ORION
WP1 requirement analysis

1 Background

The ORION project aims at the prototypic implementation of an integrated strategy for long-term consolidation and harmonization of One Health Surveillance (OHS) in the European public and veterinary health landscape. This strategy is called ORION "OHS harmonization framework" and consists of three pillars:

- An "OH Surveillance Codex" (WP1) - a high level framework for harmonised, cross-sectional description and categorisation of surveillance data covering all surveillance phases and all knowledge types.

- An "OHS Knowledge Hub" (WP2) - a cross-domain inventory of public/veterinary health data sources as well as algorithms and tools supporting OHS data generation, integration, analysis and interpretation. This includes an updateable inventory of OHS best practice solutions illustrating opportunities and challenges for European and national stakeholder.

- "OHS Infrastructure Resources" (WP3) – technical and infrastructural resources that form the basis for successful harmonization and integration of surveillance data and methods. These infrastructural resources include harmonized data standards, software libraries, ontologies, terminology mappings, software tools supporting the adoption of the "OH Surveillance Codex".

With this requirement analysis WP1 lays the foundation for the development of a high level framework for harmonized description and categorization of One Health Surveillance data. This OHS Codex is envisioned as a guidance document supporting better understanding of the specific meaning and constraints of surveillance data from the different OHS domains. As one of the results this document illustrates that the lack of resources supporting mutual understanding and interpretation of data from the different OHS domains is a major bottleneck for meaningful OHS data analysis.

The envisaged OHS Codex will therefore be composed of two components:

1. *The One Health Surveillance Codex (OHS Codex) guidance document and*
2. *A community driven "ORION glossary".*

Both components will in the future facilitate and contribute to mutual understanding between OHS sectors (i.e. animal health, public health, food safety, et cetera).

This requirement analysis report was developed within the 1st year of the ORION project and might be updated in the course of the ORION project if new findings become available.

2 Targeted community, main stakeholders and organizations

This sections answers the question, who might benefit from the envisage OHS Codex.
According to the understanding of the ORION WP1 members the OHS Codex will be useful for organizations and researchers that are

- involved in One Health surveillance implementation
- involved in One Health surveillance data reporting
- involved in One Health data harmonization and standardization

Specifically, this includes

- members of the EJP consortium
- other national authorities involved in OHS
- European authorities and institutes, as e.g. EFSA, ECDC, EuroStat
- Other stakeholders in OHS, as e.g. research organizations

3 Requirement analysis methodology

To identify the needs of the community / stakeholders and to identify existing resources that could support OHS harmonization WP1 applied the following methodologies:

- A thorough study of scientific literature and web-based information sources
- An online survey collecting information from all EJP project members
- Face-to-face interviews with domain experts
- Collection of "User Stories" from domain experts

4 Requirement analysis results

4.1 Review of scientific literature and web-based information sources

This information research was performed to identify frameworks that support the harmonization of metadata from cross-sectoral and multi-disciplinary domains. Our strategy was to look for already existing solutions for metadata harmonization. This included to search for solutions outside the OHS domains, specifically in domains that have long-term experience with cross-sectoral data analysis and rely heavily on valid and precise metadata. Once the existing frameworks were identified we highlighted the key features of the most relevant framework.

4.1.1 Frameworks for OHS data harmonization

EFSA and ECDC have made substantial efforts to harmonize data collection and reporting within their domains. These achievements are important assets that all future OHS harmonization efforts needs to consider, for example the Data Collection Framework (DCF[1]) and the SIGMA project (SIGMA[2]) from EFSA as well as the European Surveillance System (TESSy[3]) from ECDC. Other ongoing joint efforts of theses stakeholders support OHS data harmonization as well, for example the joint molecular typing database. Another

[1] https://www.efsa.europa.eu/en/supporting/pub/en-444

[2] https://www.efsa.europa.eu/en/supporting/pub/en-1428

[3] https://ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy

collaborative effort is the compilation of the yearly European summary reports (EUSRs) on *trends and sources of zoonoses, zoonotic agents and food-borne outbreaks* and on *antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food*. However, the current practice of provisioning OHS metadata is still not optimal for a truly integrated OHS data analysis and impedes an easy compilation of e.g. the EUSR (Reference: Bolaert et al., 2016 and summary reports). Moreover, there is currently no strategy for a harmonized description of metadata that could be adopted by other domains relevant for OHS analysis such as e.g. environmental science.

