D4.4 Data Management Plan

WP4 Joint Integrative Projects

Responsible Partner: Sciensano

Contributing partners: SVA

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
Starting Date	01/01/2018
Duration	60 Months

DOCUMENT MANAGEMENT

Deliverable	D4.4 Data Management Plan
WP and Task	WP4; Task 5-1
Leader	Valérie De Waele (Sciensano)
Other contributors	Sophie Giraud (Sciensano), Mickaël Cargnel (Sciensano), Hein Imbrechts (Sciensano) and Ann Lindberg (SVA)
Due month of the deliverable	M6
Actual submission month	M6
Type *R: Document, report* *DEC: Websites, patent fillings, videos, etc.* *OTHER*	R
Dissemination level *PU: Public* *CO: confidential, only for members of the consortium (including the Commission Services)*	PU

HISTORY OF CHANGES			
Version	**Publication date**	**Authors**	**Change**
1.0	30.06.2018	Valérie De Waele (Sciensano)	▪ Initial version
			▪

Table of Contents

ABBREVIATIONS

ADMS: Asset Description Metadata Schema

AMR: AntiMicrobial Resistance

DCAT-AP: Application profile for data portals in Europe

DOI: Digital Object Identifier

EC: European Commission

ECDC: European Centre for Disease Prevention and Control

EFSA: European Food Safety Authority

EU: European Union

EUCAST : European Committee on Antimicrobial Susceptibility Testing

FAIR: data Findable, Accessible, Interoperable, and Reusable

GDPR: General Data Protection Regulation (EU) 2016/679

JIP: Joint Integrative Project(s)

JRP: Joint Research Project(s)

JSON: JavaScript Object Notation

ODIP : Open Data Interoperability Platform

OH: One-Health

OH-EJP: One-Health European Joined Programme

OIE : Office International des Epizooties, World Organisation for Animal Health

OWL : Ontology Web Language

PMT: Project Management Team

RDF: Resource Description framework

RDFa : Resource Description Framework in Attributes

URI : Uniform Resource Identifier

WHO: World Health Organisation

XML: eXtensible Markup Language

OVERARCHING DATA MANAGEMENT PLAN

The One-Health European Joint Program (OH-EJP) aims at integrating the complementary expertise of partners across Europe in order to prepare common action against infectious health threats. Those threats include zoonotic infections both in animals and humans, and infections or toxin contamination in feed and food. To reach the objective, the OH-EJP consortium will develop a sustainable framework for an integrated community of research groups. Research groups are represented by reference laboratories in fields of human and veterinary medicine, food and environmental sciences. The OH-EJP will emphasis on food-borne microbial infections and intoxications, in the scope of a One-Health (OH) perspective.

To achieve those objectives, a significant amount of data will be collected, processed and generated, such as OH-EJP deliverables, scientific publications (e.g. peer-reviewed research articles) and research data. According to the European Commission (EC), "*research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion, or calculation*". In general terms, OH-EJP data will follow the "*FAIR*" principles, meaning "*Findable, Accessible, Interoperable and Re-usable*"). The FAIR principles will ensure soundly managed data, leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. The data will be made findable and accessible within the Consortium, and to the broader research community, stakeholders and policy makers. Also, data has to be compliant with national and European ethic-legal framework, such as the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679), which is applicable since May 2018.Data management plans (DMPs) describe the data management life cycle for all data to be collected, processed and/or generated by a Horizon 2020 project. It should include information on the handling of research data both during and after the end of the project; the nature of the data, the methodology and standards applied, whether data will be shared or made open access, and how the data will be curated and preserved. The present document provides information on the general OH-EJP strategy regarding data management in the form of an overarching data management plan. It defines the strategy on how OH-EJP data are managed under conditions that conform with the requirements of Horizon 2020. Adherence to the overarching DMP will be governed by the Consortium Agreement. Due to the heterogeneity of the data that will be collected, processed or generated within OH-EJP, and due to the level of detail needed, each joint research project (JRP) and joint integrative project (JIP) will also have to develop project specific DMP's, using as baseline the present overarching DMP. The first version of project DMPs are due by month 11 (November 2018), and their development will be guided by the DMP focal point of OH-EJP, i.e. the Belgian partner Sciensano.

As the OH-EJP is a co-funded program, agreements between partners and stakeholders are required to collect/process/use data. It must be acknowledged that the source of co-funding may have priority in some decisions regarding data management, i.e. that it may dictate where and how the programme output, including data, should be deposited and named. A guiding principle is also to avoid duplication of effort, i.e. data and publications should not be deposited twice. Consequently, the principles provided by the OH-EJP overarching DMP are meant to complement any requirements from individual funders, while still ensuring that the data are FAIR, as far as possible.

The DMP is intended to be a living document, and can be further modified or detailed during the OH-EJP. The information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Those changes might include new data, changes in consortium policies (e.g. new innovation potential, decision to file for a patent) or changes in composition and external factors (e.g. new consortium members joining). At minimum, the DMP will be updated in the context of the periodic evaluation/assessment of the program, but it is foreseen that the implementation of the DMP at project level will also be part of the annual reporting.

