Managing Cost and Promoting Sustainability in an Animal Health Information System: Technical Aspects of Animal Health Australia's National Animal Health Information System

Cameron, AR[1]; Madin, B[2]; Hutchison, JM[3]; Langstaff, I[4]; Sergeant, ESG[5]
[1] AusVet Animal Health Services Pty Ltd; angus@ausvet.com.au
[2] AusVet Animal Health Services Pty Ltd; ben@ausvet.com.au
[3] AusVet Animal Health Services Pty Ltd; jenny@ausvet.com.au
[4] Animal Health Australia; ILangstaff@animalhealthaustralia.com.au
[5] AusVet Animal Health Services Pty Ltd; evan@ausvet.com.au

ABSTRACT

Information requirements for managing animal diseases change rapidly, both in terms of the diseases and their parameters of interest, and analysis and presentation of information required for decision making. New approaches are required for the development of animal health information systems to meet these challenges and to minimise costs. This paper describes some technical aspects of Animal Health Australia's National Animal Health Information System (NAHIS) that aim to minimise the cost of maintenance, whilst maximising the ability of the system to adapt to the changing animal health situation. Some of the key features of the system include: a preference for powerful, widely used, open source software tools; the use of a flexible metadata database structure; a web-based interface for the creation of new system outputs (including tables, graphics and maps); and a simple web-based data submission system matching the state-based surveillance model used in Australia.

Keywords
animal health information system, database, surveillance

INTRODUCTION

Animal health information systems perform a critical role in the collection, management, analysis and reporting of animal health information for a variety of purposes, including endemic disease control, outbreak response, disease surveillance and trade support. Problems that plagued early computerised systems such as a lack of integration of different data types, and a difficulty in extracting appropriate outputs, have been progressively overcome. However, current problems with many national systems include the cost associated with their development and maintenance, and either their limited longevity, or the cost required to constantly adapt the systems to the rapidly changing requirements for surveillance, information and reporting. Animal Health Australia[1] overcomes these constraints with certain features of its National Animal Health Information System (NAHIS). The NAHIS is a sophisticated, flexible, web-based information system to support Australian animal health surveillance programs through online submission of nationally relevant data, automation of data analysis and summarisation, and provision of customised output reports.

TECHNICAL FEATURES

The NAHIS consists of a number of integrated, web-accessible applications. These include animal health applications in support of Animal Health Australia's surveillance programs (current applications are the National Animal Health Information Program application, the Endemic Disease Information System, and the National Arbovirus Monitoring Program Information System) along with ancillary applications (a mailing list manager, animal disease information sheet management system and an animal health wiki[2]). All of these applications share a common code base and use a common database for data management.

Development Software

During the development of the system, preference was given to the use of open-source software except where commercial systems were already in place. This decision was based on a number of factors including: a significantly lower cost (avoiding the sometimes very high purchase and/or licensing fees associated with commercial software); the developers' view that mainstream open-source software products are often better

[1] A not-for-profit public company established and funded by the Australian, state and territory governments and major Australian livestock industry organisations. http://www.animalhealthaustralia.com.au
[2] A wiki is a collection of web pages designed to enable multiple authorized users to contribute or modify content using a simplified markup language.

supported than commercial software (supported by an extensive and responsive developer community); the developer's view that the technical capabilities of the open-source software selected rivalled, and in some cases surpassed that of commercial alternatives; and the open-source software selected was more than adequate to meet the needs of the system.

The key software components used in the system include the operating system, database management system (DBMS), web server, web scripting language, web mapping server and statistical analysis and graphics software. At the time of development, Animal Health Australia's existing commercial software infrastructure included a Windows server operating system[3], Windows IIS web server[3] and Microsoft SQL Server[3] DBMS. These were retained.

Web scripting in the system's applications is implemented using PHP[4], which is amongst the most widely used options on the internet. Most of the system's functionality is implemented by the combination of PHP scripts and MS SQL Server database functionality. While these tools are very flexible, they are not adequate to support some of the data output requirements of NAHIS, including data summarisation and analysis, as well as graphical and map-based outputs. Accordingly, the R[5] statistical environment is used to manage many aspects of data output, including the generation of simple summary and cross tabulated tables, as well as the generation of graphical outputs (bar charts, pie charts, histograms, scatter plots and time-series charts). The generation of real-time interactive maps is managed by MapServer with the 'Google map style' user interface being managed by OpenLayers[6].

Flexible Data Structures

Traditional database design involves analysis of the data to be managed and development of table structures that closely match the data. Whilst this approach is simple and efficient, any subsequent changes in the data collected will require a change to the database structure. In a mature system such changes can be complex and costly.

In animal health surveillance, data requirements change rapidly. This may be due to the emergence of new diseases to be included in a surveillance program (requiring the collection of disease-specific information) or due to the changing needs of existing surveillance. The system was designed to provide a flexible data management structure that does not require modification of the database to accommodate different types of data.

