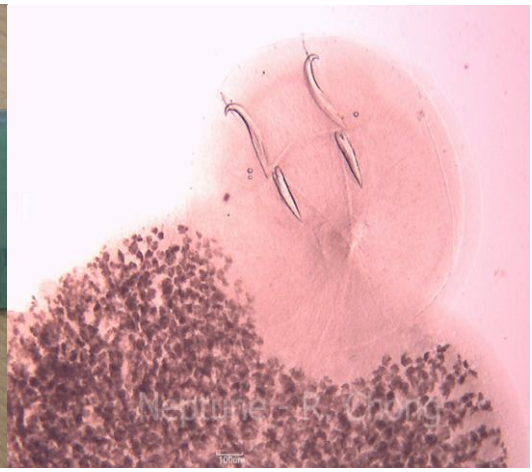
FRDC
FISHERIES RESEARCH & DEVELOPMENT CORPORATION

The Neptune project

A comprehensive database of Australian aquatic animal pathogens and diseases



Dr Marissa McNamara and Dr Robert Adlard

September 2014

FRDC Project No 2013/004

Version 1.0 1 July 2013

ISBN 978-0-9872681-2-9

The Neptune project – A comprehensive database of Australian aquatic animal pathogens and diseases 2013/004

2014

Ownership of Intellectual property rights

Creative Commons licence

Disclaimer

Researcher Contact Details

Name: Dr Marissa McNamara
Address: Queensland Museum
South Brisbane QLD 4101
Phone: 07 3842 9173
Fax:
Email: marissa.mcnamara@qm.qld.gov.au

FRDC Contact Details

Address: 25 Geils Court
Deakin ACT 2600
Phone: 02 6285 0400
Fax: 02 6285 0499
Email: frdc@frdc.com.au
Web: www.frdc.com.au

In submitting this report, the researcher has agreed to FRDC publishing this material in its edited form.

Contents

Acknowledgments **4**

Abbreviations **4**

Executive Summary **5**

Introduction **7**

Objectives **9**

Method **10**

Results and discussion **12**

Conclusion **19**

Implications **20**

Recommendations **21**

Further development 21

Extension and Adoption **22**

Project coverage 22

Project materials developed **23**

Appendices **24**

Figures

Figure 1. Screenshot of Neptune search page showing link to whole slide image library 13

Figure 2. Screenshot of an event in the Neptune database, showing hyperlink to slides 13

Figure 3. Screenshot of search page of Collibio slide library, showing the label for endemic or exotic diseases 14

Figure 4. Graph showing number of participants at Neptune webinars 15

Acknowledgments

The project manager would like to thank all the webinar speakers for their fantastic presentations.

Abbreviations

AAAPD – Australian aquatic animal pathogen and disease database

ABIN – Australian Biosecurity Intelligence Network

DA – Australian Department of Agriculture

ECU – Edith Cowan University

FRDC – Fisheries Research & Development Corporation

NCRIS – National Collaborative Research Infrastructure Strategy

PGi – Premiere Global Services, Inc

QM – Queensland Museum

Executive Summary

What the report is about

Aquatic animal health experts from the Queensland Museum (QM) have been completing work on a parasite and disease database called Neptune. Work on Neptune has taken place at QM in Brisbane since May 2013, resulting in the completion of major improvements to the database. These will allow Neptune to become Australia's most comprehensive online resource on aquatic animal health. Improvements were carried out in conjunction with IT staff from three different organisations: the Australian Biosecurity Intelligence Network (ABIN), which was based in Canberra until September 2013; Edith Cowan University (ECU), which is based in Perth; and Pixcelldata, which is based in Ireland and runs digital pathology software. The database was hosted by ABIN until September 2013, when ownership passed to ECU. This project was funded by the Fisheries Research and Development Corporation (FRDC) and the Australian Department of Agriculture (DA), with contributions from QM.

Background

Biosecurity is an increasingly important concern for Australia's aquatic animal resources. The global nature of trade and growing use of aquaculture make the introduction and spread of aquatic diseases a major issue in Australia and worldwide. The Neptune project was created in response to these concerns, and consists of a centralised web-based knowledge store with information on all aquatic animal diseases and parasites reported in Australia. Neptune was first launched in 2011, and three versions of the database were released throughout 2012. However, several key improvements were still required to achieve optimal functionality of the database. Neptune users have also had access to an online slide library that contains 180 key pathology slides of both exotic and endemic diseases. These slides were collected from pathologists around Australia, and were digitised in 2012 using Ultra-Resolution Digital Scanning. Like the Neptune database, the slide library required some upgrades to allow easier and more effective access.

Aims/objectives

The aim of the Neptune project is to enhance our understanding of aquatic diseases and parasites. Neptune does this by giving users access to detailed and easily searchable information on all Australian aquatic diseases and parasites, and by facilitating interactions between biosecurity officials from federal and state governments, as well as researchers and pathologists. Neptune was also created to reduce duplication of effort by eliminating the need for multiple, separate data sets. The objective of this project was to improve the existing version of Neptune so that it could better achieve its intended aim. In particular, eight aspects of the original database required work. These included: improving the user interface to make it more intuitive, completing data entry so that Neptune contained a comprehensive data set, integrating and updating the digital microscopy platform, and running a webinar series that included regular user training.

Methodology

The majority of work on this project was completed by the project manager, with assistance from supervisors at QM and DA. Technical work on the database and the slide library was carried out by IT staff at either ABIN, ECU or Pixcelldata. For example, the Neptune user interface was upgraded by incorporating user feedback to create a list of the most important requests, which were carried out by IT staff at ABIN. Data entry was completed by the project manager by a manual transfer of information from a source database. The digital microscopy platform was integrated more fully into Neptune by IT staff at ECU and staff from Pixcelldata. Experts from the latter company also improved the search page of the slide library. The project manager hosted free webinars and training sessions, first using Adobe Connect, which was provided by ABIN, and then using AnyMeeting. Each webinar contained a presentation from a member of the aquatic animal health community.