Insufficient data and metadata harmonization is a well known issue also in other domains not related to OHS. Especially official statistical organizations are confronted by this challenge for a long time. Efforts to address this challenge were initiated by a joint UNECE / Eurostat / OECD working group that developed the Generic Statistical Process Model (GSBPM[4]). The GSBPM is a process modelling framework as a basis for statistical organizations to agree on standard terminology and to aid their discussions on developing statistical metadata systems. Independent of that another metadata harmonization framework was developed by the Data Documentation Initiative (DDI). Their framework is called the DDI-Lifecycle model (DDI-Lifecycle[5]) and is highly complementary to GSBPM, but the scope of the model is rather focused on research than on official statistics[6].

The structure of GSBPM v5.0 is shown in Figure 1.

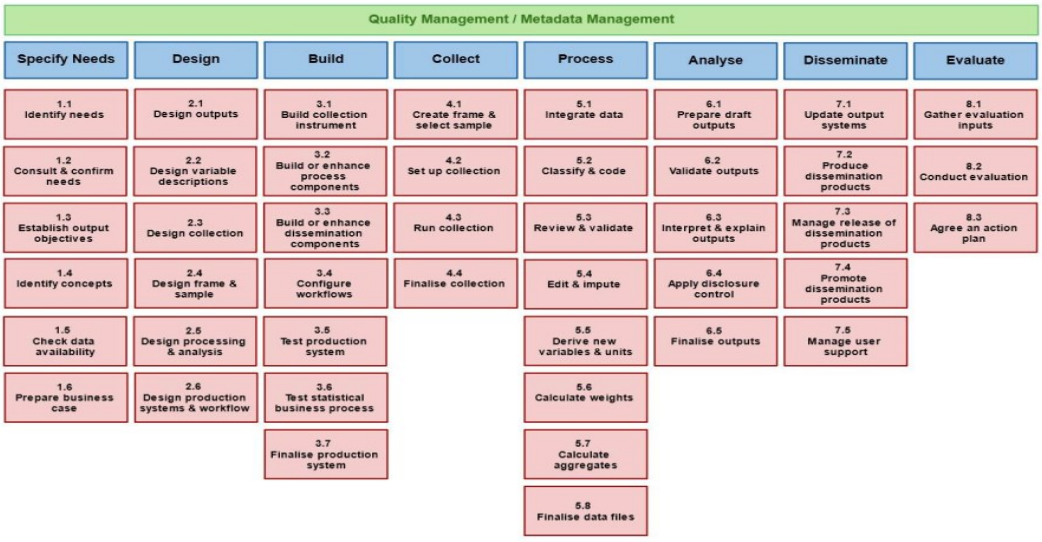

Figure 1: The GSBPM is a flexible and generic framework that can be used to provide metadata on data based on all process steps needed to produce these data. It comprises of a set of major process steps each with a pre-defined set of sub-processes.

[4] https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0

[5] http://www.ddialliance.org/Specification/DDI-Lifecycle/

[6] http://www.ddialliance.org/Specification/DDI-Lifecycle/

Below, we have outlined the key features of GSBPM that can be used to assess whether GSBPM could be a suitable model to support harmonized categorization of OHS metadataa.[4,5,7]

- Generic framework that can be used to map cross-sectoral and multi-disciplinary metadata to processes that are needed to produce data in official statistics
- Allows description of all steps influencing the data outputs, independent of the data source
- Provides integrated description of these processes
- The terminology used to name and describe each process is generic, and the terms do not necessarily correspond to the terminology used in any country (Cotton and Gillman)
- Intended to aid the development of harmonized statistical metadata systems
- It is not a linear model, meaning that the processes can be used in the order they are needed for the specific purpose
- Widely adopted by statistical organizations worldwide
- Facilitates communication and collaboration within and between organizations
- Inherent quality and metadata management
- Can be used for several different purposes → widely applicable
- Flexible framework that is neither too restrictive nor too abstract
- Can be implemented as ontology as proposed by Cotton and Gillman, 2015 [7]
- Can also be applied to the maintenance of statistical registers
- Also applies to processes where existing data are revised or or time-series are re-calculated
- Compatible with other official standards such as DDI Lifecycle and SDMX (see below for more information on DDI Lifecycle and SDMX)

As part of the literature review we also explored which technical resources exists that would support the adoption of GSBPM in the OHS domain. Specifically we explored if the DDI Lifecycle could be recommended.

