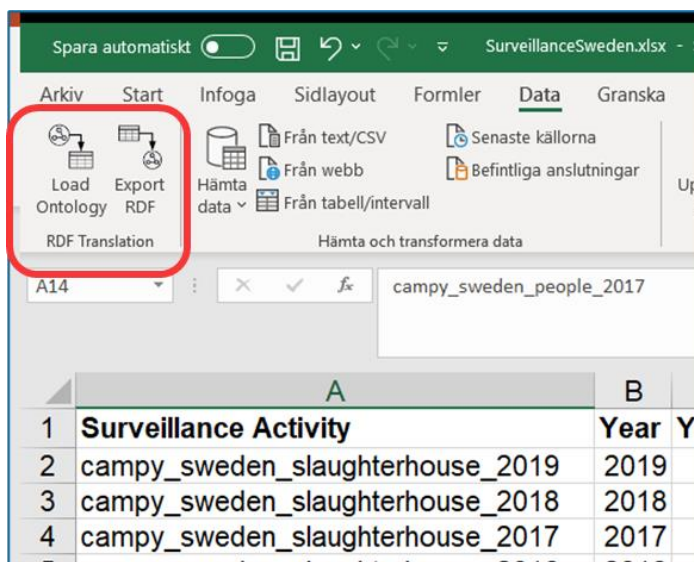


Figure 6. ExcelRDF plugin controls under the “Data” menu in Microsoft Excel.

The plugin assigns classes to columns using annotations. Users can also edit those annotations directly, rather than using “Load Ontology”. In the example shown in Figure 7, the annotations are declaring that the first column contains unique instances of “Surveillance Activity” (Unique Resource Identifier (URI): [http://purl.obolibrary.org/obo/HSO\\_0000001](http://purl.obolibrary.org/obo/HSO_0000001)); and that the second column expects data in the format of integer, which will then be assigned to the Surveillance Activity through the data property “surveillance activity year” (URI [http://purl.obolibrary.org/obo/hso#HSO\\_0000213](http://purl.obolibrary.org/obo/hso#HSO_0000213)).

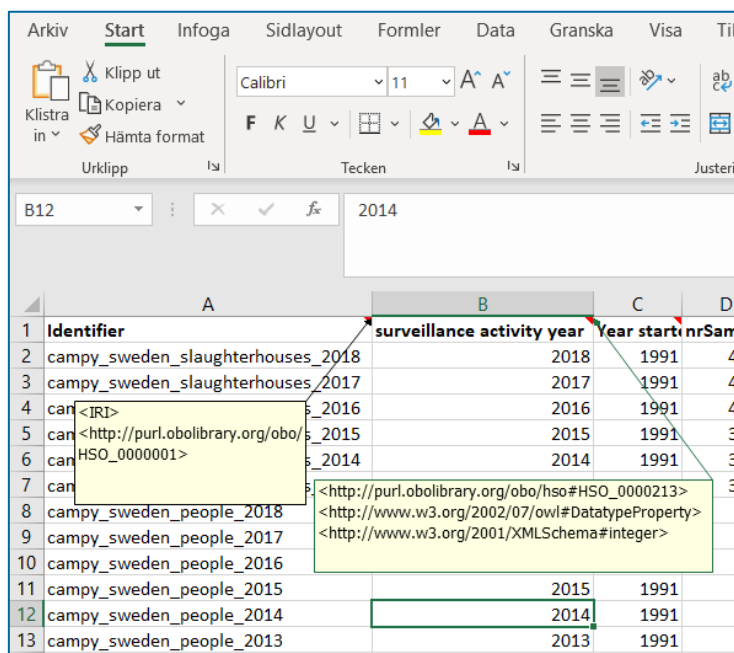


Figure 7. Annotations used by the ExcelRDF plugin to associate columns to ontology concepts.

Data is filled and manipulated in the Excel spreadsheet as usual. When a user clicks on the “Export RDF” button (Figure 6), an RDF version of the data is generated.



## 1.2. RDF and Excel converter Web applications

Web based tools were also developed- Their purpose is similar to the ExcelRDF plugin, but the conversion process is fully automated and needs less user interactions. The tool are developed with the open source Software KNIME ([www.knime.com](http://www.knime.com)). This workflow is hosted at the BfR's KNIME Web Server infrastructure, and therefore does not require any installation process by the user. They can be accessed via this link:

<https://foodrisklabs.bfr.bund.de/one-health-linked-data-toolbox/>.

### 1.2.1. Excel to RDF converter web application

The Excel to RDF workflow is separated into three steps: (i) input step, (ii) control step, and (iii) output step.

In the input step the user can provide an Excel file with the same structure described for the ExcelRDF plugin via an upload button. Further the user can decide which sheet within the uploaded Excel file is to be converted (Figure 8). An example file is provided for download under the upload button.

Figure 8. Input step of the Excel to RDF Web application. In this section the user can provide an Excel file to be converted into the RDF format.

After uploading the Excel file the user can move to the next step. The workflow then checks all the column names in the selected spreadsheet against concepts found in the ontology. In the control step (Figure 9) the user can check the coverage of the column names with the concepts available in HSO. Matches are indicated with a check mark and non-matches as red crosses. In case of concepts not found in the ontology, the user can review the file, or leave as it is, which will cause columns not identified as specific ontology concepts to be imported as annotations (textual descriptions). No information is lost in the conversion. It is also possible to suggest new concepts or labels for HSO via the HSO GitHub Page (<https://github.com/SVA-SE/HSO/issues>).



## One Health Linked Open Data Toolbox

Health Surveillance Ontology

OHEJP Glossary text mining

OHEJP Glossary

The service found columns which are currently not covered within HSO

Column Name	Class found in Ontology
surveillance activity	✓
year started	✓
has surveillance objective	✓
target country	✓
target pathogen	✓
susceptible species	✓
has surveillance purpose	✓
has surveillance context	✓
target host species	✓
target host sector	✓
applies sampling strategy	✓
has sampling unit	✓
has sampler type	✓

Figure 9. Control step of the Excel to RDF Web application. In this section the user can control the coverage of the columns within the HSO concepts.

Upon clicking next, the data table will be converted into RDF by matching the content with HSO. Column names will be automatically transformed into the corresponding HSO concept and the content of each column into the corresponding sub classes and instances of HSO.

In the output step (Figure 10) of the workflow the user gets the link to download the converted RDF and can see the content of the file. Instances in the dataset are assigned a uniquely generated URI which can be used to query the data.

## One Health Linked Open Data Toolbox

**Download file**  
[Download RDF](#)

**HSO RDF - Preview**

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY base 'http://www.w3.org/2002/07/owl#'
  <!ENTITY dc 'http://purl.org/dc
<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:hso="https://w3id.org/hso#"
  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/uberon/core#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:obo1="http://data.bioontology.org/metadata/obo/"
  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:H5O_0000001 rdf:about="campy_sweden_slaughterhouse_2019">
    <obo:H5O_0000266 rdf:resource="&obo:H5O_0000271"/>
    <obo:NCIT_C25464 rdf:resource="&obo:GAZ_00002729"/>
    <obo:H5O_0000301 rdf:resource="&obo:NCBITaxon_205"/>
    <obo:H5O_0000302 rdf:resource="&obo:NCBITaxon_9031"/>
    <obo:H5O_0000267 rdf:resource="&obo:H5O_0000277"/>
  </pre>

```

Figure 10. Output step of the Excel to RDF Web application – an RDF version of the uploaded data.



### 1.2.2. RDF to Excel converter web application

The workflow to convert RDF back to an Excel file works in a similar way and has the same user interface, but has only two sections.

In the input step (Figure 11) the user can upload an RDF file. An example RDF file is available.

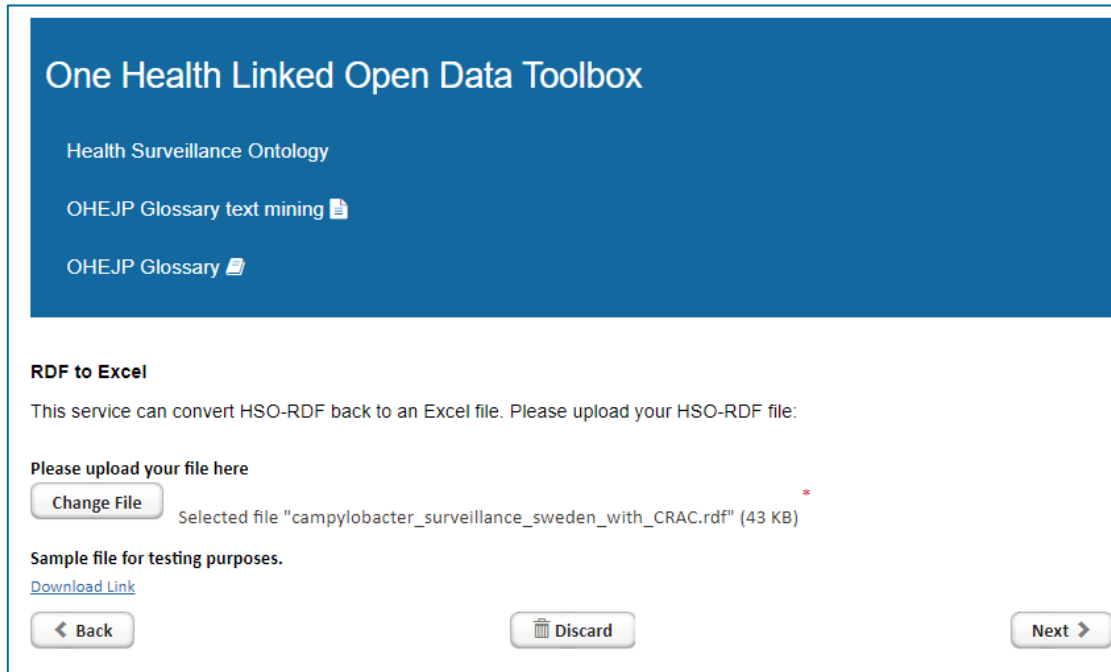


Figure 11. Input step of the RDF to Excel Web application.

Once the file is uploaded and the user presses next, the content of the RDF file will be automatically matched with the HSO. The user will be directed to the second section, the output section (Figure 12), where the content is shown in tabular format. This application is built with KNIME and deployed in the BfR's KNIME Web Server as well.

CRAC 1.1 Motivation / cause	CRAC 1.2. Requirement analysis	CRAC 1.3. Surveillance objective and constraints	CRAC 2.1. Framework design	CRAC 2.4. Sampling plan	CRAC 2.4.1. Sampled population	CRAC 2.4.2. Sampling schema
"Thermophilic Campylobacter species (spp.) are the most common cause of human bacterial gastroenteritis in many countries. A majority of infections are caused by C. jejuni followed by C. coli."	"Detection of Campylobacter spp. in food is not notifiable. From 2018 and onwards, food business operators at slaughterhouses are obliged to sample neck skins of broilers for analyses of"	"The aim was to understand the frequency and distribution of clustering isolates from humans and chicken meat in a non-outbreak situation."	"No official surveillance programme exists for Campylobacter spp. in food. National and local authorities may perform sampling as a part of"	"Since 1 January 2018, slaughterhouses are obliged to sample neck skins from poultry carcasses for Campylobacter analyses using a culture-based method (ISO"	"The programme covers more than 99% of the broilers slaughtered in Sweden."	"PH: In 2017–2019, the analysis of human isolates collected during high season has been preceded by analysis of Campylobacter in chicken meat from retail. AH: Since 2017, all"

Figure 12. Output step of the RDF to Excel Web application.





### 1.3. HSO enrichment web application

The European Food Safety Agency (EFSA) provides a public harmonized terminology for scientific research (<https://doi.org/10.5281/zenodo.344473>), which is used to collect and analyse data in a coherent way among European Member states. This terminology was considered a key resource for the development of the ontology.

This workflow is designed to automatically transform metadata and concepts from specific EFSA catalogues into the Health Surveillance Ontology. This tool was used within the project to introduce and annotate new concepts into HSO semi-automatically based on EFSA data catalogues.

The tool is based on a “point and click” interface and consists of three main steps: i) the catalogue selection, ii) the concept selection and iii) the result download.

In the “catalogue selection step” (Figure 13) the user can decide between different EFSA Catalogues (URL) by clicking on the selection button in the table. Only one catalogue can be selected. After selecting the catalogue and pressing the next button on the bottom right the user gets directed to the “concept selection step”.

Health Surveillance Ontology tool

OHEJP Glossary text mining | OHEJP Glossary

**EFSA catalogues**  
Choose EFSA catalogue

Search:

<input checked="" type="radio"/>		Sheet	
<input type="radio"/>	■	Summary	
<input type="radio"/>	■	ACTION	
<input type="radio"/>	■	AMRPROG	
<input type="radio"/>	■	ANLYMD_AMRAM	
<input type="radio"/>	■	ANLYMD_PRIVAM	
<input type="radio"/>	■	ANLYREFMD	
<input type="radio"/>	■	ANLYTYP	
<input type="radio"/>	■	COUNTRY	
<input type="radio"/>	■	COUNTRY_EUSRREPCOUNTRY	
<input type="radio"/>	■	LANG	
<input type="radio"/>	■	LEGREF	
<input type="radio"/>	■	MDACC	
<input type="radio"/>	■	MTXTYP	
<input type="radio"/>	■	NUTS_NUTS2016	
<input type="radio"/>	■	PARAM_AG	

Cancel | Next

Figure 13. Screenshot of the “catalogue selection step”

In the “concept selection step” the user can select one concept from the HSO ontology (displayed on the left table) and the entities of the EFSA catalogue (provided in the right table). The selected entities (from the right table) and the corresponding HSO concepts can then be connected and introduced from the tool into HSO. This connection is activated by the user through clicking next (bottom right corner). The algorithm translates the entities from the table into an OWL structure, which is then appended to HSO. The IDs of the EFSA data catalogue entities are incorporated into HSO as individual instances during the translation process.





**Health Surveillance Ontology tool** OHEJP Glossary text mining | OHEJP Glossary

---

**HSO Ontology**  
Please select HSO class

Clear Sorting

Search:

<input type="radio"/> @rdf:about	text
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000001">http://purl.obolibrary.org/obo/BFO_0000001</a>	entity
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000002">http://purl.obolibrary.org/obo/BFO_0000002</a>	continuant
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000003">http://purl.obolibrary.org/obo/BFO_0000003</a>	occurrent
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000004">http://purl.obolibrary.org/obo/BFO_0000004</a>	independent continuant
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000006">http://purl.obolibrary.org/obo/BFO_0000006</a>	spatial region
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000008">http://purl.obolibrary.org/obo/BFO_0000008</a>	temporal region
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000009">http://purl.obolibrary.org/obo/BFO_0000009</a>	two-dimensional spatial region
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000011">http://purl.obolibrary.org/obo/BFO_0000011</a>	spatiotemporal region
<input type="radio"/> <a href="http://purl.obolibrary.org/obo/BFO_0000015">http://purl.obolibrary.org/obo/BFO_0000015</a>	process

**EFSA catalogue entities**  
Please select EFSA catalogue entities to create subclasses

Search:

<input type="checkbox"/> termCode	TERMEXTENDEDNAME
<input type="checkbox"/> NEG	Negative/Absent
<input type="checkbox"/> POS	Positive/Present

Showing 1 to 2 of 2 entries

Back Cancel | Next

Figure 14. Screenshot of the “concept selection step”

In the “results download step”, the user can see a preview of the generated OWL snippet, with the EFSA catalogue entity(ies) integrated into the OWL schema. Furthermore, on top of the displayed schema the user finds the option to download the new OWL schema with the EFSA catalogue entities. This newly created HSO schema can then be further processed and used.

**Health Surveillance Ontology tool** OHEJP Glossary text mining | OHEJP Glossary

---

**Download file**  
[Download New OWL](#)

**Generated HSO owl**