Results/key findings

The Neptune database and slide library have been significantly improved over the course of this project. Eight changes were made to the user interface, making it easier to search for and access information. Hundreds of host and parasite species have been added to the Neptune database, along with accompanying disease and publication information, resulting in a data set that now contains 88% of the AAAPD. The digital microscopy platform is now available within the Neptune database, and contains a more detailed search page, with additional columns showing key information on every slide. Fourteen webinars have been held, as well as six training sessions, often with excellent attendance from the aquatic animal health community. During this project Neptune was changed from an Oracle platform to open source software, which reduced the cost of hosting. Long term hosting arrangements have been explored, and discussions have been held with four potential hosting bodies.

Implications for relevant stakeholders

Neptune is likely to become an indispensable national resource for Australia's aquatic animal health professional community. The user community includes researchers, government officers, aquatic animal industries, private veterinarians and consultants. The facility will also be of immense value for student training. Neptune's impact will be sustained and its benefits will far exceed its relatively small establishment costs.

Recommendations

The completion of this project has led to five important recommendations for the future. Specifically, we recommend that the FRDC considers the following conditions: that funding recipients be required to include the associated costs of adding material to Neptune in their application; that Neptune be considered the storage space for supplementary data accompanying aquatic animal health publications; that principal investigators of projects anticipated to generate important histopathology slides be required to submit specimens to the online slide library; and that input from the user community should be collected on an ongoing basis to ensure that the database continues to meet user needs. We also recommend that a copyright agreement addressing the use of whole slide images of donated slides be finalised.

Keywords

Biosecurity, database, digital microscopy, aquatic diseases, aquatic parasites, user community, online resource

Introduction

Compared to their terrestrial counterparts, the pathogens and diseases of Australia's aquatic species are in general poorly understood. Aquatic diseases can decimate aquatic animal populations and threaten fisheries and aquaculture, the sustainability of which is an FRDC strategic challenge.

It is clear that data and resources must be made broadly accessible to ensure biosecurity professionals and researchers have the information and tools to manage aquatic animal disease risks effectively. However, significant volumes of aquatic animal health data and resources have remained inaccessible due to:

- large volumes of valuable data remaining unpublished or being kept as individual or local collections
- a reliance on physical collections (e.g. microscope slides) that are not publicly accessible
- the lack of a unified platform for aquatic animal health professionals to share data and collections.

These issues were initially identified by Australia's National Strategic Plan for Aquatic Animal Health, AQUAPLAN 2005-2010. This initiative constituted the basis for the development of the Neptune project, a collaborative effort aiming to provide a centralised web-based information resource on aquatic pathogens and diseases. The initial work on this database was performed at Biosecurity Queensland, and the first three versions of the database were launched in 2012. As each version was launched, feedback was collected from the project reference group on what aspects required further work. This feedback resulted in the existence of the whole slide image library, more accurate disease maps, and better security settings. However, several key improvements were still required to achieve optimal functioning of the database. The second phase of this work, the subject of the current project, took place at the Queensland Museum with support from Dr Robert Adlard. Dr Adlard was on the steering committee for the original ABIN project and was also the Chair of the project's national Reference Group.

The Neptune project was conceived to add significant value to existing investments in aquatic animal health by building a platform for capturing and sharing information on aquatic animal pathogens and diseases. The project was initially funded through NCRIS; however, that funding expired in December 2012, and additional work was required to ensure Neptune fully addressed the needs that it was intended to address. Those needs include:

1. Ensuring the outputs of Australia's aquatic animal disease professionals are disseminated to achieve greatest effect.
2. Effectively connecting aquatic animal health professionals from government, universities, museums, diagnostic laboratories and industry.
3. Addressing geographic separation of professionals by enabling remote collaboration within the sector and with other sectors.
4. Enabling aquatic animal health biosecurity investments to impact fully by making data and resource outputs more broadly accessible.
5. Improving availability of high quality information and resources for many of Australia's most important aquatic animal diseases.
6. Providing access to readily available and aggregated distribution data through a platform capable of hosting and displaying such information.
7. Ensuring access to technology that can better capture and share physical specimens such as histological slides. These remain a key tool for aquatic animal pathologists and researchers but are not easily duplicated or shared.
8. There are currently insufficient opportunities for aquatic animal health training in Australia. Neptune will provide synergies for other FRDC training investments such as the Aquatic Animal Health Training Scheme.
9. There are currently no integrated resources to store and provide access to training material for aquatic animal health professionals.

10. Many of Australia's aquatic animal health pathologists have retired recently or are due to retire in the near future. There is a need to ensure the data and collections of these aquatic animal health professionals are captured and shared more broadly.

While the Australian Biosecurity Intelligence Network (ABIN) ran the first versions of Neptune, a change in hosting occurred during this project. On 1 September 2013, the Australian Government Department of Industry passed ownership of ABIN infrastructure and remaining funds to Edith Cowan University (ECU). ECU has aimed to provide continuity of access to Neptune during this transition and has sought to operate Neptune more cost effectively; for example, by moving to a cheaper software environment and using cheaper online meeting software. These changes have not affected performance but have significantly reduced software licensing costs.

Objectives

1. Improve the user interface of Neptune to allow more streamlined searches, better data access and more efficient data entry.

2. Complete the transformation of data from the Australian Aquatic Animal Pathogen and Disease Database (AAAPD) (FRDC project 2003/646) so that all available peer-reviewed and official reports of aquatic animal pathogens and diseases in the AAAPD are included in Neptune.

3. Establish a digital microscopy capability within the Neptune Community Space.

4. Establish a digital microscopy slide image library within the Neptune Community Space that incorporates examples of significant endemic aquatic animal diseases and links them with relevant database case entries.

5. Establish a digital microscopy slide image library for significant exotic aquatic animal diseases that is clearly differentiated from other data entries and resources for endemic diseases.

6. Implement a program of activities that engages the user community in the Neptune facility and includes monthly webinars on research project progress or special issues, online workshops that use the digital microscopy functionality (e.g. histopathology, exotic disease awareness), and regular user training (as required) on the use of Neptune capabilities.

7. Develop plans for the development of user generated mapping functionality.

8. Explore options that could provide a more cost-effective hosting environment for Neptune.

Method

The objective of this project was to improve the existing database by adding new information and capabilities. These improvements were completed by the project manager and the development staff at ABIN and ECU, with feedback from users.

