**Digivet - Digitalisation of livestock data to improve veterinary public health**

[**http://www.dcs.gla.ac.uk/~jenright/digivet\_website/index.html**](http://www.dcs.gla.ac.uk/~jenright/digivet_website/index.html)

[**https://www.sva.se/en/foka/digitalisation-of-the-veterinary-public-sector-to-improve-safety-and-preparedness/**](https://www.sva.se/en/foka/digitalisation-of-the-veterinary-public-sector-to-improve-safety-and-preparedness/)

[**https://zenodo.org/communities/digivet/**](https://zenodo.org/communities/digivet/)

**Work Packages and Milestones**

We will organise our work in 3 work packages, which cover broad parts of the data ecosystem, from a study of the enabling factors of digitalisation (WP1), to an assessment of the broader societal and end-user perspectives and impacts (WP3). In the center of the project (WP2), digital innovation is used to extract more value from data already available, and based on the lessons learned in WP1, develop workflows that improve data capture, information generation, and delivery of these information to stakeholders for decision-making (knowledge generation), and make these available to the wider community as open-source metadata libraries and software tools.

***WP1: Current Data Opportunities and Investments***

This WP will identify and compare enabling factors for data digitalisation and preservation through a series of stakeholder workshops, combined with desktop reviews of the data ecosystem performed in conjunction with WP2. The aim is to provide a comprehensive description of the present-day data eco-systems and pipelines for data analytics for animal health contingency planning that have emerged in countries under different regulatory frameworks, taking into consideration the security and privacy implications of digitalisation, including ethical, legal, social and behavioural considerations. Workshops will capture opinions of stakeholders in three of the countries involved in the proposal (UK, SE and DK), and deliverables will be enriched with the additional experience of stakeholders in NO and EE, both involved in the case studies. Note that the main governmental stakeholders in the UK, SE and DK are all involved with this proposal either directly or indirectly as part of other duties of the applicants involved.

**Task 1.1 Mapping the current data ecosystems and associated regulatory frameworks:** a desktop review will be conducted to identify actors, institutions and infrastructures, data sources and availability, data intelligence architectures, and routes of communication and data sharing and preservation. This will be informed by Task 2.1 (WP2). which will conduct a survey of technologies, formats and systems.

**Deliverable 1.1:** Technical report and policy brief compiling the results of this task.

**Task 1.2 Drivers and barriers:** a theoretical framework will be developed for identifying, comparing and contrasting enabling societal, technological, environmental, economic, and political (STEEP) drivers and barriers to data digitalisation and innovation of new technologies in public services. The theoretical framework will combine a social science approach to animal disease governance, data governance and FAIR principles of responsible data usage (that data be findable, accessible, interoperable and reusable). This theoretical framework will be applied to the workshops carried out in Task 1.3 and used to assess data digitalisation innovations in WP3.

**Deliverable 1.2:** A public report documenting the theoretical framework.

**Task 1.3 Stakeholders perception of current challenges/opportunities and knowledge gaps:** participatory workshops with key national stakeholders will be held to explore the challenges in data digitalisation and identify key opportunities for developing solutions within WP2. A planning phase will ensure identification of the key national stakeholders within each of the 3 case studies across the core partner countries (SE/DK/UK). Depending on the overlap among stakeholders identified within each of the core nations, workshops may be combined (for example a common section for introduction of the project may be followed by discussions in smaller working groups). The aim is to investigate risk perception amongst stakeholders, current use of digital data in animal health management, and the kinds of farm systems that are advantaged or disadvantaged by different approaches to data digitalisation.

**Deliverable 1.3.** Technical report and policy brief (circulated and revised by stakeholders).

***WP2: Creating Smarter Data Systems:***

This work package will focus on the practical implementation of methodological innovations that can be used to optimize the extraction of information from data, thereby ensuring that digital data is used to its fullest extent both nationally and internationally. We will explore existing and emerging digital innovations for collecting, managing, using and preserving these data over different timescales. We will use real datasets where they are in the public domain, but we do not plan to use the more sensitive data to which we have access as part of our other duties (for example involving disease records and antimicrobial prescription identifiable at the individual level). Instead we will instead generate and use simulated data with equivalent properties in order to allow us to work more easily within an open-source context. We will explore best practices for digitalisation and information extraction within each case study, and assess their impact through quantitative modelling and qualitative investigation. The research will be highly interdisciplinary, combining methods from the fields of data management, data science, semantic modeling (aiming at interoperability), mathematical modeling, statistical modeling, machine learning, and information communication.

**Task 2.1 Data Collection, Curation, Preservation:** this task will include, in conjunction with Task 1.1, a desktop survey to map the constellation of data sources, datasets, data formats and uses for data associated with each of our case studies. Besides the technical report produced in Deliverable 1.1, the survey will be published as a public repository including descriptions of formats and sources of data. Where possible and appropriate, direct links to the data sources will be provided.