It is also foreseen that the expectations from the OH-EJP on FAIR data management will be of value also for institutional development and maturation with regards to proper data management, thereby contributing to the overarching goals of alignment and integration at the EU level. To support development of good research data practice among partner institutes; guidelines and training will be provided by the joint integrative research work package to develop DMP competences.

The overarching DMP is structured according to the H2020 templates: Data management plan v1.0–13.10.2016. It includes 6 components summarized in the Table 1:

1. Data Summary
2. FAIR data
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other issues

A last section provides an action plan table (Table 1), which presents important topics requiring progress and/or update in future version of the DMP.

1 DATA SUMMARY

1.1 Explain the relation of the data to the objectives of the project

The overall goal of OH-EJP is to combine different expertise of partners across Europe in order to better address threats related to zoonotic diseases in human and animal and infections or toxin contamination in feed and food. This will allow for coordination and preparation of joint public and animal health action plans. Each OH-EJP project (JRP and JIP) is collecting or processing, and/or generating data with its own purpose and specificities to serve the common goal of integrated expertise and capacity building.

At the start of the EJP, the consortium manages 13 projects: 2 integrative projects and 11 research projects. As a step in the development of the overarching DMP, a questionnaire was distributed to project leaders to capture the current state of the art with regards to data management, and to identify needs for further development and training. Below, the relation of the data of each project with their specific objective is presented.

- Integrative projects
  - The ORION project aims at establishing and strengthening inter-institutional collaboration and transdisciplinary knowledge transfer in the area of surveillance data integration and interpretation, along the OH objective of improving health and well-being. Data collected and/or generated serve the objective of providing a prototypic implementation of an integrated strategy for long-term consolidation and harmonization of OH Surveillance solutions.
  - The COHESIVE project will collect different kinds of data to support discussion to develop guidelines for national One Health structures. This type of information will be retrieved through questionnaires to make a blue print of human-veterinary collaboration and acquire better knowledge of the present situation. Data over existing risk assessment tools will be collected in order to make a decision tree on when to use which tool. Some COHESIVE partners will be permitted to access the data with the aim of setting up their own Information System with databases harboring WGS/NGS data, metadata and epi data.
- Research projects
  - The NOVA project aims at developing epidemiological methods of investigations of potential new sources for the surveillance of foodborne diseases.
  - The ListAdapt project will explore the diversity of strains in different compartments of the farm to fork chain to better explain the adaptation capacity of *L. monocytogenes*.
  - The Metastava project will harmonize and optimize the use of metagenomics across Med-Vet partners and share methodologies.
  - The project AIR-SAMPLE will develop methods for a standardized protocol for air sampling in poultry flocks.
  - The project MoMIR-PPC aims at creating a network that will focus on the prevention of food-borne pathogens in the food chain in order to control zoonotic food-borne infections, optimize husbandry and feeding practices, and decrease the use of antimicrobials in farm industries and hospitals. Based on data obtained from animal infections and human carriers, new approaches will be developed to predict, identify, prevent and control the appearance of animal and human super-shedders based on immune response and gut microbiota composition. The data on the dynamic of super shedders and the analysis in farm conditions will result in a new mathematical model, which provides essential information to producers to support and strengthen biosecurity measures, with a cost effectiveness. This project will also lead to improve diets or additives (pre, probiotics, neutraceuticals) that better protect humans and livestock. Taken together, this will allow to reduce antimicrobial usage. The results will be

disseminated and communicated to both the public and the medical-Veterinary society, to decrease import of such bacteria in the future. Results will be disseminated through a variety of written and oral media. Primary data manuscripts will be published in peer-reviewed journals and we anticipate that this will include manuscripts in high scientific impact journals as well as those that specialize in veterinary or animal disease. All of the scientists will regularly attend and contribute to international and national scientific meetings as well as industry orientated meetings.

- The objective of the project MedVetKlebs is to develop, evaluate and harmonize methods for sampling, detection, strain typing and genome-based genotyping of Klebsiella pneumoniae, and share these methodologies across Institutions and with the scientific community in order to optimize the current practices. The purpose is to enlarge and promote a scientific network during the life-time of the project in order to involve more countries concerned by the subject and gain additional expertise. This will allow to identify gaps where furthers investigations are needed to inform current policy questions and design novel approach. The research findings will be disseminated and knowledge transferred to the diverse target audiences through training/exchanges activities at national and international level.
- The project IMPART will harmonize methods for detection of resistant bacteria (to colistin and carbapenem, and resistance of *Clostridium difficile*) and subsequent susceptibility testing. Phenotypic data (MIC-values) will be generated to be able for EUCAST to set epidemiological cut-off values to interpret future susceptibility tests of veterinary pathogens.
- The data generated by ARDIG project will help examine the dynamics of anti-microbial resistance (AMR) in different epidemiological units (human, animal, food and environment) from countries which represent significant difference in their usage of antimicrobial agents and AMR prevalence, both in the human and veterinary sectors, as well as different climate, management systems and the potential for transmission of resistance. It will also help in understanding differences and similarities between methodologies used by national institutes in different countries.
- The project RaDAR will help develop common modelling methodologies.
- The project MAD-VIR data management aims at harmonizing and optimizing the practices of identifying all virus including emerging threats and food-borne zoonosis in key institutions/laboratories throughout EU countries.
- TOX-detect is the development and harmonization of innovative methods for comprehensive analysis of food-borne toxigenic bacteria, ie. Staphylococci, Bacillus cereus and Clostridium perfringens.

