



Improving the accessibility of

Australian Agricultural Information on the Internet

**A report for the
Rural Industries Research and Development Corporation
by Jim Groves**

**June 1998
RIRDC Publication No 98/50
RIRDC Project No BDL-2A**

© 1998 Rural Industries Research and Development Corporation.
All rights reserved.

ISBN 0 642 54069 1
ISSN 1440-6845

Improving the Accessibility of Australian Agricultural Information on the Internet
Project No: BDL-2A
Publication No. 98/50

The views expressed and the conclusions reached in this publication are those of the authors and not necessarily those of persons consulted. RIRDC shall not be responsible in any way whatsoever to any person who relies in whole, or in part, on the contents of this report unless authorised in writing by the Managing Director of RIRDC.

This publication is copyright. Apart from *any fair dealing for the purposes of research, study, criticism or review* as permitted under the Copyright Act 1968, no part may be *reproduced in any form, stored in a retrieval system or transmitted* without the *prior written permission* from the Rural Industries Research and Development Corporation. Requests and inquiries concerning reproduction should be directed to the Managing Director.

Researcher Contact Details

Name Jim Groves
Organisation Baradel Consulting
Address 82 Cypress Drive
 BROADBEACH WATERS QLD 4218
Phone: (07) 5526 9698
Fax: (07) 5526 9685
email: grovesc@winshop.com.au

RIRDC Contact Details

Rural Industries Research and Development Corporation
Level 1, AMA House
42 Macquarie Street
BARTON ACT 2600

PO Box 4776
KINGSTON ACT 2604

Phone: 02 6272 4539
Fax: 02 6272 5877
email: rirdc@netinfo.com.au
Internet: <http://www.rirdc.gov.au>

Published in June 1998
Printed on environmentally friendly paper by the DPIE Copy Centre

Foreword

The Internet has become a byword for instant access to information anywhere in the world. Some estimates put the number of pages of information available at around 320 million and it is quickly becoming an important business tool.

But anybody who has tried to find specific information, knows the frustration of having to wade through the scores of "returns" that even the best, refined search engines deliver.

That frustration factor can be multiplied many times for farmers and other economic operators in regional Australia with their slow connections, STD lines and the uneven reliability of their communications infrastructure.

For these reasons, the Corporation commissioned Jim Groves, a consultant specialising in Internet issues for the regional community, to analyse the special information needs of rural Australia and how to substantially improve access to their online information needs.

He study makes a valuable contribution to the process. In this report he canvasses issues such as developing specialised search engines, more and better training in search techniques, how content providers can help through better design of their pages and supporting a range of regional communications initiatives.

The project is part of the Corporation's Human Capital, Communications and Information Systems Program which aims to facilitate innovation in the agricultural sector by improving the use of communication and education processes and appropriate information systems.

Peter Core

Managing Director

Rural Industries Research and Development Corporation

ABOUT THE AUTHOR

Jim Groves is a former economic and policy analyst for the Federal Government. He has been working on projects on Internet content issues for Australian farmers for the Rural Industries Research and Development Corporation (RIRDC) including:

- *Policy Issues in Putting Rural Australia OnLine*, May 1996
- *What's On The Web Now for Australian Farmers?* with Ms Jenny Da Rin, December 1996
- with Robert Parker, Jenny Da Rin and Rosie Simpson, *The Australian Farmers' Guide to the Internet*, December 1997

He also is working on the forthcoming papers with Jenny Da Rin:

- *The Internet Organisation: a study of the demand for and supply of Internet content in rural Australia*
- *The "Third Revolution" and Rural Australia: An Exploration of the Economic and Social Impacts of Farm Internet Use*
- *Online Education and Training for Australian Farmers: the role of the Internet in farm education and training*
- *Buying and Selling Online: the opportunities of electronic commerce for Australian farm businesses*
- future papers will be on farm management and organisational change implications.. A summary of this work program was delivered as *Internet Content for Australian Farmers - a Summary of the Issues* to Australian Farm Management Society Conference, February 1997.

He can be contacted on phone (07) 5526-9698, fax (07) 5526-9685, e:mail grovesc@winshop.com.au. His Web page is at <http://www.winshop.com.au/grovesc/>.

ACKNOWLEDGEMENTS

Respondents to the issues paper issued as part of this project and to the draft report are gratefully acknowledged. These include Ross Andrews of the Grains Research and Development Corporation, Ian Bailey of AgNet, Simon Bryant of the Department of Communications and the Arts, James Cameron of Digital Equipment Corporation, Marina Cavill of the Centre for International Research on Communication and Information Technologies, Ross Cutler of the Pig Research and Development Corporation, Irena Goltz and Jim Graham of the Department of Primary Industries and Energy, Peter Love of Macquarie University, Scott Newman of the Meat Quality Co-operative Research Centre, Bruce Ward and Greg Keegan of the NSW Department of Agriculture, and Max Zeller of @griWeb. The data from the survey of participants of the Farmwide Online Services Pilot project quoted in Chapter 2 was provided by Rosie Simpson, who was also instrumental in the commissioning of the work. Other sources are acknowledged in the footnotes - particularly prominent is Danny Sullivan of Calafia Consulting and his Search Engine Watch site. The program manager in RIRDC who has supported the project is Dr Roslyn Prinsley.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	vi
1 INTRODUCTION	1
2 THE PROBLEM OF INFORMATION RETRIEVAL ON THE INTERNET	3
3 SEARCH ENGINES.....	8
3.1 AN OVERVIEW OF SEARCH FACILITIES	8
3.1.1 DIRECTORIES.....	8
3.1.2 SEARCH ENGINES.....	11
3.1.3 DATA BASES	16
3.2 PROBLEMS WITH SEARCH ENGINES	17
3.2.1 GAPS IN COVERAGE	17
3.2.2 “SPAMMING”	19
3.2.3 OTHER PROBLEMS OF INFORMATION OVERLOAD	20
3.3 SEARCH ENGINE DEVELOPMENTS	22
3.3.1 REFINING SEARCHES	22
3.3.2 “PUSH”	23
3.3.3 SPECIALISED SEARCH ENGINES	24
3.3.4 OTHER DEVELOPMENTS	26
3.4 USER TRAINING	27
3.4.1 ONLINE TUTORIALS	27
3.4.2 OTHER TRAINING ISSUES.....	29
3.5 THE ROLE OF CONTENT PROVIDERS	30
3.5.1 DESIGN, PRESENTATION AND MARKETING	31
3.5.2 META-DATA.....	34
3.5.3 OTHER IMPROVEMENTS ON THE DIRECTORY IDEA.....	36
3.5.4 REGIONAL INITIATIVES.....	37
4 OTHER ASSISTANCE FOR FINDING INFORMATION ON THE INTERNET.....	39
4.1 NEWSGROUPS, FORUMS AND MAILING LISTS	39
4.2 INFORMATION SERVICES.....	42
5 CONCLUSION - A STRATEGY TO IMPROVE THE ACCESSIBILITY OF AUSTRALIAN AGRICULTURAL INFORMATION ON THE INTERNET	45
ATTACHMENT A - LIST OF SEARCH FACILITIES	47
ATTACHMENT B - THE ISSUES PAPER	50
ENDNOTES	52