DDI Lifecycle[8] is a feature rich XML schema and metadata framework. It was mainly developed for social sciences and the community extends their efforts to open the standard to other domains. To explore if the application of the DDI Standard could be an achievable goal within WP1 we defined at first key features which DDI should provide:

- a structured and GSBPM compliant metadata approach
- the possibilities to integrate ontologies for linking of data
- interoperability & harmonization of surveillance data
- open source tools to maintain and adopt the standard
- an adaptable feature rich schema

In our analyses we found several advantages of DDI:

- works with the FAIR data principles
- supports versioning
- possibility to link data of different data sources
- possibility to integrate ontologies with the RDF standard

[7] Cotton, Franck, and Daniel W. Gillman. 2015. "Modeling the Statistical Process with Linked Metadata." Ceur Workshop Proceedings 1551. CEUR-WS.

[8] Description of DDI Lifecycle 3.2: http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/

- active community (DDI Alliance)
- good mapping to other important metadata standards like:
    - the Dublin Core
    - ISO/IES 11179 for Data Registries,
    - DataCite metadata schema
    - SDMX
    - GSBPM (Fig.2)

However our analyses also showed several major disadvantages:

- DDI has a detailed high complex schema which is hard to adapt even with software tools
- a common understanding of terms and data structures is mandatory to apply the standard
- tools for DDI are already developed[9], but there is a lack on actively maintained open source solutions. The only "mature" software product (Collectica[10]) is a commercial product.
- DDI needs to be extended to cover specific concepts from the One Health Domains

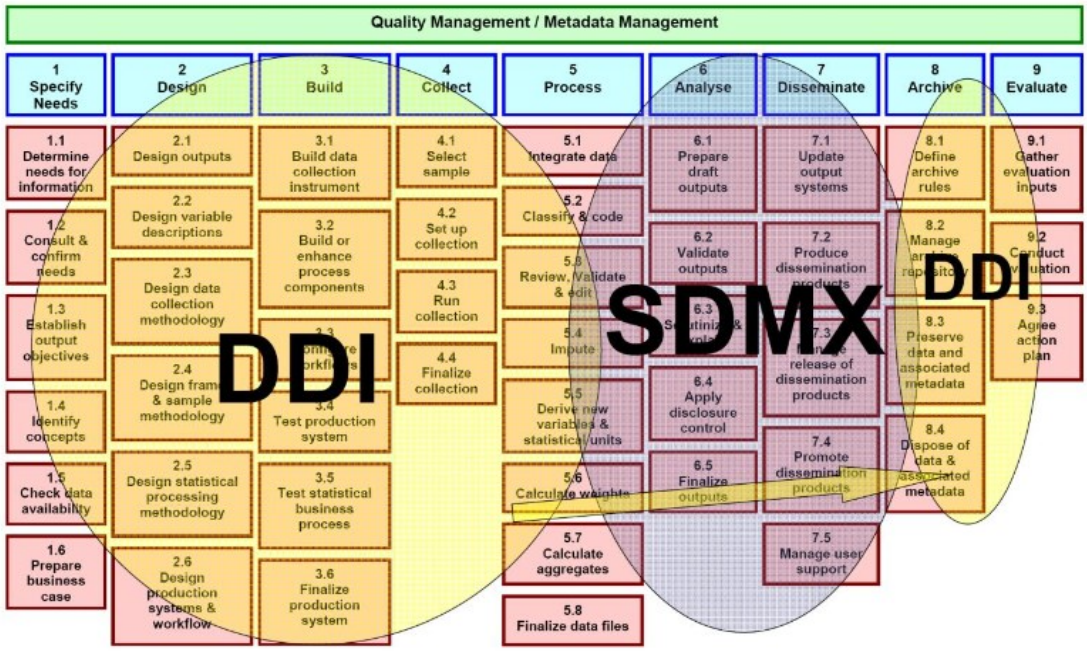

Figure 2: This graphic present the overlaps of the GSBPM schema with SDMX and DDI [8]

4.1.2 Implications for the OHS Codex

Metadata standards and formats are used to manage and maintain data collections. Metadata standards ensure also data quality and facilitate the development of software tools which supports users e.g. with data retrieval, data provisioning or data analysis. Therefore, well described metadata schema are an important resource that could ensure the right

[9] http://www.ddialliance.org/resources/tools

[10] https://colectica.com/

interpretation for e.g. surveillance data. However the implementation of newly developed harmonized metadata standards into domains where domain-specific standards already exist is difficult and practically not feasible. Therefore there is the need to develop a strategy that allows to bridge and harmonize existing metadata schema from the different domains causing minimal implementation efforts for all involved parties.

