Free internet access to databases for all agroecological regions of Australia - why not?

Jim Scott

Centre for Sustainable Farming Systems, University of New England, Armidale, NSW, 2351, www.une.edu.au/csfs Email Jim.Scott@une.edu.au

Abstract

Typically, research produces complex datasets which are specific to particular agroecological regions, soil types, seasons and/or years. Examples of the types of data collected in such experiments include climatic, soil, herbage, crop, livestock, economic and management information. This paper describes a range of databases and explores the obstacles which need to be overcome in order to make them more available. If data relating to agriculture and natural resource management could be shared more widely, there will ultimately be great benefits to land managers.

Whilst scientific findings are usually made public through journal articles and extension literature, it is rare that the data supporting those publications are made available outside the research institution. Allowing access to data would enable the creation of a powerful repository of information which future scientists and farmers could benefit from indefinitely into the future as it would accelerate and better target future research.

Quality datasets are required in order to develop generalised relationships beyond year- and site-specific data. Current problems facing sustainable and profitable farming in a risky environment require modelling to explore long-term relationships. The provision of more comprehensive datasets from all agroecological regions would provide the means of improving the predictive capacity of models and decision support tools. Since the arrival of the “Google” era, it has become obvious that it is technically feasible to assemble and access vast quantities of data and information via the Internet in a convenient and cost-effective manner. A number of existing agricultural research databases are described along with the potential benefits of linking these datasets. Further, the opportunities to link research data from experimental plots to farmlets (representing the farm system) to whole farm data warrants exploration as it could provide great benefits to the research-extension-adoption continuum by providing better access to objective information.

Key Words

Knowledge, experimental results, research collaboration, validation

Introduction

Although databases are now common in agricultural and natural resource management studies, many have been created for relatively specific purposes. Examples include databases of plant pathogens (Shivas, Beasley et al. 2006) or soil related information (Johnston, Barry et al. 2003). Currently, there are few examples of integrated national databases in Australia. One of the reasons for the lack of coherent integrated databases is that many database projects have suffered a common fate of only being supported for the duration of those specific projects.

It is becoming increasingly obvious to all those connected to the Internet, that vast databases of information and data can be created and accessed via the Internet. This applies not only to library catalogues, journal articles and telephone numbers but also to items for sale, images, and of course, databases of indexed terms extracted automatically from web pages using search ‘engines’ such as Google (http://www.google.com).

As the capacity of modern computing tools has increased dramatically over recent years, it is now feasible to contemplate organising much of Australia's agroecological data into a form that can allow easy access by any authorised user via the Internet.

It is common in research conducted within all agroecological regions to compile data relating to factors such as climate, soil, herbage, biodiversity, crop yields and quality, livestock, economic and/or management information. There is potential value in making data collected from different years available to enable results to be examined over longer timeframes than are usually covered in individual projects. In addition, considerable data is at times gathered from farmers and this can have value to them individually for purposes such as benchmarking as well as collectively for other end users such as Catchment Management Authorities.

This paper explores some of the opportunities and obstacles of sharing agricultural and natural resource data via one or more large databases that could be accessed by any authorised user via the Internet.

National research databases

In Australia, the use of databases for storing national research data has been evolving as computer technologies have improved and especially over the past decade. For example, the Australian Soil Resource Information System (ASRIS) was established by Johnston, Barry et al. (2003) and has provided on-going access to soils information across the nation since its establishment. A notable feature of this soils database is the extensive collaboration between and continuing support of the agencies that brought it about.

Several national and regional research datasets that have been organised into relational databases are depicted in Figure 1 below. These datasets represent the results of at least $35 m of research conducted over the past decade. Collectively, they provide the potential for Australian researchers to explore common issues across time and space. As suggested in the figure, the overall aim of collating quality datasets is to provide sources of trusted objective information about agriculture and natural resource management. These datasets can be derived from research trials, whole farmlet trials ('fact farms' which allow the exploration of farming systems issues at a scale credible to farmers) and from individual farms. Some examples of these databases will be described further below.

Research databases

One of the early national agricultural research databases was created for the Sustainable Grazing Systems (SGS) Key Program (Scott and Lord 2003). That database, compiled as compatible but separate databases for each site using MS Access, contained some 127 tables and hundreds of queries which allowed exploration of five theme analyses (water, nutrients, pastures, animal production and biodiversity) across 8 national research sites. The database was especially useful for distilling results for publication in a special journal issue reporting the results of all of the trials conducted in that national experiment. A survey of participants in that national experiment found that the use of a common database was valued highly and ranked second highest out of 13 innovations) (Andrew, Lodge et al. 2003). It was also useful for extracting data to compare with a model developed in parallel with the database during that project (Johnson, Lodge et al. 2003). Following completion of this project, a pilot study was satisfactorily carried out which showed the feasibility of migrating these databases from MS Access to SQL Server which could be queried via a web interface.

