CENTRAAL VETERINAIR INSTITUUT
WAGENINGEN UR

SCIENTIFIC / TECHNICAL REPORT submitted to EFSA

Epidemiology of different agents causing disease in aquatic animals: scientific review and database development[1]

Prepared by:

Barry Hill (Project co-ordinator), Allan Reese, Peter Dixon, Birgit Oidtmann, Richard Paley, Edmund Peeler, Grant Stentiford, David Stone, Keith Way and **Mike Hine.**

Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom

Paolo Calistri, Carla Ippoliti, Alessio Di Lorenzo and **Lara Savini**

Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, Italy

Olga Haenen and **Marc Engelsma**

Central Veterinary Institute (CVI) of Wageningen UR, Lelystad, Netherlands

[1] **Question No EFSA-Q-2008-05009 (CFP/EFSA/AHAW/2008/01). Accepted for Publication on 15 January 2010**

SUMMARY

Critical scientific reviews were carried out on the geographical occurrences and host ranges of the listed diseases in Council Directive 2006/88/EC as amended by Commission Directive 2008/53/EC (de-listing of spring viraemia of carp), and an expert assessment and analysis was carried out on the efficacy of current methods for their diagnosis and pathogen strain identification/differentiation. All the reviews were carried out by a systematic method, similar to the Cochrane approach, developed during the early stages of the project, and were subjected to an independent quality audit. Disease occurrence and/or pathogen detection information was entered into an Excel data table for that disease in order to facilitate the transfer of the information into a database.

A total of almost 3000 scientific publications and other documents identified by the literature searches, and other means, for the listed diseases were assessed for relevant information and of these 463 were critically reviewed.

A review summary report was prepared for each disease, including causative agent description, available tools for typing (strain identification), descriptive epidemiology including the current worldwide distribution of the host and pathogen (including different strains) and outbreak data in the EU and worldwide. The summary reports for all the listed diseases are presented at Annex B.

The quality audit of the disease reviews was carried out by an independent expert. The reviewers were asked to provide the auditor with between 5 and 10 reviewed documents for each disease/pathogen/. For each source, the entries made in each field of the Excel data entry form were checked independently by the review auditor against the information provided in the relevant publication. The entire record was checked for consistency, and particular attention was paid to how the following fields had been completed: i) classification of report, ii) report type, iii) quality assessment and iv) mortalities. Any omissions or inconsistencies were reported back to the reviewer concerned who then amended their data table accordingly. All the reviews were compared and checked for consistency. The auditors report is presented at Appendix C.

Particular attention was given to the host range and geographical distribution of the various genotypes of VHSV and of the various *Bonamia* species, addressing doubts about the identification and taxonomy of some, as well as their association with natural outbreaks. Documents containing genotypic information on VHSV were critically reviewed to establish a consensus on the most appropriate typing scheme (the internationally accepted typing scheme) and this scheme was then used to reassess the genotype of the virus isolates entered in the Excel data entry table by the VHS reviewer. In addition, data gaps for VHSV and Bonamia spp. have been identified where further research is needed to address the issues of pathogen definition and strain differentiation. A detailed report of the review is presented at Annex D. For the critical review on the identification of *Bonamia* species a concise overview was prepared on the characteristics of known *Bonamia* species (*Bonamia exitiosa*, *Bonamia ostreae*, *Bonamia roughleyi*) and more detailed information collected on the ultrastructure, phylogeny and taxonomy issues of *Bonamia* species. In addition, three phylogenetic analyses were carried out based on available GenBank sequences to demonstrate the phylogenetic relationship between the *Bonamia* species. Finally, data gaps for *Bonamia* spp. were

identified where further research is needed to address the issues of pathogen definition and strain differentiation. A detailed report of the review is presented at Annex E.

A database and GIS application was developed to fit the requirements for a geographical representation in a web application of the epidemiological data contained in the Excel spreadsheets completed for each disease. The data tables have been loaded into the GIS database that allows the geographical spread of outbreaks to be displayed as points on a map. The epidemiological data for VHSV and *Bonamia spp.* provided by the reviewers has been inserted into the database. The diseases occurrence is represented as polygon-based layers based on administrative boundaries at Country/regional/province level and as point-based layers (specific locations) geo-referenced through latitude and longitude data, when geographical coordinates of the point location are available. The occurrences of VHS and *Bonamia spp.* infections are visualized on the web-GIS application, and the related epidemiological data recorded by the reviewers are accessible through queries.

The application is accessible via the web, and is an interactive tool to navigate into the epidemiological data collected. Using the query tool, the user can retrieve data from the database, and the selected information is shown on the map, reported in summarized form on the left side of the page. The user can also access the table with the complete information related to chosen criteria in tabs over the map. A full description of the database and web-GIS application is provided at Annex G.

This project has created a data resource for EFSA that will inform expert work on each of the diseases in Council Directive 2006/88/EC as amended by Commission Directive 2008/53/EC.

TABLE OF CONTENTS

BACKGROUND

The Scientific Panel on Animal Health and Welfare (AHAW) has adopted three Scientific Opinions dealing with the risk of transmission of infectious diseases amongst aquatic animals (fish, molluscs, and crustaceans) by means of transfer of vector species (EFSA 2007a, b and c) and a Scientific Opinion on aquatic animal species susceptible to the diseases listed in the Council Directive 2006/88/EC (EFSA, 2008). During the development of these opinions, the very wide variation in strain pathogenicity of fish viruses listed in Council Directive 2006/88/EC, and the wide distribution of non pathogenic strains were identified as issues to be further addressed. It was pointed out that the lack of uniformity in case definition of listed aquatic diseases, and their causative agents, was a source for potential difficulties in relation to pathogen identification and diagnosis.