1.2 Specify the types and formats of data collected/generated

Different data types will be collected/generated, such as publications and research data, related to foodborne surveillance, AMR and emerging threats. Other types of data include questionnaire data (e.g. paper-based/online questionnaires), clinical data, biological data (e.g. measurements in biological matrices/tissues), molecular data (including data on part of or whole genome), modelling data (e.g. estimated exposure and/or effect parameters), …etc. A list of the different deliverables has been established, and this list will be further detailed to precise the type of data generated. Additionally, a comprehensive list of data collected and generated will be gathered from the different projects as the projects progress.

Data formats should be selected with the view to facilitate data storage and transfer. Therefore, data will be machine-readable format, preferably in formats intended for computers (e.g. RDF, XLM and JSON), but also in human-readable format marked-up to be understood by computers (e.g. microformats, RDFa). Additionally, it is recommended to use non-proprietary formats if possible.

1.3 Specify if existing data is being re-used (if any)

The Project Management Team (PMT) of the OH-EJP program encourages partners to make existing data available for research within the EJP Consortium. To support such data re-use, lists of datasets collected and generated during the course of the program will be made available on the OH-EJP website, and access procedures drafted for those data. If relevant in their research task, the consortium partners should be able to make use of these existing data.

1.4 State the expected size of the data (if known); handling/storage of "big data"

The expected size depends on the extent and the nature of the data that are made available, and will be evaluated during the course of the project by the Consortium partners. Big data handling and storage is expected for some projects, and adapted procedures will be described in the appropriate project DMPs.

1.5 Outline the data utility: to whom will it be useful

According to the domain of expertise, data generated within OH-EJP program / projects can be useful to:

- Other partners belonging to the OH-EJP Consortium (EJP beneficiaries);
- European Commission services and European Agencies, such as EFSA, ECDC, DG-SANCO, DG-HEALTH;
- International agencies, such as OIE, WHO; National authorities involved in animal and public health;
- European scientific community, such as European and national reference laboratories, scientific from medical and veterinary research institutions;
- Industries involved in animal management and extension services;
- General (scientific) public.

It is the objective of the Consortium to provide most of deliverables to the widest public possible; however, restrictions in the use of data might also apply. If so, the rational for such restrictions should be provided.

2 FAIR DATA

Through the life cycle of the OH-EJP data, the FAIR principles will be followed as far as possible, while ensuring compliance with national and European ethic-legal framework. The FAIR component of the DMP still comprises points to clarify, which will be addressed during the course of the programme.

Points addressed

2.1 Making data findable, including provision for metadata

2.2 Making data accessible

2.3 Making data inter-operable

2.4 Making data re-usable

2.1 Making data findable, including provisions for metadata

2.1.1 Outline the discoverability of data (metadata provision)

Because of the co-funding setup of OH-EJP, with Programme Managers receiving their mandate from Programme Owners, agreements have to be made between OH-EJP partners and relevant data national owners/providers to ensure data discoverability and identifiability. During the course of the program, the relevance and opportunity to make those co-funded data findable and accessible to other OH-EJP partners will be assessed case by case. Different considerations will be taken into account to support the decision of making those data findable, such as scientific relevance of data for other OH-EJP partners, technical feasibility, formal agreement with the data owners/providers, and compliance with national and EU ethic-legal framework.

Data discoverability can be obtained by different means, which include:

- Providing data documentation in a machine-readable format;
- Using metadata standards or metadata models;
- Providing open access (e.g. open data repository);
- Providing access through application;
- Providing online data visualisation/analysis tool for the data, to help researchers to explore data in order to determine its appropriateness for their purposes;
- Providing online links between research data and related publications or other related data;
- Providing data visibility through a communication system (e.g. social media, website).

All deliverables will be listed on the OH-EJP website (www.onehealthejp.eu), and the ways by which OH-EJP output can be accessed will be communicated via social media and other suitable channels to increase visibility of OH-EJP work. For public deliverables, a link will be available between the OH-EJP website and the appropriate open repositories where the data is submitted. Some repositories, such as Zenodo, provide also social media link.