A solution towards this objective would be the establishment and adoption of a so called "Consensus Reference Annotation Schema" that each existing metadata schema could be mapped to. In the course of the ORION project WP1 will explore if the existing GSBPM framework can be used for this. This will also be tested in dedicated WP1 pilot studies.

4.2 Survey on OHS data harmonization

The survey for WP1 on OHS data harmonization was prepared together with WP3. Questions were drafted by both WPs and finally merged. The survey was implemented in LimeSurvey (https://www.limesurvey.org/) and sent out to all scientific representatives of all One Health EJP consortium partners. The survey was open for one month before results were collected. Only complete answers were taken into consideration for the evaluation of the survey. Detailed results from the survey can be found in the annex "WP1_WP3_survey results.xlsx".

The survey was answered by six experts from six different EU countries (Sweden, Norway, Austria, Spain, The Netherlands and Denmark). The role of these experts was shared between researchers, epidemiologists and veterinarians. Three respondents were from multi-sectoral institutions, while the other three were from institutions with solely animal health or public health focus. The respondents listed only two initiatives, namely SIGMA and ComAcross, when they were asked if they were aware of any initiatives that focus on the harmonization of OHS data.

Most experts considered harmonized metadata for OHS data as the most important aspect of data harmonization (Figure 3). Other aspects such as harmonized data exchange formats, methods for data analysis and tools for cross-sectoral translation were considered similarly important, but less important than harmonized metadata (Figure 3). None of the experts were aware of the Generic Statistical Business Process Model (GSBPM) or the Data Documentation Initiative (DDI).

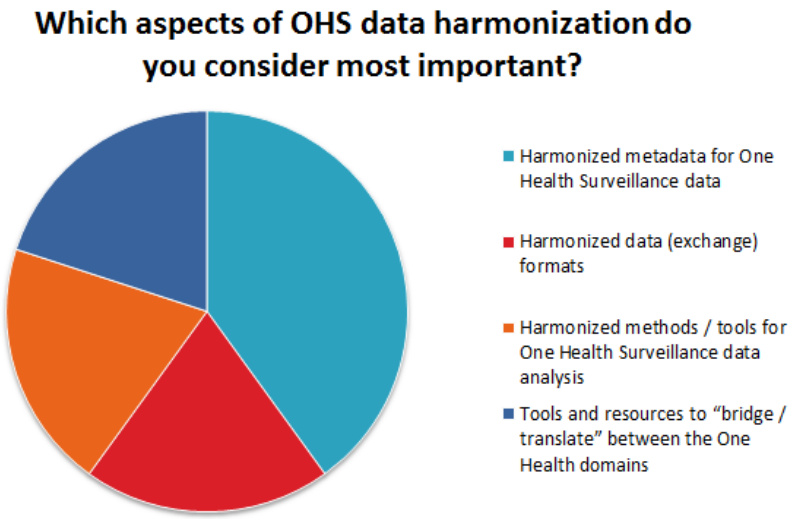

Figure 3: Presentations of results on importance of certain aspect for OHS data harmonization from ORION survey.

4.3 Interviews with collaboration partners from ARDIG and GOHI

Interviews were held with our collaboration partners from the EJP project "Antibiotic Resistance Dynamics" (ARDIG[11]) and the "German One Health Initiative" (GOHI[12]). Both projects (ARDIG and GOHI) address OH surveillance of antimicrobial resistance (AMR) in zoonotic and indicator bacteria. The ARDIG project also covers surveillance on antimicrobial use (AMU). The interviewees were asked about their respective problems when analyzing AMR and AMU monitoring/surveillance data from different domains.