Research efforts to support the development of the European aquaculture sector have focused, among others, on infectious diseases of aquatic animals as a major production limiting factor, and particularly diseases listed in the Council Directives 91/67/EEC, 93/53/EEC, 95/70/EC and 2006/88/EC. As a result of the efforts to improve our understanding of aquatic pathogens, and in conjunction with the application of new biotechnology tools, the large biodiversity of these organisms has gradually been recognized with numerous strains and types.

As an example, four genotypes of viral haemorrhagic septicaemia (VHS) virus have been described which represent different strains of the virus causing a broad spectrum of disease in various fish species under various environment conditions. According to available data, genotype IV is exotic to Europe, while genotypes II and III are marine types of the virus. Until recently, no natural infection of rainbow trout with types II, III, or IV had ever been detected. However, in late 2007, a particular outbreak of VHS in sea trout has challenged the robustness, relevance, and accuracy of the current typing systems and our approach to pathogen biodiversity and epidemiology.

Similarly, the question of the epidemiology of the two main species of *Bonamia, B. ostreae* (Northern hemisphere) and *B. exitiosa* (Southern hemisphere) was regarded as being well settled until recently. However, the description of *Bonamia exitiosa*-like organisms in Latin America and the detection of a *B. exitiosa* type in Europe, where it was believed to be exotic, are coming to question the taxonomy of the group. Beyond the question of a possible recent introduction of this species, the question of typing and its epidemiological congruence is increasingly raised and needs to be addressed.

The existence of highly pathogenic strains is a well-known fact, and the need to differentiate them from the non-significant (or non-pathogenic) strains is well recognized. The challenge for regulators and risk managers is to determine which of these strains of pathogens should be covered by control measures and which could be regarded of lower significance. However, in the absence of guidance on the matter, approaches to address this issue across the different groups of pathogens, host species and regions of the world have varied.

The Aquatic Animal Health Standard Commission of the World Organisation for Animal Health (OIE) has proposed guiding principles for an appropriate approach to pathogen differentiation. Those principles mainly put emphasis on the needs for robust taxonomy of pathogens under consideration, and strict validation of the diagnostic and typing techniques. However, the lack of knowledge in the actual epidemiology of different strains, their temporal and spatial distribution, their involvement in natural outbreaks, and their phylogenetic

interrelations, is hampering the appropriate and adequate approach to strain differentiation. Therefore such knowledge is of key importance.

TERMS OF REFERENCE

i) Provide updated critical scientific reviews on diseases of fish and shellfish listed in Council Directive 2006/88/EC, with special focus on their causative agents, methods for testing and typing of the resulting isolates;

ii) Describe the current host and geographic distribution of the various strains of VHSV and *Bonamia* spp. within the European Union, as well as their association with natural outbreaks;

iii) Establish a GIS multilayered database including reference to strains of the pathogen, dates and geographic location of isolation, host species, outbreaks description, and any other information necessary to ensure comprehensive mapping of types/strains of the relevant pathogens;

iv) Identify the data gaps for VHSV and *Bonamia* spp. and where further research is needed to address the issues of pathogen definition and strain differentiation.

ACKNOWLEDGEMENTS

This contract/grant was awarded by EFSA to:

Centre for Environment, Fisheries and Aquaculture Science (Cefas), Barrack Road, Weymouth, DT4 8UB, United Kingdom.

Contract/grant title:

Epidemiology of different agents causing disease in aquatic animals: scientific review and database development

Contract/grant number:

EFSA/AHAW/2008/01

1. OBJECTIVES, PARTICIPANTS AND PROJECT TIMELINES

1.1. OBJECTIVES

The objectives of this project as described in the Grant Agreement can be summarised as:

1. To conduct a detailed critical scientific review and update on the geographical occurrences and host ranges of the listed diseases (6 of fish, 5 of molluscs and 3 of crustaceans) in Council Directive 2006/88/EC as amended by Commission Directive 2008/53/EC (de-listing of spring viraemia of carp), and an expert assessment and analysis of the efficacy of current methods for their diagnosis and pathogen strain identification/differentiation. The diseases in question are:

Fish diseases

Epizootic haematopoietic necrosis (EHN)
Epizootic ulcerative syndrome (EUS)
Viral haemorrhagic septicaemia (VHS)
Infectious haematopoietic necrosis (IHN)
Koi herpesvirus disease (KHVD)
Infectious salmon anaemia (ISA)

Mollusc diseases

Infection with *Bonamia exitiosa*
Infection with *Bonamia ostreae*
Infection with *Perkinsus marinus*
Infection with *Microcytos mackini*
Infection with *Marteilia refringens*

Crustacean diseases

Taura syndrome
Yellowhead disease
White spot disease

In addition, there was a requirement to identify the data gaps for VHSV and *Bonamia* species and where further research is needed to address the issues of pathogen definition and strain differentiation.

2. To develop a GIS multilayered database including reference strains of the pathogen, dates and geographic location of isolation, host species, outbreak descriptions, and any other information necessary to ensure comprehensive mapping of types/strains of the relevant pathogens.