EXECUTIVE SUMMARY

With the explosive growth of the Internet in recent years, the problem of “information overload” has become acute. Even the quite sophisticated software tools (“search engines” etc) available to Internet users to find information have become overloaded.

Australian farm businesses are increasingly seeking to use the Internet as a way of meeting enhanced information and communication needs. Many, however, are not experienced users and many suffer from poor telephone lines and hence slow operating speeds on the Internet. Ease of accessibility of information is, therefore, a vital consideration if the Internet is going to achieve its potential in assisting rural Australians to overcome the “tyranny of distance”.

A significant proportion of Australian farm users express considerable frustration at the difficulty of finding information on the Internet. This is notwithstanding the

- large number of directories of World Wide Web sites, many devoted to agricultural sites
- availability of powerful general Internet search engines. These are software tools which “crawl” the Web indexing the content of sites, and then using this catalogue to return a list of sites containing keywords nominated by a user. The search engines are by far the most efficient and effective way to find information on the Internet - and are used by more than 70% as their main means of finding Web sites by frequent Internet users.

Nevertheless, the frustration arises because

- directories cannot take users to specific pieces of information from within sites
- for all their technological wizardry, search engines face three problems in servicing users
 - even the largest do not index all of the total content existing on the Web. Aside from pages that are missed in the indexing process, the search engines are unable to search data bases underlying Web sites, and some Web site design features can hamper the search engines from indexing the information
 - the problem of information overload manifests itself through search engines returning large number of “hits” in response to common keywords. Partly this reflects the sheer number of sites on the Internet, and partly it reflects the software driven nature of the robotic search engines which means that they cannot know what is in a user’s mind when specifying keywords, and so they return far too many irrelevant “hits”. Search engine developers are responding to this problem by providing users with opportunities to “refine” searches, but this has only proved partially successful to date
 - “spamming”, ie, unscrupulous content providers taking advantage of opportunities to artificially boost the prominence of their sites in response to common keywords. Search engine owners have responded through anti-spamming measures, which will never be entirely effective but also can adversely affect legitimate site operators.

Many farm users blame their own inexperience for the problems they experience in finding information on the Internet. This is not always an accurate perception - often the problem is a general one of information overload. This paper presents suggestions and recommendations on ways in which farmers can be assisted to find the information they want and ways in which the accessibility of information itself can be improved.

RECOMMENDATION 1. Create a specialised search engine.

Some developments in search engine technology could hold out significant prospect for reducing the problem of information overload facing Australian farmers if they were to be adopted by one or more of the organisations servicing their information needs.

One in particular would be the development of specialised search engines, ie, web crawlers that operate like the general Internet search engines, but are restricted to only a limited range of sites known to contain information relevant to Australian agriculture. There are surprisingly few such facilities available on the Internet, despite its technical feasibility, and none aimed at assisting Australian farmers.

Such an initiative would remove the problem of irrelevant hits. It would require some development and maintenance expense, but this does not appear to be prohibitive. This is the most strongly recommended initiative in this paper.

RECOMMENDATION 2. Provide training in search techniques.

Training needs of Australian farm users also need to be considered and met. There is considerable demand for appropriate Internet training among Australian farmers, although it needs to be recognised that it is quite possible now for people to learn how to use the Internet more than adequately without receiving any formal training. In recognition of this demand, however, user training, including in search techniques, should be given prominence in the regional initiatives supported by the Regional Telecommunications Infrastructure Fund and elsewhere.

RECOMMENDATION 3. Encourage significant content providers to adopt “best practice” in terms of the accessibility, navigability and marketing of their Web sites.

Content providers have an interest in making their sites as accessible and navigable as possible to their target audiences. Major public sector content providers also have a responsibility to consider how their content provision fits into the operation of the Internet as a whole from the point of view of their customers. Web publishing has come a long way in a short time, with a number of design, presentation and marketing rules and principles having been developed.

The more navigable individual Web sites are, and the better the more substantial sites are presented and marketed, then the easier it will be for users to find the information they want. This is particularly the case for those, like many farmers, with poor communication links. While Web publishing has come a long way in a long time, and some formal standards are beginning to be developed, it is still not entirely clear what constitutes “best practice” in this area. Some attempt could be made at identifying such “best practice”, and promulgating it among Australian agricultural content providers.

RECOMMENDATION 4. Create a directory of the search pages on sites with site-specific search engines.

Many of the more substantial sites have site-specific search engines, which provide more thorough and accurate searching of those sites than is possible from the general Internet search engines. In particular, such site-specific search engines can search data-bases underlying Web sites. Their drawback is, of course, that they are restricted to only the one site. One way of tapping this technology for users more generally would be to create a directory of the search engine pages of the more substantial sites. It may be possible to develop this into a “meta-crawler” which returns results from each of those sites in response to a single request by the user.

This Recommendation overlaps with the proposal to create a specialised search engine (Recommendation 3), but differs from that proposal in a number of respects. On the one hand, it would be limited to sites with an on-site search facility, and so could not include other sites which would be included in a specialised search engine. On the other hand, it would be able to use the ability of the on-site search facilities to get behind the HTML code and into underlying data bases to be able to draw out more substantial information in response to search queries that would be invisible to a search engine.

