WORLD ORGANISATION FOR ANIMAL HEALTH
Protecting animals, preserving our future

Original: English
June 2017

AD HOC GROUP ON HIGH THROUGHPUT SEQUENCING, BIOINFORMATICS AND COMPUTATIONAL GENOMICS[1]

Paris, 27–29 June 2017

———

The Fourth meeting of the OIE *ad hoc* Group on High Throughput Sequencing, Bioinformatics and Computational Genomics (HTS-BCG) was convened at the OIE Headquarters from 27 to 29 June 2017.

The original Terms of Reference (ToRs), including the agenda for this meeting, and the list of participants, are presented in Appendices I and II, respectively.

1. Opening

Dr Elisabeth Erlacher-Vindel, Head of the OIE Science and New Technologies Department, welcomed the participants of the *ad hoc* Group on HTS-BCG (referred to here as the 'Group') on behalf of Dr Monique Eloit, Director General of the OIE. Dr Erlacher-Vindel described the project for the creation of the OIE Genomic Platform (referred to here as the 'platform'), a web-based system for collecting, storing and distributing sequence data of pathogens that have been the subject of notification to the OIE. Dr Erlacher-Vindel also explained that the specific task of the Group was to provide options to progress the project and move onto the implementation phase. These options should include a budget estimation and a list of user requirements.

2. Appointment of chairperson and rapporteur

The meeting was chaired by Prof. Massimo Palmarini, and Dr James Watson was designated as rapporteur.

3. Background and discussion on the Terms of Reference

The Group reviewed the Terms of Reference (TOR) in light of recent technological advances in genomic sequencing, updates to the WAHIS+[2] project, and new requests from the OIE. After discussion, the Group considered whether it would be appropriate to expand the ToRs, and proposed the following modified ToRs:

3.1. Revised Terms of Reference

1. Define high level user requirements;
2. Define the minimum set of metadata and the required quality standards to be attached to the sequences;
3. Define the information flow for the submission to, and the release of sequences from the future WAHIS+ system, including authentication, authorisation and a description of sequence data governance;
4. Propose strategies to initiate and sustain sequence data sharing, which would involve promoting the active engagement of OIE Reference Centres;
5. Identify analytical tools to be integrated into the platform;

---

1 Note: This ad hoc Group report reflects the views of its members and may not necessarily reflect the views of the OIE. This report should be read in conjunction with the September 2017 report of the Biological Standards Commission because this report provides its considerations and comments. It is available at: http://www.oie.int/en/international-standard-setting/specialists-commissions-groups/laboratories-commission-reports/meetings-reports/

2 WAHIS: World Animal Health Information System (of the OIE)

6. Identify possible disease models to test the system;
7. Determine next steps for implementation;
8. Timeline and budget estimation.

4. High level user requirements

The platform is a system to collect, store and analyse pathogen sequence data associated with the reporting of notifiable diseases to the OIE by Member Countries. The purpose of the platform is to provide an accessible and transparent centralised system within the OIE that integrates these data with the information provided by WAHIS+.

The Group agreed that the following points represent the key user requirements for the platform:

- The platform must be designed for full interoperability with WAHIS+;
- Sequences should only be submitted for, and linked to, an existing disease notification;
- All sequence data must have a unique identifier generated by the platform and be linked to either the WAHIS report ID or outbreak reference ID;
- The platform should offer a simple mechanism for uploading these sequence data, including a batch upload capability;
- Access and operation within the platform must rely on an authentication and authorisation system;
- The OIE Delegate's approval and the subsequent release of sequence data should follow the workflow proposed in the agreed business process (see TOR 3);
- The platform's database should include a curated set of 'pathogen genome reference sequences' (see TOR 5) for comparative analyses;
- The platform should offer tools to download sequences;
- The platform should allow submission of either complete or partial pathogen genome sequences;
- The platform should provide an audit trail of data usage;
- The platform should automatically generate and distribute analytical reports to laboratories and Member Countries providing sequences;
- In accordance with the principles of the Nagoya Protocol, the notifying Member Country should retain ownership of the sequence data;
- Users of sequence data should explicitly agree to terms and conditions of usage at registration to the platform. This is particularly relevant to restrictions on distribution and publication of sequence data.