1.2. PARTICIPANTS AND PROJECT TIMELINES

The Grant Agreement for this 9-month project was signed by EFSA on 9th Dec 2008 but was not received by the Beneficiary until 12 January 2009. Receipt of the signed document was essential for the Beneficiary and Co-beneficiaries to commit time and expenditure on the project. Furthermore, due to difficulty in finding a date available to the Beneficiary, Co-beneficiaries and the EFSA project officer for the kick-off meeting, it was not possible to make a full start on the project until mid-February 2009. Therefore, the Beneficiary asked EFSA for a no-cost 2 months extension to the deadline so that the project end date would be 9

Nov 2009 rather than 9 September 2009. This extension was formally agreed via Amendment No.1 to the Grant Agreement signed by EFSA on 1 April 2009.

The consortium of partners in the project comprised:

Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom (Beneficiary)

Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, Italy (Co-beneficiary 1)

Central Veterinary Institute (CVI) of Wageningen UR, Lelystad, Netherlands

(Co-beneficiary 2)

A kick-off meeting involving representatives of the beneficiary, the co-beneficiaries and EFSA was held in Paris on 15 February 2009. At the meeting, the requirements of EFSA were clarified and the approaches and methods to be used to meet them were discussed and agreed.

An interim meeting was held at EFSA, Parma on 18 May 2009 at which the interim report was presented by the Beneficiary and Co-beneficiaries and clarifications sought and provided by EFSA on any adjustments needed to the output to date and the planned work for the rest of the project period.

A workshop on Database and web-GIS training was held in Teramo, at the Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale" on 21-22 October 2009.

The final meeting was held at EFSA, Parma on 30 October 2009 where the draft final report was presented by the Beneficiary and Co-beneficiaries, during and after which EFSA requested amendments.

This Final Report was submitted electronically to EFSA within the 14 days period allowed after the project end date of 9 November 2009.

2. MATERIALS AND METHODS

2. 1. THE REVIEW APPROACH

In response to the original application, EFSA advised that the critical reviews should be done in a systematic manner by use of a method such as the Cochrane approach, or similar, and details of how this would be done were explained in the second application. The Cochrane method has been developed in the context of human medicine and explicitly deals only with randomized controlled trials (RCTs) of medical interventions: the research question is explicitly "is treatment Y demonstrably better than treatment X." The medical effect has to be measurable and formal meta-analysis is used to combine results from otherwise unrelated trials. "The ultimate goal of the Collaboration is to help people make well informed decisions about health care" (eppi.ioe.ac.uk/cms/).

A side-effect of making reviews and meta-analysis more controlled and open has been the development of further standards — e.g. the Consolidated Standards for Reporting Trials or Consort (www.consort-statement.org) and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) (www.strobe-statement.org) — to encourage researchers to plan and report such studies in ways that make comparisons easier and fairer. [Side effect

is that quality of reporting can be assessed against standard, and forms part of quality assessment of study.]

All Cochrane reviews are registered with the international Cochrane Collaboration (founded 1993) which operates through national and regional centres. There is a very small number of paid staff but reviews are carried out by volunteer or project-funded effort since they rely upon specific expertise in evaluating the literature. Training material on the methods can be found through the Cochrane website, in particular the Cochrane Handbook for Systematic Reviews of Interventions (www.cochrane.org/resources/handbook).

Cochrane reviews are made public through the Cochrane database and have been used as evidence by bodies that determine policy: e.g. NICE (http://www.nice.org.uk/nicemedia/pdf/QuittingSmokingInPregnancyOverview.pdf) and EFSA ("Scientific Opinion on the substantiation of health claims related to fluoride" EFSA Journal 2009; 7(9):1212), combining the Cochrane review findings with economic and societal concerns.

Another group that develops systematic review methods in a wider field is the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) is part of the Social Science Research Unit at the Institute of Education, University of London. Their remit covers education, health promotion, employment, social care, and crime.

The current project was not appropriate to be registered with the Cochrane organisation since the project deals with animal health and the literature being studied was not restricted to a particular method. Nevertheless, we have benefited greatly from adopting the philosophy and systematic methods of the Cochrane approach (ibid). The overall aims of Cochrane are to be comprehensive, using all available data, and to avoid bias, whether generated within individual studies or by the review process.

2.2. DATA CONSIDERATIONS

The raw materials for the current review were reports, published and unpublished, of outbreaks and occurrence of named diseases. We could define an outbreak as an observation of the disease in a species or area where it had previously been absent. In contrast, an occurrence is an observation that the disease agent is present but in the knowledge or assumption that this was a continuing presence. It is, however, often difficult to distinguish these is published reports due to the sporadic nature of data collection and the well-recognised biases by authors and editors in deciding what is worth writing up and publishing.

Systematic reviews of methods for diagnosis and typing as applied to the reports were also expected. The quality assessment of diagnostic methods was left to individual reviewers, as in many cases the problem is that diagnosis and definition of the conditions are intimately linked; it is not just a question of applying specificity and sensitivity considerations to compare novel tests with a gold standard.

A concept that implicitly runs through any data capture, storage and analysis process is that of 'units'. Here the units of sampling are publications, but a single publication may report more than one outbreak or area of occurrence. The database group planned to create tables in which one unit (row of a table) would be an outbreak or an occurrence. The principle we tried to implement was for reviewers to extract as much information from each source in a natural structure that would allow audit of that stage. The mapping of source material into database structures would then be a second and distinct step. This separation allows any restructuring

and inference drawn from wider expertise – i.e. adding to what was directly stated in the source – to be identifiable and therefore open to later revision. Analysis may subsequently use other units, particularly geographical splits.

