

Discovery of Australia's research and researchers on-line

Paper presented at Network 2000 by Andrew McCredie, KBE Branch, ISR
8 December 2000

There are currently over two million documents about Australia's researchers and their research on the internet. My paper today is about what Government and the researchers themselves can do to make this information more discoverable.

The Age of the Internet

Five years ago very few of us used the internet. Today we use it regularly. We use it for banking, paying bills, shopping, doing business, or just for getting information. You don't have to be Bill Gates to realise that we have entered the Age of the Internet.

In universities and in Government the main use of the internet is to get information. We use search engines like Google, Alta Vista or Look Smart to find out about people and organisations. We also use them to locate news, documents or other information published on the internet. And as you would all have experienced, as wonderful as the internet is, it can at times be a very time-consuming and frustrating business finding what you are looking for.

There is a lot of information out there on the internet but most of it is poorly organised and low quality. Just think back to the results you obtained the last time you did a key word search. Internet search engines are good if you know exactly what you are looking for and how to fully utilise them. Then of course there is the possibility that the information you are seeking isn't there.

What applies to the internet generally also applies to Australian research organisation web site content in particular, that is:

- The internet is increasingly the primary source of information about Australian research and researchers
- There is a lot of information out there, over 2 million web documents, in fact, and most of it is poorly organised and useless
- Finding the information can often be hard work
- Search engines are good if you know exactly what you are looking for and how to fully utilise them. If you don't know about domain names, text delimiters, and boolean operators steer away from the advanced search functions

I will come back to addressing these problems later.

Government has a role in enabling discovery

In 1996-97 the Commonwealth Government invested \$3.5 billion in research and development, including \$2.3 billion in university research. The outputs from this investment are the trained

researchers and the research results. In both cases the Australian economy is the intended beneficiary of this huge investment of public money.

The trained researchers, some 91000 in 1996-97, provide the ideas and expertise for industry, business, government and the education system. The research results provide the foundation for innovation and economic growth.

Enabling the discovery of Government-funded research activity and new technologies has to be an essential component of Government investment in research and development. If innovation and technology transfer are the objectives, then money spent on making Australian industry and investors aware of the research that is being done, and the innovation opportunities that exist, has to be money well spent. On the scale of things we are talking about loose change!

Impediment to innovation and technology transfer

A long-standing impediment to the commercialisation of Australian research results and inventions has been the relative inaccessibility to information about what Australian researchers are doing and what they have discovered. The impediment still exists to a great extent. Where does industry look to find out what research expertise we have? Where do investors go to identify emerging Australian technologies in a given industry sector?

With the advent of the internet age we have the means to overcome this impediment in an affordable way. Research Finder, which I will describe shortly, aims to help overcome this impediment to innovation and technology transfer, at relatively little cost. And by relatively little cost we are talking about \$100,000 or 0.003% of Commonwealth expenditure on R&D.

Recent Government Reports

There have been three recent Government reviews which have addressed the need to make research expertise and research results more discoverable - the Innovation Summit and its Implementation Group chaired by David Miles, the Science Capability Review led by the Chief Scientist, and the Health and Medical Research Strategic Review and its Implementation Committee chaired by the Peter Wills.

The Innovation Summit held in February this year was the least explicit about discoverability of information but nonetheless stated, in the context of acting on ideas, that:

“knowledge flow in Australian innovation is vital”.

The final report of the Science Capability Review, “The Chance to Change”, recommends that:

“Universities must introduce strategies to stimulate and facilitate increased transfer of knowledge to business and society ... across all sectors of the economy...”

Finally, the Health and Medical Research Strategic Review (Wills Review). In the context of “improving mechanisms for integration of research-based knowledge into ... practice”, it recommended to:

“Develop a strategy for early and effective dissemination of research results to practitioners, policy-developers, decision-makers and consumers in Australia”

and to

“Fund research transfer infrastructure within priority-driven research organisations, including expertise in ... web sites ...”

The Wills Review Implementation Committee, in addressing these recommendations, recognised that the internet has made static, centrally-maintained, expensive directories of researchers and research capability redundant. To cut a long story short, the Wills Implementation committee endorsed:

“the establishment of Research Finder as the premier search engine for Australia’s research capacity”.

What is Research Finder?

Research Finder is an internet search tool targeted at researchers, industry and investors both here and overseas. It is also useful for Government and the general public. It enables on-line discovery of Australia's researchers, research activity and emerging technologies.