1. The user interface (UI) was improved by streamlining data searching, viewing and data entry activities. Several changes were made based on user feedback. The consensus from users was that Neptune still had clumsy or non-intuitive aspects, including: the order of host and agent species listed within events was shuffled every time an event was viewed; the total number of pages found for a given search was not displayed; and assets, such as photographs of diseases, could only be selected one at a time. The older versions of Neptune also lacked the ability to export data to Excel, which represented another issue for users. To address these concerns, eight improvements were made to key features of the database. These were completed by adding code to implement new functionality.

2. Transformation of data from the AAAPD has occurred throughout this project. This task consisted of manual data entry, in which thousands of host/parasite records from the AAAPD, a spreadsheet which totals 9365 rows, were individually entered into Neptune. Every line of data was entered into an individual event, with each event corresponding to a scientific paper or report. The data transformation often required the addition of new data fields; for example, completing the taxonomic hierarchy on the first occasion a species was entered into Neptune. In addition to scientific names, common names and synonyms were added to Neptune where possible; common names and synonyms were taken from the Australian Faunal Directory (AFD) and the World Register of Marine Species (WoRMS). In some cases, invalid taxonomic names from older papers were listed in the AAAPD, and expertise from Queensland Museum curators was required to resolve the taxonomic identification.

3. The digital microscopy capability of older versions of Neptune was not fully established and was available only through an external site. This project incorporated digital microscopy capability into Neptune more fully, and has eliminated the need for a separate username and password. This capability required development work that was planned and implemented in consultation with the IT staff at ECU and the staff at Pixcelldata, who run the digital slide viewing software (Collibio). This change was made without requiring expenditure of Neptune development funds.

4. The digital microscopy slide image library contains examples of endemic aquatic animal diseases; however, these assets were not catalogued in a systematic way, and they were not linked to relevant case records. These issues have been addressed to fully integrate the library into the database. Case records in Neptune have been linked to digital slides via URLs, which can be found in the general information listed at the beginning of events. The ability to label diagnostics features on the slides digitally is also present, and was demonstrated by Dr Stephen Pyecroft during the September 2013 webinar.

5. A digital microscopy slide image library for significant exotic aquatic animal diseases has been established separate from other data entries and resources for endemic diseases. This was done by adding additional data fields to Collibio, which are displayed in the slide library next to the thumbnails of each slide. These fields allow users to easily differentiate the two disease types. Users can also view general information about each slide, because a comprehensive catalogue has been created that lists the following fields: host species, disease type (endemic or exotic), disease/pathogen, organ, and slide contributor.

6. A program of user activities and training has been implemented throughout the project. Webinars on aquatic animal diseases and issues of interest to the user community have been organised regularly. Online user training workshops have been held to demonstrate capabilities and to show viewers how

to use Neptune. We also hosted an online interactive pathology workshop to demonstrate Neptune's digital microscopy capabilities. These webinars were originally hosted using Adobe Connect for screen sharing, with teleconference numbers from Premiere Global Services, Inc. (PGi). After the transition to freeware was made in September 2013, Adobe Connect was no longer available. From then on, the project manager used the free version of AnyMeeting to run webinars, and the free version of ScreenLeap to host online training. PGi teleconference numbers were still used for the audio component of webinars.

7. The maps currently available from Neptune are either pre-made PDFs or layers of key diseases, or user generated maps of limited scope, only able to show locations of selected disease entries when latitude and longitude is available. The project manager has investigated the potential for user generated maps that include more variables, including distributions of known hosts. The generation of more sophisticated maps was discussed with development staff at ECU to get an idea of the work required to make significant changes to the existing mapping capabilities.

8. In mid-2013, the project manager asked relevant companies to provide a quote for hosting Neptune, along with the associated costs of data protection and backup, website maintenance and use of communication technology. These quotes were based on the original parameters of the Neptune database: an Oracle ADF application on an Oracle Weblogic server, with data stored on an Oracle 11G database using 10.3.5 version of Weblogic. Since the transition to freeware occurred in September 2013, the specifications of Neptune have changed. Neptune is now running on an Intel based server with 32 GB of memory and 500 GB of disk running Linux, Java 1.6 or greater, and Glassfish 3.1.2, with another 4 core Intel based server with 16 GB of memory and 1 TB disk. Four additional servers, known as Hitachi Content Platforms, are also required to run the online slide library. The cost of this current setup with ECU has been estimated. Discussions have also been held with CSIRO since December 2013 to determine if they can provide a more cost effective hosting environment given the new specifications of Neptune.

Results and discussion

Seven of the eight objectives have been completed. Results against each objective are detailed below, followed by a general discussion and conclusion.

1. To improve the user interface of Neptune to allow more streamlined searches, better data access and more efficient data entry.

On the 12th of June, 2013, the Neptune user interface was upgraded. A total of eight changes were made: the ability to export Neptune search results to an excel spreadsheet (data access); the ability to keep the taxonomic tree open to the species being edited after saving (data entry); the preservation of a consistent order in the agent and pathogen species lists given in events (data access); the display of Genbank and CAAB index information for disease agents (data access); the ability to "Add all" or "Remove all" assets to the shopping trolley (data access); the ability of users to search for free text that has been entered into events (streamlined searches); the display of the total number of pages of results found for a given search (streamlined searches); and the insertion of default values into the new events screen (data entry). These changes were made to make Neptune easier to use and reduce the amount of time required to search for, and enter, information. In the case of Excel, data from Neptune searches can now be exported to any software that recognises comma separated value (csv) files, including Microsoft Excel and Access.

2. Complete the transformation of data from the Australian Aquatic Animal Pathogen and Disease Database (AAAPD) (FRDC project 2003/646) so that all available peer-reviewed and official reports of aquatic animal pathogens and diseases in the AAAPD are included in Neptune.

At the beginning of the project 68% of the AAAPD had been entered; 88% of the AAAPD has now been completed. Because the AAAPD spreadsheet contains 9364 rows, this means that approximately 1873 rows have been entered into Neptune since May 2013. Because each row contains information on a host species and an agent species, a corresponding 3746 host and agent species have been entered since the beginning of the project, of which about 25% were new (that is, not already present in the Neptune taxonomic lists). This means that 936 new host and agent species were added to the Neptune lists. There have been some delays to data entry due to the transition of ABIN to ECU, which resulted in Neptune being offline for several weeks.