RECOMMENDATION 5. Support regional communication initiatives.

Farmers experience information overload not only on the Internet but in other aspects of their farm business management. For this reason, information needs to be filtered and packaged. As many information requirements are local, much of this packaging is best done on a local basis.

A number of regional communities are acting to improve Internet service provision and content to their residents generally, including in respect of agricultural information. Some of these initiatives receive government support, such as through the Regional Telecommunications Infrastructure Fund. This support should continue.

RECOMMENDATION 6. Create an aus.agriculture newsgroup.

Internet applications other than passive information provision through the World Wide Web also need to be considered. These include usenet newsgroups, Web-based forums, mailing lists and chat opportunities. There is scope for the creation of an aus.agriculture newsgroup. The best way to do this would be to incorporate the existing Farmwide newsgroups into the usenet system, so that they can be accessed from all news servers without requiring users to log onto Telstra's BigPond server.

Existing forums and mailing lists have not been particularly successful, but a number are performing adequately and should be promoted. Content providers should consider greater use of mailing lists and chat sessions for communicating interactively with their customers.

The strategy embodied in the above proposals would go a long way to solving the problem of information overload on the Internet.

By reducing the barriers to information retrieval on the Internet, they will improve the viability and sustainability of Australian agriculture and regional Australia both directly and through reducing some of the barriers to Internet use in rural areas.

Since the development of the World Wide Web, initially in 1989, content growth on the Internet has been explosive. Estimates are in the vicinity of 320 million Web pages. At recent growth rates, it will not take many years before there are more pages of information on the World Wide Web than there are people on the planet.

Even with all this information, not all knowledge is available on the Internet. Indeed, at one end of the spectrum, many sites provide little useful information at all - while at the other end of the spectrum some large sites provide massive amounts of information. Yet the sheer volume of information that is available creates a problem of “information overload”.

The popular conception is of the phenomenon of “Net surfing” - that is, users following trails of hyper-text links from site to site, picking up interesting information along the way. In fact, only 8% of frequent Web users find sites via “surfing”¹ - given the volume of information available in cyberspace, “surfing” is too inefficient a process of information location and retrieval for anything other than entertainment purposes.

While, some users might confine themselves to a restricted number of sites they have picked up from other sources, most users seriously looking for information, however, use one or more of the directories and search engines available on the Web.² Indeed, a whole industry has grown up dedicated to helping people find the information they seek on the Internet. The technology involved in these search engines is very powerful and truly impressive. Nevertheless, they, too, get overwhelmed by the problem of information overload.

Indeed, it can accurately be said that

*The Internet ... has fallen victim to its own success. In a few short years, it has gone from being the communications province of scientists and engineers to a primary route of information exchange for everyone from financial analysts to fashion designers. So much clutter and traffic snarl the computer networks that the Clinton administration has announced its intention to build a new, separate system - the Internet II - just so that scientists can get some work done again.*³

This paper looks at this issue of finding information on the Internet from the point of view of Australian farmers. Farmers tend to be more serious users of the Internet than others, using it primarily for business related purposes as distinct from entertainment purposes.⁴ Many Australian farmers also have poor communication links, implying slow download times and hence an unwillingness to spend the time required to “surf” from site to site in search of particular information.

Nevertheless, the issues raised in this paper are hardly unique to Australian agriculture.

The paper is based on research undertaken by the author during 1997, particularly on the Internet, and through responses to an issues paper circulated to interested parties in early 1997. The issues paper is reproduced at Attachment B.

The paper is organised as follows. Following this Introduction, Chapter 2 discussed the problem of information retrieval on the Internet in greater detail from the perspective of Australian farmers. Chapter 3 provides more information on the existing system of search engines and directories, and how they can be better applied to helping Australian farmers acquire the information they seek. Chapter 4 discusses the potential for other solutions, including human-based help and information retrieval systems. Chapter 5 provides a concluding discussion of a possible strategy for improving the accessibility of agricultural information on the Internet in the form of a set of options for consideration by RIRDC and others.

2 THE PROBLEM OF INFORMATION RETRIEVAL ON THE INTERNET

No one knows how “big” the Internet really is. Estimates for early 1998 were for 320 million HTML pages accessible to search engines, and growing rapidly. It should be noted that the very concept of the size of the Web is flawed as the number of “pages” in many sites is elastic almost to infinity due to features such as dynamically generated documents, personalised news pages, shopping baskets etc.⁵ To say nothing of the thousands of Usenet newsgroups, mailing lists and Internet Relay Chat (IRC) channels.

Agriculture sites are, of course, only a small proportion of this total. In *The Australian Farmers’ Guide to the Internet*, some 240 Australian Web sites relevant to Australian farm businesses are listed. AgriSurf⁶ lists over 10,000 Web sites relevant to farm businesses. While this is less than 1% of all Web sites, many other sites are, of course, relevant to non-business aspects of farm families’ lives.

Included in these 240 Australian Web sites are a number of sites that perform the function of a directory. Indeed, in another RIRDC study, no less than 24 Australian sites were considered to be directories, that is, their main function is to provide direct hyper-text links to a number of other agricultural sites both in Australia and overseas.⁷ Two of the more prominent of these directories are AgNet and @griWeb. In addition, many other sites perform some of the functions of a directory by providing a large number of hyper-text links. Examples include the Farmwide site and the ABC Rural site.⁸ This leads to the conclusion that Australian agriculture is well served by such directories.

While valuable and worthwhile, directories only assist so far. Directories only list, and provide hyper-text links to, sites. They may provide short descriptions or a few keywords about sites, some provide ratings of sites, and the directory itself may be searchable.

Directories cannot, however, take users to particular content within sites; they cannot describe in much detail the content within each site; and, being compiled manually, they are not capable of keeping up with the explosive growth in Internet content.

The other major search facility available on the Internet are the search engines. These continually “crawl” the Web indexing the content of Web sites. Users type the keywords they are seeking, and are presented almost instantaneously with a list of Web pages containing those keywords, a short description drawn from the pages themselves, and hyper-text links to them.