```
<owl:Class rdf:about="https://w3id.org/hso#NEG">
  <rdf:type rdf:resource="http://purl.obolibrary.org/obo/BFO_0000003"/>
  <rdf:label xml:lang="en">Negative/Absent</rdf:label>
</owl:Class>

<owl:Class rdf:about="https://w3id.org/hso#POS">
  <rdf:type rdf:resource="http://purl.obolibrary.org/obo/BFO_0000003"/>
  <rdf:label xml:lang="en">Positive/Present</rdf:label>
</owl:Class>
```

← Back  Delete this result | Close

Figure 15. "Results/download step" view



### 1.4. HSO Dataset Browser

The HSO concept browser is a “point and click” tool for searching within a HSO based triplestore database. The tool generates automatic dashboards or tables based on the selected surveillance datasets and concepts. This tool was intended as a demonstrator to show basic functionalities of Linked Data in the context of disease surveillance. The data, which is stored in the triplestore data base, is generated with the Excel to RDF Tool.

The tool provides first an overview of stored concepts and their corresponding data that the user can select. Based on the user selection the tool will provide different responses in the next steps. Below we showcase the application of this tool for a surveillance data set.

In the first step the users get an overview of HSO concepts for which RDF data is available in the triplestore. The displayed “Level” is the hierarchical level of each concept within the HSO. The “Matches” column highlights how much data/metadata is stored in the triplestore database and connected to specific HSO concepts. For instance, if a user would select the concept “Country”, the user would see all countries that are stored in the database.

One Health Linked Open Data Toolbox

Health Surveillance Ontology OHEJP Glossary text mining OHEJP Glossary

Please select an ontology concept.

Search:

☉	Level	Concept	Matches
<input type="radio"/>	1.1.1.1.1.1	matrix type <a href="#">↗</a>	3
<input type="radio"/>	1.1.1.1.1.2	specimen from organism type <a href="#">↗</a>	1
<input type="radio"/>	1.1.1.1.2.1	sampling strategy <a href="#">↗</a>	9
<input type="radio"/>	1.1.1.1.4	surveillance sampler type <a href="#">↗</a>	2
<input checked="" type="radio"/>	1.1.1.1.5.2	data set <a href="#">↗</a>	2
<input type="radio"/>	1.1.2.1.1.1	matrix <a href="#">↗</a>	7
<input type="radio"/>	1.1.2.2.3.1.1.3	Country <a href="#">↗</a>	23
<input type="radio"/>	1.2.2.3.1	Surveillance Activity <a href="#">↗</a>	166

Showing 1 to 8 of 8 entries

Figure 16. The start page with the concept table for selection

Assuming that the user selected “data set” and clicked next the user gets a table displaying the id of the data sets available and a short description of each of them (Figure 17). The user can then decide which data they want to see in more details.

One Health Linked Open Data Toolbox

Health Surveillance Ontology OHEJP Glossary text mining OHEJP Glossary

Search:

☉	Row0	Metadata
<input checked="" type="radio"/>	"113e8ab5-0e69-41c4-8d58-9be3c677e74c"	Campylobacter surveillance in Sweden, Sweden, SVA, 2010-2019
<input type="radio"/>	"b82cf570-11fa-4244-b713-ee5c6c1c24"	Salmonella in egg, Europe, EFSA, 2020

Showing 1 to 2 of 2 entries

Figure 17. Data set selection step in the "HSO Data set browser"



After that the user gets an overview of the selected data set in a dashboard (Figure 18). On top a graph is displayed which is automatically adjusted to the data. Data sets which contain multiple years are displayed with a line chart and data sets with only one year as a bar chart. The countries where Surveillance activities has been conducted are then highlighted with a dot on map on the top right side. The size is adjusted to the number of cases. The map also has an adjustable slider to filter for different years. Under the chart and the map a download option is available and the data itself is displayed in form of a table. This table is sortable and searchable.

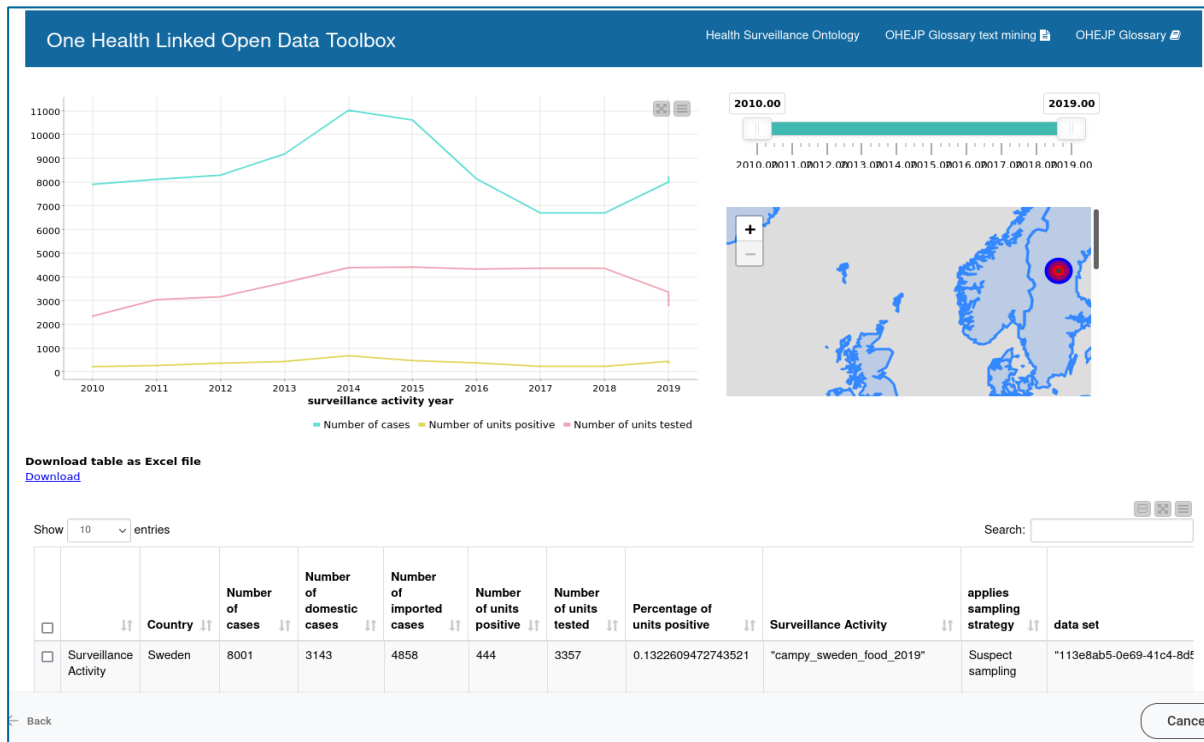


Figure 18. Output: Dashboard with charts and data table

## 1.5. EFSA ECDC data comparison application

This tool was developed with the objective of comparing data between the public health sector and the animal health sector. In this tool we connected surveillance data of *Campylobacter sp.* stored in a HSO triplestore database with the campylobacteriosis case data from ECDC surveillance atlas (<http://atlas.ecdc.europa.eu/public/index.aspx>). Users can choose surveillance data from the EFSA knowledge junction (<https://zenodo.org/communities/efsa-kj/?page=1&size=20>) and compare it with ECDC surveillance atlas. The objective of this tool was first to illustrate the use case of linking different data sets with different metadata and different related entities such as campylobacteriosis with *Campylobacter sp.* with the help of HSO and SPARQL queries. As the previously described tools, this tool is a web application with a “point and click” interface. To demonstrate different use cases we created two versions: one where surveillance data from the EXCEL to RDF Workflow can be uploaded and another one where Campylobacter Surveillance data from the EFSA Knowledge junction of 2019 is collected and can be selected. The data is then compared against the data of ECDC Surveillance atlas.

Depending on the tool in the first step users see either a selection table where they can choose between different data sets encoded in a URL or they can upload an RDF file. After the selection, the tool compares the data between the HSO triplestore and ECDC surveillance atlas. The tool will display a map and a table of the countries stored in the data set. The locations are highlighted in red and the size of the circle indicate the percentage of cases against the population of each country. The user can then choose one or more countries to compare. After selection and clicking next the tool will compare the case data in a bar plot, and will also show the data of both Surveillance activities in two different tables.



**Choose data set**

- <http://www.w3.org/2002/07/Campylobacter in RTE milk and milk products, EU>
- <http://www.w3.org/2002/07/Campylobacter in raw milk, EU>
- <http://www.w3.org/2002/07/Campylobacter in pigs, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE fish and fishery products, EU>
- <http://www.w3.org/2002/07/Campylobacter in fresh broiler meat, EU>
- <http://www.w3.org/2002/07/Campylobacter in cattle, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE other processed food products and prepared dishes, EU>
- <http://www.w3.org/2002/07/Campylobacter in fresh turkey meat, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE spices and herbs, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE sprouted seeds, EU>
- <http://www.w3.org/2002/07/Campylobacter in cats and dogs, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE meat and meat products, EU>
- <http://www.w3.org/2002/07/Campylobacter in other foods, EU>
- <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 in other animals, EU>
- <http://www.w3.org/2002/07/Campylobacter in other processed food products and prepared dishes, EU>
- <http://www.w3.org/2002/07/Campylobacter in fruits, vegetables and juices, EU>
- <http://www.w3.org/2002/07/Campylobacter in broilers, EU>
- <http://www.w3.org/2002/07/Campylobacter in fresh pig meat, EU>
- <http://www.w3.org/2002/07/Campylobacter in fresh bovine meat, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE other foods, EU>
- <http://www.w3.org/2002/07/Campylobacter in meat and meat products, EU>
- <http://www.w3.org/2002/07/Campylobacter in turkeys, EU>
- <http://www.w3.org/2002/07/Campylobacter in RTE salads, EU>

Figure 19. Data set selection step in the EFSA ECDC comparison application

The key feature of this tool lies in the underlying technology to automatically modify and combine the data collected from different sources and in different formats. The ECDC data is a CSV file and the EFSA data is produced with the Excel to RDF Converter tool (described in 1.2.1). In the background both data sets are queried and compared, applying reasoning between the data sets. This allows the generation of a flexible data pipeline, without the modification of the original data. This pipeline could be extended to more complex statistical analysis of the data. The logic is partly derived from HSO, and partly built into the SPARQL query. Logic built into the query is the intended behaviour of smart applications, where the data remain unchanged, but as knowledge evolves, the queries can be updated to combine the data and produce outputs that are up to date with current demands and knowledge.

**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"/>	"Country"
<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

Figure 20. Country selection step

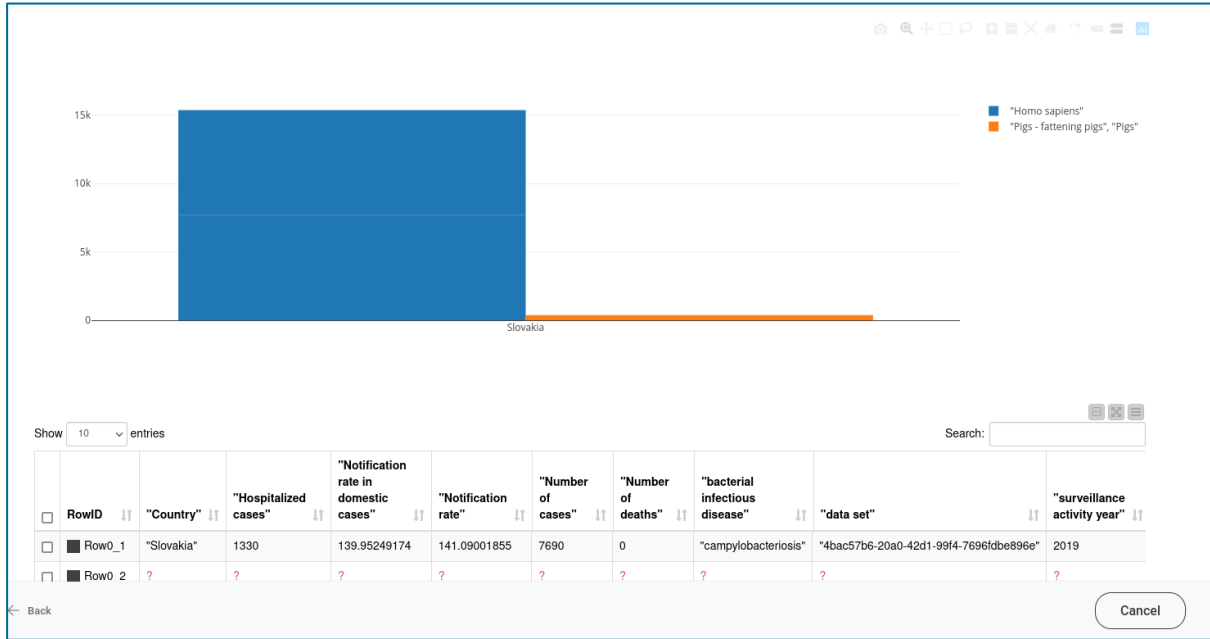


Figure 21. Data output with table and chart



## SECTION 2: USING INTEROPERABILITY TOOLS TO PUBLISH FAIR DATA

The tools presented in [Section 1](#) support semantic interoperability and the linked data model described in the [Introduction](#).

Interoperability, in turn, is only one of the main challenges in computer-mediated, data-driven knowledge discovery. In order to “assist humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflow”, the FORCE11<sup>8</sup> (The Future of Research Communications and e-Scholarship) community has described a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable – now widely known as the FAIR principles (<https://www.force11.org/group/fairgroup/fairprinciples>).

Although primarily used to drive FAIRness in the publishing of research data, the FAIR principles are applicable to any data of scientific value, where maximization of the potential of generating knowledge from data is desirable. In the case of One Health surveillance, the FAIR principles can guide the production of data that maximizes its utility for generating *evidence*.

The FAIR principles are listed in Table 1.

*Table 1. Guiding principles to make data Findable, Accessible, Interoperable, and Reusable (FAIR)<sup>9</sup>.*

<b>To be FINDABLE</b>	F1. (meta)data are assigned a <u>globally unique and eternally persistent identifier</u> . F2. data are described with <u>rich metadata</u> . F3. (meta)data are <u>registered or indexed in a searchable resource</u> . F4. metadata <u>specify</u> the data identifier.
<b>To be ACCESSIBLE</b>	A1 (meta)data are <u>retrievable by their identifier</u> using a <u>standardized communications protocol</u> . A1.1 the <u>protocol</u> is open, free, and universally implementable. A1.2 the <u>protocol</u> allows for an authentication and authorization procedure, where necessary. A2 <u>metadata are accessible</u> , even when the data are no longer available.
<b>To be INTEROPERABLE</b>	I1. (meta)data use a <u>formal, accessible, shared, and broadly applicable language</u> for knowledge representation. I2. (meta)data use <u>vocabularies that follow FAIR principles</u> . I3. (meta)data include <u>qualified references</u> to other (meta)data.
<b>To be RE-USABLE</b>	R1. meta(data) have a <u>plurality of accurate and relevant attributes</u> . R1.1. (meta)data are released with a <u>clear and accessible data usage license</u> . R1.2. (meta)data are associated with their <u>provenance</u> . R1.3. (meta)data <u>meet domain-relevant community standards</u> .