According to EC, metadata is a systematic method for describing such resources and thereby improving access to them. In other words, it is data about data. Metadata provides information that makes it possible to make sense of data (e.g. documents, images, datasets), concepts (e.g. classification schemes) and real-world entities (e.g. organisations, places). Metadata is often called data about data or information about information. **D**ifferent types of metadata exist for different purposes, such as descriptive metadata (i.e. describing a resource for purposes of discovery and identification), structural metadata (i.e. providing data models and reference data) and administrative metadata (i.e. providing information to help management of a resource). In our case, we are mainly interested to describe a resource for purposes of discovery and identification.

Each OH-EJP partner will use metadata standards or metadata models appropriate to their own data, which will be described in the individual project DMP. The DMP team will provide an inventory of metadata standards or metadata models related to OH-EJP data. The first call integrative projects, ORION and COHESIVE have already identified gaps in metadata standards in their domains of expertise and it will be part of their objectives to develop new metadata frameworks. Most research projects, for which appropriate metadata standards do not exist, will take advantage of existing metadata frameworks and adapt them to describe their data according to their needs.

To provide metadata on the web, two approaches/syntaxes exist for representing data and resources, i.e. XLM (Tree/container approach) and RDF (Triple based approach). Different metadata schemes exist for both XLM and RDF approaches. A metadata scheme is a labelling, tagging or coding system used for recording catalogue information or for structuring descriptive records. A metadata

scheme establishes and defines data elements and the rules governing the use of data elements to describe a resource.

2.1.2 Specify standards for metadata creation (if any)

Because of the lack of appropriate metadata standards, it is expected that the OH-EJP integrative projects will need to develop metadata frameworks in the course of their project. For the on-going first call projects, the following approaches were reported:

- ORION project will explore how metadata standards provided by the UNECE High-Level Group for the Modernisation of Official Statistics, like the Generic Statistical Information Model (GSIM see https://statswiki.unece.org/display/gsim/Generic+Statistical+Information+Model) or the Generic Statistical Business Process Model (GSBPM - see https://statswiki.unece.org/display/GSBPM), can be used to create a mapping between metadata standards established in the different OH sub-domains.
- COHESIVE will develop a metadata structure based on the framework of EpiJSON (Epidemiological JavaScript Object Notation). The framework provides a unified data format to facilitate the use and structured interchange of epidemiological information in an unambiguous way, linking genomic data to information on the type of disease, the sample collection (who, where, when), the source of the sample (patient, food item, animal, and their identification and biological details), the connections between the various sources of the samples to define the outbreak and the inter-relations between the various components of the outbreak.

Some criteria will be ascertained to ensure best practice in metadata management:

- Availability: metadata need to be stored where it can be accessed and indexed so it can be found;
- Quality: metadata need to be of consistent quality so users know that it can be trusted;
- Persistence: metadata need to be kept over time;
- Open License: metadata should be available under a public domain license to enable their reuse.

2.1.3 Outline the identifiability of data and refer to standard identification mechanism

The assignment and management of persistent identifiers to the data will be assessed in the course of the project and will be described in the project DMPs. It is recommended to use Uniform Resource Identifier (URI) to facilitate links between different data. Most repositories are providing automatically persistent identifiers such as DOI, e.g. the functionality provided by Zenodo platform.

2.1.4 Outline the approach towards search keyword

To facilitate the queries by keywords, metadata elements need to be aligned across the OH-EJP. Therefore, the metadata elements must include the term “OHEJP”, to facilitate finding of OH-EJP data. The selection of the appropriate repository for the OH-EJP deliverables and data should provide filtering system based on the metadata elements, e.g. SPARQL system, which is a standardised language for querying RDF data, able also to query linked data.

2.1.5 Naming conventions and clear versioning

The naming convention for deliverables was stated in the OH-EJP Grant Agreement of September 2017, which is in the format: “D Name of deliverables”. For other data generated by OH-EJP Consortium, the recommended naming convention consisting in 3 mandatory parts separated by an underscore:

- A prefix with a short and meaningful name of data

- A root composed by:
  - the acronym of the project
  - the acronym of the program “OHEJP”
- A suffix indicating the date of the last upload into the repository in YYYYMMDD format.

Because of the co-funding setup of the programme and because some repositories have their own naming conventions, the above naming convention should be regarded as recommendation but is not compulsory.

2.2 Making data openly accessible

The data and metadata of OH-EJP should by default be made openly available to European Commission services and European Agencies; EU National Bodies; OH-EJP consortium; and the general public. According to H2020 online manual, open access refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. In the context of research and innovation, 'scientific information' can mean: peer-reviewed scientific research articles (published in scholarly journals), or research data (data underlying publications, curated data and/or raw data). Open access to scientific publications means free online access for any user. The costs of open access publishing are eligible, as stated in the Grant Agreement. Open access to research data refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement. Users should normally be able to access, mine, exploit, reproduce and disseminate openly accessible research data free of charge.