The result is by far the most efficient and effective way of finding information on the Internet. Yet, notwithstanding this technological marvel, two sets of problems remain (see section 3.2 below for details).

First, not all Internet content is indexed by search engines. Even the largest index perhaps one third of the Web pages in existence, and they cannot search beyond Web pages into data bases that may lie behind them. In short, they leave gaps.

Second, there is what users perceive as the opposite problem, viz, information overload. Searches can result thousands of “hits” being recorded. Many are irrelevant to the subject sought, many are repetitive, and some of the links are no longer active. In many cases, the sites listed may be relevant in general terms to the subject, but still not provide the precise information the user may be looking for.

The following example is a comparison search of 8 search engines for army worms on 5 August 1997.⁹ The results are shown in Table 2.1.

Table 2.1 - Searching the Internet for “Army Worms”

Search Engine	Number of “hits”	Number of Relevant Hits in the First Ten
AltaVista (simple query)	58,970	1
Excite	412,819	0
InfoSeek	317,042	7
Lycos	12,131	0
WebCrawler	21,771	2
Yahoo	Reverted to AltaVista ¹⁰	
HotBot	4,595	0
Answers	1,459	1

As can be seen from Table 2.1, the search engines can find literally tens of thousands of sites in a search. Only the first 200 or so are presented, usually in groups of ten. It is a common experience for there to be no relevant sites in the first ten or more sites presented - this was the case in half the searches quoted in Table 2.1.¹¹

This problem of “information overload” is not merely academic. Many Australian farmers complain about the problems they have searching. Many blame their own lack of knowledge on search techniques, but to many the problem is the inherent limitations of the technology.

Table 2.2 below displays the relevant section of a survey of participants of the Farmwide Online Services Pilot project conducted in early 1997.

Table 2.2 - Experience with Searching for Information (%age of respondents)

	Grain	Beef-Sheep	Other	Total
Have you searched for information?				
Yes	95	92	91	93

No	3	5	7	5
No answer	2	3	2	2
Did you find the answer?				
Yes	69	77	88	74
No	31	23	12	26
Are you happy with search engines				
Yes	49	38	39	43
No	2	2	5	2
Reasonably	28	33	37	31
Not using regularly enough to say	20	23	18	22
No answer	1	3	2	2

As expected, almost all respondents had tried to find some information on the Internet. A substantial majority generally succeed in finding what they are looking for, but a significant minority - 26% - do not.

While few (only 2%) of respondents said they were not happy with the search engines, only a minority (43%) expressed complete satisfaction. Around half of all responses were qualified in some way.

A separate online survey of Australian farmers' Internet requirements was also conducted as part of the project on user needs. 58% of respondents to that survey indicated that they had had difficulty finding information on the Internet, with little difference between commodity groups or according to Internet experience. The most common problem was that of information overload (mentioned by 33% of those who had had problems finding information), followed by their own lack of experience (25%), absence of the information required (19%), and the relatively small amount of Australian information being overwhelmed by American sources (8%). Given the differences in the questionnaires, the results of the two surveys are broadly comparable. Moreover, 6% of respondents to our survey nominated "time wasting" as a problem with the Internet - a problem that would partly reflect search problems.

Interestingly, the 26% of Farmwide participants who generally fail to find what they are looking for is very close to the proportion of all people - approximately 25% - of all people who try the Internet in one year but not in the next.¹²

The Farmwide Online Services Pilot Project conducts online training sessions each Monday evening, including on search techniques. Complaints about the difficulty of finding information are frequently raised in these, and other, chat sessions. The example of the farmer "bumbling around" looking for information on army worms was quoted above. Other comments include:

I have searched the Web on Native title and got 3 million documents, where do I start?

*I find AltaVista gives too broad an area if you know what I mean.*¹³

A Farmwide chat session was conducted on 27 November 1996 as part of the user needs project,¹⁴ during which participants were asked for their search experiences. Examples of comments made were:

our searches have only been partially successful and quite frustrating for specific ag information

we have tried to access all sorts of things, newspapers, weather, farming online and continually get lost

we have great difficulty finding useful info on the net

we spend a long time searching for very little results - disappointing!

We spend a lot of time looking and then the content is usually American or only an advertisement.¹⁵

Problems with searching are not the prime concern of Australian agricultural Internet users - infrastructure issues take that particular prize. This is shown in Table 2.3, which shows Farmwide respondents' answers to a general question on what is required to improve their experience with the Internet. Infrastructure issues are the major concern (nominated by 55% of respondents), followed by user knowledge (32%), and content issues (12%).

Table 2.3 - Key Factors to Improve Experience with the Internet (%age)

	Grain	Beef-Sheep	Other	Total
More stable connection	17	9	11	13
Faster speed	39	49	37	42
More knowledge of system	29	32	33	32
Better search engines	1	2	4	2
More commercial information	13	5	13	10
More government information	1	0	0	1
Ability to buy and sell online	0	2	2	1

Only 2% nominated better search engines as their key requirement. This, however, should not be interpreted as indicating that searching problems are trivial - only that other problems are regarded as of greater significance. Moreover, the significant proportion of respondents indicating greater knowledge as their key requirement would include an (unknown) proportion for whom the greater knowledge required is in relation to search techniques - and who once gaining that greater knowledge would then find that part of the problem is the system itself rather than their knowledge of it. Finally, there is an inter-relationship between infrastructure problems and search problems - the more difficulty people have in finding information, the more significant will be the infrastructure barriers they face in taking advantage of it.

As a result of this difficulty in finding information, many users end up spending considerable time "surfing the 'Net" searching for a relatively simple piece of information. Compared with other, more traditional, means of gaining information the Internet can appear slow and cumbersome (*a fortiori* to people with poor

communication links). In a context where most farmers are relatively satisfied with existing information sources,¹⁶ this can constitute a significant barrier to Internet usage.

