WP 1, Milestone 2: Data collation, formatting and automatic updating tools

Author: Robin Simons, Verity Horigan, Amie Adkin
Institute: APHA
Country: United Kingdom
Phone: +44(0) 01932 357837
e-mail: Robin.Simons@apha.gsi.gov.uk
Due date: 01/09/15
Completion date: 31/12/15
Version: 2

List other involved partners: N/A

(Name, institution, country, phone, e-mail)

No.	Name	Institute	Country	Phone	e-mail
1					
2					
3					
4					

SPARE is funded by the European Commission Animal Health and Welfare

Contents

1. Executive Summary

The purpose of this task was to format input data for the risk assessment model and assess the possibility of real time updates of datasets using automated macros. There are a wide variety of relevant datasets required to assess input data; (i) the pathogen presence in each country globally, and (ii) the scale of transmission in classes such as legal imports of meat, live animals, and vectors. An assessment was also made on the quality of data within each dataset and whether proxy data would need to be assigned for any pathway. Table 1 summarises the key data required and whether the model can currently automatically update the data or manual input is needed. A qualitative assessment of the data quality based on assessor's opinion is also provided. The quality assessment is based on factors such as whether there was a quality assurance statement and whether it was thought better data could be realistically obtained leading to a ranking of quality rather than an independent value per factor. For example, while there are almost certainly data gaps in the OIE disease prevalence data, it is assessed to be the most comprehensive data source available at the current time across the range of case study pathogens, and therefore assessed to be of high quality. Finally we provide an indication of whether the data are actual observations or estimated based on other proxy data (e.g. habitat suitability). Further details for each factor are provided in the subsequent sections of this report.

Table 1: Summary of required data, whether the SPARE model can automatically update the data or manual update is necessary, an assessment of data quality and whether the data are a primary source

Factor	Dataset	Data update	Data quality	Proxy data?
Disease Prevalence	Wahid/OIE	automatic	High	No
	Handistat II	N/A (historic)	High	No
	EMPRES-i	manual	High	No
Trade	Comext	automatic	High	No
Transport	Eurostat	automatic	High	No
Live Animal imports	TRACES	manual	Medium	No
Vector prevalence	ECDC	manual	Medium	Yes
	VectorMap	manual	Low	Yes
	SPARE	manual	High	Yes
Wild birds - range	IUCN red list/birdlife	manual	High	No
Wild birds - population	Eionet	manual	Medium	No
Wild birds- migration	various	manual	Medium	No
Pathogen specific data	various	manual	N/A	No

2. Disease prevalence data

2.1 OIE/Wahid reports

The key dataset for animal disease presence is incidence data reported to the OIE, with data from 2005 available via WAHID, and with prior data from HandistatII (1996-2004), by exporting country for the relevant time period. For certain pathogens/time periods this dataset can be combined with others including FAO: EMPRES-I, and USDA. Rules to account for missing data and data quality issues have been collated and based on previous risk assessment assumptions.

The OIE immediate and follow up reports are available from the Wahid database section of their website: http://www.oie.int/wahis_2/public/wahid.php/Wahidhome/Home
The format of the database is not helpful for accessing data. However all reports from the last year are listed on a separate page as weekly disease information:
http://www.oie.int/wahis_2/public/wahid.php/Diseaseinformation/WI

Annual data for each country are also available by year, and can be accessed from the OIE reporting history page;
http://www.oie.int/wahis_2/public/wahid.php/Countryinformation/Reporting

The SPARE model has functionality to access these WAHID reports and extract the relevant information. Code has been written in R to automatically obtain historical data from 2005 (e.g. pathogen, country, species affected, number of outbreaks, number of animals susceptible, cases, deaths, destroyed and slaughtered) from the annual reports. The data pre 2005 (i.e. from Handistat II) currently need to be obtained manually. These data can then be augmented with recent data from the weekly disease information page. There is also code to manipulate these data into an Excel spreadsheet ready for use by the SPARE model. This formats the data into monthly cases by country, with an ISO3 code for the country and a total number by country.

2.2 Empres-i database

The Empres-i database is available online: http://empres-i.fao.org/eipws3g/
This data is unable to be obtained automatically due to the formatting of the website. New records must be downloaded manually. Code has been written to manipulate these data into an Excel spreadsheet ready for use by the SPARE model. This formats the data into monthly cases by country, with an ISO3 code for the country and a total number by country.

3. Legal Trade data

3.1 Comext

The scale of legally imported trade is based on data obtained from the EU Commission

COMEXT-EUROSTAT database by mode of introduction; for example, products of animal origin, and products capable of sustaining vectors. Trade data from COMEXT can be found on the Eurostat download page: http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?sort=1&dir=comext%2F201506
These files include data by month as well as summary data for the whole year, where there are completed years of data (most recent 2014).
An R script has been written to determine the location of the most recent data, download it and unzip into a local file in the SPARE directory. The code may need to be amended to specify which files and which directory.

Other databases on trade data include FAOstat: http://faostat3.fao.org/home/E and ComTrade: http://comtrade.un.org/. However, these are not set up for easy updating and there are differences in the type and level of detail of the product groupings. Such data could be used to supplement data gaps within the COMEXT data.