5. Define the minimum set of metadata and the required quality standards to be attached to the sequences submitted to the platform

Dr Natalja Lambergeon, project manager of the OIE World Animal Health Information and Analysis Department (WAHIAD), joined the Group for this part of the meeting.

Epidemiological metadata relating to the disease notification should remain in WAHIS+ and not be duplicated in the platform. Detailed information on methods and quality controls associated with the generation of the submitted sequences should remain in the originating laboratory. However, submission of the sequence should require a statement of compliance to the standards defined in Chapter 1.1.7 *Standards for high throughput sequencing, bioinformatics and computational genomics* of the OIE *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* (*Terrestrial Manual*).

An example of the metadata scheme is defined in Appendix III.

6. Define the information flow for the submission to, and the release of sequences from the future WAHIS+ system, including authentication, authorisation and a description of sequence data governance

The notification, validation and dissemination of epidemiological information on an animal health issue follow a formal process established by the OIE, and endorsed by OIE Member Countries.

The submission of pathogen genomic sequences to the platform will follow a similar process. This process includes the development of a standard operating procedure governing the approval for the upload and release of sequences by the competent OIE Delegate (or Focal point). The approval should also specify the level of disclosure of sequence data (e.g. open access, restricted to eligible partners only, or fully embargoed).

The diagram in Appendix IV shows the draft proposal for the link, submission, and approval for the inclusion of a sequence in the platform. There are four potential stakeholders that are involved in this process. These may include (i) the OIE Delegate (or Focal point), (ii) the national reference laboratory undertaking the initial diagnosis, (iii) an OIE Reference Centre, and (iv) the OIE staff responsible for managing the platform. Sequences are submitted by either the national reference laboratory that performed the diagnosis or an OIE Reference Centre. The OIE Delegate (or Focal point) will have the final decision on the release of the sequence. The schematic diagram shows the submission process for a sequence to the platform. It is important to reiterate that sequence submission is possible only after the confirmation of the notifiable disease (either immediate notification, six-monthly report or annual report on wildlife) by the OIE Delegate.

Submission of pathogen sequences related to six-monthly and annual reports on wildlife have not yet been considered by this Group, but would need to be addressed in the future.

7. Propose strategies to initiate and sustain sequence data sharing, which would involve promoting the active engagement of OIE Reference Centres

Pathogen sequence data, including partial and whole genome sequence information, provides unprecedented diagnostic opportunities as reflected in Chapter 1.1.7 *Standards for High Throughput Sequencing, Bioinformatics and Computational Genomics* of the *Terrestrial Manual*, adopted by the World Assembly of Delegates of the OIE in May 2016. The Group recommended that a strategy to encourage data sharing should demonstrate to the sequence providers the added value and mutual benefits of sharing information. For example, the platform should:

- Provide to the submitter a set of minimum information on the data, e.g. pathogen typing;
- Provide an overview of the global epidemiological context of the pathogen and its associated history;
- Allow diagnostic reagents to be modified to identify emerging pathogen variants;
- Lead to improved ability to track disease outbreaks;
- Allow the development of improved national and international strategies for disease diagnosis, control and prevention;
- Provide advanced access to non-publicly available pathogen sequence data.

Furthermore, data sharing will improve the control of animal diseases and contribute to international public good and the United Nation Sustainable Development Goals.

The Group recognises there are concerns surrounding data sharing including intellectual property rights, data ownership and publication. The Group recommends all data should be submitted and used in accordance with the principles of the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization (ABS) to the Convention on Biological Diversity, and in compliance with common scientific standards and practice.

To foster a sense of active participation in the platform, the Group recommends that OIE reports (annual or six-monthly) include metrics on provision of pathogen-specific sequence data to the platform by Member Countries. This should include highlighting any contributions to reports on global disease trends, in-depth analyses by OIE Reference Centres and any new findings.

8. Identify analytical tools to be integrated into the platform

Dr Paolo Tizzani, Veterinary Epidemiologist of WAHIAD, joined the Group for this part of the meeting.