Any barriers to Internet usage by rural residents have the potential to restrain demand for Internet services, and hence reduce the commercial incentive to provide services. Such barriers, therefore, reduce the opportunity Australian agriculture has of using the Internet to gain competitive advantage.¹⁷

3 SEARCH ENGINES

3.1 AN OVERVIEW OF SEARCH FACILITIES

It is fair to say that the explosive growth of Internet content has been accompanied by the development of an entire industry devoted to helping users find information. These help facilities fall into four broad categories - directories, search engines, data bases and tutorials.

There are hundreds of such search facilities available. Linkaway lists 292 “search engines” (mostly directories and data bases). Yahoo’s “searching the Web” page lists 132. Worldwide Search Engines Annotated lists 957 such sites and even this includes only one agriculture site¹⁸ of the dozens available - indicating that the total number of search facility sites must number in the thousands.

3.1.1 DIRECTORIES

Directories (or subject catalogues) are manually-compiled lists of sites in various categories, including hyper-text links. Some provide short descriptions or keywords pertaining to the content of sites. A handful provide a “rating” of sites. Some are large and complex, covering a wide variety of subjects; others are small and specialised aimed at a particular group of users.

Yahoo is the best known directory on the Web. It lists an estimated 1 million pages in a subject directory structure. A sentence or so of information is provided on many - but not all - of the sites included. The directory is searchable in a manner that to the user appears similar to the search engines. Where no match is found, Yahoo pipes into a search engine - AltaVista by default.

The combination of such a large directory and the reversion to a search engine makes Yahoo one of the, if not the, most popular search facilities on the Internet. Indeed, it is understood to be one of the few major search facilities returning a profit. It is the most favored search facility among US farmers.¹⁹

One problem with a complex catalogue structure like Yahoo’s is that once users start following one branch of the tree, they lose sight of other branches. Software, called an Information Visualiser, is being developed to allow a “three-dimensional” view of the catalogue structure to appear on the computer screen and so assist users follow the structure.²⁰ This is, of course, much less of a problem for the simple catalogue structures involved in the specialised agriculture directories.

More fundamentally, the categorisation of topics is problematical. It is a moot point whether it is even possible to categorise all knowledge - and regardless of philosophical niceties, categorisation systems imply that the categoriser and the user have the same understanding of the categories.²¹

Increasingly, search engines are becoming hybridised, with a directory included alongside the search facility. A particularly important example of this is Magellan, whose directory includes ratings of many of the substantial sites.

This idea of rating sites in a directory can be taken further, to include articles on topics included in the directory and access to interactive features such as forums and mailing lists on those topics.²²

There are many other directories, mostly covering a specialised subject or region. As noted above, 24 Australian agricultural directories were found in late 1996 - plus other sites fulfilling the function of a directory. Some are listed in Table 3.1.

Table 3.1 - Australian Agricultural Directories

Directory Site	URL
Agnet Australian Agriculture	http://agnet.com.au
Agriculture on the Web	http://www.comcirc.com.au/ag/
AgriWeb	http://www.ruralnet.com.au/AgriWeb/
Australian Agricultural Resource Centre	http://www.comu.net.au/gipsmirror/agri/
Australian Agricultural Sites	http://www.icconnect.net.au/farmers.html
Australian Agriculture & all things related	http://www.benet.net.au/farm.html
AAA Agricultural Announce Archive	http://www.aaa.com.au/Farm_Radio.html
Australian Indexes - Agriculture	http://www.moreinfo.com.au/ausindex/agriculture.html
Australian Web Servers - Agriculture	http://www.csu.edu.au/links/agriculture.html
Cowley's Farm Web	http://www.cowleys.com.au/public/farmweb.htm
Horsenet - Horse Information Service	http://www.horsenet.com.au/
Landcare Web	http://www.agfor.unimelb.edu.au/Lcweb/Lcweb.html
Silo Home Page	http://www.wagga.net.au/~silo
Weather Information	http://atmos.es.mq.edu.au/weather/

Table 3.1 (ctd) - Australian Agricultural Directories

Some Other Sites with over 50 links	URL
ABC Rural page	http://www.abc.net.au/rural/
Adelaide University Faculty of Agricultural and Natural Resource Sciences	http://www.waite.adelaide.edu.au/
AgInfo	http://www.aginfo.aust.com/aginfo.htm
Charles Sturt University Faculty of Science and Agriculture	http://www.csu.edu.au/faculty/sciagr/
CSIRO	http://www.csiro.au/
DPI Queensland	http://www.dpi.qld.gov.au/
Elders	http://www.elders.com.au/Elders/welcome.html
Environmental Resources Information Network	http://www.erin.gov.au
Farmwide Online	http://www.farmwide.com.au
Meat Research Corporation	http://www.dpie.gov.au/mrc/
National Farmers' Federation	http://coombs.anu.edu.au/SpecialProj/NFF/NFFHomePage.html
Pomare Poll Dorset Stud	http://users.hunterlink.net.au/~ddjrs
Rural Press Limited Online	http://www.rpl.com.au/farming/
The Pig Page	http://www.ozemail.com.au/~rroutley/
UNE School of Rural Science	http://www.une.edu.au/~rural-science/RurSciHP.html
University of Western Australia Soil Science and Plant Nutrition Group	http://www.general.uwa.edu.au/u/soilweb/welcome.html

Source: J Da Rin and J Groves, *What's on the web now for Australian farmers?* RIRDC, 1996.

Typically, directories list sites, provide hyper-text links to them, and include a few keywords or a sentence or so on site content. In its "topics" listing, AgNet's arranges the keywords in a way that facilitates a search of the directory using the Edit-Find function in the browser. AgNet does not, however, provide a description of site content due to the time and cost of producing such descriptions and keeping them current, the subjective nature of any review, and the fact that it is relatively simple for users to click on a link and see the site for themselves.²³

Some other sites (eg, Farmwide's) have an on-site search engine searching the directory and other content on the server. Others consider such a facility to be unnecessary because of the logical way in which the subject directory is structured.

Australian Indexes provides a rating of sites, such as "excellent", according to the criteria of usefulness, comprehensiveness and currency. Unfortunately, its agriculture section included only 8 entries as at 5 August 1997 - mostly other directories.