3. Establish a digital microscopy capability within the Neptune Community Space

On the 2nd of April, 2014, ECU deployed the first stage of a single sign-on scheme for the Neptune database, in which the Neptune slide image library was available via a tab in Neptune (figure 1), but a separate login was still required. On the 9th of September, 2014, work on the second and final stage was completed. Users can now access the slide library by clicking on the same tab in Neptune without having to login a second time. Collibio software has been used because it provides an established digital microscopy platform that is highly functional. This functionality has been integrated within Neptune by the single sign-on scheme, and links from cases directly to slide assets. This approach has provided benefits in terms of cost, functionality and ease of use.

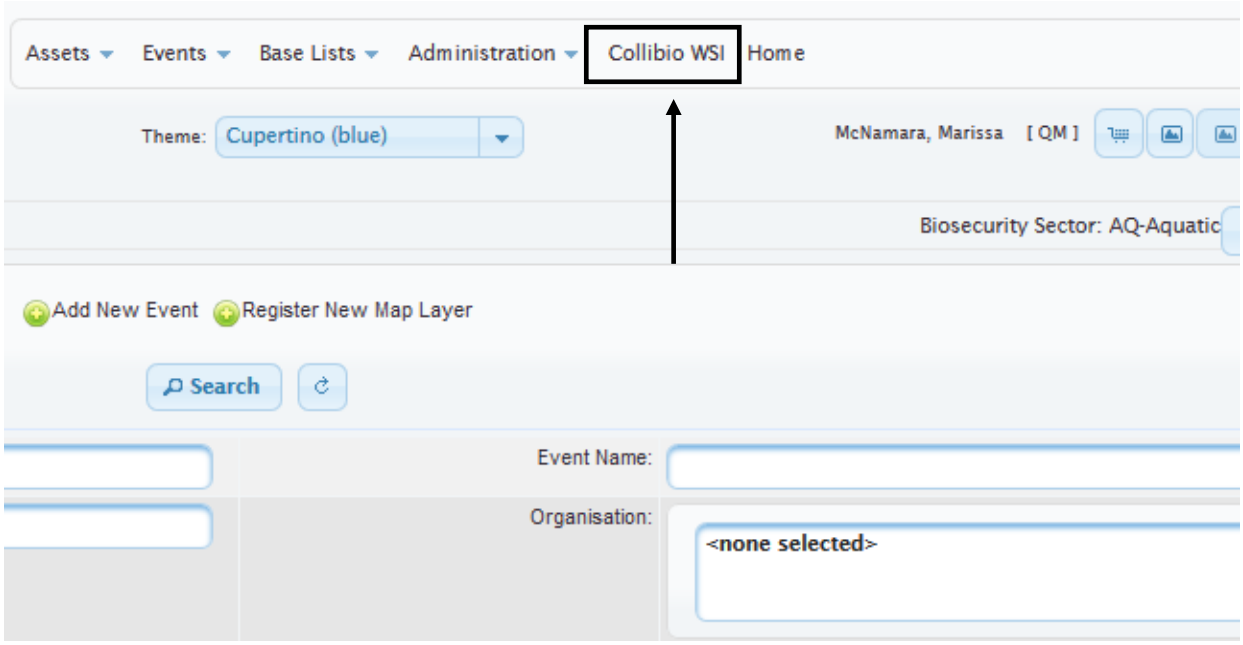


Figure 1. Screenshot of Neptune search page showing link to whole slide image library.

4. Establish a digital microscopy slide image library within the Neptune Community Space that incorporates examples of significant endemic aquatic animal diseases and links them with relevant database case entries.

Links to relevant slides have been pasted into Neptune events so that users can click on the links to view slides (figure 2). A consistent and documented cataloguing system has been created and is under review.

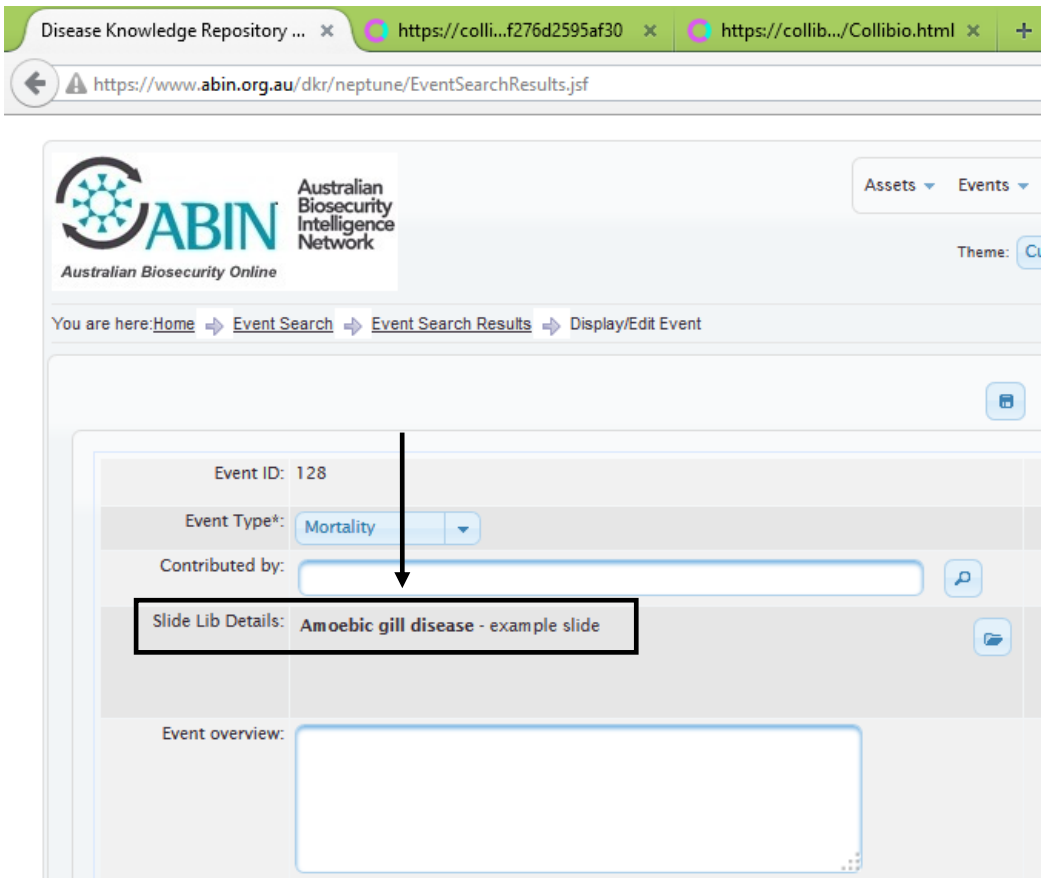


Figure 2. Screenshot of an event in the Neptune database, showing hyperlink to slides.