2.2.1 Specify which data will be made openly available; if some data is kept closed provide rationale for doing so

Data, including deliverables, produced in the course of the project should be made openly available as the default, while respecting compliance with European and national ethic-legal framework on personal data protection.

Depending on the deliverables, restrictions might apply for specific reasons that will be stated in the overarching DMP and in project DMPs for each research or integrative project. Similarly, restrictions can be foreseen for other scientific data used/generated during the projects and will be described in specific project DMPs. The rational for keeping data closed might include:

- Open access is incompatible with rules on protecting personal data: protection of the personal right needs to be ascertained either by avoiding open access to sensitive and personal data, or by anonymizing the data if relevant and feasible.
- Open access is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited: In general, open access does not affect the decision to exploit research results commercially, e.g. through patenting. The decision on whether to publish through open access must come after the more general decision on whether to publish directly or to first seek protection.
- Open access is incompatible with the need for confidentiality in connection with data from external owners/providers: Because of the co-funding setup of OH-EJP, partners might use data collected or generated by or with co-funders. If relevant for other research partners, agreements with co-funders will be discussed to make those data accessible to other OH-EJP partners, while respecting compliance with European and national ethic-legal framework.
- Open access is incompatible with the need for confidentiality in connection with security issues
- Open access would mean that the project's main aim might not be achieved.

To help partners in their decision to use open access, restricted access or keeping data closed, the DMP team will provide a decision tree. So, the access to publications or research data, will be data specific. The decision to select a specific type of access (open, restricted or close) will be under the responsibility of the individual project partners which collected/processed/generated the data, and the rational to keep data closed will be described in the project DMPs.

2.2.2 Specify how the data will be made available

Deliverables will be made findable and accessible through the OH-EJP platform. Some deliverables will be kept confidential, but most will be made publicly available. Public deliverables will be linked to the open repository where they were deposited in machine-readable format. For example, data in machine-readable format (e.g. JSON) will be uploaded in the sub-community One-Health EJP on OpenAIRE platform hosted by Zenodo, and data can be found through a web browser and downloaded by a potential interested user. Regarding peer-reviewed publications, the OH-EJP Grant Agreement provides a gold open access opportunity. Similar accessibility processes are available for other research data collected or generated during the program.

2.2.3 Specify what methods, codes or software tools are needed to access the data

For most data, only standard software, e.g. web browsers, pdf-file readers, and text readers, will be needed. However, certain data, such as genomic data, might require specialised tools and languages to access the data. Specialised tools , such as FoodChain-Lab, might also be required to generate data. Additionally, one of the goals of COHESIVE project is to develop a new tool that is in itself a web-based data collection and analysis tool; documentation for newly developed tools will be provided. Where non-standard tools are used, procedures to access the data will be documented in the project DMPs.

2.2.4 Specify where the data and associated metadata, documentation and code are deposited

Data should be submitted to an appropriate repository (i.e. a place where digital information (publications, reports, data, metadata) can be stored.). Partners of OH-EJP consortium consider this is the best means of making these data FAIR. The DMP team recommends to submit data to discipline-specific or community-recognized data repositories where possible, and otherwise to a generalist repository, (such as Dryad Digital Repo, figshare, Harvard Dataverse, Open Science Framework, GitHub). Besides making data FAIR, criteria to select appropriate repositories include:

- Be broadly supported and recognized within the scientific community
- Ensure long-term persistence and preservation of datasets
- Provide expert curation
- Provide stable identifiers for submitted datasets
- Allow public access to data without unnecessary restrictions

Recommended data repositories can be filtered and accessed through OpenAIRE portal, and the Scientific Data FAIRsharing collection. The OpenAIRE services provide tools to validate repositories/journals and register them in the OpenAIRE network. However, the filtering system provided by OpenAIRE is limited to data source type (such as publication repository, institutional repository), compatibility, and country, but so far it is not possible to filter by topic. In areas where well-established subject or data-type specific repositories exist, partners should submit their data to the appropriate resources. To facilitate the selection of the repositories, the DMP team will develop a list of repositories in collaboration with partners experts in the different fields. This list will be

evaluated with regard to the criteria above and to FAIR requirements, and will have a filtering tool on topics relevant for OH-EJP data. A preliminary example is shown below:

- Biological sciences: nucleic acid sequence (eg:European Nuclotide Archive- ENA, GenBank), functional genomics, which bridge disparate research disciplines (European Genome-Phenome Archive – EGA), metabolomics (MetaboLights), proteomics (PRIDE),
- Modelling: mathematical and modelling resources (BioModels Database, Kinetic Models of Biological Systems – KiMoSys), Network Data Exchange – NDEx),
- Health Sciences: immunology (ImmPort), pathogen-focused resources (Eukaryotic Pathogen Database Resources – EuPathDB, VectorBase), repositories suitable for restricted data access (Research Domain Criteria Database-RDoCdb),

Additionally, the DMP team has set up a sub-community One-Health EJP on OpenAIRE platform. Some projects (e.g. ORION) have developed their own system, such as the Virtual Resource Environment (VRE), which is hosted on D4Science.org.