The platform should have some relatively simple functions and analytical tools that facilitate subsequent handling of the data and enhance user experience. In particular the Group recommends:

8.1. Mandatory functions

- Initial "identity check" of the submitted sequence. This tool will prevent pathogen sequences being submitted by mistake (e.g. wrong file uploaded by the user or spurious sequences);

- The results of the identity check described in the previous bullet point should be reported back to the user with options to make corrections. The system report on identity pass/failure status will be particularly important when users upload batches of sequences;

- Users should have the possibility to override the identity check. Occasionally, there may be instances when the genetic sequence of a specific pathogen deviates substantially from what is known at the time of submission. This sequence should still be submitted despite failing the identity checks. The authorisation levels needed to override the system needs to be carefully evaluated.

8.2. Recommended functions

- *Phylogenetic tool*

  The platform should hold a representative number of pathogen sequences relevant for each specific pathogen ("reference sequences"). These sequences will likely represent, albeit not exclusively, existing "reference strains" for each pathogen. The tool will automatically create a phylogenetic tree where the submitted sequences are compared with the reference sequences for that particular pathogen. Other molecular epidemiological tools may also be considered.

- The Group suggested that these reference sequences are reviewed by new or existing *ad hoc* Groups similarly to the diagnostic tests and vaccines in the disease-specific chapters of the *Terrestrial Manual*.

- *Geographical mapping of submitted sequences*

  Each submitted sequence is linked to a specific notified event. Hence, it should be possible to visualise the geographical and temporal distribution of similar sequences.

9. Identify possible disease models to test the system

As mentioned below, the Group recommended the development of the platform in two stages. The first stage will concentrate on a small number of diseases (3–4). Due to the diversity of the genetic structure of each pathogen (e.g. viruses with a segmented or non-segmented genomes), and the inherent characteristics of different diseases, the platform will need to contain "pathogen-specific" features.

The Group recommended that the expertise and existing work carried out by the OIE Collaborating Centre for Virus Genomics and Bioinformatics (MRC-University of Glasgow, Centre for Virus Research, Scotland, UK), the OIE Reference Laboratory for Avian Influenza and Newcastle Disease (Istituto Zooprofilattico Sperimentale delle Venezie, Padova, Italy), the OIE Reference Laboratories for Avian Influenza and Bluetongue (Australian Animal Health Laboratory, Geelong, Australia) and the OIE Reference Laboratory for Rabies (Animal and Plant Health Agency, Weybridge, UK), and other OIE Reference Centres if needed, be used to expedite the initial stage of the project. Hence, the Group recommended focusing initially on the following diseases: avian influenza, bluetongue and rabies. The Group also recommended considering peste des petits ruminants as a possible additional model disease. The additional advantages of these diseases are that they are caused by viruses with a different genomic structure and global geographical distribution.

10. Determine next steps for implementation of OIE Genomic Platform

The Group considered the steps necessary to advance development of the platform, focusing on developing a preliminary, or beta, version of the system with the disease models proposed in TOR 6.

The Group, having been advised of the planned timeline for WAHIS+ development, proposed developing the beta version of the system in parallel with Stage 1 of the WAHIS+ project. This pilot phase should be completed by the commencement of STAGE 2 of the WAHIS+ project. The human resources and budget required for implementation will depend on whether the system will be newly built or an adaptation of existing software.

The platform and WAHIS+ project have to be developed in close coordination.

11. Timeline and budget

11.1. Timeline

The draft timeline of the pilot phase is reported in Appendix V. The final timeline will be developed by the appointed business analyst.

The platform must integrate seamlessly with WAHIS+. For operational purposes, the platform must be thought of as having two components, a data management and reporting component (core architecture) and the sequence analysis component (the bioinformatics components).

The list of key milestones in the draft timeline is reported in Appendix V:

- Business analyst appointed (Time 0);
- Core architecture of the platform designed and agreed in conjunction with the developers of WAHIS+, in order to assure interoperability of the two systems (Time 3 months);
- Planned core architecture implemented (Time 12 months);
- Bioinformatic component agreed between the developers of the core architecture and bioinformaticians (the Group recommends a plug-in architecture) (Time 6 months);
- Integration of bioinformatic component with the core architecture (Time 12 months);
- Modules for specific diseases developed with the help of OIE experts (Time 18 months);
- Interface module (Graphic User Interface) to allow web access developed (Time 12 months);
- System for reporting back to the user implemented (Time 12 months).