From the point of view of scientific review, something to avoid is to design a database too early and too rigidly. In such situations information is lost because codings are forced into the categories provided, and do not reflect a full and natural interpretation of the source material.

Reviewers were probably not previously conversant with the distinctions in data units, and source documents had certainly not been structured with recognition of the needs of data collation. Hence it was expected that there would be difficulties in interpreting rules and that reviewers would make individual decisions, or ask the coordinator make decisions, that would need later consideration for consistency.

For both the above considerations, although it was agreed from the outset that it was important to define concepts and rules, it was also recognised that these would generally emerge from the data, and that making the application consistent across numerous reviewers would be an iterative process. Following a structured and systematic process enables a review to be paused at any stage and resumed by the same or another reviewer without the need to backtrack and repeat stages.

It was also considered desirable to make copies of all sources as used in the current review and deposit these with EFSA as an archive. Taking an archive of materials used had appeared to be part of Cochrane method. An enquiry to the Cochrane organisation revealed that this is far from standard and it is usually assumed only that individual reviewers will keep the material they personally reviewed. That would be consistent with the definition of "personal research" applied in UK copyright law. There appears to be issues with regard to sharing of copyrighted material that are outstanding, and would be resolved only by case law.

2.3. INSTRUCTIONS TO REVIEWERS

Reviewers, as experts in the individual diseases, generally felt they were aware of current literature for their areas. Prior to the kick-off meeting for the project, there had been informal discussions that identified key fields to record for each outbreak or occurrence. These were proposed at the kick-off meeting as column headings in a data-capture table to be implemented in Excel as the software that was available and familiar to all reviewers.

After further discussions, an operating procedure, guidelines (see Annex A) and a workbook were distributed by the review process manager (Allan Reese) to each reviewer. The Excel workbook comprised three sheets: a data table with column headings, a codebook documenting the column use, and a metadata table. There was one workbook for each disease.

The Guidelines instructed the reviewer to carry out searches and record the places and terms searched in the metadata table. The expectation was that online-search results could be downloaded to a personal bibliographic database and further consideration of each item flagged by notes or keywords within that list. Extra sources found other than by online searching were to be added to the list manually. Reviewers were also asked to compile, as they reviewed, a glossary of terms that might need definition. This might be the individual reviewer providing the definition as had been used, or identifying a term whose semantics might vary between sources or between reviewers.

Source documents that came through the sift were to be individually scrutinized and information extracted to the data table. The link from a table row to a bibliographic item was required but the form was not defined. In practice, reviewers have copied the bibliographic description (authors, title, etc) into the citation column as text.

When a source described a single outbreak in one species at one location, the mapping to a row in the data table was natural. When a source described more species, age groups, or locations, the reviewer had to decide whether to put the information into one row or split the information across extra rows. The advice offered was that information that was simply lists with no cross-linkages should be entered on one row, listed items separated by semi-colons; lists that were cross-linked and therefore describing separate and distinct events should be entered as separate rows. For example, a paper reporting that a condition had been found in trout and salmon in the UK, Norway and Poland would be one report with hosts "trout; salmon" and locations "UK; Norway; Poland", but a paper reporting the condition in salmon in the UK and Norway but trout in Poland would need two rows. It will often be difficult to make this distinction because reports are not standardized, and in reality the relationships (in database sense) may be many.

A very clear instruction was that the data capture at this stage should copy the source author as closely as possible, even including mistakes and misspellings. Editorial 'correction' of the source must be a secondary stage so that any changes can be identified later.

Another point that arose repeatedly during the data capture was the treatment of apparent repeat reports. The recommendation was to reply upon the most basic source. An outbreak should therefore be cited by the first report, if necessary cross-referencing to subsequent confirmatory reports. Reports from review articles should be included only where the original source was not available, but in that case the quality of the report relied upon the judgment of that review author, and the report might be downgraded as hearsay.

2.4. THE REVIEWERS

The disease reviews were conducted by a total of 10 experts as follows:

Epizootic haematopoietic necrosis (EHN) – Barry Hill (Cefas)
Epizootic ulcerative syndrome (EUS) – Birgit Oidtmann (Cefas)
Viral haemorrhagic septicaemia (VHS) - Peter Dixon (Cefas) and David Stone (Cefas)
Infectious haematopoietic necrosis (IHN) – " "
Koi herpesvirus disease (KHVD) - Keith Way (Cefas) and Olga Haenen (CVI)
Infectious salmon anaemia (ISA) - Richard Paley (Cefas)

Infection with *Bonamia exitiosa* - Marc Engelsma (CVI) and Mike Hine (Cefas)
Infection with *Bonamia ostreae* - " "
Infection with *Perkinsus marinus* - " "
Infection with *Microcytos mackini* - " "
Infection with *Marteilia refringens* - " "

Taura syndrome - Grant Stentiford (Cefas)
Yellowhead disease - "
White spot disease - "

2.5. SELECTION OF INFORMATION SOURCES

The reviewers individually searched for publications and other reports on the occurrence of 'their' disease(s) by location and by host and on methods for their diagnosis or detection and typing of the pathogen. Searches for published and unpublished studies and other sources of information on each disease were conducted on several occasions during 2009. For the fish and crustacean diseases, ASFA, ProQuest Deep Indexing: Aquatic Sciences, and Scopus: Natural Sciences were the main scientific publications databases searched. Searches were also made of ProMED mail and Google scholar for other sources of information. For the mollusc diseases, Ovid-CAB Abstracts, Ovid-Biological Abstracts, Ovid-Current Contents were the main publication databases searched. On-line disease databases searched included OIE World Animal Health Information Database (WAHID) for 2005-present, OIE HandiSTATUS II for 1996-2004, OIE Regional Aquatic Animal Disease Reports (Asia-Pacific region) for 2002-2009, and the Defra/OIE International Database on Aquatic Animal Diseases (www.collabcen.net) for all first reports of occurrence by location or by host. All relevant data was compared with the occurrence and distribution information for each disease in the OIE Manual of Diagnostic Tests for Aquatic Animals (2009) and the EFSA report on aquatic species susceptible to diseases listed in Directive 2006/88/EC' (EFSA 2008).