**Deliverable 2.1.** Repository of data sources and descriptions relevant for the different case studies.

**Task 2.2 Data FAIRness:** this task will aim to improve data FAIRness within the case studies in three steps: (1) scoring of key datasets identified in Task 2.1 (and listed in the D2.1 repository) based on the set of 14 metrics defined to quantify levels of FAIRness; (2) identification of tools freely available to improve data FAIRness, with particular emphasis to content mapping, knowledge modeling, and annotation of datasets contents and metadata for different formats and technologies used by the identified sources of data; (3) development of data workflows that make use of the tools identified to improve data FAIRness. These workflows will describe steps and freely-available tools to annotate data and all of its contextual meta-data using existing knowledge models for data annotation (ontologies), to ensure that these data are machine readable and interoperable across sectors. Where possible and appropriate, ontology models will be developed to complement those already publicly available (e.g. <https://bioportal.bioontology.org/>). The results of the first step (scoring of datasets) will be added to the data descriptions in the data repository (D2.1), and the suggested workflows also added to the repository as an additional resource.

**Deliverable 2.1.** Open and reproducible workflows to produce FAIR data within the 3 case studies identified, and a conceptual diagram of the knowledge graph for each of the 3 case studies.

**Task 2.3 Data analytics/new technologies for data processing:** The creation of a conceptual knowledge graph for each of the case studies will allow the construction and optimization of analytical workflows to produce information from the available data. Testing of analytical approaches to improve data capture will be performed in parallel to the previous task, which aimed that captured data should be FAIR. Analytical methods will be developed to maximize the information extracted - and therefore the value - of available data. Different approaches (including predictive analytic methods, spatial statistics, statistical surveillance methods) will be assessed for their usefulness in providing timely information to support decision making in each of the case studies, given their defined objectives, and considering the data availability and limitations identified so far for each of the case studies. Topological data analysis uses insights from theoretical mathematics to characterise abstract properties of datasets, and will be used to compare datasets and timepoints with the aim of identifying persistent (and therefore potentially important) structures. We expect implementation of our methods to be in a mixture of Python, R, & C++, and the tools produced will be open source and developed using good programming practices including unit testing and extensive documentation in order to facilitate uptake in the wider community. We anticipate potential limitations might include differences in data coverage and spatio-temporal resolution, and the size (and therefore computational feasibility) of approaches that are suitable for smaller datasets.

**Deliverable 2.3.** Publicly available and reproducible holistic data analysis workflows for the case studies. All software and code produced within this task will be made public along with documentation, allowing re-use in other settings.

**Task 2.4. Information delivery and communication:** Bespoke dashboards will be developed to allow end-users, within each case study, to visualize the data and the information generated as a result of the analytical methods applied in Task 2.3, aiming to support decision-making. These will be user-friendly visualization interfaces for users to *interact* with the information available, written in programming languages used widely by veterinary epidemiologists and data scientists (e.g. R, Python, Julia & JavaScript). As with task 2.3, all code will be made open source and with good programming practice in mind so that the tools can be reused within other cases and settings. When appropriate, application programming interfaces (APIs) will be used to connect developed dashboards to complementary applications available for each of the case studies. We envision the dashboards being part of an “ecosystem of solutions”, where the tools produced in this project to digitalise and analyse epidemiological data can be further developed to connect, for instance, to other dashboards and databases available at diagnostic laboratories, whole genome sequencing analytical platforms, and supra-national outbreak tracking systems. In conjunction with WP3, training will be delivered to end users, and recorded demos and instructional videos will be made available online (see Task 3.2).

**Deliverable 2.4.** Dashboards to support decision-making by relevant stakeholders within each case study.

***WP3: Assessing broader societal and end-user perspectives and impacts:***

This WP will assess the broader societal and end user impact of data digitalisation interventions developed in WP2 and produce further recommendations to address remaining challenges and gaps. This will involve exploring whether existing practices of data collection, sharing, communication and preservation are sufficient to address the gaps in data rights and needs of end-users and beneficiaries and whether these practices will be ‘resilient’ to future system shocks (such as weakened infrastructures and institutions, and scarcening resources) such that they are fit for the purposes of future decision makers.

**Task 3.1 Ethical, social and behaviour consequences of data digitalisation:** Participatory workshops will explore the ethical, social and behavioural consequences of data digitalisation including skills requirements and user training needs; labour market requirements; security, privacy and risk; and longer-term impacts on the public sector. International workshops involving key stakeholders will be held for each of the case studies to explore the benefits and limitations of the interventions developed in WP2 and compare findings across countries. They will also explore challenges within the existing data landscape for each case study and produce recommendations about how these could be addressed in future. The international component of the workshops is crucial to drive the international collaboration necessary to address cross-border challenges associated with exotic and highly contagious diseases such as ASF, and to harmonise national approaches to regulating and reporting antimicrobial usage and foodborne illness. These workshops are part of the participatory methodological approach of the project to co-produce outputs with stakeholders. International workshops will include up to 20 key stakeholders involved in different domains relevant to the digitalisation of animal health data including farmers, veterinarians, policy makers, academics, and industry stakeholders. One workshop will be conducted in each of the 3 main partner countries (SE/DK/UK). Documentation of the workshops will be thematically analysed to produce a report of results.

**Deliverable: 3.1** A report and policy brief of results from each of the international workshops, assessing the interventions designed in WP2 and exploring future challenges (three reports in total).

**Task 3.2 Best practices to future-proof data intelligence and risk communication:** This task overlaps with WP2 task 4, in which training will be delivered to end-users of the dashboards constructed. Based on their feedback during training, as well as feedback gathered during the international workshops, guidelines for the use of the workflows produced in this project will be made available.

**Deliverable 3.2** Best practice guidelines on the use of tools produced in WP2.