5. Establish a digital microscopy slide image library for significant exotic aquatic animal diseases that is clearly differentiated from other data entries and resources for endemic diseases.

Work to make this change was completed by Collibio on the 2nd of October, 2013. The Collibio library now contains a category field for images called "disease type" that is displayed as a column in the library. The description of "endemic" or "exotic" disease can be added to every image, and is visible to users at all times (figure 3). In addition, several other fields were added to the Collibio search page. These include host, disease/pathogen, organ, source and case history. These fields are displayed as columns, and are also displayed on the right-hand side of the screen when a slide is clicked (figure 3).

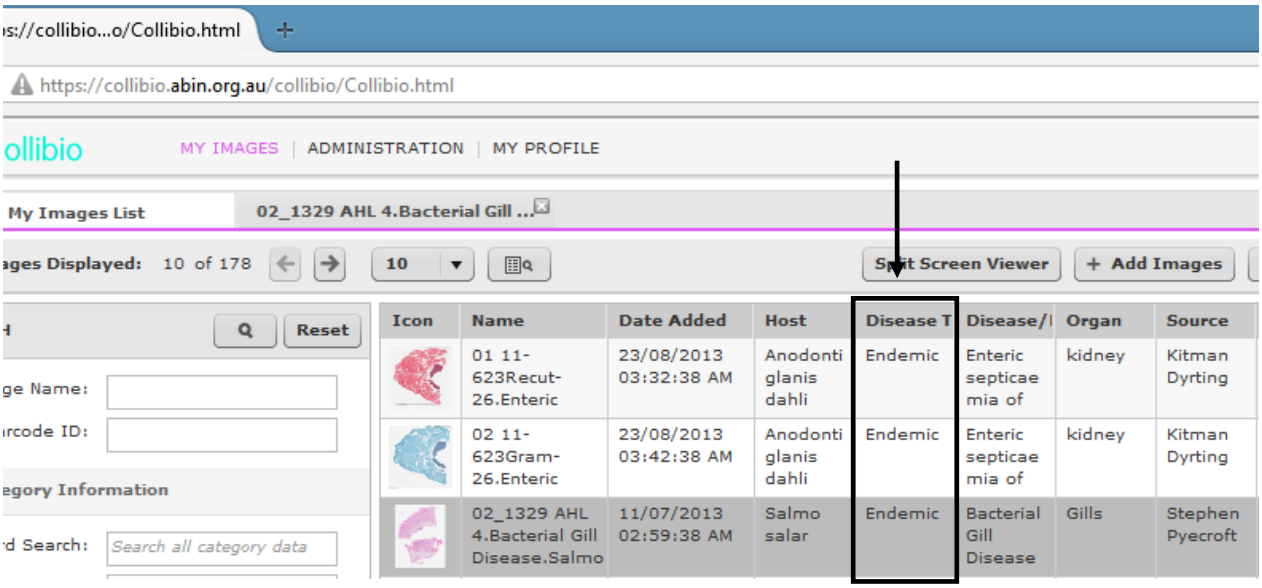


Figure 3. Screenshot of search page of Collibio slide library, showing the label for endemic or exotic diseases.

6. Implement a program of activities that engages the user community in the Neptune facility and includes monthly webinars on research project progress or special issues, online workshops that use the digital microscopy functionality (e.g. histopathology, exotic disease awareness), and regular user training (as required) on the use of Neptune capabilities.

The Neptune Project has hosted regular webinars and user training. The following webinars have been held:

- On June 19th, 2013, Kate Hutson discussed "Infectious Diseases of Barramundi and Health Management."
- On August 20th, 2013, Katie McMahon discussed the DAFF-FRDC Aquatic Animal Health Training Scheme. This webinar was originally scheduled for July.
- On September 11th, 2013, Stephen Pyecroft discussed the Neptune slide library, reviewed the pathology of several cases in the archive, and demonstrated the capabilities of the Collibio software. This webinar was originally scheduled for August.
- On September 19th, 2013, Shane Roberts discussed "South Australian marine mortalities – summer 2013."
- On October 23rd, 2013, Marty Deveney discussed "Integrated Management for Fluke Infections in Fish."
- On November 27th, 2013, Ian Anderson discussed "Marine hatchery bacterial microflora; their effects and role in disease; and microflora management to get healthy larval cultures - with an emphasis on prawn/crustacean aquaculture."

- On January 15th, 2014, Jo Bannister discussed "Detection and characterisation of viruses of the genus *Megalocytivirus* in ornamental fish imported into an Australian border quarantine premises: an emerging risk to National biosecurity." This webinar was originally scheduled for December.

- On January 28th, 2014, Brian Jones discussed "Fish Dermatopathology."

- On February 26th, 2014, Charles Caraguel discussed "Welfare and aquaculture."

- On March 31st, 2014, Ben Diggles discussed "Infection of wild queenfish with sea lice (*Lepeophtheirus spinifer* and *Caligus* spp.) in Australia."

- On May 26th, 2014, Jeffrey Go discussed ""Experimental spread of megalocytivirus between freshwater and marine environments using a model euryhaline vector species."

- On June 25th, 2014, James Harris discussed "Summer mortality in abalone- progress at last."

- On July 29th, 2014, Jeff Cowley discussed "RNAi to reduce GAV infection loads in black tiger prawns."

- On August 20th, 2014, Steve Webb discussed "Some cases encountered during pathology surveys of New Zealand marine organisms."

A graph of attendance for the above webinars is shown in figure 4. Six online training sessions have been held: on the 13th of August and the 8th of October, 2013; and the 21st of January, the 16th of April, the 19th of June, and the 28th of August, 2014. In these sessions the project manager demonstrated how to use the Neptune database to search for information and enter information, and how to search and use the Neptune slide library.