<sup>8</sup> <https://www.force11.org>. FORCE11 is a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. Individually and collectively, we aim to bring about a change in modern scholarly communications through the effective use of information technology.

<sup>9</sup> <https://www.force11.org/group/fairgroup/fairprinciples>



As seen in Table 1, the production of FAIR data relies on addressing not only datasets contents (*data*), but also their *metadata*. As it can be found in the OHEJP OH Glossary<sup>10</sup>, **metadata** is:

*“literally, “data about data”; data that defines and describes the characteristics of other data, used to improve understanding of data and data-related processes.”*

While metadata is intended to be used to document the context associated with specific data, the definition above implies that what constitutes metadata is, on itself, context dependent. For an animal sample being subjected to a laboratory test, for instance, information about how that sample was collected – the animal species, who collected the sample, where and when – can be considered by the laboratory personnel as metadata. Epidemiologists analysing results of a large number of collected samples may consider all of those details to be data, and their interpretation of the context of these data would require further “metadata” about the process that generated that sample collection, such as whether it was a result of a specific surveillance program, and what was the goal of that surveillance activity.

OHS depends on the collation of data collected at many different points of the farm-to-fork continuum, by different sectors and many different agents, each with different interests and objectives for the process of data collection/generation. All information collected and stored, which contributes to preserving the context of data collection is useful to promote data interoperability and reusability. *The discussion of whether people perceive the collected information as data or metadata, is not relevant. From a technological perspective, both can be expressed in the same formats and read by machines interchangeably.*

When sharing data however, for instance at the occasion of (cross-sectoral) collaborations, when publishing data, and in particular when publishing results of data analyses, it is more common to “package” defined sets of data – *datasets*. Datasets of common occurrence are spreadsheets (in formats such as Excel – \*.XLS, \*.XLSX, or plain comma-separated-values – \*.CSV); editable documents (such as Microsoft Word files - \*.DOC, \*.DOCX); or Portable Document Format (\*.PDF). In this section, we will use **data** to refer to any information contained **within** these datasets. No matter how many levels of information can be recorded – for instance data about samples in one sheet, and contextual information about the collection of these samples in another sheet in the same file – everything in these shared files will be called data. We use the term **metadata** to refer to **information about these datasets**, which is not contained inside the dataset itself. If a dataset is made publicly available in a link, for instance, the access link is metadata of the dataset. The title of the file, a description of its content, the owner of the dataset, its license, etc, are the things we will refer to as metadata.

Data and metadata play complementary roles in achieving FAIRness.

### **Findability and Reusability – principles associated with publishing data with appropriate metadata (F\_\_R data)**

Within OHS, not all data we produce can be shared in formats that are accessible and interoperable. In particular, in the case of information meant for dissemination of results, or communication and collaboration among people, the priority is to publish the information in formats that are “human-friendly”, not “machine-readable”.

It is however important that those information are findable by their potential target users, and re-usable. We argue that datasets can be packaged in Findable and Re-usable ways writing FAIR metadata, even when data are, themselves, not FAIR. We will describe an example based on the Swedish One Health pilot.

<sup>10</sup> <https://aginfra.d4science.org/web/orionknowledgehub/catalogue>





All countries involved in the ORION project have carried at least one “One Health pilot” to test, in practice, the implementation of the tools and principles developed during the course of the project. All pilot reports are available publicly as ORION deliverables. A full report for the OH-pilots carried out in Sweden and the UK are appended to this report (see Annex I and Annex II).

We will focus on the Swedish pilot (Annex I) in this section.

Every year since 2009 a national report on the outcome of surveillance activities of infectious diseases in animals and humans is produced in Sweden. The report is produced with contributions from the animal health, public health and food safety sectors. The Swedish National Veterinary Institute (SVA) coordinates the production of the report. The purpose of the pilot was to strengthen the “One-Healthness” of the process of collaboration across the three sectors, and the lessons learned in the process are detailed in Annex I.

Here, we focus on the final product – a PDF report – and the workflow developed to publish this report in a findable and reusable (F\_\_R) format. In this case, the entire content of the report – text, figures, charts, etc – is considered the data, which is available in the format most suitable for its target audience of humans (mostly the general public), and not meant for publishing in machine accessible or interoperable format. Before the ORION OH-pilot, this PDF was published yearly online by SVA, but the lack of metadata information and, in particular, the lack of an explicit license, meant that this report was not findable nor re-usable.

Knowledge models exist for the annotation of “data about data”, metadata. In the European Union, public sector information is made available through the EU Open Data Portal (<https://data.europa.eu/>). To publish datasets in this portal, data owners must specify metadata using the DCAT Application Profile for data portals in Europe (DCAT-AP<sup>11</sup>), which is a specification based on the Data Catalogue Vocabulary (DCAT) developed by the World Wide Web Consortium (W3C, w3.org). Sweden, as many countries, has its own chapter of the open data portal for public sector information (PSI), hosted at [dataportal.se](http://dataportal.se). A Swedish adaptation (profile) of DCAT-AP, the DCAT-AP-SE<sup>12</sup> is specified for publishing datasets in this portal.

We used the ExcelRDF plugin introduced in Section 1 to write the metadata for the Swedish surveillance reports published between 2006 and 2019 using the DCAT-AP-SE specification. We followed this workflow to publish it in the dataportal:

- 1) The metadata about the surveillance report was collected in an Excel file. Table 2 shows all the information collected. Title, description and publisher are mandatory fields.
- 2) These metadata were converted to RDF using the ExcelRDF plugin described in Section 1, generating the RDF file shown in Figure 22 Inspecting the RDF result, note that:
  - a. Some fields have textual entries, such as title and description, and therefore the text provided by data owners was directly used.
  - b. Some fields have a set of values available for choice, similar to a “Pick list”. This is the case for instance for themes, accrual periodicity and language. In these cases the DCAT-AP(-SE) specification provides the URI for the choices available.
  - c. In some cases, the metadata needs to point to another metadata set, describing specific related resources. For instance the publisher and the contact, in this case, must be declared separately with their own set of metadata. All information about SVA as an organization, and as the publisher of the dataset were stored in their own RDF file. It is the URI for this file that is then provided as metadata to the dataset. The links provided in Table 2 for publisher and contact point can be used to access these RDF files directly.

<sup>11</sup> <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe>

<sup>12</sup> <https://docs.dataportal.se/dcat/en/>



- 3) Metadata specifications for a dataset assume that the same data can be provided in multiple formats. These are called distributions. We will describe distributions in better details with the fully FAIR example presented further below. In the case of the surveillance report there is only one data format, one distribution – the PDF file. A second set of metadata needs to be written for that distribution specifically. The only mandatory field for the metadata of a distribution in the access URL. For the example shown here, the metadata for this distribution was written in a separate RDF, then given a unique identifier (<https://data.sva.se/dcat/surveillancereport/sr2019pdf.rdf>), and this identifier is provided to the dataset in the field “distribution” (see Figure 22).
- 4) These RDF files were uploaded to an SVA catalogue to create their unique identifiers. This is actually a recursive process with steps 2) and 3). For instance, it is saving the RDF file generated in step 2 that creates its unique place in the web: <https://data.sva.se/dcat/surveillancereport/sr2019.rdf>. But this address for the file is then used within the RDF file as its unique identifier.
- 5) As SVA is a public organization in Sweden, it has the right to its own catalogue in dataportal.se. Once the metadata files are uploaded to SVA's catalogue, they are automatically uploaded to the dataportal overnight. For this example, the result can be found at [https://www.dataportal.se/en/datasets/59\\_1643/surveillance-of-infectious-diseases-in-animals-and-humans-in-sweden-2019](https://www.dataportal.se/en/datasets/59_1643/surveillance-of-infectious-diseases-in-animals-and-humans-in-sweden-2019), and a screenshot is provided in Figure 23.

Table 2. Metadata collected for the report “Surveillance of infectious diseases in animals and humans in Sweden, 2019”. Fields shaded in green are mandatory.

Metadata field	Data collected
<b>Title</b>	Surveillance of infectious diseases in animals and humans in Sweden, 2019
<b>Description</b>	Surveillance of infectious diseases in animals and humans is the annual report describing the surveillance activities carried out in Sweden during the year. The report covers surveillance for important animal diseases and zoonotic agents in humans, food, feed and animals, carried out and compiled by experts from several Swedish governmental agencies, university and private industry with surveillance mandates along the entire food chain, from farm to fork.
<b>Publisher</b>	SVA
<b>Distribution</b>	PDF distribution, see text
<b>Theme</b>	Agriculture, fisheries, forestry and food
	Government and public sector
	Health
<b>Keywords in English</b>	surveillance; disease
<b>Keywords in Swedish</b>	övervakning, sjukdom
<b>Contact Point</b>	SVA
<b>Periodicity</b>	Annual
<b>Date issued</b>	2020-06-16
<b>Language</b>	English
<b>License</b>	CC-4
<b>Related</b>	Related resources: the dataset of Campylobacter surveillance (presented further below in this report) and a page with a list of all surveillance reports published previously