Cowleys FarmWeb is unique in Australia in that it provides a facility for user rating of sites. Users can ascribe "gold", "silver" or "bronze" medals against each of the criteria of content, useability and presentation. The default is silver. However, only a proportion of the sites included in the directory have been rated at all. Moreover, the ratings do not show how many people provided the rating - making the system vulnerable to content providers giving their own sites artificially good ratings.

The only other agriculture directory to provide user ratings is the Linkranker page at the US @griculture Online site. It allows users to rate sites on a scale of one to five, with the average score displayed. It also shows the number of users to rate a site, giving some indication of the credibility of the ratings. However, it only includes those sites which users have submitted for rating, which are only a small proportion of all agricultural sites. 166 sites were listed at 2 July 1997 - none from Australia - many of which had been rated by only one user.

Internationally there are, of course, many more agricultural directories. Examples include AgriSurf, the University of Florida's Agrigator and the University of Guelph's CAFFEINE.

As noted above, it can reasonably be concluded that Australian agriculture is well served by directories. Of course, the people responsible for their development are to be commended. The directories certainly appear to be quite popular with users.²⁴

A variation on the directory theme has been developed by David White of the Bureau of Resource Sciences in Canberra. This looks much like other directories, but rather than being a Web site, it is a piece of software to be downloaded onto a user's hard disk. When online, the user can use this directory to go straight to sites of interest without having to remember their URL, bookmark the site, or go to another search facility to look up the site. Development of this system and making it generally available would be a useful adjunct to existing directory services.

3.1.2 SEARCH ENGINES

Search engines come in three types

- general Internet search engines
- site-specific search engines
- meta-search engines, ie, search engines of search engines.

General Internet Search Engines include AltaVista, Excite, HotBot, InfoSeek, Lycos, Northern Light (the new kid on the block) and WebCrawler. Following Calafia Consulting²⁵, these seven are described as the "major search engines" in the following discussion - although WebCrawler is less "major" than the others.

There are many other search engines, making for quite a competitive market. Some are front-ends to one of the majors (eg, Search.com is a front-end to InfoSeek). With few of the major search engines returning a profit, a rationalisation of the ownership of the major search engines is underway - Excite, for example, has taken over WebCrawler (and Magellan) but so far has continued to operate them as separate facilities.²⁶ One view is that there soon only be two major search engine companies.²⁷ On the other hand, with the daddy of all software marketers, Microsoft, about to enter the market with a search engine currently dubbed "Yukon", the market is about to become even more competitive.²⁸

These search engines use special software, called “spiders”, “crawlers” or “bots” to continually visit Web sites and index their content. Users enter keywords into a search box, and the search engine almost instantaneously returns a list of pages containing those keywords, including a short description drawn from the pages themselves and hyper-text links. This is undoubtedly the most powerful, effective and efficient way of searching for information on the Internet.

Search engines “find” sites to index by following links from other sites and from site submissions by content providers. Once included in the index, most search engines regularly revisit sites for content updating - some visit those sites which frequently change more often than other sites.

From a user’s point of view, the crucial difference among search engines is that an identical search can throw up very different results. While some of these differences can be explained by the different features listed below, the main reason for the differences are the proprietary (and hence secret) algorithms which the search engines use to decide which sites from their index they will display in response to a particular query. An example of the differences is in Table 2.1 above - where the number of relevant “hits” in the first ten ranged from zero for a number of search engines up to 7 for InfoSeek.

In general, keywords contained within a site’s title are rated the most highly; followed by keywords near the top of the page such as in the headline.²⁹ Frequency of keyword use is also important (within limits imposed for anti-“spamming” reasons). Words contained in “meta-tags” (see section 3.4.2 below) are important to some engines, unimportant to others, and completely ignored by yet others. Some “boost” Web pages according to the number of hyper-text links pointing to them; some give a boost to sites reviewed for its directory. All the major search engines downgrade - or even exclude - sites where they detect unacceptable “spamming” techniques used by content providers in an attempt to upgrade the prominence given to the site by the search engines (see section 3.2.2 below).

The list of “hits” displayed to users includes the page title, a description of the site drawn from the page itself, and then the URL. The title and the URL, or both are in hyper-text allowing the use to go straight to the site concerned.

It is not the place of this paper to compare search engines,³⁰ other than to list some of the other differences among search engines, including

- whether all words on a page are indexed. The major search engines provide “full-text” searches, ie, they index every word on a Web page (excluding so-called “stop words”, ie, frequently occurring words that add little to a keyword search); some index only abstracts drawn from the page
- the size of the search engine, ie, the number of pages it indexes. The largest (currently HotBot) is believed to have around 110 million pages
- pages crawled per day - the more pages indexed per day, the more the search engine can keep up with constantly evolving content. The best claim up to 10

million pages per day. Some also display the date when a Web page was added, which gives users an idea of the freshness of the listing

- speed with which submitted pages are indexed after being submitted by content providers. This ranges from within minutes to up to 3 weeks. Most search engines index other pages in a site at a later date, ranging from 1-2 weeks later up to 3 months
- the depth search engines go within sites. Some try to index every page they can find in a site; the others only sample pages in sites. A number cannot index content confined to frames, and some cannot follow links contained within image maps. Both these restrictions limit the number of pages indexed from within sites.

Each of these differences could be put together to deduce which is the “best” search engine. However, as noted above, from a user’s point of view these differences are overwhelmed by differences in the algorithms used to choose which sites to list from the index in response to a particular query. These algorithms are secret, and frequently changed.

While InfoSeek performed the best in the query submitted in the exercise reported in Table 2.1 above, on a number of other occasions AltaVista has returned the most useful responses. For example, in the conduct of the RIRDC user needs study, the question was asked “What do Australian farmers want from the Internet?” AltaVista’s and HotBot’s algorithms focused on the phrase “Australian farmers”, and returned a number of Australian agricultural sites. Most of the others, however, focused on the word “Australian” and so returned a wide range of sites pertaining to something Australian (eg, scouting). Some, such as InfoSeek and Yahoo, returned a particularly eclectic range of sites, often with no apparent connection with any of the words used in the search phrase.