Printed copies or pdf files were obtained for all relevant publications. Only documents or other sources that contained original information on detection of the disease and/or the pathogenic agent were critically reviewed. A bibliography of all the relevant documents and other sources of relevant information to the reviews was compiled for each disease.

2.6. CRITICAL REVIEWS

The reviewers critically scrutinised the information provided in each document or other source they identified from their searches. Publications, unpublished documents or other sources containing no relevant information to the objectives of the project were dismissed. After this first sift, the relevant information in all the remaining documents was critically examined and its quality assessed. Occurrence information was entered into the Excel data table for that disease. Information on methods for diagnosis of each disease and/or methods for detecting and typing the pathogen was critically assessed and compared with the methods recommended in the current edition of the OIE Aquatic Manual (OIE, 2009).

2.7. QUALITY AUDIT OF THE DISEASE REVIEWS

The quality audit method was developed by the expert auditor (Dr Edmund Peeler) in consultation with both the review-process manager and the project co-ordinator. Reviewers were then asked to provide the auditor with between 5 and 10 reviewed documents for each disease/pathogen/. For each source, the entries made in each field of the Excel data entry form were checked independently by the review auditor against the information provided in the relevant publication. The entire record was checked for consistency, and particular attention was paid to how the following fields had been completed: i) classification of report, ii) report type, iii) quality assessment and iv) mortalities.

The list of countries where the pathogen was reported was checked against OIE and non-OIE reports of occurrence of the pathogen in i) the Defra/OIE International Database on Aquatic Animal Diseases (http://www.collabcen.net/idaad/) and ii) archived reports from ProMed (http://www.promedmail.org/pls/otn/f?p=2400:1000:). When these searches indicated

countries or regions not included in the datasheets, the relevant reviewer was informed and asked to investigate.

Reviewers were invited to provide feedback about any problems with the use of the record form, in particular ambiguity and interpretation.

2.8. SUMMARY REPORTS OF REVIEWS

Reviewers were provided with a template (see Annex B) for a summary report on diagnostic methods, organism typology, and their findings on the disease occurrence by host, geography and time. These were completed for each disease by the expert who conducted the review of that disease. To follow Cochrane principles, reviewers were instructed to keep these very succinct: one or two sides of A4.

2.9. CRITICAL REVIEWS OF VHSV GENOTYPING AND OF IDENTIFICATION OF *BONAMIA* SPECIES

Information sources were identified as described in Section 2.5 and those sources containing no information on either the genotyping of VHSV or on sequence data were dismissed. Documents containing genotypic information were critically reviewed to establish a consensus on the most appropriate typing scheme (the internationally accepted typing scheme) and this scheme was then used to reassess the genotype of the virus isolates entered in the Excel data entry form. In many of the early studies, the analysis was not considered sufficiently robust because it was performed using a limited sequence data set. In most cases the limited sequence data used in the analysis resulted in not all recognised genogroups being represented and therefore, it was difficult to make an accurate assessment on the validity of the genotyping scheme. In such cases the analysis was repeated to include a more comprehensive dataset. Briefly, sequence information described in the document was obtained using European Molecular Biology Laboratory (EMBL) database browser tool (http://srs.ebi.ac.uk). Additional published sequences were obtained in a similar manner and the sequences were truncated to ensure that there were of comparable length to those in the publication of interest. Multiple alignments were then performed using Clustal W and the phylogenetic analyses were conducted using Neighbour-joining methods using MEGA version 3.1 (Kumar, Tamura & Nei 2004). The genotype of a virus was then redefined based on the internationally accepted typing scheme and entered on the Excel data entry form against the appropriate virus entry. Where an information source had already defined the genotype of a virus using the internationally accepted typing scheme, the genotyping data was used without further analysis.

For the critical review on the identification of *Bonamia* species a concise overview was prepared on the characteristics of known *Bonamia* species (*Bonamia exitiosa*, *Bonamia ostreae*, *Bonamia roughleyi*) and more detailed information collected on the ultrastructure, phylogeny and taxonomy issues of *Bonamia* species. The review was based upon publications retrieved by the Ovid literature searches for the disease reviews. In addition, three phylogenetic analyses were carried out based on available GenBank sequences to demonstrate the phylogenetic relationship between the *Bonamia* species. The decision to exclude *Bonamia roughleyi* in the review on *B. exitiosa* was based upon the lack of information on this species and consultation with Ryan Carnegie (VIMS, USA). For the ultrastructure data, unpublished information from co-reviewer Mike Hine (Cefas) was included. The information on the taxonomic status of the host species was based upon publications retrieved by multiple Ovid searches on this topic.

2.10. DATA TRANSFER

The second stage of data interpretation and cleaning was first attempted using a data spreadsheet that would copy each cell from the data-entry sheet into a read-only cell and have an adjacent copy in a cell that could be edited. This was not made to work, so a system was substituted of simply taking a copy of the data-entry sheet and working on that copy within each workbook.