```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:adms="http://www.w3.org/ns/adms#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:odrs="http://schema.theodi.org/odrs#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<dcat:Dataset rdf:about="https://data.sva.se/dcat/surveillancereport/sr2019.rdf">
  <dcterms:title xml:lang="en">Surveillance of infectious diseases in animals and humans in Sweden,
  2019</dcterms:title>
  <dcterms:description xml:lang="en">Surveillance of infectious diseases in animals and humans is the
  annual report describing the surveillance activities carried out in Sweden during the year. The report covers
  surveillance for important animal diseases and zoonotic agents in humans, food, feed and animals, carried out
  and compiled by experts from several Swedish governmental agencies, university and private industry with
  surveillance mandates along the entire food chain, from farm to fork.</dcterms:description>
  <dcterms:publisher rdf:resource="https://data.sva.se/dcat/main/svaAgent.rdf"/>
  <dcat:distribution rdf:resource="https://data.sva.se/dcat/surveillancereport/sr2019pdf.rdf"/>
  <dcat:theme rdf:resource="http://publications.europa.eu/resource/authority/data-theme/AGRI"/>
  <dcat:theme rdf:resource="http://publications.europa.eu/resource/authority/data-theme/GOVE"/>
  <dcat:theme rdf:resource="http://publications.europa.eu/resource/authority/data-theme/HEAL"/>
  <dcat:keyword xml:lang="en">surveillance</dcat:keyword>
  <dcat:keyword xml:lang="en">disease</dcat:keyword>
  <dcat:keyword xml:lang="sv">övervakning</dcat:keyword>
  <dcat:keyword xml:lang="sv">sjukdom</dcat:keyword>
  <dcat:contactPoint rdf:resource="https://data.sva.se/dcat/main/svaOrg.rdf"/>
  <dcterms:accrualPeriodicity
  rdf:resource="http://publications.europa.eu/resource/authority/frequency/ANNUAL"/>
  <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2020-06-16</dcterms:issued>
  <dcterms:language rdf:resource="http://publications.europa.eu/resource/authority/language/ENG"/>
  <dcterms:license rdf:resource="http://creativecommons.org/licenses/by/4.0/">
  <dcterms:relation rdf:resource="https://old.sva.se/en/reports-and-publications-in-english/disease-
  surveillance/disease-surveillance-reports"/>
  <dcterms:relation rdf:resource="https://www.dataportal.se/en/datasets/59_1684/campylobacter-
  surveillance-in-sweden"/>
</dcat:Dataset>
</rdf:RDF>
```

Figure 22. RDF version of the metadata for the Surveillance of infectious diseases in animals and humans in Sweden, 2019, presented in Table 2.

The screenshot shows a web browser displaying the Swedish data portal (Sveriges dataportal) for the dataset 'Surveillance of infectious diseases in animals and humans in Sweden, 2019'. The page includes a search bar, navigation links, and a detailed description of the report. A 'Use data' section offers a PDF download and a 'Download data' button. An 'About dataset' sidebar provides contact information, keywords (sjukdom, surveillance, disease, övervakning), categories (Regeringen och den offentliga sektorn, Hälsa, Jordbruk, fiske, skogsbruk och livsmedel), publication date (16 juni 2020), language (engelska), and update frequency (årlig).

Figure 23. Screenshot of the Swedish data portal for Public Sector Information showing the published surveillance report.



Note that the actual PDF report was never uploaded to the dataportal or to any other location than its regular publication address at SVA's website. The writing of proper metadata, and publishing of these metadata using an open data portal allowed us to make this dataset reusable by linking the license information directly to the file, and by making it findable to indexing engines capable of searching datasets based on the specification used. Here we used DCAT as a global specification used for the description of datasets. In section 3 we discuss how the availability of OHS specifications could improve findability in the specific cross-sector health surveillance context.

### **Accessibility and Interoperability – principles associated with producing smarter data (\_AI\_)**

The tools and workflows presented in Section 1 were used to create an accessible and interoperable OHS dataset from the Swedish OH pilot described in details in Annex 1.

The “Surveillance of infectious diseases in animals and humans in Sweden” described in the OH-pilot (Annex 1) is published yearly, containing results for all diseases subjected to surveillance in the country, for one specific year. We created a dataset that contains results presented for only one specific hazard – *Campylobacter* – but covering multiple years. The dataset, from here on referred to as the “*Campylobacter* surveillance dataset” is available at [https://data.sva.se/opendata/surveillance/campy/campylobacter\\_surveillance\\_sweden.csv](https://data.sva.se/opendata/surveillance/campy/campylobacter_surveillance_sweden.csv), and a screenshot is presented in Figure 24.

All data in the dataset is already available publicly through the annual surveillance reports (years 2010 to 2019), but in the RDF version the specific concepts are mapped to the Health Surveillance Ontology.

ORION has also developed a One Health Consensus Report Annotation Checklist (OH-CRAC<sup>13</sup>) to promote OHS report harmonization. The specific steps in OH-CRAC are available in HSO as annotation properties, that is, any surveillance activity declared using HSO can be annotated with textual information for each of the surveillance steps recommended in the OH-CRAC. To create the “*Campylobacter* surveillance dataset”, we have extracted all surveillance methods and results from the surveillance reports published from 2010 to 2019 which could be translated into HSO concepts or properties. Moreover, for the main year of the OH-pilot (2019), we have pasted all text from the *Campylobacter* surveillance chapter into the relevant OH-CRAC fields (columns not shown in Figure 24, but accessible in the public file). The resulting file is therefore a tabular representation of information, but also preserves all text published in the original report, and indexed by the specific step of the surveillance process using the One Health Consensus Report Annotation Checklist.

To create a proof-of-concept of how surveillance data could be made available in accessible and interoperable formats, this dataset was converted to RDF using both available workflows presented in Section 1: the ExcelRDF plugin, and the KNIME workflow. In both cases, the result is an RDF file which provides a machine readable version of the human-friendly data presented in Figure 24.

---

<sup>13</sup> available in the OHS Codex: <https://oh-surveillance-codex.readthedocs.io/en/latest/5-the-dissemination-principle.html>



	A	B	D	F	G	H	W	X	Z	AA	AB
1	<b>Surveillance Activity</b>	<b>Year</b>	<b>has Surveillance Objective</b>	<b>Target Pathogen</b>	<b>Susceptible species</b>	<b>has Surveillance purpos</b>	<b>Number tested</b>	<b>number positive</b>	<b>Total cases</b>	<b>Domestic cases</b>	<b>Imported cases</b>
2	campy_sweden_slaughterhouse_2019	2019	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	4363	230			
3	campy_sweden_slaughterhouse_2018	2018	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	4331	377			
4	campy_sweden_slaughterhouse_2017	2017	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	4419	474			
5	campy_sweden_slaughterhouse_2016	2016	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	4389	678			
6	campy_sweden_slaughterhouse_2015	2015	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	3759	437			
7	campy_sweden_slaughterhouse_2014	2014	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	3162	363			
8	campy_sweden_slaughterhouse_2013	2013	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	3046	267			
9	campy_sweden_slaughterhouse_2012	2012	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	2346	217			
10	campy_sweden_slaughterhouse_2011	2011	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	2788	357			
11	campy_sweden_slaughterhouse_2010	2010	prevalence estimation	Campylobacter spp	Gallus gallus	detect changes	3357	444			
12	campy_sweden_people_2019	2019	case detection	Campylobacter spp	Homo sapiens	control			6693	2865	3828
13	campy_sweden_people_2018	2018	case detection	Campylobacter spp	Homo sapiens	control			8132	3645	4487
14	campy_sweden_people_2017	2017	case detection	Campylobacter spp	Homo sapiens	control			10608	6023	4585
15	campy_sweden_people_2016	2016	case detection	Campylobacter spp	Homo sapiens	control			11021	6893	4128
16	campy_sweden_people_2015	2015	case detection	Campylobacter spp	Homo sapiens	control			9180	4709	4471
17	campy_sweden_people_2014	2014	case detection	Campylobacter spp	Homo sapiens	control			8288	3709	4579
18	campy_sweden_people_2013	2013	case detection	Campylobacter spp	Homo sapiens	control			8114	3305	4809
19	campy_sweden_people_2012	2012	case detection	Campylobacter spp	Homo sapiens	control			7902	3155	4747
20	campy_sweden_people_2011	2011	case detection	Campylobacter spp	Homo sapiens	control			8214	3275	4939
21	campy_sweden_people_2010	2010	case detection	Campylobacter spp	Homo sapiens	control			8001	3143	4858
22	campy_sweden_food_2019	2019		Campylobacter spp							
23	campy_sweden_onehealth_2019	2019		Campylobacter spp							

Figure 24. Excel version of the "Campylobacter surveillance dataset".

## All the way FAIR publishing

The workflow to annotate metadata presented earlier was also applied to the Campylobacter surveillance dataset to publish it in the Swedish PSI data portal, making it a FAIR resource which can be accessed here: [https://www.dataportal.se/en/datasets/59\\_1684/campylobacter-surveillance-in-sweden](https://www.dataportal.se/en/datasets/59_1684/campylobacter-surveillance-in-sweden). A screenshot is provided in Figure 25.

Home / Search data & APIs / Campylobacter surveillance in Sweden

## Campylobacter surveillance in Sweden

Statens veterinärmedicinska anstalt

Årlig BY

"Campylobacter surveillance in Sweden is carried out in humans by the Public Health Agency of Sweden (Folkhälsomyndigheten - FoHM), in slaughterhouses executed by Svensk Fågel, and analysed by the National Veterinary Institute (Statens veterinärmedicinska Anstalt - SVA), and in retailers by the Swedish Food Agency (Livsmedelsverket - SLV). Results are consolidated yearly by SVA, and published in the Surveillance of infectious diseases in animals and humans in Sweden (see related resources). This dataset is updated annually with the number published in that report."

### Use data

**CSV**

CSV

Access URL ▼

**RDF**

RDF

Access URL ▼

**About dataset**

**Kontakt**  
Statens veterinärmedicinska anstalt

**Nyckelord**  
Statens veterinärmedicinska anstalt  
sjukdom  
surveillance  
disease  
övervakning  
campylobacter  
campylobacteriosis

**Kategori**  
Regeringen och den offentliga sektorn  
Hälsa  
Jordbruk, fiske, skogsbruk och livsmedel

**Utgivningsdatum**  
5 september 2019

**Språk**  
engelska

**Ingångssida**  
<http://datadrivensurveillance.org/campylobacter-surveillance-in-sweden/>

**Uppdateringsfrekvens**  
årlig

Figure 25. Page at the Swedish PSI data portal for the Campylobacter surveillance dataset.

In this case, we have an example where a dataset is made available with two different distributions: a CSV distribution, which a user can access through the direct link, and open in any software that handles spreadsheets, or any statistical analysis software; and the RDF distribution, meant for use by smart applications capable of handling semantically explicit, linked data.

All the metadata for this dataset, as well as its specific distributions, were also written using the workflows presented in this report. They can be downloaded through the dataportal.se, or directly



This meeting is part of the European Joint Programme One Health EJP.  
This project has received funding from the European Union's Horizon 2020  
research and innovation programme under Grant Agreement No 773830.



through

the

dataset

URI:

[https://data.sva.se/dcat/surveillancereport/campylobacter\\_surveillance\\_sweden.rdf](https://data.sva.se/dcat/surveillancereport/campylobacter_surveillance_sweden.rdf).

This FAIR dataset allows the results of campylobacter surveillance in Sweden, across all sectors, not only to be findable and reusable, but also available for both human and software agents. Having a machine-readable version of this dataset allows smart software applications to be built to consume these data. In this case, it is particularly important to have one permanent link to the dataset ([https://data.sva.se/opendata/surveillance/campy/campylobacter\\_surveillance\\_sweden.rdf](https://data.sva.se/opendata/surveillance/campy/campylobacter_surveillance_sweden.rdf)), and as new surveillance activities are carried out yearly, this dataset grows, allowing application which consume its data to be automatically updated. This is in contrast with yearly reports, for which a new link is produced every year.





## SECTION 3: OPPORTUNITIES TO CONNECT DATA IN OHS

So far, we have discussed what semantic interoperability is, and how it can be operationalized using specific tools. We illustrated the use of these tools with specific examples implemented in the ORION project.

In this Section, we discuss why semantic interoperability is relevant for OHS specifically, as identified during the ORION project, and describe a vision for the future of OHS supported by semantic interoperability tools.

Early in the ORION project we have drawn the *surveillance pathway*, structuring our understanding of the steps involved in collecting, processing, and transforming data into information to support decision in disease prevention, control and eradication. This is depicted in Figure 26 as a linear sequence of steps, but we see this pathway as one iteration of a cycle that repeats itself in a continuous feedback loop, with results from one cycle informing design, adjustment and optimisation in the next cycle.

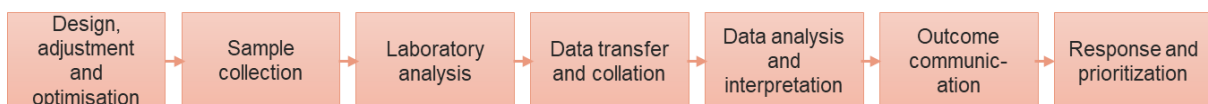


Figure 26. Surveillance pathway depiction used in ORION to anchor discussions throughout all work packages.

In a OHS context, if we consider the specific example of surveillance against foodborne zoonoses (FBZ), we can generalize this as represented by three parallel surveillance pathways, carried out within the animal health, public health and food safety sectors. This is schematically depicted in Figure 27.

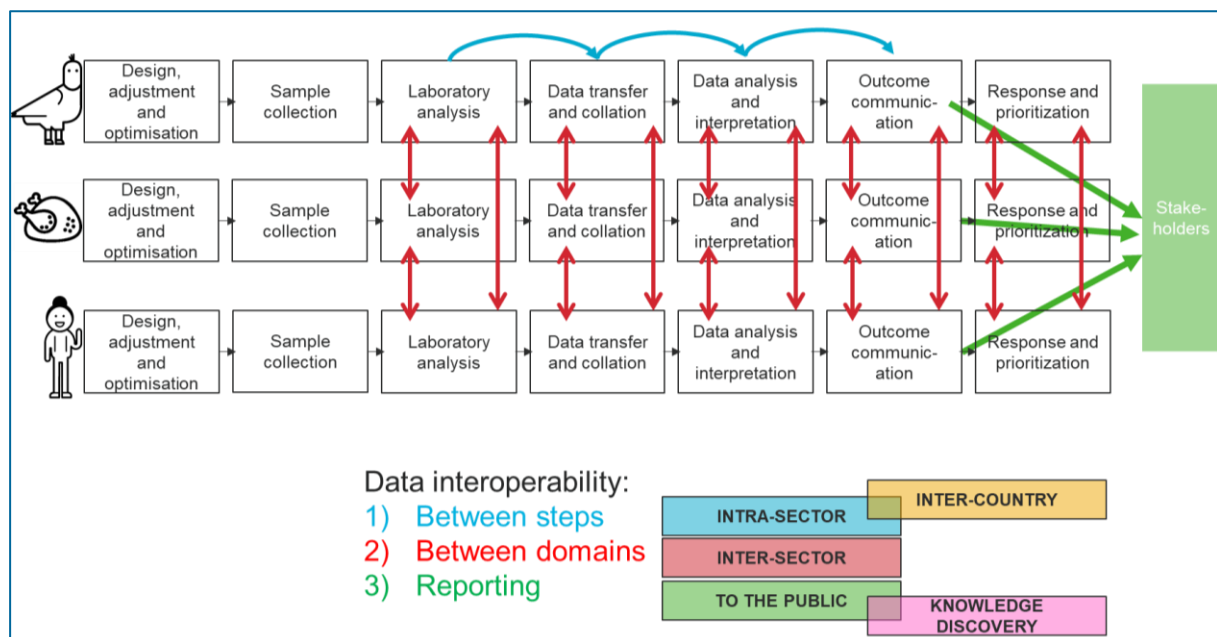


Figure 27. Data interoperability needs in One-Health Surveillance (OHS).

Figure 27 represents the data interoperability needs in OHS highlighting three main flows of data and information: 1) between steps in the surveillance pathway; 2) across health surveillance domains; and 3) when reporting results to stakeholders, including the general public. We will generally speak of “data interoperability”, since we focus here on the development of computational tools. But we can also see





this as an issue of **transparency**, where there is a need to explicitly annotate data with its context and inherent assumptions for reuse or for communication purposes.

Two additional levels of interoperability are identified in the bottom part of Figure 27: inter-country interoperability; and knowledge discovery. Addressing interoperability in these levels is not always a requirement to be able to carry out OHS activities within a region. However, they are an added benefit, allowing even more value to be generated from data. If the primary three levels of interoperability are addressed through semantic interoperability, then the knowledge model based on which interoperability is achieved can be extended (and even made available in several languages) to enable inter-country interoperability.

Knowledge discovery is the process of extracting useful knowledge from data<sup>14</sup>. The process requires a lot of data, and typically a knowledge model documenting assumptions and logic rules over which a machine can reason when analysing and validating the data. It is highlighted here in two contexts where interoperability would enable linking of large amounts of data to improve the power of knowledge discovery. This first is *accumulating evidence* on disease situation awareness to provide actionable information to decision makers in OHS. The second is enabling surveillance data to join the corpus of *research data*. Research data is increasingly made public and semantically explicit due to increased awareness among researchers of the power of data-driven discovery, and more and more even as requirement of funding agencies.

In the Introduction we have described the process of semantic interoperability as three steps: 1) creating a knowledge model; 2) annotating data with this model; and 3) consuming annotated data in smart (semantically aware) applications. The tools developed in ORION and introduced in Section 2 address specifically the two first steps. They are tools to enable the third step, where the full power of the linked data model is materialized. Tools for data consumption were not explicitly developed in ORION because they, ultimately, depend on the data workflows in use within institutions. It is the behaviour, structure and needs of the data consumers that will determine whether they can interact with RDF data.

In year 1 of ORION we carried out a requirement analysis reported in the deliverable JIP1-3.1<sup>15</sup>. The lessons learned from studying data workflows within institutions allowed us to conclude that earlier in the surveillance pathway, where data is mainly flowing within an institution or between institutions tasked with a specific surveillance activity, the data governance challenges are greater than technological challenges. While the application of the linked data model earlier in the pathway would potentialize its benefits, there would be little gain in demonstrating the benefits of the linked data model in steps where adoption is not likely to happen.

It was beyond the possibilities of ORION to change report generation processes and tools within institutions. These workflows were developed to maximize utility within the institution, and can understandably not be tailored to serve the purpose of interoperability with external actors. It was not our goal, for instance, to impose changes from the currently widely adopted relational database management systems (RDBMS), queried through Structured Query Language (SQL), into graphical databases or triple stores where data can be stored as RDF and queried through knowledge formulated queries expressed in SPARQL (SPARQL Protocol and RDF Query Language) or user-friendly graphically constructed queries.

Instead, we chose to focus on areas where data can already be shared, and therefore focused our pilot and tools on data for dissemination and reporting. That is, we focused our “linked-data” model in the specific case of “linked-open-data” (LOD). This is shown schematically in Figure 28.