11.2. Budget for the pilot phase

The Group recommended that the platform's developers should access the core architecture of WAHIS+ to adopt compatible technology for the platform.. This would ensure interoperability and guarantee that the two systems will exchange and make use of mutual information. In addition, existing bioinformatic tools will be adapted for the platform, when possible.

The budget for the pilot phase includes the following items (in brackets is the estimated time for the completion of the assigned tasks):

- One general project manager for 18 months (*already appointed*);
- One business analyst for 12 months;
- Meeting of a small 'Technical Group' (general project manager, business analyst, and experts) to develop system specifications);
- At least one software engineer and half-time (0.5) system administrator for 18 months;
- At least one bioinformatician for 18 months;
- At least one *ad hoc* 'Focus Group' to discuss the development of modules for the selected diseases;

- At least one independent consultant to revise and budget the entire project including the steps beyond the beta version for 3 months;
- Travelling costs ;
- Other costs (hardware, software, cloud computing, etc.).

The estimated budget is approximately 500,000 euros, without the general Project Manager, based on average costs. The Business Analyst could help to generate a more accurate budget and timelines for the development of the entire platform.

12. Other considerations

- Particular attention needs to be given to the periodic review of Chapter 1.1.7 *Standards for High Throughput Sequencing, Bioinformatics and Computational Genomics* of the *Terrestrial Manual*, given the rapid advances in genomic technologies and evolution of genomic standards. In addition, bioinformatic processes should require validation or accreditation.
- Disease-specific chapters of the OIE *Terrestrial Manual* need to incorporate the methods and recommendations arising from the use of sequencing technologies for diagnostic purposes.
- Other technologies may be introduced in the future that are not based on genomic sequencing *per se* (e.g. mass spectrometry).
- The Group envisaged that OIE Reference Centres could facilitate the adoption of specific sequencing and bioinformatics approaches for individual pathogens.
- The platform may be exploited for resource mobilisation and capacity development in developing countries. Advocacy at the national level beyond the Agricultural Authorities, with, for example, the Ministries of Finance, Science and Technologies and other stakeholders, may be advantageous for animal disease control.

13. Any other matters

None.

14. Finalisation and adoption of the draft report

The Group finalised and adopted the draft report.

_____

…/Appendices

Appendix I

MEETING OF THE OIE *AD HOC* GROUP ON HIGH THROUGHPUT SEQUENCING, BIOINFORMATICS AND COMPUTATIONAL GENOMICS

Paris, 27–29 June 2017

Background

In 2012, an *ad hoc* Brainstorming Group met at the OIE Headquarters to identify opportunities and challenges of new sequencing technologies in the diagnosis, surveillance, and control of OIE listed animal diseases.

In 2013, the first *ad hoc* Group on High Throughput Sequencing, Bioinformatics and Computational Genomics (HTS-BCG) was convened to address specific tasks such as the drafting of OIE standards for HTS-BCG, and the blueprint of an OIE web-based platform to collect sequence data of pathogens that are the subject of notifications by Member Countries. To date, the outcomes include (i) a new chapter of the OIE *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* on standards for HTS-BCG, (ii) an outline of technical specifications of the OIE web-based platform, and (iii) a wireframe website of the platform to display main functionalities and possible developments.

The meeting of the *ad hoc* Group in June 2017 will discuss three main issues, namely, i) which specific metadata and sequence data will be stored in the OIE web-based platform, (ii) sequence data governance, and (iii) ways to promote a culture of sequence data sharing among the OIE Reference Centres.

