Drivers for the development of an animal health surveillance ontology (AHSO)

FC DÓREA[1], F VIAL[2], K HAMMAR[3,4], A LINDBERG[1], E BLOMQVIST[4], CW REVIE[5]
[1]Department of Disease Control and Epidemiology, National Veterinary Institute, Sweden; [2]Epi-Connect, Skogås, Sweden. SE-142 51; [3]Department of Computer Science and Informatics, Jönköping University, Sweden; [4]Department of Computer and Information Science, Linköping University, Sweden; [5]Atlantic Veterinary College, University of Prince Edward Island, Canada

*fernanda.dorea@sva.se

Abstract
Comprehensive reviews of veterinary syndromic surveillance have pointed out the hindrances to integration and interoperability among systems using different data sources. Discussions with syndromic surveillance experts in the fields of animal and public health, as well as computer scientists from the field of information management, have led to the conclusion that the solution needed is an ontology. Here we describe the advantages of the approach, and the methodological steps planned around the launch, in 2017, of a community of developers and users of the Animal Health Surveillance Ontology (AHSO). Development is expected to be community driven and the final product will be open-access.

Keywords: *syndromic surveillance, classification, vocabulary, terminology, standards.*

Introduction
In 2011 Dórea *et al.* (1) provided the first comprehensive review of veterinary syndromic surveillance (VSS), highlighting ongoing initiatives and opportunities for automated extraction of surveillance information from a fast growing amount of computerised animal health data. An update of that review in 2016 (2) indicated remarkable growth in the field, but concluded that automated analysis and interpretation of animal health data is still hindered by a number of limitations. In particular, the lack of syndromic classification standards prevents integration and interoperability among systems using different data sources.

The issue of compatibility becomes more and more relevant as the number and type of animal health data sources grows, increasing also the opportunities – and pressure – for surveillance officials to gather (timely) evidence from multivariate surveillance systems (2).

Nomenclature systems do exist in veterinary medicine; for instance the Systematised Nomenclature of Medicine (SNOMED). These nomenclatures are, however, rarely used, and their adoption would only solve the terminology problems prospectively, while historical data collected without these standards would still remain uninterpretable outside the system where they were collected. Considering these points, we have previously suggested that the solution is not to recode data, but to develop harmonised rules to interpret data, that is, to translate health data into syndromic representations (3).

Following the second ICAHS, in 2014, a workshop was held gathering interested researchers to discuss syndromic classification standardisation in VSS. Discussions since then, including other workshops with syndromic surveillance experts in the fields of animal and public health, as well as meetings with computer scientists actively involved in the field of information management, have led to the conclusion that the solution needed is an animal health surveillance ontology (AHSO).

Ontologies – what and why?
An ontology, in the information science context, "*defines a common vocabulary for researchers who need to share information in a domain* (4)". However, compared to vocabularies and agreed terminologies, ontologies further provide, "*machine-interpretable definitions of basic concepts in the domain and relations among them* (4)".

To understand the importance of "relationships", consider as an example a terminology of geographical locations. We could start at the country level, and drill it down into smaller political units. We would for instance know that the city of Geneva is located in the Canton of Geneva, in the country Switzerland. We would also know that it is in the same country as Zurich, since both cities stem out of that country's hierarchy. We would not know, however, that Geneva is actually closer to the city of Lyon, in France, than to Zurich, or even that Switzerland borders France. What we are missing is a model that includes other concepts – such as geographical coordinates – and in particular, that is capable of handling relationships such as "has coordinates" (or geographical polygon) and "is neighbor of" (or even "has North neighbor", "has South neighbor", etc.). When such relationships are modeled, the location of events can be given as geographical coordinates or as political regions, and the events can be queried based on these relationships, for instance: all events which happened in a 3km radium from X-Y coordinates; or all neighboring cities to the city where an event happened.

Ontologies are data models which capture, in a transparent way both for humans and for computers, the knowledge structures needed to address tasks in a specific context. In the case of AHSO, medical knowledge and the structure of animal production, among other concepts, are needed to perform the task of querying animal health data to identify (syndromic) events relevant for surveillance.

Examples of such concepts include pathogens, clinical signs and organ systems; and their associated relationships, such as, "pathogen causes clinical signs", "pathogen affects organ system", etc.