The first Excel spreadsheet, with an example of few data, was sent by the reviewer to the database managers (Carla Ippoliti and Paolo Calistri) in mid June 2009. It started the process of creating the decoding functions and of implementing MS Excel formulas to facilitate the "translation" of data reported by the experts into the rules and tables of the database.

The process of data interpretation was designed to be carried out through a number of supporting tables for the key fields, in which, for each value written in data-entry by the reviewers, the normalised value to be reported in the final database was indicated.

Visual basic procedures were implemented into the MS Excel workbook to split the listed items (separated by semi-colons) into multiple rows. The procedures work as follows: if in a row it shows salmon; trout as host species and the disease occurred in UK; Poland, then the complete initial row was copied into 4 rows each one identical to the others apart from species - location: salmon - UK, trout - UK, salmon - Poland, trout - Poland.

The splitting procedure was applied to some key fields (host species, pathogen, geographical location, year of occurrence) where multiple values were reported.

Decoding functions have also been implemented to split the latitude/longitude coordinates of location occurrences reported by the experts.

During the training workshop of 21-22 October 2009, some minor changes were requested by EFSA and these were implemented subsequently.

All data for *Bonamia* and VHS diseases have been uploaded into the database, on which a mapping system is based: the values that need an interpretation by the reviewers have been indicated as "999" code (as for not found, not indicated, not reported values) and the year of occurrence not clearly evident is indicated as 1900.

2.11. DATABASE AND GIS DEVELOPMENT

The database scheme was designed to fit the requirements for a geographical representation in a web application of the epidemiological data contained in the MS Excel spreadsheets completed for each disease.

The scheme comprises one main table (where the epidemiological data are stored) and some supporting tables, where values for standardized variables are included. Another table containing epidemiological data reported in specific locations is also available.

The diseases are coded in a table, and this allows inclusion in the same database information on other diseases, sharing the same data structure. The same philosophy is applied to the other supporting tables, where new codes can be added without changes in the whole database, or descriptions can be modified at any time.

The database scheme details, as well as the description of the tables and relationships, are described fully in Annex G.

The alphanumeric data collected are linked with the geographical layer of the administrative units, to be shown on the web-GIS.

The diseases occurrence is represented as:

- polygon-based layers based on administrative boundaries at Country/regional/province level, and

- point-based layers (specific locations) geo-referenced through latitude and longitude data, when geographical coordinates of the point location are available.

The occurrences of VHS and *Bonamia spp.* infections are visualized on the web-GIS application, and the related epidemiological data recorded by the reviewers are accessible through queries.

The application is accessible via the web, and is an interactive tool to navigate into the epidemiological data collected. Using the query tool, the user can retrieve data from the database, and the selected information is shown on the map, reported in summarized form on the left side of the page. The user can also access the table with the complete information related to chosen criteria in tabs over the map.

A complete description of the web application, its development and details on the final product is provided in Annex G.

The relational database, the geographical data, the software needed to publish the maps on the World Wide Web and all other technical components have been chosen according to EFSA technical requirements.

Together with Paolo Calistri, Carla Ippoliti, Alessio Di Lorenzo and Lara Savini, other personnel who collaborated in the technical development of the database/GIS included: Daniela Cioci (Oracle database development), Daniele Zippo and Roberta Cicconi (Remote Management).

3. RESULTS

3.1. REVIEWS

Almost 3000 scientific publications and other documents identified by the literature searches and other means were examined for information relevant to the review. Those with relevant information were critically reviewed. Table 1 shows the number of documents obtained for each disease and the number selected for critical review in each case.

Table 1. Numbers of documents examined and the numbers critically reviewed

Disease	Number of documents examined in first sift	Number of documents critically reviewed for occurrence data
Epizootic haematopoietic necrosis (EHN)	29	7
Epizootic ulcerative syndrome (EUS)	300	27
Viral haemorrhagic septicaemia (VHS)	~ 700	70
Infectious haematopoietic necrosis (IHN)	> 600	27
Koi herpesvirus disease (KHVD)	49	49
Infectious salmon anaemia (ISA)	> 200	15
Infection with *Bonamia exitiosa*	213	64
Infection with *Bonamia ostreae*	”	20
Infection with *Perkinsus marinus*	557	22
Infection with *Microcytos mackini*	24	8
Infection with *Marteilia refringens*	85	39
Taura syndrome	~ 150	20
Yellowhead disease	> 500	25
White spot disease	~ 200	70
Total	2907	463

All the disease reviews were completed and the relevant information entered into the Excel data forms under the appropriate headings. After quality checks of random examples of data entries for each disease by the review-process manager and the expert auditor (see 3.2.), the data forms were forwarded to the database/GIS developer for transfer of the information into the database.

Summary reports of each disease review were written by the relevant reviewer. There were two reports for each disease:

(i) disease detection, pathogen identification and typing

(ii) disease occurrence and distribution.

The review summary reports are provided at Annex B.

3.2. QUALITY AUDIT OF DISEASE REVIEWS

3.2.1. Pathogen distribution

The key information from all papers audited (e.g. correct identification of the pathogen and its location) was accurately recorded in the datasheets. The audit found additional reports of disease occurrence which had not been identified by the reviewers for only two pathogens: ISAV and *Mikrocytos makini*. The audited ISAV report had omitted recent reports detailing the widespread distribution of the virus in Chile. The audited *Mikrocytos makini* report made no reference to its presence in the US. The reviewers altered their reports accordingly. A comparison the worldwide distribution of pathogens derived from the OIE database listed a number of diseases not identified in the reviews (see Annex 3 for details).