For this reason, user preferences among search engines will be a highly subjective thing, including the extent to which a search engine’s algorithm most closely mirrors users’ individual thought processes, their experience with different engines, familiarity, understanding, appreciation of the various differences across search engines, and the way in which search engines present their results. And these preferences will vary from time to time and even search to search.

This has not stopped some people from attempting to rate the various search engines with a view to providing a recommendation for users.³¹

Because of the ultimate subjectivity of such testing, it would be more useful to directly gauge the opinions of Australian farm users as to the best search engine. Online surveys - which by their nature will be dominated by North Americans - show Yahoo as the most popular search facility (nominated by over 30% of respondents), but with different rankings among the major search engines, ranging from around 5% to 20% each.³²

The only information on the preferences of Australian farmers available to date is the Farmwide survey cited in Chapter 2 above. Table 3.2 compares the level of

satisfaction of AltaVista users with those using Yahoo and other search engines as their preferred search facility.³³

Table 2.3 - Satisfaction With Search Facilities (%age)

	AltaVista	Yahoo	Other
Proportion using first ³⁴	73	13	18
Proportion happy			
Yes	45	47	36
No	2	6	3
Reasonably	31	31	26
Not using regularly enough to say	22	16	36

The overwhelming majority of Farmwide participants use the AltaVista search engine. This was the search engine used during the training participants received, and is the search engine which appears on the Farmwide home page. Hence this cannot be considered an unbiased sample.

There is little evidence here of a significant difference in user satisfaction across the various search facilities. While AltaVista and Yahoo record a higher “yes” vote than the others, this appears to reflect differences in experience levels rather than in satisfaction.

On the other hand, as mentioned above, only 16% of American farmers use AltaVista (plus a proportion who effectively use it via Yahoo) - placing it in first or second place, possibly behind WebCrawler, with InfoSeek third

- the search engines page at *LinkRanker* mentioned above includes six of the major search engines, plus Yahoo and 15 others. Of these major search facilities, the highest rating is given to WebCrawler (average score of 4.2 out of a possible 5), followed by AltaVista (4.04), Lycos (3.8), Open Text (3.57), Yahoo (2.64), InfoSeek (2.6) and Excite (2.5)

In any case, the standard recommendation by experts is to use a number of search engines. For example, Mr Ron Middleton, who conducts the Farmwide training chat sessions each Monday evening, recommends that farmers use three or four engines - AltaVista, Lycos, InfoSeek and HotBot. He regards Yahoo as being for “fun, but not of practical use for serious research” and Magellan as “too restrictive”.³⁵

Site-specific search engines are used on many servers to assist users find the information contained in sites located on that server. These are particularly important for data bases - indeed, the general Internet search engines cannot search data bases that lie behind Web sites.

Some sites provide both site-specific search facilities and a link to general Internet search facilities. An example is the Farmwide site, which uses search software from Digital to provide a search facility for its site, including the farm organisations’ home pages contained on its server, as well as providing a link to Digital’s AltaVista search engine for searching the entire Internet. Digital even makes available a “personal” version of AltaVista for users to download into their PC to search their hard drive(s).

Some of the site-specific search engines are a cut-down version of one of the major search engines - Farmwide's use of AltaVista is an example.³⁶ Others have developed their own search facility from scratch - an example appears to be the Agdex search engine at the site of the Alberta (Canada) Department of Agriculture, Food and Rural Development, while yet others use other software available on the market. An example of the latter is the Harvest system, discussed further in section 3.3.3 below.

Site-specific search engines are, of course, an essential feature helping users to find information on the Internet. Their particular strength is their comprehensive coverage of the site concerned - indeed, they generally go beyond the particular site to include all the sites contained on the server. Features such as frames, image maps, and large data bases lying behind the Web site, are no barrier to an appropriately-designed site-specific search engine - unlike the general Internet search engines.

The ability to search all sites on a server adds to the case for content providers to seek to use the same server as other similar content providers. Examples include the farm organisations use of the Farmwide server, the use of the Department of Primary Industries and Energy's PIENet site³⁷ by a number of Commonwealth primary industry statutory authorities (including most of the research and development corporations, including RIRDC³⁸), and the Dairy Australia site created jointly by the Australian Dairy Corporation, the Dairy Research and Development Corporation and the Australian Dairy Industry Council.³⁹

The problem from the point of view of this paper, however, is obvious - to take advantage of the power of the search engine, the user must know of the site's existence, how to get to it, and that it does contain the information being sought. For this reason, this paper does not consider them further.

Meta-search engines (or meta-crawlers) are search engines that search other search engines. These come in two types - Web-based and personal.

Web-based meta-crawlers operate like ordinary search engines from a user's point of view. Keywords are entered, and a list of "hits" returned. In this case, however, the meta-crawler has no index of sites of its own - it merely automates the process of conducting searches on multiple search engines and, hopefully, removing duplicates.⁴⁰ Agscape Unified Search Interface (which is a meta-crawler as well as a directory) rated relatively highly (4.25 out of a possible 5) at LinkRanker. FireFly is an example of an intelligent agent, which builds up a personal profile of a user's preferences, and can use this to assist other users.⁴¹

While search engines return results to users remarkably quickly, it takes significantly more time for most meta-search engines to search a number of search engines. For this reason, they have not proved popular with users. Moreover, they work best, of course, when the search expression used is common to all the search engines concerned.

Personalised meta-search engines are software that individual users can purchase for use from their own desktop. Like Web-based meta-search engines, users enter their

search query and the software automatically searches a number of the general Internet search engines. Other features include file management to facilitate off-line browsing of Web pages and intelligent agents to build up a profile of the user's preferences. Demonstration versions are generally available for downloading. The most common price for the full package seems to be \$US49.95.⁴² Over 100 products are believed to be available⁴³.

Like Web-based meta-search engines, personalised meta-search engines can be slow to return results. Moreover, they can be difficult to set up,⁴⁴ and \$US49.95 (approx \$A69) appears a large amount to outlay for a service that can be accessed for free. The number of products available, however, would seem to indicate that there must be a market for them.

3.1.3 DATA BASES

From a user's point of view, any site that provided a significant amount of information could be regarded as a "data base". A more narrow definition would be that an Internet data-base is a collection of information, contained in specialised data-base software, lying behind a Web page and accessible via a search facility from that Web page (or the server on which it is contained), but not accessible via a general Internet search engine.