Current limitations that hinder integrated OHS analysis identified during the interviews with these domain experts were:

- Differences in definition of surveillance and monitoring systems
- Differences in data sources (e.g. reimbursement data, sales data, pharmacy data, prescription data…)
- Differences in antibiotic panels between human and animal sectors
- No standardization panels between drug / pathogen in human sector
- Different Antimicrobial Susceptibility tests (AST) such as dilution methods, disk diffusion methods, ETEST…
- Different guidelines used in AST (e.g. DIN, EUCAST, CLSI…)
- Different criteria (ECOFFs or CBP)
- Different AMU units (e.g. Tonnes of antibiotic, DDD, DDD/kg, DDD/1000 patients…)

[11] https://onehealthejp.eu/structure/jrp2-ardig/

[12] https://www.gohi.online/GOHI/EN/Home/Homepage_node.html

4.4 "User Stories" for OHS requirement analysis

"User Stories" from ORION consortium members were collected during the ORION Pilot Study Workshop in Copenhagen (November 2018). In a "User Story", each person describes their role within the OHS community, as well as their needs to achieve a specific aim. We have also extracted some user stories from meetings with EFSA and ECDC. Detailed results from the user stories can be found in the annex "User_stories.xlsx".

4.4.1 Personas of the OHS community

A persona is a name for a group of peoples that perform similar activities in the light of the aim in scope. Within this analysis the following personas were identified:

- Researcher/ data scientist
- Epidemiologist/ microbiologist
- Data provider
- Advisor on control of zoonosis and antimicrobial resistance (AMR)
- Stakeholder (e.g. EFSA/ ECDC)

4.4.2 Detailed user stories

- A researcher/ data scientist wants to perform integrated analysis of OHS data with the general aim to improve health and well-being. To accomplish this, resources that support the mutual understanding of surveillance data (specific meaning and constraints) from the different OHS domains are needed.
- A researcher/ data scientist wants to perform analysis of OHS data to study e.g. disease dynamics. To accomplish this, data needs to be fair, easily available, timely and available as open linked data.
- A researcher/ data scientist wants to use specific OHS datasets for analysis and/or link OHS data to other datasets (e.g. transportation data, lad use data). To accomplish this, un-aggregated or raw OHS data needs to be accessible (e.g. in databases)
- A researcher/ data scientist wants to spend more time analyzing data and less time cleaning and managing data. To accomplish this, data needs to be provided in a harmonized format with metadata that is easily understandable and complete.
- A researcher/ data scientist wants to enhance research and collaboration for improved OHS surveillance. To accomplish this, knowledge sharing and efficient communication has to be supported.
- An epidemiologist wants to exchange data and knowledge with experts from other OH domains for more informed outbreak investigation and risk management from farm to fork. To accomplish this, the epidemiologist wants a platform/ framework for knowledge exchange, communication and training.
- An epidemiologist wants to perform high-quality outbreak investigation. To accomplish this, easy to follow guidelines on the software/ hardware needed and on the interpretation of OHS data in an epidemiological context are needed. In addition it would be valuable to have a test dataset that can be used to validate the used software/ hardware and interpretation criteria.
- An epidemiologist wants to enhance the interpretation and use of OHS data to support better decision making and evaluation of control measures. To accomplish this, the mutual understanding of the constraints and meaning of data from different OHS domains needs to be improved.
- A data provider (e.g. meat producer) wants to be provide data to support reliable OHS without being afraid of data not being used for purpose. To accomplish this, it has to be

ensured that the data providers are protected against data misuse and wrongful data interpretation that lead to unsubstantiated corrective actions.

- A data provider (e.g. fish processing company) wants the provided data to be anonymous and protected against data fraud. To accomplish this, it has to be guaranteed that sensitive data is only saved in official databases and not shared through unofficial channels.
- An advisor wants to enable better decision making and evaluation of control measures. To accomplish this, OHS data from farm to hospital needs to be available and reliable.
- An advisor wants to protect animal and public health by using OHS data to the best possible extend. To accomplish this, an integration of all available expertise needs to be given.
- A stakeholder (e.g. EFSA employee) wants to perform valid analysis of temporal and spatial trends. To accomplish this, EU harmonised monitoring schemes are needed that provide valid and precise data without biases.
- A stakeholder (e.g. ECDC/ EFSA employee wants to collectively analyse and compare data from different databases (e.g. EFSA's DCF and ECDC's TESSy) without data transformation. To accomplish this, data and metadata in databases from different sectors have to be interoperable.