3.2.2. Consistency in data entry

The main issues highlighted by the audit were completion of fields for mortality, prevalence, quality and report classification. Some reviewers had been inconsistent in within the review of a single pathogen in completing some of these fields. There is also no consistency between reviewers in the format of the entries. Prevalence and mortality have not been recorded as numeric variables so no analysis of these data are possible. Descriptors of quality and report classification have not been consistently applied. Currently it is not possible to use these fields to produce summary statistics or cross-tabulations (see Annex 3 for a full description).

3.3. CRITICAL REVIEWS OF VHSV GENOTYPING AND OF IDENTIFICATION OF *BONAMIA* SPECIES

In addition to the individual disease reviews, separate critical reviews were also completed on methods for identifying species, strain and genotype (as appropriate) of VHSV and *Bonamia* spp., and reports prepared. The reports also identify the data gaps for these pathogens and where further research is needed to address the issues of pathogen definition and strain differentiation. The VHSV and *Bonamia spp.* reports are provided at Annex D and Annex E respectively.

3.4. DATABASE AND GIS DEVELOPMENT

The main outcomes in developing the database and web-GIS can be summarized as follows:

Database analysis has been performed based on the type of information collected into the MS Excel spreadsheet and the structure of the geo-database has been designed in Oracle 10g software.

A number of procedures have been implemented to facilitate the transfer of MS Excel spreadsheet data into the Oracle tables.

The epidemiological data for VHSV and *Bonamia spp.* provided by the reviewers has been inserted into the database tables; the consistency of the data has been only partially verified with the reviewers.

A unique geographical layer with World extent has been created, with various levels of representation for the disease occurrence: Nomenclature of Territorial Units for Statistics – NUTS - Level 0, 1, 2, 3, for EU countries; ESRI shapefile Level 2 for the rest of the World; for North Europe and North Atlantic Seas, International Council for the Exploration of the Sea (ICES) polygons.

The relational database and geographical available software have been considered and the most updated releases have been chose to fulfill the project needs and the EFSA technical management requirements;

A web-GIS application has been developed to geographically access the disease distributions and perform standard spatial interactions. From the technical point of view, the web-GIS application is composed by a server-side (ArcGIS Server 9.3 Standard Enterprise for the Java Platform) and a client-side (ArcGIS Server 9.3 Javascript APIs v. 1.5 from the online ESRI serverapis). The client is platform-independent and is composed by html and javascript code (includes Dojo toolkit objects and syntax. Dojo is the base on which ArcGIS 9.3 Server Javascript APIs are built on).

A workshop on database contents and structure, and web-GIS application was held in Teramo, Italy on 21-22 October 2009 and included experts from the Community Reference Laboratories for fish, mollusc and crustacean diseases respectively.

A pen drive with database scheme and data, geographical base layer, web-application components, technical solution and installation instructions document was provided to EFSA during the final meeting held in Parma on 30 October 2009.

Assistance during the phase of installation on EFSA server has been assured by IZS A&M technical personnel to EFSA data managers.

A full technical description of the structure of the database and web-GIS, with user guidelines, is provided in Annex G.

4. DISCUSSION

Various authors over several decades (eg Cochrane 1979; Sargeant et al 2005; Waddell 2009) have noted the need for domain experts to avoid the bias of preconceptions when carrying out systematic reviews. Authors of conventional reviews may have made honest attempts to summarize all information that they considered reliable in their field, but there was generally no way of repeating or validating their work. We need a method that encourages a comprehensive and consistent review and, as Waddell et al (op cit) note, "when methods are not described, those methods have not been used."

The Cochrane-like approach was novel to most of the reviewers and as a result they did not apply it with full understanding from the start. The Cochrane principles of collaboration, minimizing bias, ensuring quality, and striving for relevance matched exactly the needs of the project. The principles of continuity, keeping up to date, promoting access, and enabling wide participation indicate scope for continuing work. In extending the Cochrane method to disease occurrence and strain identifications, we were conscious that the research questions were not as clear-cut as in a study compatible with the Cochrane Collaboration. As a result, the current work had to be approached as a "design process", a task in which the goal could be defined only through the process of moving towards it. Parnas and Clements (1986) articulate the view that this process is often disguised in research and "faked" with hindsight to appear as rational from the start.

Within the time and resource constraints we have attempted to proceed at each stage with review and revision. The list of information fields to be extracted from each source was discussed and revised, but cannot be considered final or definitive. Each reviewer was asked to extract information from each source into a table, and then revise as necessary to check that they had been consistent over sources from first to last. The audit process should then have checked for consistency across the panel of reviewers, but due to insufficient time remaining it was then difficult to go through a further round of data revision. From the process, we have learned much about the planning, costing and execution of this type of review.

As a result, the project has delivered all the requested reviews to a quality that would not have been achieved without the formal review process, but with the complement that we are also more conscious of the shortcomings. For example, several reviewers did not complete the metadata sheet to record their search strategy as it evolved. That said, it is a fact that any review is a summary only of the best state of knowledge at a certain point in time. The

systematic data extraction highlighted to reviewers the disparate nature and quality of disease reports. Facts that were accepted by one review may still be questioned subsequently as new knowledge, techniques or interpretation emerge. Changes to taxonomy or identification of strains will compel revisions to the current database, and new observations will need to be added.