2.2.5 Specify how access will be provided in case there are restrictions

OH-EJP deliverables and data can either be public or confidential. Some results might be restricted in their use. Sensitive and personal data can be made accessible only following the GDPR requirements. The aim is to reach the highest level of GDPR compliance, amongst others by:

- Relying on the EU authentication platform and security protocols for data sharing.
- Applying a strict policy in granting and revoking access to the data.
- Logging of user identity during data access, download, and upload, including version control.

As several repositories will be used to store data, the policy on how to grant access to restricted results will be developed over the course of the project and described in project DMPs.

By default, data generated with OH-EJP co-fund and accompanying metadata are directly accessible for use within OH-EJP. For sensitive data, the data owner/data provider shall agree in the transfer of the data at high level of granularity to an OH-EJP defined repository, using appropriate measures to anonymise data. Prior to generation of the data, the data owner/data provider shall confirm ethico-legal compliance of the study in which new data are generated.

For existing data, not generated with OH-EJP co-fund, the data owner/data provider specifies the level of granularity that data will be stored and/or transferred: anonymised single measurement data; pseudonymised single measurement data; or aggregated data. The data owner/data provider indicates for each level of granularity whether the data are directly accessible for use within the OH-EJP. In case the data owner/data provider indicates that the data are not directly accessible for use within OH-EJP, the data owner/data provider will be asked approval when consortium members request access to the data to meet the goals of a particular objective.

2.3 Making data interoperable

To generate interoperable data, the OH-EJP consortium will liaise with Joinup platform. Joinup is a collaborative platform created by the European Commission and funded by the European Union via the Interoperability solutions for public administrations, businesses and citizens (ISA^2) programme. It offers several services that aim to help e-Government professionals share their experience with each other. And it offers also support to find, choose, re-use, develop and implement interoperability solutions.

2.3.1 Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability

At present, no specific data and metadata vocabularies are available for the One-Health surveillance domain. A common vocabulary, code lists and mapping of pre-defined values for harmonising the descriptions of metadata and data will be defined in the course of the program, specifically through the on-going integrative projects, i.e. ORION and COHESIVE, in collaboration with all OH-EJP partners.

In brief, the steps to obtain interoperable data that will be evaluated during the project include:

- Harvesting metadata standards from different Open Data portals. Different metadata standards exist, such as
    - DOI for published material (text, images),
    - DataCite for data archives,
    - CERIF for scientific data sets,
    - FGDC/CSDGM for biological profile,
    - Genome Metadata, ISA-Tab, or GEO for genome data,
    - INSPIRE for geographical data,
    - FOAF for people and organisations,
    - SKOS for concept collections,
    - ADMS for interoperability assets,
    - Data Catalog Vocabulary DCAT,

  The metadata DOI will be available for the OH-EJP sub-community platform and, therefore, the default metadata standard for OH-EJP publications. A comprehensive list of metadata useful to OH-EJP will be developed to facilitate consortium partners to select appropriate metadata for their specific need. Most repositories provide an interface to enter metadata.
- The metadata will be transformed to an appropriate syntax, such as Resource Description framework (RDF). RDF is a syntax for representing data and resources in the web. RDF breaks every piece of information down in triples: subject, predicate, and object.
- Harmonise the RDF metadata produced in the previous steps with DCAT-AP.
- To allow exchange between systems, metadata should be mapped to a common model so that the sender and the recipient share a common understanding on the meaning of the metadata. On the scheme level metadata coming from different sources can be based on different metadata schemes, e.g. DCAT, schema.org, CERIF, own internal model. On the data (value) level, the metadata properties should be assigned values from different controlled vocabularies or syntaxes, e.g.: Dates: ISO8601 (“20130101”) versus W3C DTF (“2013-01-01”), with Zenodo, it is possible to specify subjects from a taxonomy or controlled vocabulary, ie to link term to appropriate ontologies (e.g. GACS).
- The last step is to publish the description metadata as Linked Open Data. Data should be published on a repository offering a data catalogue with filtering functionality based on metadata elements. It is also recommended to create linked data. Linking data to other data will provide further context to the data. Data can be linked to URIs from other data sources, using open standards such as RDF (without being publicly available under an open licence). The linked data foundations are using Uniform Resource Identifier (URIs) for naming things, Resource Description framework (RDF) for representing data and resources, and SPARQL for

querying linked data. SPARQL is a standardised language for querying RDF data. Some examples of SPARQL initiatives at EU level are EU Open Data Portal SPARQL endpoint and DG SANTE SPARQL endpoint.

2.3.2 Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability. If not, will you provide mapping to more commonly used ontologies?

As mentioned in the previous section, there is a lack of metadata standards for the One-Health surveillance domain. A common vocabulary, codes list and mapping of pre-defined values for harmonising the descriptions of metadata and data will be defined in the course of the program.