<sup>14</sup> <https://www.sciencedirect.com/topics/computer-science/knowledge-discovery>

<sup>15</sup> <https://zenodo.org/record/3754615#.YGwFH-gzY2w>

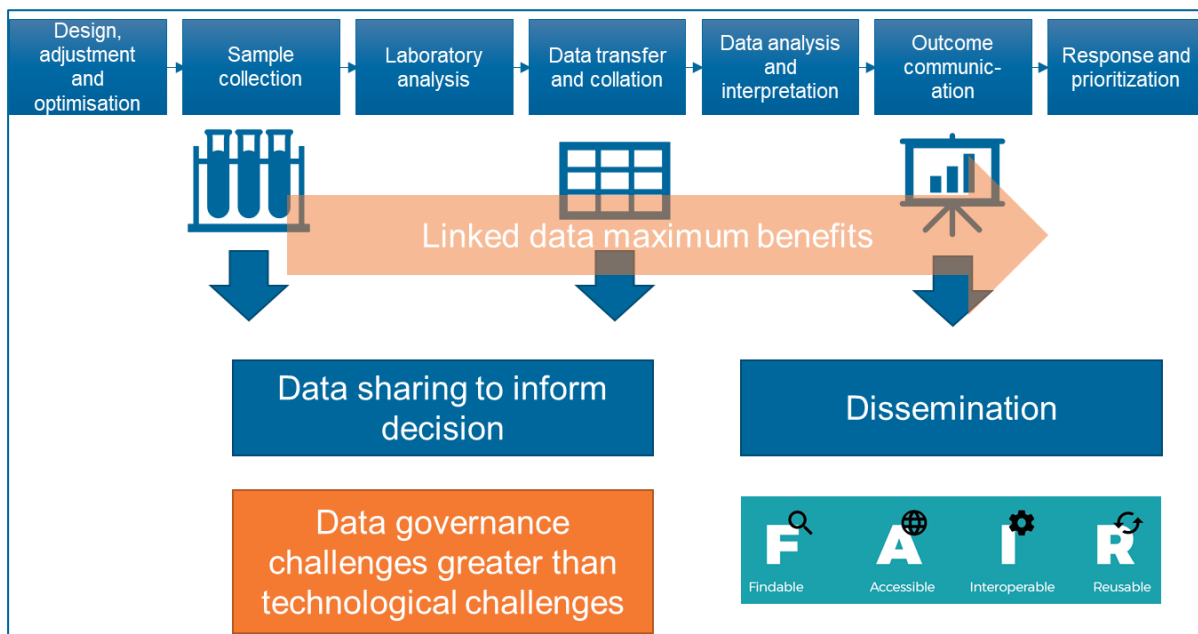


Figure 28. ORION focused on data dissemination, and the application of the FAIR framework to data already made public by institutions.

By choosing to focus on data that are already open, we hope to demonstrate the benefits of the linked-data model, and encourage its adoption trickling down from the end stages of the surveillance pathway, into earlier and earlier stages, enabling the OHS systems of the future.

In the OHS systems of the future, we do not foresee centralization of data. Instead, data owners are part of an “ecosystem of data”. The ecosystem recognizes the data sources available and their content. When a surveillance actor needs information or evidence to support decisions, they pose queries to the ecosystem, and data that can contribute to answering the questions are shared on a “need-to-basis”. *The ecosystem respects, rather than overcomes, existing governance restrictions*, and various models of data sharing (“code to data” versus “data to code”, for instance) are in place to allow the query to be answered without centralization of data.

The goal of ORION was to support this vision of the OHS systems of the future as an ecosystem of interconnected data and tools in two main ways:

- 1) By allowing OHS to become a knowledge model that connects existing data, without imposing any (re)coding. Data are preserved in the context where they are created, and OHS is explicitly represented as a context of data usage, which can evolve in time.
- 2) By connecting data through a layer of semantic interoperability that does not require any centralization of data. Smart applications can be built to operate in this layer, creating the OHS ecosystem.

### One-Health Surveillance as a knowledge model

Semantic interoperability is, as explained in earlier sections, based on the principle of separating data from knowledge. Interoperability achieved with the adoption of common terminologies or data standards results in the creation of a silo of coded data (Figure 29). This model has been used within health sectors, where the context of data generation and data usage is not expected to differ greatly, and data (re)coding is meant to solve syntactic (structural) differences, including language differences.

In OHS surveillance, however, the main barrier to overcome is preserving the context of data collection within each health domain when data are being (secondarily) used for inference in conjunction with data from other sectors, and differences exist not only in the data structure, but in the meaning (semantics).



The creation of the Health Surveillance Ontology (HSO), in particular its creation as a member of a large family of biomedical ontologies (The Open Biological and Biomedical Ontology (OBO) Foundry<sup>16</sup>), aimed to demonstrate how the context of data reuse can be captured in an explicit model separated from the data. Data is preserved in their original format, which is designed to best address the goals of a specific sector or institution, while all the transformation and links needed to connect those data to OHS questions are stored in a layer of interoperability. This is shown in Figure 29.

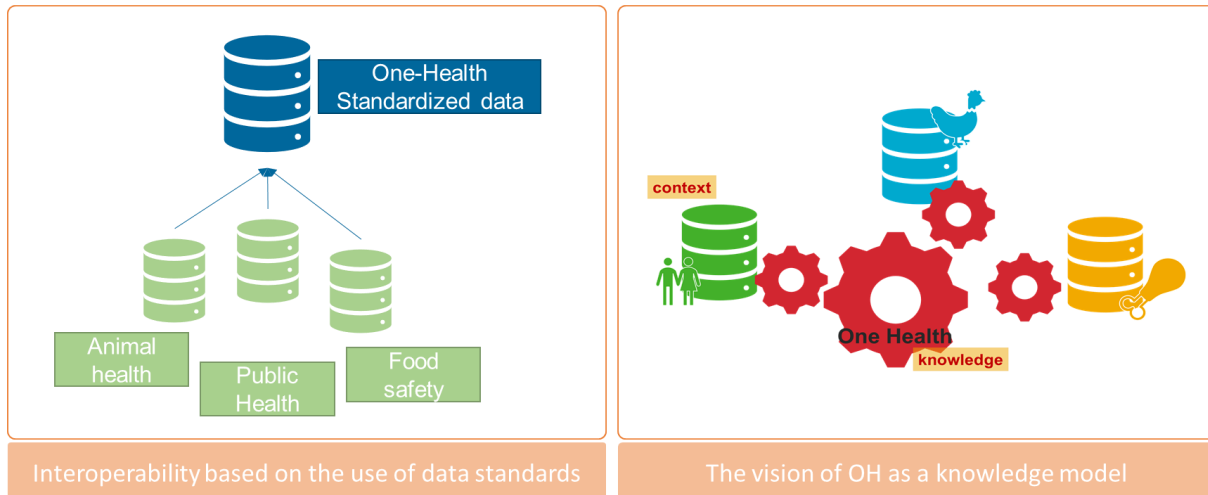


Figure 29. Two visions of OHS – based on structural interoperability, versus semantic interoperability.

In the knowledge model scenario, users – humans or smart applications – can query the data based on “knowledge questions”, rather than data questions. Moreover, the model can evolve in time, without losing compatibility with past or future data.

HSO has the potential to be the building block for the “ecosystem of OHS solutions” depicted as cogwheels in Figure 29. This vision allows institutions to continue collecting, producing, storing and sharing data according to their current practices, as long as the knowledge model that connects all data sources and tools in the ecosystem is *maintained*. While ORION provided a proof of concept for the construction of such knowledge model, reaching the vision depicted in Figure 30 will depend on HSO being adopted by a community of users committed to maintaining the ontology, and curating ever evolving versions. Figure 30 details the responsibilities and timelines associated with this vision.

STEP	WHO	HOW	WHEN
Build/maintain HSO	Small curation group	PUBLIC repositories	Continuously
Collect/Produce data	AH - PH - FS surveillance agencies	Keep their current practices	Surveillance cycles
Annotate data		Chosen workflow to RDF (Excel plug-in available)	When produced, no extra workflows needed
Store / share data		Keep their current practices	
USE / Re-USE data	All who take decisions in OHS	Shared repositories or partnerships	When needed

Figure 30. Responsibilities and timelines in a vision of semantic interoperability across health sectors supported by the Health Surveillance Ontology (HSO).

<sup>16</sup> <http://www.obofoundry.org/>



## An ecosystem of connected data

The second fundamental building block of this vision of the OHS systems of the future is the ability of the ecosystem to connect data sources. Again, data would not be centralized, or even integrated, but the ecosystem would be aware of all existing data sources and their contents. If a user posts a query to the ecosystem, the ecosystem would be able to pose queries to the data source on a need-to basis, or, depending on the data governance rules in force, simply inform the user of the existence of a data source that can contribute to their question. The user would have access to information about the data owners and could contact them directly, but data sharing would not need to be guaranteed through the ecosystem.

Let's consider again the specific example of FBZ surveillance. The surveillance pathways involved were depicted in Figure 28 as three parallel pathways in order to highlight the connections between steps across sectors. In reality, these activities are not carried out at the same time in parallel. Surveillance activities would be carried out at different times and at different steps of the food production chain, generating sparse data as shown in Figure 31.

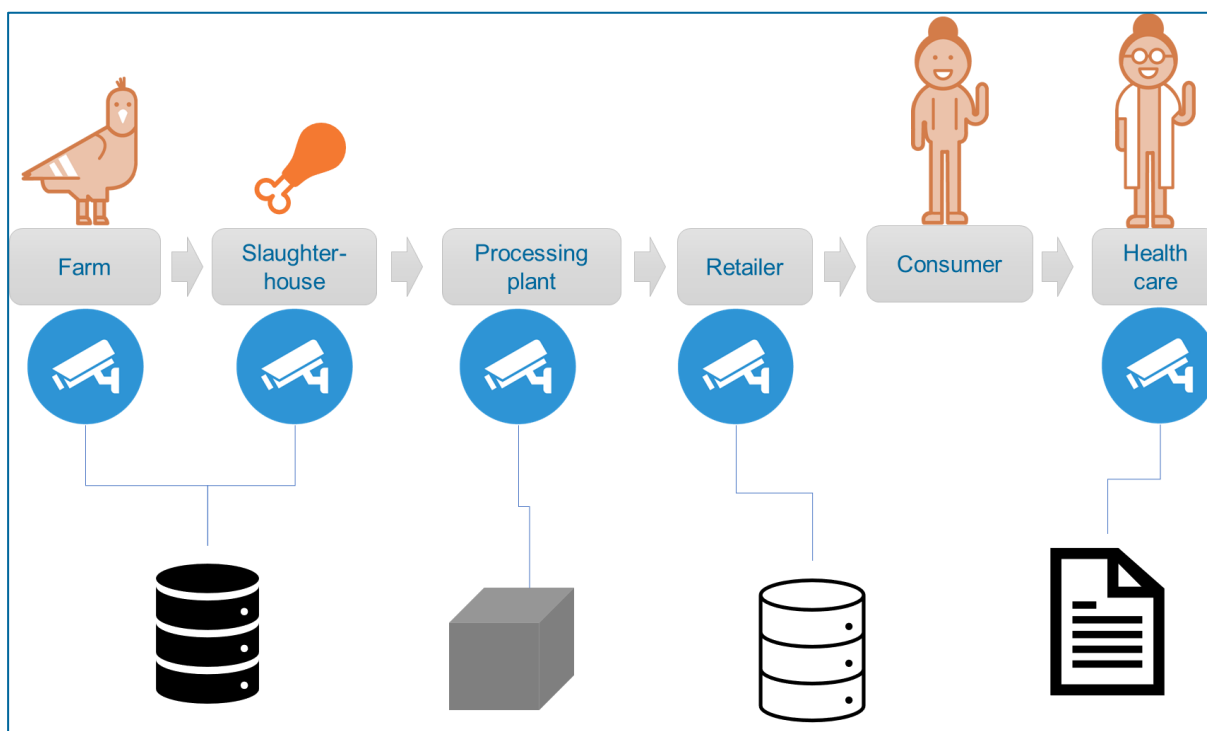


Figure 31. Surveillance in the different points of the food chain.

Currently, these data sources are stored under different formats, and from a technical integration perspective, they are “black boxes” which all other data sources cannot access. In Section 2, we exemplified how the tools developed in ORION can be used to annotate data and metadata, and publish FAIR data, or simply to annotate metadata, making data findable and reusable<sup>17</sup>. If all data owners published data in “FAIR” or “F\_\_R” formats, the OHS ecosystem could easily be built as a “hub” that connects these data, as shown in Figure 32.

<sup>17</sup> Reusable as described in the FAIR principles presented in Section 2 does not mean public or open. Different desired levels of restrictions and permissions of usability are possible, but they should be expressed through specific licensing options explicitly stated in the metadata

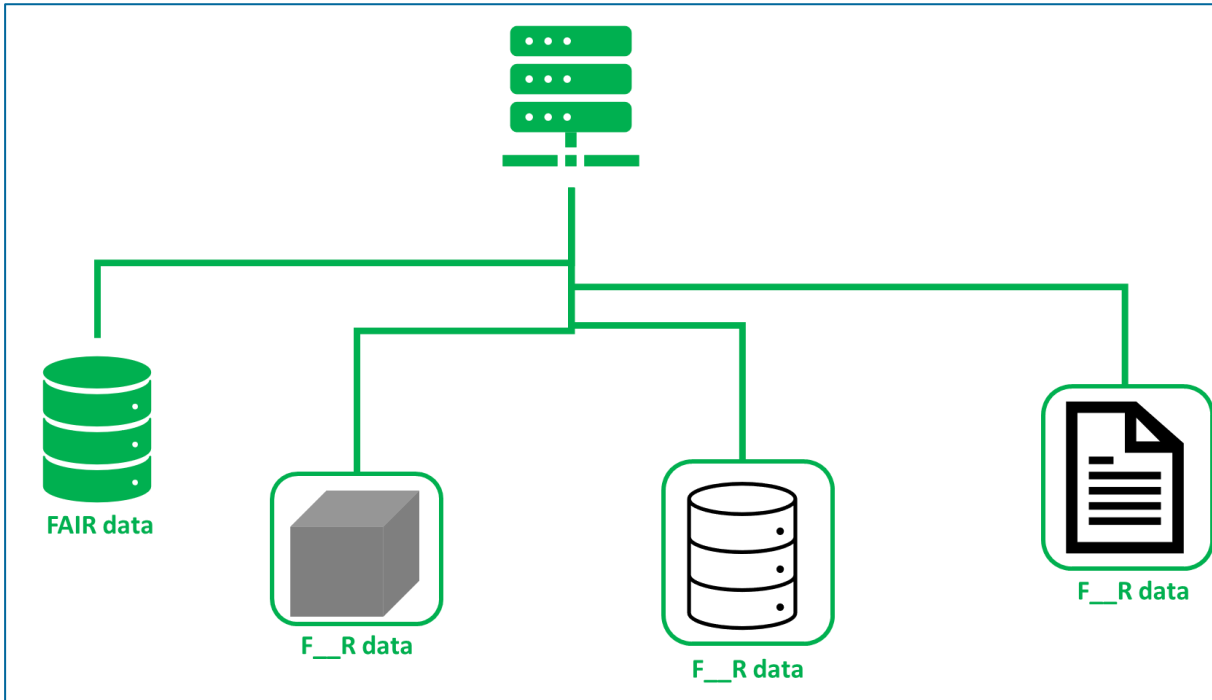


Figure 32. An ecosystem of connected data where not all data are FAIR, but all data are F\_R.

In this report we have focused on HSO as a tool for data annotation, and demonstrated the use of tools that can attach HSO tags to both data and metadata. In Section 2 we have mentioned that ORION has also developed a One Health Consensus Report Annotation Checklist (OH-CRAC<sup>18</sup>) to promote OHS report harmonization. As we highlighted then, the specific steps in OH-CRAC are available in HSO as annotation properties. The properties can be used not only to annotate data, but also metadata.

In Section 2 we described the publication of the campylobacter surveillance in Sweden dataset as a fully FAIR resource, which URI is: [https://data.sva.se/dcat/surveillancereport/campylobacter\\_surveillance\\_sweden.rdf](https://data.sva.se/dcat/surveillancereport/campylobacter_surveillance_sweden.rdf)

Applications pointed to this URI access the metadata of the dataset as RDF. In this RDF, CRAC annotations have been used to annotate the dataset. That is, CRAC has been used as a particular set of information to make this dataset findable by OHS applications.

<sup>18</sup> available in the OHS Codex <https://oh-surveillance-codex.readthedocs.io/en/latest/5-the-dissemination-principle.html>



## SECTION 4: DATA-RELATED FINDINGS FROM ORION OH PILOTS

Every partner involved in ORION carried at least one national “One Health pilot” (OH-pilot) in order to either test the implementation of an ORION developed tool in practice, or test an improvement in the OH-ness of a specific surveillance activity within the country. An extra OH-pilot, called the supra-national pilot, was carried out to investigate whether tools developed in ORION could be applicable in workflows to produce and/or disseminate surveillance results by EFSA and ECDC.

A full final report for all pilots is available at [onehealthjep.eu](http://onehealthjep.eu), in the group called “ORION knowledge Hub”. During the final full consortium meeting of the ORION project, all pilots summarized their results, and presented their lessons learned. Specific data-related findings brought up during this workshop are summarized in this section – and referred to as lessons-learned from the “OH-pilots workshop”.

Table 3 below lists all the OH-pilots, identifying whether any data issues were addressed in the pilot. This table gives a brief overview of the issues addressed in each pilot. Further below we detail and discuss some of these issues, with particular focus on the discussions held on the OH-pilots workshop, and the lessons learned in the pilots attached to this deliverable.

Table 3. One Health pilots carried out in the ORION project, and specific data issues addressed.

Country	Pilot name and overall goal	Data issues notes
BE	Promoting the implementation of One Health collaborative approaches between disciplines in Belgium with a focus on the elaboration of the future national One Health Antimicrobial report using qualitative research methodologies	Through questionnaires and discussions with stakeholders, the pilot concluded that there is no aligned strategy for data collection, data analysis and communication, across sectors and across the country. While the majority of stakeholders expect data on AMR and AMU from the different sectors (human, veterinary, environment) to be compiled and presented in a centralized report that provides a global picture of the situation in Belgium, resistance from the part of the data providers exist to collaborate and integrate data.
DE	Toxoplasma gondii in livestock and humans – the pilot aimed at describing the OHS system, analysing available data about the surveillance, and incorporating the results into the ORION surveillance inventory.	Data integration issues were not specifically addressed, but the need for this pilot arose from the identification of a lack of documentation on surveillance methods and results for zoonoses which are not reported regularly to EFSA. To address this need, WP2-epi developed an inventory of surveillance systems, which is the focus of an own deliverable.
DK	Evaluation of sequence based surveillance of Campylobacter – integration of whole-genome sequencing analyses across agencies to improve surveillance	This pilot demonstrated the specific gain for surveillance of foodborne diseases (using Campylobacter as the use case) achieved when WGS data are shared across agencies. The technical solutions for the data integration were the focus of other pilots, in particular UK and NO.
DK	Does One Health data integration and interpretation across sectors Improve surveillance and disease control? Investigate whether OH integration of surveillance data from different sectors was possible and could improve surveillance and campylobacter control in the food chain to reduce risk to public health.	While the pilot above focused on the integration of WGS data, this pilot, developed in the same country and using the same case disease (Campylobacter) investigated the benefit of integrating epidemiological data across surveillance actors. The benefits for surveillance were demonstrated, and the work results in proposed improvements to the Danish National Campylobacter Action Plan. The exercise to aggregate surveillance data across sectors brought up several challenges in terms of both structural and semantic interoperability. But most particularly, it highlighted that a lot of the surveillance context is missed because documentation of metadata at the surveillance activity level (as opposed to the epidemiological unit/sample level), in particular surveillance methods, is not thoroughly documented, and especially not documented with the view of being reusable by other sectors.





DK	Increasing the One Health interpretation of AMR and AMU surveillance. Explore how the zoonoses chapter of the DANMAP report could identify One Health objectives, and whether further integrated analyses and interpretations would enhance the One Health conclusions for decision-makers and other users of the report.	This pilot focused on the OH-ness of AMR and AMU surveillance reporting implemented in DANMAP. It promoted closer OH collaboration between actors and improved usefulness of report for users.
NL	Towards a One Health surveillance of hepatitis E in the Netherlands. Stimulation of collaboration between the institutes and perform joint analyses of available whole-genome-sequence (WGS) data.	Integration of the different data flows into OHS. This pilot provided and assessment of the current state of OH collaboration by mapping the institutes, projects, and collaboration. Meetings and collaboration strategies were defined to set joint goals to add and combine data into HEV-net, and perform joint analyses to gain more insight in the OH aspects of hepatitis E.
NO	Building a One Health bioinformatics analysis IT platform. This aimed to allow sharing of data between the AH and PH institutes in Norway.	This pilot focused specifically on WGS data sharing between agencies, and all its associated metadata. The chosen platform allows implementation of many of the structural and semantic interoperability issues brought up here. The full report is available at <a href="http://onehealthejp.eu">onehealthejp.eu</a> , in <a href="#">this link</a> .
SE	Increasing the OH-ness in surveillance reporting and dissemination.	Attached as <a href="#">Annex I</a> . The data interoperability opportunities explored in this pilot were already extensively described in previous sections.
UK	Salmonella data sharing pilot project in the UK - aimed at establishing and strengthening inter-institutional collaboration and knowledge transfer in the area of surveillance data integration and interpretation, along the One Health objective of improving health and well-being.	While in this deliverable we are focusing on the specific challenges of integrating the contents of different datasets, the UK pilot focused on how to operationalize the data sharing across surveillance actors in the current framework of surveillance.  Their full pilot report is attached as <a href="#">Annex II</a> .
WP1	WP1 OH pilot - Aimed at testing and improving two specific ORION resources developed in WP1: OHEJP Glossary and OH-CRAC	CRAC is already directly linked to the HSO, as described in earlier sections.
Supra-national	Supra-national pilot – aimed at testing the functionality and usability of the ORION solutions such as OHEJP Glossary, OH-CRAC and HSO with the support of international stakeholders such as EFSA and ECDC	The supra-national pilot supported the testing and development of the tools described in Section 1.

Some pilots' lessons demonstrate the need for some of the tools developed in ORION. The DK One Health data integration pilot, for instance, demonstrated that the context of surveillance within a sector needs to be better captured if data are going to be integrated with other sectors. This is a need that CRAC tried to address (WP1). Other pilots focused on testing the implementation of these tools.

In the pilots workshop, pilot leaders discussed what we called “the chicken and the egg problem”: it is hard to convince stakeholders of the benefits of data sharing before having done it. As it can be seen by reading through Table 3, the pilots covered different dimensions of the surveillance data sharing process, and complemented each other in the continuum from technically being able to share data files, all the way to being able to interpret datasets across sectors, generate joint results, interpret joint results, and use those results to feed surveillance design.

In the case of WGS data sharing, for instance, a pilot in DK demonstrated the benefits that could be achieved in surveillance if data were shared, while pilots in the UK and NO dealt with the technical implementation of data sharing capabilities. Still in this specific case of WGS data, it was demonstrated





that sharing on a demand bases may miss the detection of clusters, while aggregation by one agency on the behalf of all sectors has implications for real-time surveillance.

Another common problem is that the needs and benefits of data sharing often become very clear during a zoonotic outbreak, but the structure (technical and legal) for data sharing needs to be built in “peace time”. Participants suggested that the common goal during the outbreak is the main facilitator of the data sharing process, and that inter-sectoral collaboration needs to focus on identifying these shared goals and addressing the structures needed to reach them.

It was a clear learned lesson that the issues related to “data harmonization” are not as important as the challenge of data interoperability. But with “data interoperability” we do not mean just the technical aggregation of joining data. We mean most especially the difficulties in semantically aggregating data form different sectors, which carry surveillance with different goals, different sampling methods, aimed at different epidemiological units, and which are used to documenting their surveillance methods in different formats and in different depths. As data from each sector is designed to attend the surveillance goals within that sector, documentation of surveillance methods leaves out key assumptions and methods that may be perceived by that section as “obvious”, or which are inherently implicit in the way that sector designs surveillance. On a technical level this can only be addressed with knowledge modeling, as addressed in WP3 and extensively detailed in this report. But the technical level only addresses communication across computers – a similar step of semantic interoperability targeting humans needs to be incorporated into the way we collaborate across sectors. In particular, a need was identified to better document the *gaps in knowledge*.