Current practices for animal diseases can be compared with the systematic approach now expected in those designing or reporting medical randomized, controlled trials (CONSORT op cit) and even the rigorous standards applied to amateur reports of rare birds. A review published in 2006 removed an 1837 record of a bird (Bulwer's Petrel) from the "British List" (BOURC 2006) 32nd report / October 2005 / Ibis 148: 198–201 (January 2006). Errors in data are usually assumed to have been committed in good faith, but occasional cases of scientific fraud are also known, e.g. Colonel Richard Meinertzhagen who stole specimens from museums and relabelled them (Knox 1993).

The data tables have been loaded into a GIS database that allows the geographical spread of outbreaks to be displayed as points on a map. More problematical, the spread of occurrence can be mapped by country or province, but geopolitical boundaries may not correspond to meaningful zoological areas, particularly as rivers or marine inlets are commonly used as boundaries. While the maps are useful presentational devices, having only predefined fields and selection criteria limits the analytical use of the data. Open-minded analysis of the occurrence and typology of each disease will require drilling through the map to the captured data to allow alternative selection, grouping and summaries. The layered approach of original wording, standardized wording and database tables facilitates such work.

The expectation that a database would emerge without effort from the data elided the need to clarify the data structure. None of the sources had been written in the expectation of being tabulated, so there were issues of identifying what constituted an outbreak and deciding whether to split or lump reports. It is also a consequence of the accumulation of knowledge that one disease event may described in several sources spread over many years. Some concepts emerged, or were clarified (or clouded!), during the design process of the project, for example, the identification of "excess" mortality when reports were vague. The structure of the disease database will therefore evolve as well as the content, and the content will need to migrate to new platforms as technology changes.

This project has created a data resource for EFSA that will inform expert work on each of the diseases in the call. It also indicates the need for further standardization of review methods. In a subsequent review, we would recommend specific training for the reviewers to ensure they understood each of the steps and the rationale underpinning it, and had suitable and robust tools for recording their own work contemporaneously. We also see scope for EFSA to lead in recommending a skeleton for data extraction and developing the standard vocabulary needed, initially, for interpreting existing sources but subsequently as a recommendation for authors to use in reports. Distinctions such as between an "outbreak" and "detected occurrence" need to be clarified. Journal editors and peer-reviewers have a role in promoting transparent communication, and recommendations may lead to standards.

5. REFERENCES

BOURC (2006) Records Committee: 32nd report (October 2005). Ibis 148: 198–201

Cochrane AL (1979). 1931-1971: a critical review with particular reference to the medical profession. In: *Medicines for the year 2000.* London: Office of Health Economics, pp 1-11.

EFSA, 2007a. Scientific Opinion of the Panel on AHAW on possible vector species and live stages of susceptible species not transmitting disease as regards to certain fish diseases. The EFSA Journal (2007) 584, 1-163.

EFSA, 2007b. Scientific Opinion of the Panel on AHAW on possible vector species and live stages of susceptible species not transmitting disease as regards to certain mollusc diseases. The EFSA Journal (2007) 597, 1-117.

EFSA, 2007c. Scientific Opinion of the Panel on AHAW on possible vector species and live stages of susceptible species not transmitting disease as regards to certain crustacean diseases. The EFSA Journal (2007) 598, 1- 91.

EFSA, 2008. Scientific Opinion of the Panel on AHAW on aquatic animal species susceptible to diseases listed in the Council Directive 2006/88/EC. The EFSA Journal (2008) 808, 1-144

ESRI ArcGIS Server Javascript API Concepts

ESRI ArcGIS Server Javascript API Reference

ESRI ArcGIS Server Blog

http://blogs.esri.com/Dev/blogs/arcgisserver/

http://www.ices.dk/aboutus/icesareas.asp

http://resources.esri.com/arcgisserver/index.cfm?fa=JSAPIs

http://resources.esri.com/help/9.3/arcgisserver/apis/javascript/arcgis/help/jshelp_start.htm

http://resources.esri.com/help/9.3/arcgisserver/apis/javascript/arcgis/help/jsapi_start.htm

Knox, A.G. (1993) Richard Meinertzhagen—a case of fraud examined. Ibis 135: 320-325.

Kumar S., Tamura K. & Nei M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5, 150-163.

OIE (2009). Manual of Diagnostic Tests for Aquatic Animals 2009. OIE, Paris, pp.222

David Lorge Parnas, Paul C. Clements (1986) *A Rational Design Process: How and Why to Fake it.* IEEE Trans. Software Eng. 12(2): 251-257; available at web.cs.wpi.edu/~gpollice/cs3733-b05/Readings/FAKE-IT.pdf accessed 6/11//09

Sargeant, J.M., M.R. Amezcua, A. Rajić, et al. 2005. *A guide to conducting systematic reviews. in agri-food public health*. Available via: www.fsrrn.org.

L. Waddell, A. Rajić, J. Sargeant, S. Parker, A. Deckert, S. McEwen (2009) *The Methodological Soundness of Literature Reviews Addressing Three Potential Zoonotic Public Health Issues*. Zoonoses and Public Health. Published Online: Jan 19 2009 4:57AMDOI: 10.1111/j.1863-2378.2008.01194.x

6. ANNEXES

6.1. Annex A: Guidelines for reviewers

6.2. Annex B: Disease summary reports

6.3. Annex C: Report of quality audit of reviews

6.4. Annex D: Critical review of VHSV genotyping

6.5. Annex E. Critical review of identification of *Bonamia* species

6.6. Annex F: database and GIS workshop program and participants

6.7. Annex G: Characteristics of database and GIS application