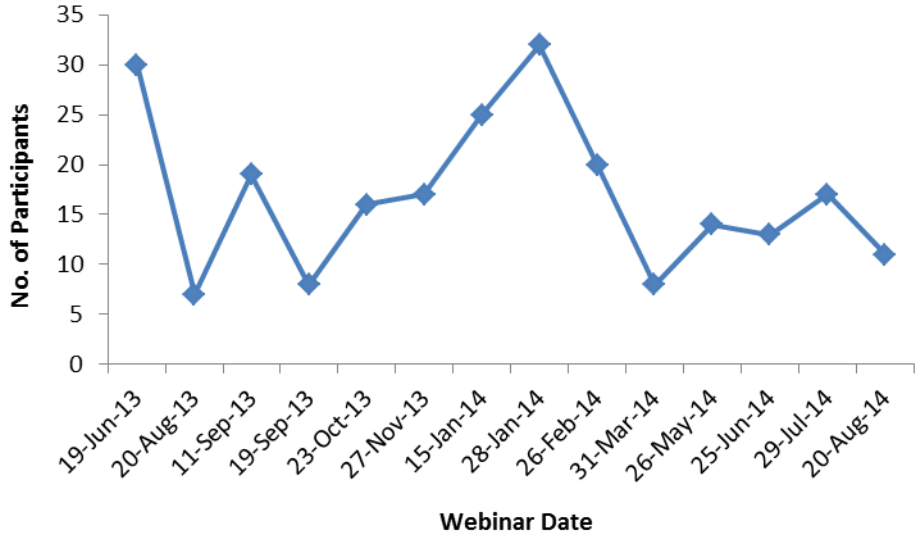


Figure 4. Graph showing number of participants at Neptune webinars.

7. Develop plans for the development of user generated mapping functionality.

A draft of mapping plans has been made. The draft summarises the current status of Neptune mapping capabilities and what additional functionality is being sought. Essentially, Neptune allows users to plot points on a map layer and view more complex pre-made map layers. The ability to download and display host distribution data, and save and print user-generated maps, are two improvements being considered. The cost of making significant changes to Neptune's mapping capabilities has been verbally estimated by staff at ECU at $25,000.

8. Explore options that could provide a more cost-effective hosting environment for Neptune.

In the first four months of the project, twelve companies were approached about providing hosting for Neptune. The majority of these companies were not able to make an offer due to their inability to host Oracle based platforms. However, two companies have provided quotes: Melbourne IT and UXC Red Rock Consulting.

In September 2013, ECU moved to an open source platform that is not dependent upon Oracle. Discussions are being held with ECU to determine if it can provide a long term hosting environment for Neptune. Discussions have also been held with CSIRO since December 2013 to see if they can provide hosting alternatives. A report addressing this possibility has been written and is under review.

Discussion

The Neptune database has been significantly improved throughout this project, and now represents Australia's key online resource for aquatic animal health. Going forward, however, there are several issues that need to be addressed regarding Neptune's future use. These include funding, maintenance of data quality, growth of the slide library, and user engagement.

Since it was launched in 2011, Neptune has gone through several stages of financial support: it was first publicly funded via NCRIS, then transferred to a subscription scheme, and finally institutionally funded via ECU. It is now at a crossroads where its long term financial stability remains uncertain.

Consideration of a subscription scheme model: Funding questions have been faced by other data sharing enterprises. For example, the Swiss-Prot databank is an internationally used online resource that was supported by the Swiss government beginning in 1986. In 1996, facing a funding crisis as two grants were not going to be extended, the founders resorted to a subscription scheme (Abbott, 1998). Substantial public funding was eventually secured in 2002 from the United States National Institutes of Health, and Swiss-Prot (now UniProt) is again free for all users. The lesson to be taken from Swiss-Prot is bipartite: first, that even widely used databases can face financial straits; and second, that user fees were only a stopgap measure, and not a long-term solution (Chandras et al., 2009). In fact, successful subscription schemes seem to be the exception rather than the rule- Ball et al. (2004) note that commercial models for data sharing have not been very successful, and Chandras et al. (2009) found that, while some databases have charged user fees to supplement grant money, no biological resource "has been able to recover *all* costs… directly from individual end-users [italics added]." Given this reality, it is not surprising that some new databases fall by the wayside when their initial grants expire. A recent example is the CONTRAST project, which was launched in 2006 with funding from the European Commission. This project created a database to help monitor schistosome parasites in Africa and ran for four years. However, the authors noted that the web interface is no longer available due to lack of financial support (Saarnak et al., 2013). In light of these trends, institutions and governments may need to be more aware of the value of databases, and more willing to fund them for an extended period of time. In fact, some recent examples demonstrate that this has already begun to happen.

Public funding model: Scientific culture has begun to acknowledge the importance of data sharing measures and their underlying support. For example, the National Institutes of Health have been requiring applications for $500,000 or more to include a data sharing plan since 2003 (National Institute of Health, 2003). Similarly, a report by the US Committee on Responsibilities of Authorship in the Biological Sciences stated that authors of scientific publications should be required to deposit their data sets into suitable databases. This report also stated that "funding organisations should provide the recipients of research grants and contracts with the financial resources needed to support dissemination of publication-related data and materials" (Cech, 2003). Schofield et al. (2009) concluded that journals and funding bodies are the most important enforcers of data sharing policies, and added that, for some organisations, compliance with these conditions is a condition for additional funding. It is important to note that the existence of permanent, functional databases has become even more critical over the last decade as more and more material is generated during research studies. Biologists have entered the world of big-data, and are increasingly dependent on massive online data sets (Marx, 2013). However, some of that information is not stored

reliably. For example, many authors upload their data on journal websites as supplementary material, but Anderson et al. (2006) found that only 83% of the links to supplementary data were functioning one year after publication. Given these developments, we recommend that the FRDC considers the following condition: that funding recipients be required to include the associated costs of adding material to Neptune in their application. We also recommend that Neptune be considered the storage space for supplementary data accompanying aquatic animal health publications.