All that said, participants highlighted that data sharing needs to be fit for purpose – not a catch all goal. Data sharing is indeed not always needed, as long as the issues identified above can be tackled through efficient collaboration.

Issues of data governance and ownership were brought up not only in respect to raw data, but along the entire continuum of data analysis and output generation. Similarly, meta-data documentation to retrace things retrospectively and reconstruct/compare results needs to be addressed throughout the surveillance process, not just for raw data.



## CONCLUSIONS AND LESSONS LEARNED: CONSIDERATIONS FOR A FAIR-ER OHS FUTURE

---

One current definition of One-Health surveillance (OHS) is “the systematic collection, validation, analysis, interpretation of data and dissemination of information collected on humans, animals and the environment to inform decision for more effective, evidence-and system-based health interventions<sup>19</sup>”. Bordier et al, 2018<sup>20</sup> pointed out that other concurrent definitions of OHS all emphasize the role of cross-sectoral collaboration in the improvement of health management. (For definitions of One-Health and related terms visit the OHEJP glossary at <https://aginfra.d4science.org/web/orionknowledgehub/catalogue>).

Barriers to data sharing are often listed when evaluating challenges to the establishment of such cross-sectoral frameworks. The challenge of extracting information from data coming from such heterogeneous contexts goes far beyond the simple access and aggregation of data. Ammon and Makela (2010)<sup>21</sup> described in detail the integrated collection and analysis of data on zoonoses in the European Union (EU), first established in 1992, and currently a joint task of the European Centre for Disease Prevention and Control (ECDC) and the European Food Safety Agency (EFSA). Despite this opportunity for joint data analysis, the authors pointed out several challenges to data comparability, from methodological differences between countries, and challenges of data quality and validation, to differences of population structure and population reporting level among sectors.

The experiences reported highlighted the complex data and meta-data structure needed to capture and take into account all the contextual information about the data collected and the data collection processes. As Beauté and colleagues (2020)<sup>22</sup> more recently pointed out, an accurate description of structural elements of surveillance systems is essential for interpretation and evaluation, but even when these descriptions exist, confusion can remain on their interpretation.

Within individual countries, activities of surveillance in public health (PH), animal health (AH) and food safety (FS) all generate data which can contribute to OHS. Converting those data into valuable information for decision requires not necessarily that those data are aggregated, but that they are interoperable, so that sharing can be performed on demand, for specific problems, respecting various models of data disclosure. Interoperability focuses on cooperation among systems, referring to their ability to continuously communicate and exchange information, and use the information that has been exchanged<sup>23</sup>.

In this deliverable, we have described ORION's contributions to data interoperability in One Health Surveillance (OHS). We have focused on the specific case of sharing surveillance outputs across sectors to enable joint surveillance evaluation, prioritisation and design. Our developed tools are therefore most applicable to the case of publishing reports and data at the end of specific surveillance cycles, such as for instance yearly reports published individually by countries and, in the specific case of the European Union, jointly by EFSA and ECDC.

As we have centred tool development around the development of a knowledge model for surveillance, the achievements in terms of data and meta-data structuring, and harmonisation and disambiguation

---

<sup>19</sup> Stärk KDC, Arroyo Kuribreña M, Dauphin G, Vokaty S, Ward MP, Wieland B, et al. One Health surveillance - More than a buzz word? *Prev Vet Med.* 2015;120(1):124–30.

<sup>20</sup> Bordier M, Uea-Anuwong T, Binot A, Hendrikx P, Goutard FL. Characteristics of One Health surveillance systems: A systematic literature review. *Prev Vet Med [Internet].* 2018;(October):0–1. Available from: <https://doi.org/10.1016/j.prevetmed.2018.10.005>

<sup>21</sup> Ammon A, Makela P. Integrated data collection on zoonoses in the European Union, from animals to humans, and the analyses of the data. *Int J Food Microbiol [Internet].* 2010;139(SUPPL. 1):S43–7. Available from: <http://dx.doi.org/10.1016/j.ijfoodmicro.2010.03.002>

<sup>22</sup> Beauté, J., Ciancio, B.C., Panagiotopoulos, T., 2020. Infectious disease surveillance system descriptors: proposal for a comprehensive set. *Euro Surveill.* 25. <https://doi.org/10.2807/1560-7917.ES.2020.25.27.1900708>

<sup>23</sup> Definition of Interoperability. In: HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations. 2nd editio. 2010. p. 190



through semantic expression are achievements on themselves, which are preserved for future use in the form of a publicly available ontology – the Health Surveillance Ontology (HSO). As shown in Figure 30 (Section 3), the development of an ontology is not a “one-time” process, and ontology maintenance will require a community of users committed to keeping the knowledge model evolving. One of the main advantages of a knowledge model is its ability to adapt to new knowledge, but this of course requires the work of dedicated curators. We have, during the lifetime of ORION, succeed in making HSO a member of the Open Biological and Biomedical Ontology (OBO) Foundry to ensure that it can be part of a larger community of ontology developers and users.

## A FAIR OHS future

The FAIR principles have been discussed in Section 2. The FAIR principles were also highlighted in EFSA's technical report “Publication of scientific data from EU-coordinated monitoring programmes and surveys” in 2019<sup>24</sup>.

The Health Surveillance Ontology (HSO) is a FAIR model, which therefore enables annotation of data – and meta-data – in fulfilment of the interoperability principle. To implement data annotation, the existence and public availability of HSO is of course not enough. Data owners must have access to data *annotation* workflows. While these workflows can only be implemented by individual institutions, and a “one workflow fits all” does not exist, the ORION project has provided a number of proof-of-concept workflows using country specific and EFSA and ECDC publicly available data, as extensively described in Section 1 and Section 2.

While we have focused on publicly available data, it is important to note that being “open or publicly available” is not a requirement for data to be FAIR. FAIR data is findable by those who must find it, and accessible to the software agents who will process it. In turn, “publicly” available data which does not have an explicitly declared license, is not reusable, and therefore not FAIR.

The European Food Safety Authority Advisory Forum Task Force on Data Collection and Data Modelling delivered their conclusions<sup>25</sup> on September 2020 focused on four key priority areas: data collection and reporting processes, data models, IT infrastructure, and data analysis. The implementation of data interoperability through semantic annotation enables (and in fact provides the foundation for) many of the recommendations made by the task force for a future ideal EU food safety system, including a specific recommendation to initiate and promote the development of ontologies.

In the introduction, we have presented the linked data model as enabling knowledge to be stored in applications, so that data is consumed by the right applications on demand. This preserves data context while at the same time making it evolvable and sharable without the need for data coding or centralization. This is in agreement with the task force's recommendation 4.5 to develop IT architectures centred in *building services, not websites*. The key element of a future food-safety system specifically, and OHS generally, is to avoid data centralization, and indeed even data analysis centralization, moving towards an “**ecosystem of solutions**”. In this ecosystem, data providers do not transfer their data, rather choose to make the ecosystem aware of their existence, content and structure. Data analysis tools are also developed independently by many actors who choose to add their tools to the ecosystem. Data analysis tools and actual data meet on demand, and data access is negotiated for purpose. This view of an ecosystem meets yet another important recommendation from the task force – (2.3) “Good practice aimed at data interoperability should be sought through collaborative data governance”.

## A FAIR-ER future

Surveillance is a cyclical process, that happens in a continuous feedback loop. Surveillance execution happens in cycles, usually annually, and the surveillance pathway introduced in Section 3 is a good representation of the activities within one cycle. However, we should not lose sight of the fact that the

---

<sup>24</sup> doi:10.2903/sp.efsa.2019.EN-1544

<sup>25</sup> doi:10.2903/sp.efsa.2020.EN-1901



main reason to share surveillance outputs across sectors is for surveillance evaluation and action, in particular for the redesign and prioritisation of activities in the next cycle.

In Section 2 we have demonstrated the creation – and FAIR publishing – of a dataset of surveillance outputs for one specific system (Campylobacter surveillance in Sweden), but across several years. As opposed to the annual surveillance report system, which contains results for all surveillance systems in the country but only for one year, the Campylobacter dataset is “***Extendable***”.

In this view of an ecosystem of solutions, the Campylobacter dataset would be a resource in the ecosystem with a fixed resource identification. As more results become available – and indeed even as data may be updated or corrected from previous years – applications that use those data can always be sure to be accessing the most up to date data available. Surveillance methods themselves can be changed, but these changes are also documented in the extended data. Extendable datasets are how we see data within institutions, but the idea is not yet built into the cycles of surveillance data publishing.

A last idea we would like to add to the FAIR principles is the idea of “***Reproducibility***”, which is also directly related to “***Transparency***”. In an ecosystem of solutions, data analysis is performed within the ecosystem using codes that are themselves part of it. Any user is free to reproduce the analyses, and more importantly, check the inherent assumptions made within.

When we publish data in yearly cycles as independent resources, in aggregated formats, and without enough surveillance descriptors to track the surveillance methods that generated those results, we lose reusability even within the system, as compatibility with historical and future values becomes fraught.

Developers of data analytical methods are used to the concepts of reproducibility and versioning, but we miss these same concepts in the way we store surveillance methods and results. ORION has provided two key resources to advance solutions that make surveillance data inherently associated with its methodological context: the One Health Consensus Report Annotation Checklist (OH-CRAC<sup>26</sup>), and the Health Surveillance Ontology.

Moving towards a FAIR-ER future we suggest that OHS actors in general, and data providers in particular, investigate workflows that allow them to capture all methodological details in force in a given surveillance system and year, as established in OH-CRAC. If these data are to be shared, we then recommend their semantic annotation with HSO and following the FAIR principles, in particular making sure that data have enough meta-data to be findable, and proper licensing to be reusable. If FAIR data publishing is achievable, then we urge data providers to think about whether unique resource identifiers can be given to specific datasets individually but perpetually, so that these datasets are Extendable through the years, rather than continuously replaced by a new yearly batch. Lastly, we suggest versioning any changes to both the dataset and its metadata, and ensuring that assumptions and analyses performed to aggregate raw data into the published data are reproducible.

ORION's contribution to this future have been documented in this deliverable finalized in June 2021, but the ORION tools will live through its community in the links provided throughout this deliverable.

We highlight in particular the [One-Health Surveillance Codex](#) for a list of all ORION resources.

---

<sup>26</sup> <https://oh-surveillance-codex.readthedocs.io/en/latest/5-the-dissemination-principle.html#one-health-consensus-report-annotation-checklist-oh-crac>



# ANNEX I – ONE-HEALTH PILOT REPORT: SWEDEN

## Increasing the OH-ness in an existing surveillance report – pilot WP3 Sweden

### 1. Background

Every year since 2009 a national report on the outcome of surveillance activities of infectious diseases in animals and humans is produced in Sweden. The report is produced with contributions from the animal health-, public health- and food safety sector. The Swedish National Veterinary Institute (SVA) coordinates the production of the report. Authors from the different sectors are every year asked to review their chapters and update the numbers and change the text accordingly. This work has been completed by the authors without having to interact much with the other sectors. A joint analysis between sectors of the outcome of the surveillance has not regularly been performed before writing the report. The validated data from the three sectors are published in a pdf-report but are not made available in any other format.

A need for improved One-Health (OH) collaboration in conjunction with the production of the report had previously been identified. This need was fed into WP3 of ORION by the Public Health Agency and it was decided to focus on the Swedish surveillance report in the pilot of WP3. The purpose of this pilot was to strengthen the OH-focus of the report and to create and implement a work-process supporting the collaboration between sectors that will persist after the project ends. Moreover, the agencies will through the process gain a better understanding of each other's activities, data sources and results.

### 2. Objectives

- 1) Promote collaboration among all agencies involved in surveillance of foodborne diseases
- 2) Deepen the understanding of surveillance outputs across agencies and collaborate in the process of decision making to improve surveillance based on these results
- 3) Inform the technical development of tools for data FAIRness (findable, accessible, interoperable and reusable)<sup>27</sup>

### 3. Expected outcomes

- 1) Improved collaboration among health surveillance agencies involved in zoonoses surveillance, so that surveillance in each sector can be informed by all relevant dimensions of the One-health problematic
- 2) Production of surveillance data that is FAIR (findable, accessible, interoperable and reusable), has the necessary metadata to allow data interpretation to take into account the context of data collection, and is interoperable and readable by both humans and machines across sectors.
- 3) A revised surveillance report, where One-Health issues are specifically lifted

### 4. Performed activities

#### 4.1. Initial Workshop in 2018

To start a discussion on how to change the way of working within and between sectors into a more one-health oriented approach, and also discuss how the data in the report can be made available and easily accessible, a workshop was arranged in December 2018.

<sup>27</sup> The FAIR data principles: <https://www.force11.org/group/fairgroup/fairprinciples>





We decided initially to focus on three important chapters of foodborne zoonotic agents; *Salmonella*, *Campylobacter* and VTEC/STEC. Representatives from the public health, animal health and food safety in Sweden that have been responsible for producing the report and/or contributing to one of these chapters were invited to the workshop.

During the workshop the three sectors got the opportunity to present and discuss the data that are used today to produce the report. Potential sources of data that have not been used in the report previously were brought up for discussion as well as the lack of data that are not collected today. Furthermore, the actual process of working together in new ways between sectors when preparing the report was discussed as well as how the report itself could be improved in different ways.

Following this discussion, the participants were given a presentation about the possibilities of data interoperability through the ontology and technology work in ORION-WP3.

It was concluded that there is a need to increase the collaboration between sectors in the production of the report. It was a good opportunity for representatives from all three sectors to come together to present and discuss their surveillance data. By working more closely together it will be possible to improve the overall understanding of the results of all ongoing surveillance activities and hence enable a common interpretation of these results.

All participants were willing to meet and discuss surveillance results from 2018 in a workshop as a preparation for the work with the coming issue of the report. To be able to focus on the outcome of the surveillance, and also on how to improve the presentation of the results, it was decided to have separate workshops for each chapter. Workshops were scheduled to take place after the yearly data had been collated and preliminary results were ready to be shared.

#### **4.2. Pre-pilot workshops in 2019**

Pathogen specific workshops were organized for *Salmonella*, *Campylobacter* and VTEC/STEC in the spring, 2019. During the workshops the texts from the previous surveillance report were scrutinized and discussed in detail to identify how the information can be presented in the best way. Tables and visualizations were also examined to see how they can be improved or if it is possible to produce visualizations with data from more than one sector. Data that had previously not been used for the surveillance report were discussed and sometimes introduced in the text or in a new table after agreement.