If there is a lack of metadata standards, the consortium will reuse existing controlled vocabularies for providing metadata to resources as far as possible. A controlled vocabulary is a predefined list of values to be used as values for a specific property in your metadata schema. In addition to careful design of schemas, the value spaces of metadata properties are important for the exchange of information, and thus interoperability. Controlled vocabularies for reused can be found on Joinup (http://joinup.ec.europa.eu) and Linked Open Vocabularies (http://lov.okfn.org) platforms.

If there is no suitable authoritative reusable vocabulary for describing data, conventions will be used for describing the vocabulary: RDF Schema (RDFS) and/or Web Ontology Language (OWL). The best practice when new terms are required, is to define their range and domain. A range states that the values of a property are instances of one or more classes. A domain states on which classes a given property can be used. The new vocabulary should be published within a stable environment designed to be persistent. Existing resources from previous EU projects, EFSA and ECDC will serve as the basis for this work. ORION project will create a data and metadata knowledge model for surveillance data, in the form of the «Animal Health Surveillance OntologyThis will aggregate existing ontological models, and further model concepts needed to connect the multi-disciplinary sources of information needed in disease epidemiology and surveillance. An example of another interesting ontology is the Global Agricultural Concept Scheme (GACS), which is multilingual and includes in its pool of interoperable concepts the identities related to agriculture from AGROVOC, CAB and NAL Thesauri, which are maintained, respectively, by FAO of the United Nations, Centre for Agriculture and Biosciences International (CABI) and US National Agricultural Library (NAL).

2.4 Increase data re-use (through clarifying licences)

2.4.1 Specify how the data will be licenced to permit the widest reuse possible

For public data, the reuse of the data will be possible through the open repositories where they will be stored. In addition, the integrative project COHESIVE will develop tools and software, which will be distributed as open source software ensuring their widest reuse.

2.4.2 Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

The specific decision on an embargo for research data will be taken by the responsible OH-EJP partners. Scientific research articles should have an open access at the latest on publication if in an

Open Access journal, or within 6 months of publication. For research data, open access should by default be provided when the associated research paper is available in open access.

2.4.3 Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Public data will be available from open repositories, and therefore reusable by third parties, even after the end of the project. For confidential data, access to personal data will be compliant with GDPR, while data concerning intellectual property will be discussed between relevant partners, and decision will be taken according to the European and national rules. This section will be further detailed in the project DMPs.

2.4.4 Specify the length of time for which the data will remain re-usable

Regarding data stored on the sub-community One-Health EJP on OpenAIRE platform, all files stored within the repository shall be stored after the project to meet the requirements of good scientific practice. A strategy for storage of the files after the project will be included in the DMP in the course of the program.

For data stored on other repositories, researchers, institutions, journals and data repositories have a shared responsibility to ensure long-term data preservation. Partners must commit to preserving their datasets, on their own institutional servers, for at least five years after publication. If, during that time, the repository to which the data were originally submitted disappears or experiences data loss, the partners will be required to upload the data to another repository and publish a correction or update to the original persistent identifier if required.

2.4.5 Describe data quality assurance processes

For the OH-EJP consortium, it is essential to provide good quality data. This will be ensured through various methods. Firstly, some partner institute have existing data quality assurance processes, which can be described in their quality manual. Secondly, publications will be disseminated using peer-reviewed journals, and similarly, research data will be deposited on repositories providing curation system appropriate to the data. The development of a curation system for the sub-community One-Health EJP on OpenAIRE platform will be discussed by the PTM.

Additionally, it is part of some projects objectives to develop guidance documents to assess data quality. These guidelines will be tested and optimised over the course of these specific projects, and will be validated using appropriate approaches. For example, the OH Surveillance Codex, which is developed by the ORION project, intends to serve as quality assurance tool for One Health data in the future, and this codex will be validated through pilot studies.

2.4.6 Specify the data update approach (section not present in H2020 template)

Important datasets often grow and evolve, and we need to ensure that datasets can be updated while also maintaining a stable version of the data as published. If no versioning mechanism is available in the data repository, it might be appropriate to deposit a static version of the data to an appropriate

repository, while hosting in parallel a dynamic version in a project-specific resource. Both versions of the dataset should be findable.

3 ALLOCATION OF RESOURCES

3.1 Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Costs related to open access to research data are eligible as part of the Horizon 2020 grant if compliant with the Grant Agreement conditions.

3.2 Clearly identify responsibilities for data management in your project

To ensure best practices and FAIR principles in the data management of each project, specific project DMPs will complement the present overarching DMP. For the overarching OH-EJP DMP, Sciensano (OHEJP-DMP@sciensano.be) is the focal point regarding DMP and will liaise with the project management team and integrative and research projects. Each partner institute will be responsible for managing the data that they use, process or generate in the project. Additionally, each partner institution will transmit the names of a task leader and a deputy task leader from their IT and/or epidemiology departments to OH-EJP DMP team. Those designed leaders will have the responsibilities for the development of the DMPs in which their institution is involved. Guidelines and training will be provided by the joint integrative research work package to develop DMP competences within OH-EJP partners.