The benefits of this approach are:

- It provides a transparent and common understanding of the concepts documented in the ontology, including but not restricted to syndromes.
- Data sources do not need to be coded according to specific standards. Institutions may continue to use their own individual coding practices. Data can then be marked up to allow querying through the ontology.
- Since the data do not need to be coded into specific syndromes, the parameters of the search are defined based on current needs. Today it may be syndromes, but tomorrow it could be a search focused on specific clinical signs known to be associated with an emerging disease.
- It can accommodate knowledge change or new knowledge, which is especially important in the case of emerging disease detection.
- It allows information to be queried based on relationships between concepts, for instance "all diseases which can cause clinical sign A", or "all clinical signs associated with disease X".
- Reusing knowledge – not only among VSS initiatives, but especially existing knowledge already contained in other ontologies.

AHSO development

Modular development: from data to a data model

The ontology will be developed using a data-driven approach, to ensure that the immediate needs of those working with automated analysis of animal health data for surveillance are addressed.

The ontology will be developed in modules – each existing module being functional, with successive units added to expand the ontology. The expectation is that each module will be informed by examples from a specific data type – laboratory data, animal movement data, clinical data, etc.

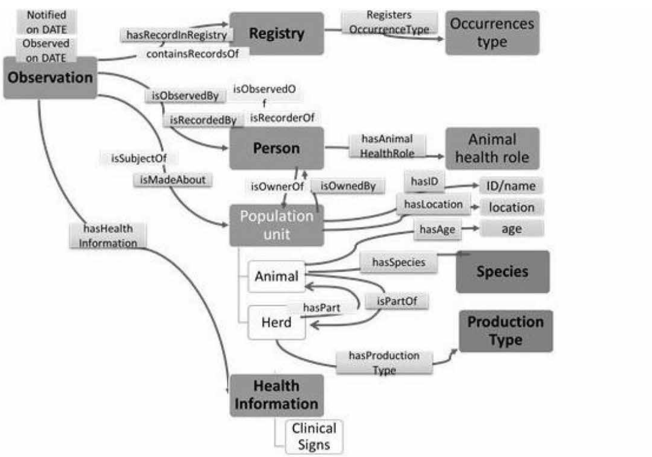
The modular development will follow the guidelines of the eXtreme Design method (XD) (5). The method is based on an iterative workflow to create new ontology modules. Each iteration is triggered by small specific examples of data which need to be modelled, and the process is focused on a test-driven and collaborative approach. In particular, the method is based on the use of ontology design patterns (ODP), which are "*reusable modeling solutions that encode modeling best practices (6)*". In other words, each cycle of development aims to solve a very specific modeling problem, and a catalogue of design patterns is searched to look for non-domain-specific modeling solutions which may be reused or adapted.

In an inventory of interested partners, we have gathered documentation of 27 animal health data sources from 8 countries. Data-sharing is not always expected to be possible, but this should not hinder ontological development. The XD protocol accounts for development based on "requirement stories", which are narrative examples of the events contained in a specific dataset, and therefore examples of the events that need to be modelled. Data owners can provide either a data sample or a user requirement story, and would not need to have any understanding of the ontology building methodology.

Below is an example of a requirement story built to exemplify records of interest from a cattle movement registry:
Farmer Nilsson, during his morning visit of his stables in Skåne on the 10th of June 2015, notices that his cow Daisy gave birth during the night to a calf that was dead-at-birth. Farmer Nilsson notifies the abortion to the electronic cattle register.

The example brings up the need to model several concepts regarding the structure of animal production (which would then apply to other animal health data types and could be reused), such as animal clustering in herds, herd location, ownership, and the fact that health information can be recorded at the animal or herd level. An expected data model to address this requirement story is shown in Figure 1. Concepts are shown in round-cornered boxes, and relationships as arrows with rectangular labels.

Figure 1. Data model to address the events recorded in cattle movement registries.



While the data model looks relatively complex, it highlights that information marked up according to these concepts could be queried according to many different dimensions, for instance "all herds belonging to the same owner", or "all herds in a specific radius from the event".

Reusing knowledge

Once the data model is in place, the concepts in the ontology need to be filled with specific available information – for instance filling in the "species" and "clinical signs" boxes in Figure 1.

The reuse of content is facilitated by the growing number of biomedical ontologies available. We performed an inventory of existing ontologies and identified over 20 ontologies written in the common language OWL (Web Ontology Language) which could contribute, in greater or lesser extent, to the animal health surveillance ontology, these include: infectious disease ontology (IDO); Epidemiology Ontology (EPO); Symptom Ontology (SYMP), among others.