4. Transport data

4.1 Eurostat

The movement of people (and international catering waste, potential illegal items or vectors on flights) will be based on data from Eurostat, which holds data on national and international transport into EU countries: http://ec.europa.eu/eurostat/data/database. These databases can be downloaded from the bulk download facility. An R script has been written that has a function, *Eurostat,* that will download files if you input the appropriate file name, it is based on the one in the *Eurostat* library[1]. However, further cleaning is necessary for some files before use in the full model. The relevant Eurostat files include:

- *avia_paexcc:* Data on commercial passenger flights with an arrival or departure in an EU MS, with the following columns.
- *avia_goexcc:* This dataset is in the same format as *avia_paexcc*, but the data are for freight aircraft and the measurements are also by month and quarter as well as year.
- *mar_pa_qm_xx:* Maritime transport data; quarterly for passengers going to MS xx (different file for each relevant, i.e. not landlocked, MS, e.g. *mar_pa_qm_be* for Belgium).
- *mar_go_qm_xx:* Maritime transport data; quarterly for goods going to MS xx (different file for each relevant, i.e. not landlocked, MS, e.g. *mar_pa_qm_be* for Belgium)
- *rail_pa_intgong*: International rail passenger transport from the reporting country to the country of disembarkation (1,000 passengers).
- *rail_go_intld*: International monthly railway transport from the loading country to the reporting country.

[1] http://cran.r-project.org/web/packages/eurostat/index.html

- *road_go_ia_rc*: International annual road freight transport by country of loading and unloading with breakdown by reporting country (lack of available data in this one, many marked as confidential).

5. Live animal data

5.1 TRACES

The scale of legally imported trade of live animals is based on data held within the EU Commission Trade Control Experts System (TRACES). Live animal data, including those not for trade purposes such as pets and animals for scientific research, are all available from the TRACES database: https://webgate.ec.europa.eu/sanco/traces/. The data used is based on veterinary certificates for imports into the EU and includes the country of origin, the importing country and also the location of the border inspection post it went through. In some cases the importing country may not be in the EU, but the border inspection post is.

Previously an extract from TRACES has been obtained by APHA and formatted for use in a similar model. As the database requires user log in, it is not possible for the SPARE model to automatically update these data.

5.2 Livestock Density data

Livestock density data are also available for cattle, pigs, poultry, sheep, goats and buffaloes can be obtained from the FAO gridded livestock density maps.
http://www.fao.org/ag/againfo/resources/en/glw/GLW_dens.html

6. Vector data: mosquitos/ticks/midges

6.1 ECDC

The ECDC has vector maps on some species for Europe and information on global distribution on the species specific webpages.
http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/vector-maps.aspx
http://ecdc.europa.eu/en/healthtopics/vectors/Pages/vectors.aspx
These data are not updateable

6.2 VectorMap

ESRI grid data and Ecological niche modelling (maxent) results for a number of different tick species (i.e. spatially explicit vector presence data, latitude and longitude, and maps of vector presence according to a ecological niche modelling algorithm):
http://www.vectormap.org/Tick_Metadata.htm#vec169

6.3 WP3 Vector competence

Data on habitat suitability of *Hyalomma marginatum* will be collated as part of SPARE project and maps produced.

7. Wild bird data

7.1 Range – IUCN red list/Bird life

Data on the range (in terms of countries present) of a number of wild birds are available from the IUCN red list. These data are compiled by Bird Life:
http://www.birdlife.org/datazone/
A list of wild birds reported to be in the EU has been compiled by the European Commission and is used in this model to define the birds present in the EU
http://ec.europa.eu/environment/nature/conservation/wildbirds/eu_species/voous_en.htm
The 'List according to Voous' is used here. A .csv file of the country range for these bird species is included in the model, this is not automatically updateable. As is an R script that can download the country data for a particular bird species from the IUCN red list website.

7.2 Bird population numbers – Eionet

Data on wild bird populations in EU countries are available from the Eionet website:
http://bd.eionet.europa.eu/article12/report?period=1&country=

These data give estimates for the minimum and maximum population size by season (winter, breeding, passaging) and population type (number of pairs, number of calling/lekking males, number of individuals or number of males). These data are provided as .csv input files to the model. An R script was written to download these files, but did not work for a few EU MSs whose data had to be obtained manually. It is not expected that they will be updated, so automatic updating is not necessary.

7.3 Migration pathways

Data on exact migration routes are limited. As such it is difficult to determine which countries European birds will migrate to; just because a species is present in two countries doesn't mean that individuals are travelling between them.

Various data sources to do with bird ringing exist, such as EURING:
http://www.euring.org/index.html, but data tend to be limited to a confined geographical areas (e.g. Europe) and/or data are poor for areas like Africa, suggesting the dataset is unlikely to be useful for our purposes. Additionally, these datasets are not freely available.

General migration flyways have been proposed, which are geographical regions where it considered that wild birds will migrate in. Example maps:

http://www.birdlife.org/worldwide/programme-additional-info/migratory-birds-and-flyways

http://ag.udel.edu/research/allenlab/info/AvianFlu.htm

http://n2gf.com/wp-content/uploads/2015/05/Flyways-migratory-waterbirds.jpg

https://nioo.knaw.nl/en/news/ecofactsheet-migratory-birds-bird-flu

Thus a simplistic methodology is proposed to use migration flyways as a proxy; it would be assumed that bird species travel between any countries in the same flyway, where data suggest they are present and do not travel between any other countries. These data are not updateable.

8. Pathogen specific data

The data collection for pathogen specific data, other than prevalence, (e.g. duration of infection, dose response etc…) is ongoing, but an initial assessment of the availability of the data was conducted as part of WP4.1. It is anticipated that these data will be based on focussed research studies, e.g. articles in peer reviewed journals, and as such will not be able to be updated automatically.