Currently, the DMP team is responsible to assess sustainable strategy and planning regarding development of the most appropriate OH-EJP repository. Once it will be clearly identified, responsibilities for data management in regards to the OH-EJP repository will be defined using the RACI model (Accountable Responsible Consulted Informed). The responsibilities might encompass the initial set-up of the data repository, its maintenance, security assessment, creation of repository structure (folders/sub-folders for each user group), development of instructions and support to OH-EJP partners regarding data repository structure, creation and management of users and user groups database, assignment of access, upload and download rights for each user group, ensuring compliance with personal data protection rules, and timely communicate with OH-EJP partners any possible compliance issue.

3.3 Describe costs and potential value of long term preservation

Currently, no need for additional resources is envisaged beyond the duration of the project to handle data. However, different strategies for data storage are under investigation and will be included in the DMP later.

4 DATA SECURITY

Point addressed:
Address data recovery as well as secure storage and transfer of sensitive data

To be fully compliant with GDPR or any additional national legislation, the OH-EJP will develop an appropriate security protection strategy as the project progresses. For instance, data confidentiality

and integrity will be implemented to secure data storage and transfer, by means of tamper-proof logging mechanism, and/or pseudo-anonymization techniques, and by means of secure data transfer mechanisms, such as TLS or FTP. Apart from the GDPR, the consortium partners regard privacy and data protection as a fundamental principle and hence apply a strict policy on this matter.

5 ETHICAL ASPECTS

Point addressed:
Ethical or legal issues that can have an impact on data sharing and that were not covered in the ethics review

Ethical aspects are largely covered in the context of the ethics review, the ethics section of the Description of the Action and the ethics deliverables. The storage and transfer of data on human subjects to the repositories used by the consortium are only considered in case of informed consents, ethics approval, compliance with GDPR and – when applicable - approval by local data protection authorities.

Partners are expected to describe in detail any controls or limitations on access to or usage of human data in the ethic section of the Project DMP. The process by which researchers may apply for access to the data, and the conditions under which such access may be granted, should similarly be described. The ethics self-assessment for each JRP and JIP has been evaluated by ethics advisors. Partners will follow recommendations received from the ethics advisors, as described in the Description of the Action.

6 OTHER

6.1 Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Some partner institutes might have existing data management processes, that will be followed to ensure OH-EJP data quality and security. Additionally, each OH-EJP project will develop its own DMP that will complement the present overarching DMP, and will provide further details regarding specific data collected and/or generated in the course of the project. The development of the project DMPs will support the development of good research data practice among partner institutes.

7 ACTION PLAN

This table 1 provides a summary of the actions to perform to address unresolved issues of the present DMP.

ACTION TABLE 1

FAIR Data Management at a glance: issues to cover in Horizon 2020 DMP and related actions to perform

DMP component	Issues to be addressed	Actions
1. Data summary	1. Explain the relation to the objectives of the project 2. Specify the types and formats of data generated/collected	• Detailed data type in the list of deliverables • List of data collected/generated

	3. Specify if existing data is being re-used (if any) 4. Specify the origin of the data 5. State the expected size of the data (if known) 6. Outline the data utility: to whom will it be useful	
2. FAIR Data 2.1. Making data findable, including provisions for metadata	1. Outline the discoverability of data (metadata provision) 2. Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? 3. Outline naming conventions used 4. Outline the approach towards search keyword 5. Outline the approach for clear versioning 6. Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how	• URL of OH-EJP website • Inventory of relevant metadata standards and models
2.2 Making data openly accessible	1. Specify which data will be made openly available? If some data is kept closed provide rationale for doing so 2. Specify how the data will be made available 3. Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? 4. Specify where the data and associated metadata, documentation and code are deposited 5. Specify how access will be provided in case there are any restrictions	• Adding to deliverables and data tables two field, one public/confidential and one rational for confidentiality • Developing a decision tree to choose between data open access, restricted access to data or keeping data closed • list of repositories with filtering system based on topics
2.3. Making data interoperable	1. Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.	• Liaise with appropriate support to ensure sustainability? • List of metadata standards useful to OH-EJP

	2. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	
2.4. Increase data re-use (through clarifying licences)	1. Specify how the data will be licenced to permit the widest reuse possible 2. Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed 3. Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why 4. Describe data quality assurance processes 5. Specify the length of time for which the data will remain re-usable	• Set up a curation system for the sub-community One-Health EJP on OpenAIRE platform
3. Allocation of resources	1. Estimate the costs for making your data FAIR. Describe how you intend to cover these costs 2. Clearly identify responsibilities for data management in your project 3. Describe costs and potential value of long term preservation	• List of managers for project DMPs and institutional DMPs
4. Data security	1. Address data recovery as well as secure storage and transfer of sensitive data	•
5. Ethical aspects	1. To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former	•
6. Other	1. Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	•