Data maintenance: Data maintenance is a second critical aspect to consider when planning Neptune's future. Maintenance includes such services as: keeping the information in the database up to date, dealing with user queries and running user training, correcting errors, and performing quality control. These tasks are of paramount importance- in fact, the director-general of the European Molecular Biology Laboratory, one of the institutions supporting UniProt, told potential subscribers that they would be paying for the "very labour-intensive expert curation" (Abbott, 1998). However, there is often inadequate support for these activities in the scientific community- in a recent international survey of 1700 researchers, 80% of respondents stated that there wasn't sufficient funding for the curation of their data (Science Staff, 2011). The fact that Neptune contains manually curated data is one of its greatest assets. For example, in a study of four major protein databases, Schnoes et al. (2009) found that UniProt was the only one without high levels of annotation errors. It was also the only manually curated resource, the others using automated methods. The validation system created for Neptune represents another asset, as it helps to ensure data quality for current and future entries. This is a two tiered system which interacts with user profiles. Neptune users are either viewers (least powerful), submitters, validators or administrators (most powerful). Individuals belonging to the latter three groups can enter new information into Neptune, but only validators or administrators can validate this information. Information that has been entered into Neptune, but not validated, is not visible on the Neptune search page to viewers or submitters, who constitute the majority of users. Therefore, only information that has been reviewed by an expert is accessible to most users.

Another key aspect of data maintenance is the ongoing entry and revision of data by users. This is especially important in the absence of a project manager. However, members of online communities can be reluctant to perform this type of task. For example, Mons et al. (2008) called for community annotation on the website WikiProteins, an online workspace connecting concepts about diseases from other databases, including PubMed. Given that information from more than one million PubMed authors was present in WikiProteins, Mons et al. (2008) suggested that each of these "million minds" could review the few entries relevant to their expertise, and stated that "community annotation" was the primary goal of WikiProteins. However, the movement on WikiProteins was ultimately unsuccessful due to lack of engagement (Masseroli et al., 2014). Similarly, Pennisi (1999) described the existence of errors in both sequences and sequence labels stored on GenBank, and found that submitters were often loath to correct entries. The lack of "reward" for this activity was cited by one researcher as a disincentive. Given that at least some Neptune users are likely to share this reluctance, we would doubly recommend that the FRDC considers making entry of data into Neptune a condition of funding new projects. We also recognise that, as the aquatic animal health community is relatively small, there is probably a need for ongoing work to be performed by a paid expert from time to time.

Online slide facility: The online slide library is a third facet of the Neptune project that requires consideration. The library, which presently contains 180 slides of key endemic and exotic diseases, is already an invaluable asset for the aquatic animal health community. As it is anticipated to grow steadily in coming years, a plan is required for managing its future development. For example, Prof Richard Whittington has obtained several hundred slides from Dr Miyazaki, an eminent Japanese researcher, showing pathology found in fish from Japan. These slides will be scanned and added to the Neptune library in the future. In addition, pathologist Stephen Pyecroft has expressed an interest in contributing more slides from his collection to the Neptune library. The Subcommittee on Animal Health Laboratory Standards (SCAHLS) has also begun sending Aquatic Slide of the Quarter specimens to the Neptune project manager, and many other potential additions to the library are sitting in local collections, waiting to be scanned. Given the present value and future capacity of the slide archive, we recommend that the FRDC considers the following condition: that principal investigators of projects anticipated to generate important histopathology specimens be required to submit specimens to the slide library. It is important to note here that the copyright of the donated slides is still being determined. Currently an agreement is being considered in which contributors would grant the Australian Department of Agriculture a perpetual, non-exclusive, royalty free

licence to use the whole slide images generated from donated slides. A draft copyright contract has been written, but expert legal advice is yet to be obtained on this issue. This agreement should be finished before future donations are scanned.

User engagement: A fourth challenge for Neptune's future lies in the user community; users often need regular encouragement to stay engaged. Existing members may sign up once and then never log in again, and potential members may never sign up at all. For example, Joyce and Kraut (2006) studied six online communities, and found that the majority of new users made one post and then never posted again. In addition, the user community itself may be stratified into more and less active contributors. A common phenomenon found in online communities is the 1% rule, which states that 1% of users contribute the majority of content (McConnell and Huba, 2006). This hypothesis was recently tested and confirmed across four health social networks with a combined user community of 63,990 individuals. The authors found that 75% of activity was generated by the top 1% of users, with 24% of activity being created by the next 9% of users. The remaining 90% of users, called "lurkers," were responsible for 1% of activity (van Mierlo, 2014). In light of these findings, it is particularly encouraging to note that Neptune users have already been searching the database, and have found the extent and quality of the information in Neptune helpful. For example, in response to a survey given by ECU, one user wrote that Neptune "has the potential to add considerably to the diversity and quality of my teaching and research." Good user engagement has also been shown during the webinar series. Webinars often have had excellent turnout, with a record of 32 attendees, and these online presentations represent a definite success of the project. Individual users have given positive feedback to the project manager about the webinars- for example, one person wrote "thanks for organising such great topics for webinars" and stated that they "learnt a lot."

Governance: Finally, a brief comment on governance is necessary. We recommend that input from the user community should be collected on an ongoing basis to ensure that the database continues to meet user needs. It would be efficient to use an existing structure for this purpose, such as the Subcommittee on Aquatic Animal Health (SCAAH), or the FRDC Aquatic Animal Health Subprogram.

Conclusion

With the error of big-data unfurling, scientists are in a different world than they were one or two decades ago. Every aspect of scientific research has become more intimately linked with technology, from performing a literature review to publishing one's own findings, and this trend is only going to increase in the future. In this world, comprehensive and easy to use databases are valuable aggregators of a wealth of otherwise disseminated data, and there is no doubt that Neptune will only grow in value and use over time. This resource will undoubtedly facilitate the goal of all aquatic animal health professionals: a better, more accurate understanding of the aquatic diseases and parasites around us.

Implications

a) Neptune is likely to become an indispensable national resource for Australia's aquatic animal health professional community. The user community includes researchers, government officers, aquatic animal industries, private veterinarians and consultants. The facility will also be of immense value for student training. Neptune's impact will be sustained and its benefits will far exceed its relatively small establishment costs. Neptune will become a world-leading facility that has a data set, resources and capabilities well exceeding any comparable aquatic animal health facilities elsewhere in the world.

b) The return on investment provided by Neptune is difficult to estimate due to the breadth of its application and the intangible value of some outcomes. However, the examples below are provided to demonstrate how Neptune can provide a very high potential return on the investment in this project.