4.4.3 Current frustrations extracted from user stories

The following frustrations were identified for the OHS community:

- Unavailability of OHS data (no access, not timely, unavailability of un-aggregated data)
- Inconsistency of OHS data from different domains (no harmonization)
- Incomplete metadata (difficulties in understanding data)
- Data sensitivity/ data misuse
- Insufficient guidelines for data analysis/ tools and interpretation
- Lack of understanding between domains (e.g. different terminologies, how surveillance systems work, etc.)
- Lack of communication, collaboration, and knowledge sharing

5 Conclusions

5.1 Frameworks for OHS data harmonization

Process modelling frameworks have been developed to support the development of harmonized metadata between and within statistical organizations. In ORION, such a framework can be used to address the requirement of harmonized and complete metadata to support mutual understanding of OHS data from different domains. One suitable framework, the GSBPM, supports the two key aims of WP1, namely the categorization of cross-sectoral and multidisciplinary OHS metadata and the mutual understanding between and within different OHS domains. The ability to adapt the GSBPM to all processes that lead to the production of official statistics should also facilitate the adoption to OHS and e.g. the production of the EUSRs. The genericity and flexibility also entails that the GSBPM can also be adopted by other OHS domains that are not part of the ORION consortium, e.g. environmental or social sciences. Other valuable features of the GSBPM are the compatibility to other standards and the possible implementation of GSBPM as part of an ontology. The implementation of GSBPM into and ontology branch is a collaborative effort between WP1 and WP3 in ORION that is envisioned to support OHS data interoperability. Based on the presented features it was concluded that the OHS Codex will propose to adopt the GSBPM as a basis for the Consensus Reference Annotation Schema.

5.2 Implications for the OHS Codex

Metadata standards are means to address the need for data harmonization and interoperability. One of the existing metadata standards - the DDI Lifecycle - was evaluated for its uptake potential in the OH domain. Our analysis showed that the framework behind DDI Lifecycle gives a good general conceptualization of data life-cycles. However our analysis and discussion with DDI experts revealed that the required efforts to adapt DDI to specific needs of the OH domain would most likely exceed the benefits. This is caused to some extend also from the lack of flexible, open source software solutions that could be used to maintain an OHS compliant DDI version.

We also concluded that the development of such a standardized metadata schema is outside the scope of the high level harmonization framework developed within WP1. However, WP3 will develop an ontological framework for OH surveillance that will address OHS data interoperability. This proposed framework by WP3 will also aim at implementing GSBPM as an ontology branch.

5.3 Survey on OHS data harmonization

The survey results emphasized the need for harmonized metadata standards within the OHS community. Results also showed that respondents were not aware of any initiatives or projects that could be considered duplication of work when compared to ORION. Only two projects that thematise OHS data harmonization were mentioned namely SIGMA and ComeAcross. The SIGMA project was initiated by EFSA to harmonize existing data models and improve technical developments for data submission, validation, analysis and reporting across the different data collection activities which are focused on the animal populations affected by disease outbreaks or subject to disease surveillance. The SIGMA project can therefore be considered as an animal health focused project, rather than a One Health initiative.

The ComAcross project focuses on sharing, collaborating and exchanging One Health experience in South East Asia. Although the project focuses on cross-sectoral integrations of OHS actions, there seems to be no specific action taken towards harmonization of surveillance metadata or data. One expected outcome of the project is the improvement of data management procedures and methods, however it is unclear how this will be achieved or if this also includes approaches to categorize or describe surveillance metadata.

5.4 Interviews with collaboration partners from ARDIG and GOHI

The results from interviews with our collaboration partners from ARDIG and GOHI emphasized the need for harmonized laboratory methods, standards and units for OH surveillance/monitoring of AMR and AMU. The harmonization of OHS methodologies will not directly be addressed by WP1, but could potentially be facilitated through a more thorough description of a GSBPM-based materials and methods section in future surveillance reports. The availability of different methods and tools within the OHS domains will further addressed in ORION WP2, where an inventory of best practice methods will be created.

5.5 User stories for OHS requirement analysis

The user stories provided the most detailed information about current frustrations and the consequent needs within the OHS community. From the provided results it became apparent

that the lack of understanding between the different OHS domains is a current source of frustration. Since it is a central aim of WP1 to support mutual understanding between the different OHS domains it was crucial to address this frustration by the creation of an ORION glossary. To facilitate the understanding on One Health Surveillance terminology between the ORION members (including Animal Health, Food and Public Health domains), we carried out a literature review to document definitions to develop the ORION-Glossary. Terms for the ORION Glossary were suggested by domain experts and definitions were adapted from stakeholders such as EFSA and ECDC, if available. The list of terms was extended to include terms used within the inventory created by WP2 and the ontological framework developed by WP3.

The frustration regarding the lack of knowledge exchange and communication between members of the OHS community from different domains will be addressed in ORION WP4. WP4 aims at developing a Knowledge Hub for knowledge generation and sharing.