Research Finder Stage 1

Stage 1 Research Finder is a high quality keyword search capability of the web sites of Australia's research organisations. It is similar to internet search engines like Google or Alta Vista but there are important differences:

- Search results are only from Australian research organisation web sites
- Results are fresher and more relevant
- Drill-down into web sites is greater

At present the Research Finder index consists of almost 2 million web documents. It covers 192 government-funded research and research-support organisations, institutions, departments and agencies. The current list includes:

22 CSIRO Divisions

64 Cooperative Research Centres

41 universities

30 technology transfer organisations

15 Research and R&D Councils and Corporations

12 Department of Industry, Science and Resources portfolio agencies

17 other government organisations involved in research

13 medical and biotechnology research institutes and organisations

Research Finder Stage 2

Stage 2 Research Finder aims to classify the 2 million web documents by Research Field and Industry Sector. Given the impossible task of manually classifying such a large number of distributed web documents, and the constantly evolving nature of these data, automated classification software will be the primary means of classifying content.

Stage 2 Research Finder will enable classification, or metadata, searches. That is, you will be able to find out for a specific field of research, who is working in the field, what research is being done or has been done. You will also be able to find emerging Australian technologies in a specific industry sector.

Classification Scheme to be used

Following consultations it is clear that a Research Field classification scheme should be based on the Australian Standard Research Classification (ASRC) 1998 produced by the ABS. In particular, we propose to use the Research Fields, Courses and Disciplines Classification (RFCD) within ASRC 1998 as the basis for the Research Field classification scheme. Both the ARC and NHMRC use the RFCD in their granting programs and reporting processes.

The RFCD classification is a hierarchical structure with 24 Divisions, 139 Disciplines and 898 Subjects. Here is an extract of the RFCD at the start of the Medical and Health Sciences Division.

The coarsest classification is Division, within each Division we have Discipline, and the finest classification is Subject.

Based on the advice we have received from text retrieval and classification experts, accurate automated classification should be feasible to the level of Discipline, but not Subject.

ISR is about to undertake a pilot project, with the help of NHMRC, to test the feasibility of accurate automated classification of Medical and Health Sciences research web content.

On-line data already classified using RFCD

While current automated classification software will only be able classify to the level of Discipline, systematically published and classified data will be discoverable by Research Finder to the level of Subject. A current example of such on-line data are successful ARC large grant recipients whose projects are fully classified by RFCD.

The overall strategy therefore is to enable Research Field classification searches to the level of Subject (the finest level) of content which is systematically published, and to classify the rest of the on-line information to the level of Discipline using automated classification software.

Potential uses of Stage 2

Expediting research funding processes

If researchers' Home Pages and on-line research activity reports are classified by Field of Research, this will serve two functions that will help streamline research funding processes. First, it will enable research funding agencies to identify and contact experts and referees in a given research field in Australia. Second, it will enable agencies, referees and researchers themselves to identify current research activity in their specific field and cognate fields in Australia.

University libraries' on-line thesis project

As university theses are progressively published on university web sites they will be discoverable by Research Finder. With appropriate meta-tagging the graduate theses could be searched by key word or metadata (e.g. title, author, publication year, field of research) as a subset of the Stage 2 Research Finder index.

Universities reporting their research activities on-line

In the future universities may be required to report their Government-funded research activities on-line. Many research groups and departments already do this. If these reports are meta-tagged at time of creation according to an agreed research classification scheme this will provide a powerful method of discovery by Government, industry and researchers.

What researchers and research organisations can do

There are two key areas where researchers and research organisations can contribute.

First, they should publish their researcher profiles, contact details, research reports and potential technologies on-line, in search engine discoverable formats.

Second, for content which is published systematically, such as annual reports and research plans, they should include research classification metadata in the web documents.

What Government agencies can do

Government needs to respond in the following ways.

First, it needs to provide researchers with incentives and assistance to publish their research profiles, contact information, research activity and results, and, where appropriate, include research classification metadata.

Second, relevant agencies should add Research classification metadata to its on-line research and research-related content, using agreed metadata syntax.

Third, key Government stakeholders will need to help fund Stage 2 development and ongoing maintenance of Research Finder. Development of Stage 2 is likely to cost \$100,000 to \$150,000, and steady state maintenance approximately \$75,000 per year.

Conclusion

The collective wisdom of three recent Government reviews has identified the need to make Australia's research expertise and research results more discoverable. Action is what is needed now if the benefits are to flow through to our economy and society in the future.

One of these actions is Research Finder. If it is fully implemented it will be a powerful discovery tool; for researchers, industry, investors, Government and the general public. Your support is needed.