Terms of Reference

1. Define the minimum set of metadata and the required quality standards to be attached to the sequences
2. Identify possible disease models to test the system
3. Define the information flow for the submission to, and the release of sequences from the future WAHIS+ system, including authentication, authorisation and a description of sequence data governance
4. Propose strategies to initiate and sustain sequence data sharing and active engagement within the network of OIE Reference Centres
5. Identify analytical tools to be integrated into the platform
6. Determine next steps for implementation

Agenda

1. Opening
2. Designation of the chair and rapporteur
3. Adoption of agenda
4. Consideration of Terms of Reference
5. Other business
6. Adoption of report

Appendix II

MEETING OF THE OIE *AD HOC* GROUP ON HIGH THROUGHPUT SEQUENCING, BIOINFORMATICS AND COMPUTATIONAL GENOMICS

Paris, 27–29 June 2017

———

List of participants

MEMBERS

Prof Massimo Palmarini
Director, Centre for Virus Research
MRC - University of Glasgow
OIE Collaborating Centre for Viral Genomics and Bioinformatics
464 Bearsden Road, Glasgow G61 1QH, Scotland
UNITED KINGDOM
massimo.palmarini@glasgow.ac.uk

Prof Gary Van Domselaar
Chief Bioinformatics, National Microbiology Laboratory, Public Health Agency of Canada, Canadian Science Centre for Human and Animal Health, Winnipeg R3E 3R2
CANADA
gary.vandomselaar@phac-aspc.gc.ca

Dr Isabella Monne
Istituto Zooprofilattico Sperimentale delleVenezie (IZSVe)
Viale dell'Università, 10
35020 Legnaro (Padova)
ITALY
imonne@izsvenezie.it

Dr Vish Nene
Director, Animal and Human Health Program, co-Leader (ILVAC), International Livestock Research Institute (ILRI)
Old Naivasha Road, P. O. Bo 30709, Nairobi 00100
KENYA
v.nene@cgiar.org

Dr Garbiele Casadei
Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "Bruno Ubertini" (IZSLER)
Via Bianchi, 9, 25124 Brescia
ITALY
gabriele.casadei@izsler.it

Dr James Watson
Research Director - Diagnosis, Surveillance and Response (Acting), Veterinary Investigation Leader, Australian Animal Health Laboratory
CSIRO
Private Bag 24 Geelong Victoria 3220
AUSTRALIA
james.watson@csiro.au

BIOLOGICAL STANDARDS COMMISSION REPRESENTATIVE

Prof Anthony R. Fooks
Head of Wildlife Zoonoses and Vector-Borne Diseases Research Group
Virology Dept
Animal and Plant Health Agency (APHA)
Woodham Lane, New Haw
Addlestone. Surrey. KT15 3NB
UNITED KINGDOM
tony.fooks@apha.gsi.gov.uk

OIE HEADQUATERS

Dr Matthew Stone
Deputy Director General, International Standards and Science,
12 rue de Prony, 75017 Paris
FRANCE
oie@oie.int

Dr Elisabeth Erlacher-Vindel
Head of the Science and New Technologies Department
e.erlacher-vindel@oie.int

Dr Antonino Caminiti
Chargé de mission
Science and New Technologies Department
a.caminiti@oie.int

Dr Natalja Lambergeon
Project Manager
World Animal Health Information and Analysis Department
n.lambergeon@oie.int

Dr Paolo Tizzani
Veterinary epidemiologist
World Animal Health Information and Analysis Department
n.lambergeon@oie.int