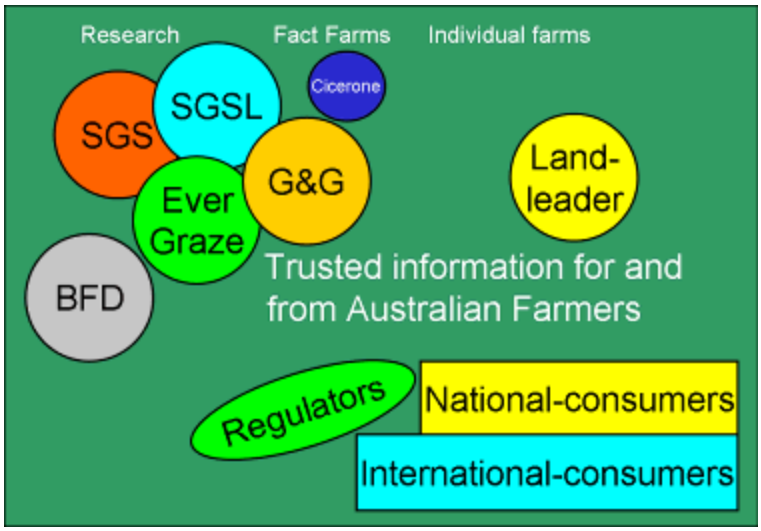

Figure 1. Diagram summarising the contribution that some of the national and regional research databases might be able to make towards providing a trusted source of information for and from Australian farmers. These databases include: Sustainable Grazing Systems (SGS), Sustainable Grazing on Saline Lands (SGSL), Better Fertiliser Decisions (BFD), EverGraze, Grain & Graze (G&G), Landleader (EMS data derived from farmers) and Cicerone (regional farmlet study).

This project was followed by the Sustainable Grazing on Saline Lands project which also employed a common database approach, using MS Access, for assembling data from across 5 sites located across southern Australia. Datasets included climate, soil moisture, runoff, soil nutrients, salinity, botanical composition, biodiversity, livestock performance and wool production. It also allowed the calculation of complex derived data such as metabolisable energy, ionic balance, a waterlogging index and transpiration calculations. This project also included a large producer network which sought to communicate results via a web-based database (Hardy, Collins et al. 2006).

The Better Fertiliser Decisions project was a broad national project which brought together some 5000 experimental years of fertiliser response and soil test data from approximately 3000 sites (Gourley, Melland et al. 2005), again in a MS Access database with a facility for creating HTML reports. Although this database is described on the ASRIS web site (http://www.asris.csiro.au/themes/nutrient.html), there is as yet no facility for direct access to the database via the Internet.

The research and adoption program, EverGraze (http://www.evergraze.com.au/index.htm), is focusing on the evaluation of perennial pastures across three States in the high rainfall zone. This program has also adopted a modelling approach and a common database structure based on MS Access for storing and analysing data. Again, there are common elements with some of the other databases as measurements include soil water, soil health, biodiversity, pastures and livestock. This program also has a sizeable component focused on adoption by carrying out many investigations on-farm using common data collection procedures.

The Grain & Graze national database provided a valuable national data repository of information relating to soil, crop, pasture, livestock and biodiversity information within mixed farming systems. Because of the challenges of supporting different databases in each of the nine Grain & Graze regions, with a complex array of over 60 participating organisations, it was decided that a national database would be created as a single comprehensive database with access provided via a web portal. In order to allow data sharing whilst protecting datasets from unauthorised use, a secure portal was created with encryption of all activities (https://rdu-online.une.edu.au/dnn3/Projects/GrainGrazeCurrent/tabid/55/Default.aspx). Once logged in, authorised users are provided with access to a members' only page where users can download data templates and upload datasets. Users are also given access to the online database with a direct link to the back-end SQL server database according to their user role assigned by the database manager on

recommendations from project leaders. In this way, some users can have access to data from just one site whilst others can have access to any or all datasets. Where available, metadata is also provided via the online database together with flexible graphing, advanced graphing and tabular views of data. The owners/custodians of each subset of data is displayed on the screen together with a link to the appropriate contact details should a user wish to obtain permission to download sub-sets of the data.

Fact farm data

The Cicerone Project (http://www.cicerone.org.au/) involved the study of three whole-farmlet management systems which were monitored intensively by researchers and farmer members over 6 years. Data arising from the project was stored in a MS Access database with links to queries made available via a secure web portal so that participants could share data relating to soils, pastures, livestock, wool quality, economics and labour.