The workshops were followed up by skype or telephone meetings to discuss the tables and visualizations in more detail and finalize the text. The surveillance report of 2018 was printed in August 2019.

#### **4.3. Real pilot in 2020**

To prepare for the pilot in 2020 new instructions were produced to clarify the role of the main responsible author of each chapter. It was also decided that all zoonotic chapters/diseases that are produced by authors from more than one sector should follow new guidelines in order to make all the zoonotic chapters more alike. To improve the OH work in all zoonotic chapters it is mandatory for the main authors to contact the other authors for a meeting to discuss the results of the previous year's surveillance and what to highlight in the report.

The process for the surveillance report started already in December 2019 with meetings to plan for the real pilot. Workshops were scheduled in January for the *Salmonella*, *Campylobacter* and VTEC/STEC chapters. Possible topics for the "In focus" sections were discussed at the meetings.

A list of authors to all the chapters in the report was sent out in December to everyone involved to verify that it was correct. Along with it the new instructions to the main responsible authors were sent out.





The process of the real pilot started in January, the texts and tables were made available for editing on a shared platform (OneDrive) and more detailed instructions on how to edit the texts and data was sent out to all authors.

Despite all extra work that was added (especially to the public health agency) because of the covid-19 pandemic the work with the report went according to schedule and the surveillance report of 2019 was printed in June 2020.

#### 4.4. FAIR data publication

To improve Findability and Reusability of the report, a permanent, unique identifier for the PDF report, with a clear license associated was needed. We identified that this could be provided by properly registering the report in a metadata portal. For data generated by public organizations in Sweden, we identified that the appropriate catalogue would be the Swedish data portal (dataportal.se).

Moreover, we chose *Campylobacter* spp. to test the publication of the data in the report in formats that are also Accessible and Interoperable. For all years for which the surveillance report is available (2006 – 2020), we collated data on the number of human cases of campylobacteriosis, as well as the number of chicken slaughter batches tested and positive for the pathogen. This dataset was then annotated with the Health Surveillance Ontology (HSO) and converted to RDF (Resource Description Framework), a standard model for data interchange on the Web. RDF is also the format used to annotated metadata in open data catalogues such as dataportal.se.

### 5. Results

#### 5.1. Achievements of the pre-pilot

During 2019 we tested the concept of working together across sectors during the production of the report. Workshops were arranged to facilitate the collaboration and resulted in a joint effort and the production of three OH-chapters (*Salmonella*, *Campylobacter* and VTEC) with partly new layout, changes in texts, tables and graphs.

A new section “In focus” was introduced in the OH-chapters. This highlighted section provides the possibility to describe (more in detail) some interesting or unusual finding from the previous year with a OH-focus. A slightly new structure to the text was also tested for the OH-chapters; the results section (of the previous year's surveillance) was moved to the beginning of the chapter and the section about legislation was moved to the end of the chapter. The changes were made to emphasize the results.

#### 5.2. Pilot in 2020

After the pre-pilot we decided to include all zoonotic agents in the OH work. This was communicated to all authors that were responsible for a zoonotic chapter with contributing authors from more than one sector. They were all obliged to have a meeting or at least check with the other sectors if they should highlight something from the previous year (outbreak or unusual findings) in an “In focus” section. The report for 2019 had “In focus” sections for the following chapters: *Salmonella*, *Campylobacter*, *Cryptosporidium*, STEC and Tularemia.

During 2020 it was decided to go back to the previous structure of the text (before the pre-pilot) to make all chapters look the same and make the report more readable. The entire report is published on SVA:s website (www.sva.se). In addition to that, separate pdf-files for each chapter were produced and are now available on the web together with information about specific diseases or surveillance activities (for both the animal health and public health sector).

#### 5.3. New Process

The new process that was introduced in the pre-pilot was be tested and evaluated in the real pilot 2020. The process can be described as follows:



- 1) A team should be formed consisting of at least an editor in chief, a project coordinator and technical support for creating figures and building the report. It is also good to have one contact person at each authority.
- 2) Preparation of instructions to authors involved in producing the report. Specific instructions should be prepared for main responsible authors of each chapter/disease to clarify their role.
- 3) Identification of relevant participants from each sector including appointing main responsibility for each chapter/disease (main responsible authors).
- 4) An initial meeting is recommended (to start the new process) where the involved parties are invited to share and explain their surveillance process and available data.
- 5) A web-meeting is held to plan the work and schedule times for workshops (separate meetings for each chapter/disease).
- 6) Workshops are held with a predefined agenda. Some examples of questions to discuss:
  - Major findings in the surveillance of each sector
  - Outbreaks
  - Unusual findings
  - Suggestions of topics to highlight in the "In focus" section
  - Improvements of last year's report, new figures or graphs to include, new available data
- 7) Online-meetings to finalise the text, tables and visualizations for each chapter.
- 8) Main author submits the chapter/disease to the editor team.
- 9) Evaluation of the process by sending out a questionnaire.
- 10) Adjust the process according to result of evaluation.

Examples of what should be provided by an editor-team:

- Instructions to authors and main authors (sent out in January in a mail to all authors)
- A template for chapters (for new chapters)
- Agendas for meetings and a checklist of what to prepare for meetings
- Shared platform between sectors to facilitate cooperative work on documents

#### 5.4. FAIR data publication

All available surveillance reports (2006 to 2020) were published with unique identifiers and license (CC-BY 4.0, creative commons, <http://creativecommons.org/licenses/by/4.0/>) in the Swedish data portal for open public information. Although it is not possible to support machine accessibility and Interoperability in PDF formats, this ensured that the surveillance reports, in their totality, are findable and reusable according to the FAIR principles.

One dataset compiling all surveillance data available in the reports, including surveillance methods annotated according to the CRAC principles<sup>28</sup> developed in ORION (consensus report annotation checklist), was also published in the Swedish data portal in both human readable format (comma separated values, CSV) and machine readable format (RDF). The latter is explicitly annotated with machine readable surveillance concepts from the Health Surveillance Ontology.

All permanent unique identifiers created are listed at: <http://datadrivensurveillance.org/campylobacter-surveillance-in-sweden/>

## 6. Implementation and impacts

The new process is already implemented in the work with the surveillance report and is not dependent on resources from the ORION project to continue. The planning work for the report of 2020 started already in December 2020.

<sup>28</sup> <https://oh-surveillance-codex.readthedocs.io/en/latest/5-the-dissemination-principle.html>



## 7. Reflections on the OH perspective

### 7.1. OH evaluation matrix

The grey shaded cells represent the level of integration in Sweden before the pilot, and green shading highlights the improvement after the pilot.

Steps in the surveillance pathway	Levels of integration			
	Level 1	Level 2	Level 3	Level 4
<b>Design, adjustment and optimisation</b>	Undertaken separately in each sector	Undertaken by a single sector for all surveillance components	Cross-sectoral consultation but undertaken separately in each sector	Undertaken by a cross-sectoral working group for OH objectives
<b>Sample/data collection</b>	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation across sectors	Joint activities across sectors
<b>Laboratory analysis</b>	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation of methods across sectors	Joint activities across sectors
<b>Data transfer /sharing</b>	No data exchange	Notification of unusual events only or when needed	Data exchange at regular intervals (e.g. yearly)	Ongoing data exchange; joint database and/or open access
<b>Data interoperability</b>	Unstructured data	Internal harmonisation (organization own coding practices)	Structural interoperability* across sectors	Semantic interoperability* across sectors
<b>Data analysis/interpretation – COLLABORATION</b>	Undertaken separately in each sector	Undertaken separately and collated by a single sector	Undertaken separately and then combined by a cross-sectoral working group	Jointly undertaken by multi-sectoral working groups
<b>Data analysis/interpretation – DATA STREAMS</b>	Interpretation of each data stream individually in each individual sector to sector specific objectives	Interpretation of multiple, sector specific data streams in each sector to sector specific objectives	Interpretation of multiple data streams from multiple sectors to sector specific objectives with cross-sector consultation	Interpretation to joint cross-sector objectives of multiple data streams from multiple sectors in cross-sector collaboration
<b>Outcome communication</b>	Undertaken separately in each sector	Joint dissemination in separate sectoral activities	Joint dissemination by a single sector	Joint cross-sectoral dissemination
<b>Prioritization and response</b>	undertaken separately in each sector	undertaken by a single sector for all surveillance components	cross-sectoral consultation but undertaken separately in each sector	undertaken by a cross-sectoral working group

### 7.2. Lessons learned 2019

The evaluation of the pre-pilot concluded a few things to continue with or to improve for the next year:

- Good experience with the “In focus” sections. We will continue with them.
- Good feedback overall on the OH-collaboration. Good to meet and have workshops. Good also to present to each other (between sectors) what type of data that are available, their limitations etc.
- Separate pdf-reports of each chapter should be produced next year. In that way the chapters can be used for more purposes than the surveillance report (e.g. on the institutions web sides next to disease information for a specific disease).



- A better solution for how to share and work together in documents between authorities is needed. The solution that was used did not enable all authors to use the track changes feature.
- It was not always clear who was responsible for what during the process. This uncertainty delayed the work sometimes and made it difficult to make decisions for a specific chapter. When more people are involved in the process of writing and deciding on layout of figures and graphs it is very important that there is someone that fully takes the responsibility of leading the work and communicating with the whole group and delivering text and data as expected.
- Experiences of shortage of time. It was difficult to manage all the new ideas and agree on all changes within the time limit for the report. Important to remember that changes take time.
- Sometimes difficult to agree on what data to present and how to present data from different sectors together in a good way.
- Important to remember that it takes time to cooperate. The expectations need to be at the right level for the first year.

### 7.3. Lessons learned 2020

A few changes were made according to the lessons learned in 2019:

- A team was created with a coordinator to keep track on the overall work, deadlines etc.
- The chapters were divided into groups with different deadlines. This was introduced to facilitate the reviewing process and worked out well.
- Responsibilities for main authors were made clear. Specific instructions were written.
- The work started earlier. A mail with information and link to all documents was sent out to all authors in January.
- Track changes worked when sharing files in OneDrive.
- More In-focus sections were produced, and all zoonotic chapters had some sort of collaboration between sectors.
- Separate PDF:s of each chapter were produced and are now available on the web together with information about specific diseases or surveillance activities.

### 7.4. SWOT-like considerations for

<b>Process</b>	
Things that worked very well during the study	The collaboration between sectors. New process and "In focus" sections.
Things that were difficult or didn't work well during the pilot study	Sometimes shortage of time. Especially towards the end of the process when the text of a chapter sometimes needs to be scrutinized by someone before it can be submitted to the editor. Problems of getting data from external sources in time.
<b>Outcome/product</b>	
Prospects for implementation of the pilot study outcome and further development opportunities	This pilot is already being implemented in the work with the Swedish surveillance report. We continue to work the same way as we did in the pilot.
Expectations that were not fulfilled and/or barriers for uptake	The production of FAIR datasets is not fully incorporated into the reporting process yet. The proof of concept <i>Campylobacter</i> dataset was compiled manually, and its regular updating at the time of every new report publishing is still foreseen as being a manual task.



## ANNEX II – ONE-HEALTH PILOT REPORT: UK

---

# Salmonella data sharing

### Background

The ORION project aims at establishing and strengthening inter-institutional collaboration and knowledge transfer in the area of surveillance data integration and interpretation, along the One Health objective of improving health and well-being. In the UK the Animal and Plant Health Agency (APHA) and Public Health England (PHE) collaborate directly and together with partner agencies across a number of One Health areas, including production of surveillance data outputs and through cross-government, multi-disciplinary committees and working groups. Regarding risk management and risk communication of especially zoonotic foodborne disease, collaboration has however historically been generally reactive around a specific zoonotic disease outbreak at the point where a suspect food vehicle or source of transmission of infection has already been identified through public health agency investigations. In order to improve knowledge and data sharing both within outbreak scenarios and to inform general surveillance and risk analysis of zoonotic disease outbreaks, a more proactive approach is considered necessary with more timely joint sharing of intelligence, including laboratory confirmed surveillance data.

*Salmonella* was chosen as the most suitable pathogen to focus on in a pilot study as not only is this pathogen an issue across the animal, food and human domains but it is also the area where we most often have outbreaks affecting animal and human populations therefore we have experience in the limitations of current processes. Most importantly, there is still a notable burden of salmonellosis in the UK associated with both imported and domestically produced food products so the opportunity to improve public health protection within the UK through enhanced collaboration between partner agencies involved in food safety was justified.

PHE implemented routine whole genome sequencing for characterisation of *Salmonella* isolated from all human laboratory confirmed cases in 2014, with application of this approach in place of traditional phenotyping methods for routine surveillance of *Salmonella* in 2015. This initially resulted in a 'gap' in ability to compare surveillance outputs from the UK human and animal population due to the introduction of WGS and lack of direct comparability with *Salmonella* isolates characterised by traditional phenotypic methods. Over time, APHA have started implementing whole genome sequencing for *Salmonella* in the livestock sector, based on the existing surveillance systems in place, however this still involves the challenges in harmonised analysis/interoperability of the data which currently, based on the approach used in the UK, involves the use of a shared single nucleotide polymorphism analysis pipeline and single linkage clustering and an integrated approach to interpretation of the surveillance data outputs. Whilst APHA has not fully implemented WGS for all isolates of *Salmonella* yet, an ongoing targeted approach is used to investigate some incidents in livestock/animal feedingstuffs. The molecular techniques that have been developed and are available to both institutions should allow for improved outbreak detection and more timely elucidation of epidemiological links between animal and human *Salmonella* incidents.



## Objectives

To develop an improved framework for sharing *Salmonella* surveillance data between APHA and PHE by:

- I. Identifying the aims and objectives of data sharing for each agency
- II. Identifying the metadata requirements for each agency.
- III. Developing an agreed framework for formally requesting and sharing *Salmonella* surveillance data between APHA and PHE and the subsequent communication of the outputs of the integrated approach.
- IV. Sourcing a suitable joint data sharing platform.
- V. Developing a joint analysis function between APHA and PHE and interoperability between datasets and interpretation outputs.

## 1. Expected outcomes

- a) Increased awareness of the surveillance data collated and used by each organisation, the data caveats and limitations for interpretation of the data and the levels of sensitivity around the respective data and implications for data sharing.