———————

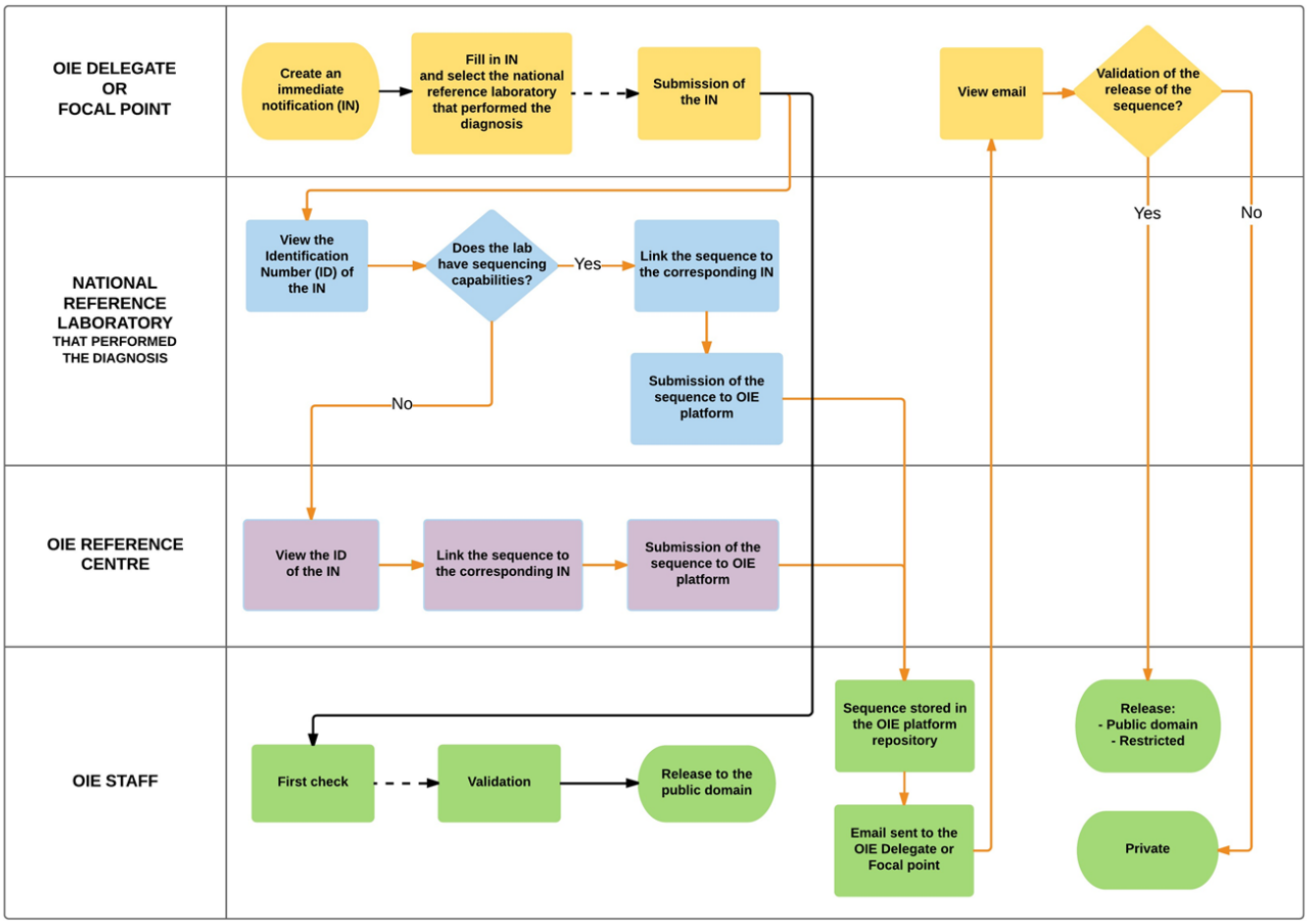
Metadata scheme for sequence data

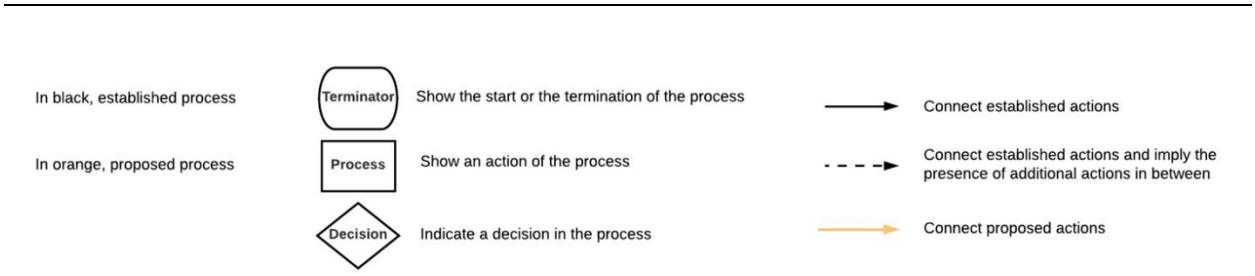
	Name	Description	Mandatory	Repeatable*	Valid Entries	Notes
1	**SourceLab**	Laboratory	Y	N	OIE laboratory identifier	
2	**SourceID**	Unique identifier from laboratory	Y	N	Unique identifier for the biological sample from the source laboratory database	
3	**SampleSource**	Source of sample – clinical, environmental etc.	Y	N		
4	**SampleType**	Primary sample or culture	Y	N		
5	**SampleHistory**	Passage history	N	N		
6	**OIEDisease**	OIE disease (species, subtype)	Y	N	Reference to the OIE Codes	
7	**Taxon**	Taxonomic Identifier	Y	N	Select from predefined list of species to be defined for all OIE listed diseases	
8	**Strain**	Strain / Type	N	N	Organism-specific strain ID	
9	**SubStrain**	Subtype / Clade	N	N	Organism-specific sub-strain ID	
10	**SeqDesc**	General description of the material	Y	N	Free text notes	
11	**SeqType**	Whole or partial genome	Y	N		
12	**SeqQuality**	Quality assessment	Y	N	Flag whether sequence meets QC guidelines	Three level flag: - Meets OIE standards - Substantially meets OIE standards - Does not meet OIE standards
13	**SeqContam**	Contamination check	Y	N	Flag to show check performed	
14	**SeqHost**	Host	N	N	OIE identifier for the species from which the sample has been collected	