Individual farm data

The Landleader pilot project (https://www.landleader.com.au) is collecting data on environmental and livestock management practices on farms through paper- and web-based surveys. The database used for the pilot project was SQL Server which was accessed via a secure web portal. In this way, individual land managers could provide data in a secure way and receive a customised report showing their relative performance against a range of criteria. Based on the results, land managers can thus be provided with customised links to information about the most appropriate Best Management Practices.

Discussion and Conclusions

Science has traditionally been open in its communication of research results through publications. One might ask therefore, why is it not possible to develop more open systems of data sharing? It appears that the sharing of data is hampered somewhat by the competitive nature of research funding. This can result in constraints brought about by the needs of researchers and their agencies to be acknowledged as the originator(s) and/or owner(s) of particular results. Where the data have been contributed by a large number of scientists and/or their agencies, there can be concerns about the ownership of these research datasets (Scott 2001). These concerns need to be overcome if Australia's land managers are to benefit from modern database facilities which will enable the efficient provision of vast amounts of well structured data which will enable more focused research efforts to fill in the gaps in data as well as accelerate research progress.

Limited availability of datasets constrains the testing of models which often seek to generalise relationships beyond year- and site-specific data. Current problems facing sustainable and profitable farming in a risky environment require modelling to explore long-term relationships. The provision of more comprehensive datasets from across many agroecological regions would provide the means of improving the predictive capacity of models and decision support tools, thereby enabling improved management.

If the culture of data sharing can be changed and facilities for storing and analysing large datasets improved, not only will access to objective information be improved for existing datasets but, in addition, it is likely that many more datasets will be offered up for long term data storage. This is especially the case when experienced researchers approach retirement and may wish to leave their datasets for long-term storage. In the future, the term 'publishing' may come to mean more than just writing a journal article; it may include also supplying the data and metadata related to that publication to a managed data 'warehouse'.

Whilst it is now obvious to all that it is feasible to enable vast databases to be accessed by the Internet, there has been insufficient commitment to ensure that this happens for research data. It is time for all those who can benefit from the availability of quality datasets to overcome the obstacles and focus on the opportunities that such a facility will create. Enabling the common storage of and providing authorised access, via the Internet, to many of the databases described in this paper would be a good start.

Acknowledgements

The creative efforts of the database developers of several of the research databases referred to in this paper, Colin Lord, Jim Cook, Dion Gallagher and Ed Campbell, are gratefully acknowledged. The development of these same databases has been variously supported by Dairy Australia, Meat and Livestock Australia, Australian Wool Innovation, Land and Water Australia, Grains Research and Development Corporation, NSW Department of Primary Industries, Victorian Department of Primary Industries, the Cicerone Project and the University of New England. Data have been provided by scientists working for many State and Federal organisations and Universities.

References

Andrew, M. H., G. M. Lodge, et al. (2003). "The Sustainable Grazing Systems National Experiment. 2. Scientific outcomes and effectiveness of the research and development processes." Australian Journal of Experimental Agriculture **43**(7-8): 993-1013.

Gourley, C. J. P., A. R. Melland, et al. (2005). Fertiliser responses and soil test calibrations for grazed pastures in Australia. XX International Grassland Conference, Dublin, Ireland, Wageningen Academic Publishers, The Netherlands.

Hardy, J., J. Collins, et al. (2006). Farmer driven innovation – the backbone of the SGSL Producer Network. Practice change for sustainable communities: Exploring footprints, pathways and possibilities. Proceedings of APEN International Conference. Beechworth, Victoria, Australia, The Regional Institute Ltd.

Johnson, I. R., G. M. Lodge, et al. (2003). "The Sustainable Grazing Systems Pasture Model: Description, philosophy and application to the SGS national experiment." Australian Journal of Experimental Agriculture **43**(7-8): 711-728.

Johnston, R. M., S. J. Barry, et al. (2003). "ASRIS: the database." Australian Journal of Soil Research **41**(6): 1021-1036.

Scott, J. M. (2001). Co-ownership of agricultural research data – a difficult issue in the digital age. Proceedings of the 10th Australian Agronomy Conference, Hobart.

Scott, J. M. and C. J. Lord (2003). "SGS Database: use of relational databases to enhance data management for multi-site experiments." Australian Journal of Experimental Agriculture **43**(7-8): 729-743.

Shivas, R. G., D. R. Beasley, et al. (2006). "Specimen-based databases of Australian plant pathogens: past, present and future." Australasian Plant Pathology **35**(2): 195-198.