- b) Development of an agreed protocol for formally requesting and sharing *Salmonella* whole genome sequence based surveillance data between APHA and PHE, which may also form the basis for sharing of sequence data related to other pathogens.
- c) Development of easily transferable methodologies for shared analysis of *Salmonella* sequence based surveillance data.
- d) Elucidation and application of joint/integrated *Salmonella* whole genome sequencing cluster identification and prioritisation criteria.
- e) Identification of opportunities to expand and provide more actionable outputs on One Health aspects of existing surveillance reports through incorporation of integrated whole genome sequence and epidemiological data from human and non human *Salmonella* surveillance systems.

The pilot project had also aimed to identify a suitable, sustainable shared platform for future real time and streamlined sharing of zoonotic pathogen surveillance data however, due to organisational restructure, diversion of resource to outbreak investigations and the COVID-19 response as well as the subsequent initiation of the BeONE project in tranche 2 of the EJP which focussed on this specific aspect and therefore would have resulted in duplication of effort, this planned objective was not delivered.

## Performed activities

- a) Routine monthly meetings to discuss project progress
- b) Interim system implemented for all *S. Enteritidis* and *S. Typhimurium* sequences obtained through isolations from UK poultry to be routinely sent by APHA to PHE for analysis through the PHE pipeline. Return by PHE of assigned SNP address for each isolate sequence and, where relevant, indication of number of human cases genetically associated at the 5-SNP level and timelines of the case reports.
- c) Process for selection of un-sequenced isolates to send to PHE for sequencing where potential or confirmed epidemiological link with a salmonellosis outbreak or a specific microbiological cluster of interest detected according to agreed criteria relevant to the situation.





- d) Joint selection of specific initiatives to document and production of case studies, including defining joint prioritisation criteria for microbiological clusters consisting of human and animal/food isolates 'of potential public health concern'.
- e) Data Sharing Protocol produced and consultation undertaken.
- f) Memorandum of Understanding produced by data protection teams in both organisations and consultation underway.
- g) Wider stakeholder engagement/ discussions held.

## Results

### 1. Standardised data sharing framework developed

A Memorandum of Understanding (MoU) between APHA and PHE is in development. This is an organisation-level agreement which will outline the policy agreed between organisations that data on any pathogen can and should be shared for disease control purposes, and will harmonise the data protection arrangements around shared data.

A data sharing protocol has been developed to outline the agreed process to be followed for requesting and sharing *Salmonella* sequence data and associated metadata at defined levels of detail/granularity between PHE and APHA. The protocol is an instructional document that is specific to *Salmonella*, but may form the basis of similar protocols for other pathogens in future.

The protocol defines the data sharing procedure to be followed depending on the specific objective of a data sharing activity. Six data sharing objectives were jointly identified as follows:

- 1) **Routine sharing of target *Salmonella* serovars from surveillance under Regulation (EC) No. 2160/2003 in poultry and from passive surveillance in other animal species where available**  
Sharing of sequence data relating to all isolates of *S. Typhimurium* and *S. Enteritidis* from samples received by APHA where these are sequenced (all poultry isolates and in some instances in other livestock species) to facilitate early identification of human-animal clusters and potential epidemiological links of the serovars causing the highest burden of salmonellosis in the UK.
- 2) **Ad-hoc characterisation of isolates**  
For initial comparisons between animal and human isolates detected within a specified relevant frame of reference (e.g. time-period, broad location, sampling setting and / or antimicrobial resistance profile).
- 3) **Incident detection and risk assessment**  
Enhanced sharing of data relating to human-animal-food whole genome sequence cluster(s) to further assess epidemiological links/ carry out risk assessment for potential outbreak events.
- 4) **Outbreak investigations**  
To facilitate/inform incident investigation and management once a human-animal-food epidemiological link is highly suspected or confirmed.
- 5) **Research-related requests**  
To provide necessary evidence to answer a clearly defined research question of relevance to animal and/or human health.
- 6) **Routine surveillance reporting**



To facilitate harmonised reporting of human and animal surveillance data within existing surveillance reports/outputs.

The protocol outlines the type of data, including level of detail of metadata required for interpretation/action, to be shared and the process to request and share these data, depending on the primary data sharing objective as detailed above because frequency of data sharing and the type of data shared must be proportional to the needs of the current sharing objective so that potentially sensitive data are not unnecessarily shared.

Data sharing objectives may evolve over time and according to specific situations/objectives, and the protocol has been developed to be suitably flexible to account for this. If the data sharing objective changes, then so too will the associated procedure for data sharing, including the types of metadata shared. This ensures a tiered approach to data-sharing and avoids the sharing of more sensitive metadata until it becomes essential to the requirements for the protection of public health. Crucially, the protocol outlines the requirement for joint interpretation of the analysis of shared data at all stages by APHA and PHE and the need for multi-agency risk assessment to determine the need for escalation of public health protection measures.

To harmonise data requests and the format of shared data, standard request and data templates have been developed. The protocol will be implemented following final sign-off by APHA and PHE.

## 2. Joint data analysis function

Currently, to ensure sensitivity in the surveillance system, a cluster of human disease cases that may indicate a potential outbreak event is defined as a whole genome sequencing cluster of two or more cases for which the whole genome sequences of the isolates are within a defined single nucleotide polymorphism (SNP) threshold (usually the 0, 5 or 10 SNP level). Clusters can then be characterised by the number of SNP differences between isolates, geography, case demographics, cluster size and growth rate etc. On average around 60 (range 12 – 120) discreet 5-single nucleotide polymorphism clusters, based on single lineage clustering, either newly reported or previously detected but with new cases reported, are assessed by the Public Health England Gastrointestinal Pathogens Unit on a weekly basis over a 12 month period. Due to the large volume of data, a prioritisation algorithm is used to identify priority human clusters for more detailed assessment and a clear understanding of the priorities and definitions used between organisations to define clusters and prioritise for investigation (risk factor based) is required.

Well defined criteria utilised in the PHE prioritisation algorithm include, but are not limited to, cluster size, cluster growth rate, severity of infection through proxy indicators such as proportion of non-faecal samples, proportion of cases in vulnerable age groups, indications of cross border/multi-country involvement etc. APHA are in the process of moving away from serotyping and using WGS more routinely, working with PHE to develop prioritisation algorithms that are complimentary to PHEs in anticipation of a larger number outbreaks being identified. Similar criteria will also be used in defining circumstances that a national control team might be triggered by APHA.

## 3. Case studies developed

- *Salmonella* Enteritidis in UK poultry production: routine sharing for risk assessment, outbreak detection/management/resolution and improved understanding of transmission and persistence at primary production.
- *Salmonella* Typhimurium outbreak investigation: source attribution and targeting of outbreak investigation efforts in complex multi-source outbreaks with presentation of outputs in format readily understandable by non-expert stakeholders (including presentation of phylogenetic



information incorporated with epidemiological information in minimum spanning trees as indicated best approach by stakeholders).

- *Salmonella* Infantis: phenotypic testing, epidemiological criteria and AMR based criteria for data sharing.
- Raw pet food study: research study defining criteria for OH research initiatives and also as case study for 'vehicle' based criteria for data sharing.

## Implementation and impacts

The improved data sharing process developed within the ORION project will be implemented in ongoing *Salmonella* surveillance activities involving APHA and PHE after the ORION project has finished. The principles of interpretation of the data and communication channels as well as joint prioritisation criteria have been included in an updated version of the internal PHE document 'Guidance for management of WGS clusters of GI infections'. A workshop with food safety authorities is planned to discuss interpretation of the analyses and translation of outputs into outbreak detection/management/resolution and assessment of public health/food safety policy.

Utilisation of the *Salmonella* data sharing protocol will be accompanied by ongoing evaluation of its effectiveness, initially an evaluation after 6 months of full implementation and thereafter a full review after 12 months where further consideration will also include the data sharing platform. Refinement where needed, under different data sharing scenarios as they arise will be implemented.

Opportunities to expand on One Health aspects of the surveillance outputs already routinely produced jointly by APHA and PHE have been identified as a result of increased awareness of surveillance processes between APHA and PHE. This will be taken forward into future versions of the annual reports in a stepwise process.



## Reflections on the One Health perspective

### 1. The OH evaluation matrix

The grey shaded cells represent the level of integration in the UK before the pilot, and green shading highlights the improvement after the pilot.

Steps in the surveillance pathway	Levels of integration			
	Level 1	Level 2	Level 3	Level 4
<b>Design, adjustment and optimisation</b>	Undertaken separately in each sector	Undertaken by a single sector for all surveillance components	Cross-sectoral consultation but undertaken separately in each sector	Undertaken by a cross-sectoral working group for OH objectives
<b>Sample/data collection</b>	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation across sectors	Joint activities across sectors
<b>Laboratory analysis</b>	Undertaken separately in each sector	Undertaken by a single sector for OH objectives	Harmonisation of methods across sectors	Joint activities across sectors
<b>Data transfer /sharing</b>	No data exchange	Notification of unusual events only or when needed	Data exchange at regular intervals (e.g. yearly)	Ongoing data exchange; joint database and/or open access
<b>Data interoperability</b>	Unstructured data	Internal harmonisation (organization own coding practices)	Structural interoperability* across sectors	Semantic interoperability* across sectors
<b>Data analysis/interpretation – COLLABORATION</b>	Undertaken separately in each sector	Undertaken separately and collated by a single sector	Undertaken separately and then combined by a cross-sectoral working group	Jointly undertaken by multi-sectoral working groups
<b>Data analysis/interpretation – DATA STREAMS</b>	Interpretation of each data stream individually in each individual sector to sector specific objectives	Interpretation of multiple, sector specific data streams in each sector to sector specific objectives	Interpretation of multiple data streams from multiple sectors to sector specific objectives with cross-sector consultation	Interpretation to joint cross-sector objectives of multiple data streams from multiple sectors in cross-sector collaboration
<b>Outcome communication</b>	Undertaken separately in each sector	Joint dissemination in separate sectoral activities	Joint dissemination by a single sector	Joint cross-sectoral dissemination
<b>Prioritization and response</b>	undertaken separately in each sector	undertaken by a single sector for all surveillance components	cross-sectoral consultation but undertaken separately in each sector	undertaken by a cross-sectoral working group

\* Structural or syntactic interoperability refers to the format of data exchange – datasets are standardised into a common, agreed format, for sharing purposes. Semantic interoperability is concerned with ensuring the integrity and meaning of the data across systems. Data is semantically interoperable when it is annotated with the relevant context of data collection, in machine readable formats (schemas or ontologies). Semantic interoperability is particularly important in OHS in order to allow data reuse across sectors, and reuse of data for research and knowledge discovery, while preserving the original context of the data



## 2. Lessons Learnt

- 1) Data sharing agreements are complex and require plenty of time to set up. It is important not to underestimate how long it takes for even relatively straightforward agreements to be put in place, even when these are between national partner agencies.
- 2) Different institutes have different internal policies regarding data protection and 'data ownership', and it is important to take the time to understand similarities and differences fully before developing a formal framework for data sharing. Considerations are needed with respect to the ownership of results of Whole Genome Sequencing analysis. It is not clear who owns the analytical results – do they belong to the curator of the comparison data, the sample provider, the sample funder etc.? There are additional considerations regarding data obtained from samples collected and tested by industry.
- 3) Surveillance systems differ markedly within institutes for different pathogens in different species and between institutes for the same pathogens. A full understanding of all parties entering a data sharing framework that there is the need for consideration of these differences and incorporation of the strengths/ caveats and data limitations in all joint outputs is essential. All sectors involved in the analyses of surveillance data (human, animal, food) should be involved in writing up outputs that will be disseminated internally within government and externally to stakeholders to ensure context is preserved and interpretation is accurate/robust. Problems can arise when trying to communicate the uncertainties/unknowns surrounding the results of a joint analysis – surveillance is based on collecting samples and we do not always know what we don't know.
- 4) It is vital to have clearly defined genetic threshold parameters and epidemiological parameters (geographic, temporal etc.) in order to carry out the optimum analysis and draw appropriate conclusions from shared analyses.
- 5) Dynamic/ real time data sharing is key to successful use of the data for the protection of public health, especially in outbreak related contexts.
- 6) Risk communication, including with other relevant government agencies and bodies involved in the protection of public health requires a clear agreed communication framework and the tailoring of outputs for non-expert audiences



### 3. SWOT-like considerations

<b>Process</b>	Max 2-3 points in each
Things that worked very well during the study	<p>Sharing of pathogen expertise (microbiological and epidemiological) by experts from the animal and public health sectors to inform approach to outbreak investigations and risk management</p> <p>Identification of criteria for data sharing with analysis through a single pipeline and interpretation facilitated through integrated WGS/epidemiological data and use of annotated phylogeny trees</p> <p>Demonstrable enhancement of source attribution through integration of sequencing and epidemiological data from both the human and animal surveillance systems including related surveillance data from slaughterhouse and food sampling – facilitated by the access to multiple isolates to inform phylogenetic and epidemiological interpretation</p>
Things that were difficult or didn't work well during the pilot study	<p>Didn't work well: Involvement of data protection / information management teams in the two organisations – addressing the sharing of whole genome sequencing data was difficult to effectively communicate the complexities which fell outside experience and available templates held by these teams for data sharing</p> <p>Difficult: The selection of data to share where only un-sequenced data available, but this in large numbers/ amounts – criteria for selection</p> <p>Difficult: Presentation of outputs in a format understandable to wide range of stakeholders</p>
<b>Outcome/product</b>	
Prospects for implementation of the pilot study outcome and further development opportunities	<p>Case studies demonstrating the utility of real time sequence data and information sharing</p> <p>Routine sharing process now in place for target serovars but to be extended to non-target serovars</p> <p>Expert workshop to confirm approach to joint analyses and joint prioritisation criteria</p> <p>Development of criteria to be used in defining circumstances that a national outbreak control team might be triggered by APHA in response to defining an outbreak in animals without associated human disease cases</p> <p>Data sharing platform to be considered for future implementation where shared analysis tools including prioritization algorithm can then be developed</p>
Expectations that were not fulfilled and/or barriers for uptake	<p>Due to other resource pressures and then announcement of re-structure of one participating agency in the pilot, consideration and adoption of shared platform not undertaken</p> <p>In absence of shared data platform development of joint analysis tools not feasible so been put on hold – for future development</p>