15	**SeqDate**	Collection date	Y	N	The date the sample was collected	
	SeqMethod	Method or platform used to generate sequence	Y	N		
	SeqLibrary	Library preparation method	N	N		
16	**SeqRef**	Sequence	N	Y	Link/s to sequence information (e.g. GenBank)	Allows multiple references to be listed
17	**Reference**	Literature	N	Y	Citation/s for relevant literature describing the sequence	Allows multiple references to be listed
17,1	**ReferenceDOI**	Literature	N	N	Digital object identifier (DOI) for reference	
17,2	**ReferenceURL**	Literature	N	N	Uniform Resource Locator (URL) for reference	

* Repeatable: the field accepts more than one entry

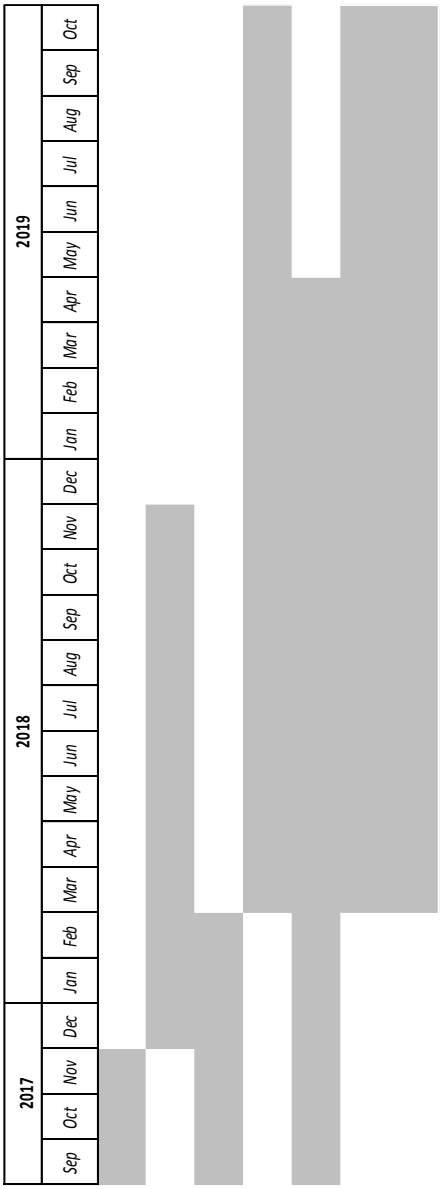
Appendix IV

Draft proposal for the link, submission, and approval for release of a sequence in the OIE Genomic platform





Timeline of the pilot phase



ID	Task Name	Start	Finish	Duration
1	Core Architecture Planning	01/09/2017	01/12/2017	13.2w
2	Core Architecture Implementation	01/12/2017	30/11/2018	52.2w
3	Bioinformatics Planning	01/09/2017	01/03/2018	26w
4	Bioinformatics Integration	01/03/2018	28/10/2019	86.6w
5	Disease Module Development	01/09/2017	30/04/2019	86.6w
6	Interface Module Development	01/03/2018	28/10/2019	86.6w
7	Reporting Development	01/03/2018	28/10/2019	86.6w