The database consists of two distinct parts. The first contains metadata, or definitions of the structure of the data to be stored. This is a description of a traditional data table or tables, and includes information on the fields to be included. The second part stores the actual data in a series of generic tables.

To the user, the database operates like a traditional database; there appear to be a number of different tables, each with a structure specifically suited to the surveillance data to be managed. These may be thought of as 'virtual tables'. The system administrators are able to modify the structure of existing virtual tables and create any number of new ones by using the web-based administration interface, and without requiring any changes to the underlying database structure. This gives database administrators the ability to rapidly respond to changing data requirements. For example, if a new surveillance program is required to gather data about an emerging disease appropriate data structures can be defined in minutes, and the system is immediately ready to handle data input, management and analysis. Similarly, if the information needs of a surveillance program change, the structure can be quickly modified or extended. This ability makes the full data management, analysis and reporting power of a mature information system available as soon as data gathering starts, instead of the more common situation where ad-hoc spreadsheets or one-off desktop databases are used until a more permanent home can be found for the data.

Data Capture Systems

In Australia, state and territory governments are responsible for disease control and eradication within their own jurisdiction. Surveillance is therefore largely conducted at the state and territory level, using a variety of local data management systems. Whilst a range of efficient options for data capture were proposed during the development of the system, it was decided that the best approach was to consider people before technology. The aim was therefore to ensure that those contributing to the system would be able to do so with minimum change to their existing data management systems and without the need to develop new

[3] http://www.microsoft.com
[4] http://www.php.net
[5] http://www.r-project.org/
[6] http://openlayers.org/

skills; all of the people responsible are familiar with the ubiquitous spreadsheet (e.g. Microsoft Excel®, OpenOffice Calc), and many use it routinely for local data management. Alternatively, where state and territory coordinators have local databases, these can normally export data to this format. Data submission therefore involves transferring data from Excel to the database via a web interface. Users copy a data block from Excel and paste it into a blank text box on one of the animal health application web pages. The data is sorted and undergoes detailed consistency checks, and feedback is provided instantly to the submitter in the event of problems. Whilst this approach may be considered relatively rudimentary, it is appropriate for the task, and allows surveillance data for new diseases to be rapidly incorporated into the system.

Output Definitions

Surveillance information analysis and reporting requirements also change rapidly. A common feature of information systems is that the more people use it, the more they want out of it. The approach to implementing system outputs in the NAHIS was designed to offer the greatest possible flexibility and growth of the system without requiring complex coding of new outputs. As with virtual table structure, output definitions are stored as metadata and generic routines use this metadata to create the required output. The outputs available consist of simple filtered data listings, tabular summaries including cross tabulations, graphical data summaries, and maps.

The output definition consists of two components: a Structured Query Language (SQL) query defining the data to be extracted, and a set of display commands defining how the data will be manipulated and displayed. For tabular and graphical outputs, the display commands take the form of an R script; for maps they take the form of a MapServer[7] map file definition. In both cases, most common output formats can be quickly defined through the application administrator's web interface, which automatically generates the required SQL and R or MapServer code. For highly specialised outputs the administrator has the option of manually refining the SQL and R code. Unlike the web interface, this requires a detailed knowledge of the SQL and R programming languages, but offers almost unlimited flexibility for complex custom designed outputs without the need to change any underlying system code.

Extending Functionality

The NAHIS makes use of a common code base for managing functions shared between the different applications. Common core data tables (for instance, user access control and disease lists) mean that the system is well placed to serve a number of functions beyond surveillance data management. A simple example of this is the development of a mailing manager application, which handles the distribution lists for much of Animal Health Australia's communication material.

A more complex example is the implementation of a customised Wiki within the system, based on PmWiki[8]. At its simplest, this is used to maintain documentation for the information system and surveillance programs. The wiki has been customised for use as a management tool for the development, revision, approval and distribution of disease information sheets. The wiki's features of allowing collaborative document development and revision have been combined with customised business and user access rules to handle a formal cycle of revision, editing, approval and publication.

CONCLUSION

Animal Health Australia's NAHIS is based on a range of standard software tools, both commercial and open source. However, these have been implemented in a way which aims to maximise the longevity and functionality of the system whilst minimising maintenance costs. The ability for the applications' inputs and outputs to be extended and modified via the web-interface has allowed the system to grow without the need for expensive redevelopment. However, this flexibility comes at a cost. The use of open-source software sometimes means higher initial development costs than may be expected with commercial software. More importantly, the flexible database structure and generic data handling functions are all more complicated to develop than would be the case in a more rigid system. Similarly, the data structure imposes a number of data processing overheads and may be slightly slower than a system based on a dedicated data structure. Despite these drawbacks, the NAHIS has enjoyed heavy use and continual expansion since its development four years ago, which is a testament to its success in meeting its design objectives.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the valuable feedback provided by Animal Health Australia staff and many Australian animal health stakeholders during the development and growth of the system and its associated applications.

[7] http://mapserver.org/
[8] http://www.pmwiki.org