Public tools exist which allow specific pieces (or even entire) ontologies to be imported for reuse. We highlight the web-based tool Ontofox (http://ontofox.hegroup.org).

Filling missing pieces through expert elicitation

Not all required knowledge will be available in existing ontologies. We anticipate field experts to be an important source of knowledge for the ontology, but no ontological design knowledge will be required for contribution. Tools exist to collect information from experts in simple formats, such as Excel spreadsheets, from which the knowledge can be integrated into the ontology.

Project structure and community involvement

The project is currently funded by Sweden's innovation agency (VINNOVA) and the funding is being used to train and maintain a group of ontology developers, in partnership with the Department of Computer and Information Science from the University of Linköping (Sweden). This group will serve as the "ontology curator group", driving iterations around ontology development and the public release of versions.

Following the example of successful community efforts to develop common standards, such as "schema.org", AHSO development is public and open, using two main tools:

- **GitHub** (https://github.com/SVA-SE/AHSO) is being used to publicly store ontology codes, allowing any interested parties to: see and download the current version of the ontology; suggest improvements and corrections; submit requirement stories or other issues that need to be addressed in future development stages; access a wiki with relevant references regarding ontologies in general, and the project in particular.
- A discussion forum (Google group) has been created to establish conversations with a community of users particularly interested in influencing ontology development. Please send an e-mail to the first author if you are interested in joining this group, or visit https://groups.google.com/forum/#!forum/ahsontology.

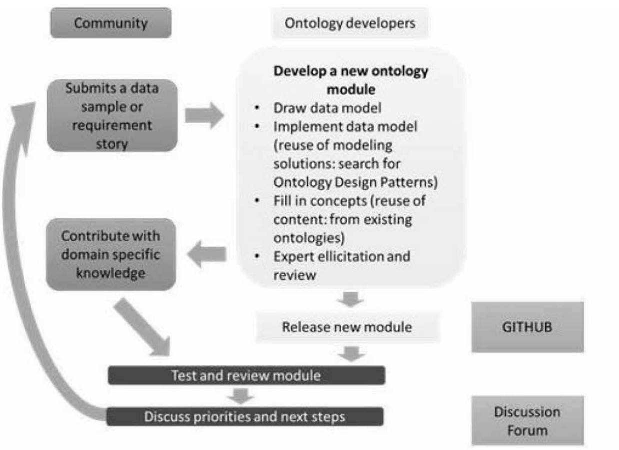
Any interested members of the VSS community are encouraged to become engaged through interacting with these tools. As explained, no level of ontological knowledge will be needed to be an active member of the community, as the ontology curator group will oversee the task of translating the community input into iterations of ontology development, as well as translating the results back to the community.
The development process is summarised in Figure 2.

Next steps

The project is currently funded for a 2-year developmental stage which started in January 2016. In 2017 we expect to mature the tools for community involvement, and progress with development of at least 3 ontological modules.

In 2018 the project is expected to evolve into a cycle of implementation. At that stage, the role of the curator group will be to provide help on the practical matters of incorporating the ontology into specific VSS systems.

Figure 2. Ontology development and community participation.



Discussion and conclusions

We have discussed the development of an animal health surveillance ontology (AHSO) to promote efficient translation of animal health data into (syndromic) surveillance information. The adoption of an ontology will maximise knowledge sharing and interoperability among VSS systems. While the development of any ontology is a long-term task, the growing number of biomedical ontologies and open access tools for ontology construction allows reuse of both knowledge and modeling solutions, and the development of AHSO will build on achievements evident in other ontologies. The methodology proposed is problem-oriented, collaborative, and promotes community involvement.

Acknowledgements

This work is funded by Sweden's innovation agency (VINNOVA). Help and training from Eva Blomqvist and Karl Hammar (Department of Computer and Information Science, University of Linköping, Sweden) have been essential to the development of the team.

References

1. **Dórea FC, Sanchez J, Revie CW.** *Prev Vet Med* 101, 1–17, 2011
2. **Dórea FC, Vial F.** *Vet med: Res Rep* 7 157-170, 2016
3. **Dórea *FC et al.*** *ICAHS 2 proceedings.* 2014
4. **Noy NF, McGuiness DL.** http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
5. **Blomqvist E *et al.*** In: Hitzler P *et al*. *Ontology Engineering with Ontology Design Patterns*. 2016
6. **Presutti V et al.** In: Suárez-Figueroa MC *et al. Ontology Engineering in a Networked World*. 2012