Example 1: Online histopathology workshop for 12 participants. Costs savings over a traditional laboratory-based work shop will include lower course fees, no travel and accommodation costs, and lower capital costs to the course provider (e.g. microscopes, teaching laboratories). Online workshops offer flexible scheduling (laboratory facilities are not required), access to more extensive teaching material (a shared microscope slide image library would be available) and cost savings of up to $30,000 (based on participant costs of $2500 for a traditional workshop). For example, the online workshop held on September 11th, 2013, had 19 participants.

Example 2. Evidence for Australia's disease status. Australia's trading partners may require advice regarding our disease status for some non-listed diseases. Official information is not collected for non-listed diseases, so other sources of information may be required. Neptune will collate comprehensive published information on the occurrence of aquatic animal diseases in Australia, so that determining whether an agent has been recorded previously will involve a single search of the database. The alternative would be an extensive literature search that may take hours or days.

Recommendations

We recommend that the FRDC considers the following conditions:

- that funding recipients be required to include the associated costs of adding material to Neptune in their application.
- that Neptune be considered the storage space for supplementary data accompanying aquatic animal health publications.
- that principal investigators of projects anticipated to generate important histopathology slides be required to submit specimens to the slide library.
- that a copyright agreement addressing the use of whole slide images of donated slides be finalised.
- that input from the user community should be collected on an ongoing basis to ensure that the database continues to meet user needs. It would be efficient to use an existing structure for this purpose, such as the Subcommittee on Aquatic Animal Health (SCAAH), or the FRDC Aquatic Animal Health Subprogram.

Further development

We propose that the Aquatic Animal Health Subprogram Steering Committee consider the recommendations provided above, particularly where they are relevant to the operation of the committee.

The data entry for objective 2 is at 88%, as stated in the results. Given that 12% remains to be entered, additional work will be required to complete data entry. This work can be done using the AAAPD Excel spreadsheet as a guide, which contains information from all of the events in the original Australian aquatic animal pathogen and disease database. It is important to note that the final stages of data entry consist of the events that are most time-consuming to complete, including monographs that consist of a series of descriptions of new species.

The need for an expert to complete ongoing work on the Neptune database was mentioned in the discussion; it is possible that this person could also perform data entry for the information remaining in the AAAPD spreadsheet.

Extension and Adoption

The project was extended and communicated to the end user in several ways. The webinars have been well attended and were advertised to all Neptune users. Invitations were also sent to Health Highlights subscribers via Joanne Slater. Notices about Neptune have also been posted on the FRDC's Facebook page.

Project coverage

An article describing Neptune was released online on the 15th of August 2013 in Austasia magazine. A presentation about Neptune was also given by the project manager at the Second Australasian Scientific Conference on Aquatic Animal Health, which took place in Cairns in July 2013. The project manager gave a similar presentation at the Australian Society for Parasitology conference in Canberra in July 2014.

Project materials developed

The Neptune database and online slide library are the most important materials that have been developed during this project. Both can be accessed at https://www.abin.org.au. A manuscript describing the Neptune database and whole slide image library has also been written and is under review.

Appendices

Project Staff

Project manager: Dr Marissa McNamara

Project supervisor: Dr Robert Adlard

This project has been implemented in close consultation with the Australian Department of Agriculture.

Intellectual Property

Existing intellectual property is covered by agreements between the Australian Department of Agriculture, CSIRO, and Biosecurity Queensland. Intellectual property developed during this project is vested in the Australian Department of Agriculture.

References

Abbott, A. 1998. Swiss databank to start charging for use. Nature 394, 214.

Anderson, N.R., Tarczy-Hornoch, P., Bumgarner, R.E. 2006. On the persistence of supplementary resources in biomedical publications. BMC Bioinformatics, 7:260.

Ball, C.A., Sherlock, G., Brazma, A. 2004. Funding high-throughput data sharing. Nature Biotechnology 22, 1179-1183.

Cech, T.R., Eddy, S.R., Eisenberg, D., et al. 2003. Sharing publication-related data and materials: Responsibilities of authorship in the life sciences. Plant Physiology 132, 19-24.

Chandras, C., Weaver, T., Zouberakis, M., et al. 2009. Models for financial sustainability of biological databases and resources. Database- The Journal of Biological Databases and Curation, bap017.

Joyce, E., Kraut, R.E. 2006. Predicting continued participation in newsgroups. Journal of Computer-Mediated Communication 11, 723-747.

Marx, V. 2013. The big challenges of big data. Nature 498, 255-260.

Masseroli, M., Mons, B., Bongcam-Rudloff, E., et al. 2014. Integrated Bio-Search: challenges and trends for the integration, search and comprehensive processing of biological information. BMC Bioinformatics 15, S2.

McConnell, B., Huba, J. 2006. The 1% Rule: Charting citizen participation, accessed 15 May 2014. <http://web.archive.org/web/20100511081141/http://www.churchofthecustomer.com/blog/2006/05/charting_wiki_p.html>.

Mons, B., Ashburner, M., Chichester, C., et al. 2008. Calling on a million minds for community annotation in WikiProteins. Genome Biology 9, R89.

National Institutes of Health. 2003. Final NIH statement on sharing research data, NOT-OD-03-032, accessed 29 May 2014. < http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

Pennisi, E. 1999. Keeping genome databases clean and up to date. Science 286, 447-450.

Saarnak, C.F. L., Utzinger, J., Kristensen, T.K. 2013. Collection, verification, sharing and dissemination of data: the CONTRAST experience. Acta Tropica 128, 407-411.

Schnoes, A.M., Brown, S.D., Dodevski, I., et al. 2009. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. PLoS Computational Biology 5, e1000605.

Schofield, P.N., Bubela, T., Weaver, T., et al. 2009. Post-publication sharing of data and tools. Nature 461, 171-173.

Science Staff. 2011. Dealing with data. Challenges and opportunities. Introduction. Science 331, 692-693.

van Mierlo, T. 2014. The 1% Rule in Four Digital Health Social Networks: An Observational Study. Journal of Medical Internet Research 16, e33.