Many of the supposed search facilities on the Internet are, in fact, data bases. Defined in this way, there would be many thousands of data bases - including, for example, most online shopping sites.

Some data bases relevant to Australian agriculture include

- the Environmental Resources Information Network (ERIN)⁴⁵ provides access to 15 data bases on environmental issues. Users can, for example, find what's on the Register of the National Estate in their area or a list of endangered species in their local area
- the Queensland Department of Natural Resources' *The Long Paddock* site⁴⁶ is a data base of climatic information, particularly on the Southern Oscillation Index and medium-range weather forecasts deriving therefrom
- the Sydney Futures Exchange⁴⁷ provides a data base on the prices of the commodities it trades, including greasy wool, wheat, 90 day bank bill rates, 3 and 10 year government bonds, and shares. The Australian Stock Exchange site also provides a searchable data base on share prices
- Wesfarmers Dalgety's site⁴⁸ provides access to a data base of property listings
- Streamline⁴⁹ provides access to the Australian Rural Research in Progress and Australian Bibliography of Agriculture data bases.

As noted above, this narrow definition of data base does not include even quite substantial information provided in HTML format (or other formats such as Adobe's pdf and Microsoft Word) in a number of other sites.

In response to the surveys mentioned in Chapter 2, some farmers expressed a view that the problem of information overload could be overcome through the creation of a “one-stop-shop”, with all information wanted by farmers available at the one site.

This would be a monumental, expensive and ultimately futile task. The diverse and detailed nature of farmers’ information needs are such that it would simply not be possible to collect all relevant information - governing agronomy, marketing, management, animal husbandry, the weather, etc etc - into a single place. An idea of the cost can be gained from the Australian Legal Information Institute, which received \$492,000 in funding to create a data base of Australian statutes and High Court and Supreme Court decisions⁵⁰ - a much simpler task than a “one-stop-shop” for agriculture. Even if such a data base could be created, the resources required to keep it up-to-date would also be enormous.

Rather, what is required is for content providers to provide as much information as possible in as accessible a manner as possible, and for existing and improved search techniques to be available to make the entire Internet into, in effect, the data base so sought. Of course, individual content providers will continue to regard a data base as the most cost effective way of providing information.⁵¹ As one respondent to the issues paper pointed out,⁵² long-term beef experiments, for example, are basically a thing of the past - but there is a tremendous amount of data that has already been collected to develop predictive models. The role of content providers is discussed in section 3.5 below.

3.2 PROBLEMS WITH SEARCH ENGINES

This section discusses some of the problems the search engines have in more detail. These problems relate to incomplete coverage of the Internet, the potential for abuse via “spamming” and the problem of information overload.

3.2.1 GAPS IN COVERAGE

As noted above, even the largest search engines with around 100 million pages indexed only cover around one thirds of the 320 million HTML pages believed to exist on the Internet.

Pages may not be included in search engines for a number of reasons

- some sites may not have been submitted to the search engines and, in the absence of links from other sites trawled by the search engines, may not therefore be found at all. Moreover, content providers can bar search engines from all or part of their site. Content providers may act in this way for a number of reasons: they may regard the site as only in “draft” form not ready for a general “launch”; they may

simply overlook some or even all search engines in their site marketing; they may prefer more targeted marketing;⁵³ they may regard the effort required to submit their site to the search engines as not

- each search engine has a “submit URL” facility for content providers. This requires, however, each search engine to be visited and the form used. There are software aids available which assist somewhat,⁵⁴ but the freeware versions do not cover all search engines, the full version involves a cost, and only some of the time and difficulty of submitting to all search engines is reduced
- some search engines restrict the number of submissions they will accept in this way from any one domain on anti-spamming grounds⁵⁵
- it is impossible for search engines to keep up with the constant evolution of the Internet. Only InfoSeek offers instant indexing - the others take between one day and 3 weeks, and some limit the number of pages from the one site that can be submitted in any one day on anti-spamming grounds. Other pages in the site are then listed for indexing by the “spider” on the normal “schedule” - taking between one week and 3 months. Similarly, sites which are not submitted but are found via links from other sites can take up to three months to be found
 - in addition, overseas “mirror” sites such as AltaVista’s are usually some time behind their main US site. AltaVista’s European site, for example, is up to one week behind the US site⁵⁶ - the Australian mirror site would be similar
 - another limitation to their ability to handle constantly changing sites arises from their “politeness”. A visit to a site by a search engine is very much like any other visit from the point of view of congestion on a server - indeed, they can constitute a significant proportion of server traffic.⁵⁷ For this reason, search engines voluntarily restrict their demands on any one server to, say, one page per minute. This is a good thing from the point of view of minimising congestion, but has the downside that where there are a large number of pages on a server (hundreds of thousands in some instances), it is simply not possible to crawl everything⁵⁸
- some search engines only sample pages within sites - they do not index every page⁵⁹
- a number of search engines cannot index the content contained within frames, while others cannot follow client-side image maps. These include a large number of important sites. There are ways around this for content providers, including the use of meta-tags, no-frames tags and text-based links⁶⁰
- some search engines cannot enter password protected sites, while others can if the content provider provides them with a user name and a password. Content providers may wish to do this in order to let people know what is in their site, even if they cannot access it without a password.⁶¹ As “pay-per-view” sites become more common, this will become a bigger issue

- search engines cannot search data bases lying behind Web sites,⁶² nor non-HTML documents such as Adobe PDF files.

For all these reasons, it is possible for there to be useful content on the Internet which is not found by search engines. However, the major complaint from users is the apparent opposite - too many “hits”. As a result, the search engines have deliberately decided to focus on service - improving their algorithms and maintaining a rapid speed of response - rather than attempting to do the impossible and index every single page on the Web.

3.2.2 “SPAMMING”

Another problem search engines have is that of content providers seeking to artificially boost the placement of their site in response to common keywords. The returns to a commercial site from being in the first ten “hits” recorded compared with, say, the tenth ten “hits” can be very substantial indeed. There is, therefore, a significant commercial incentive for many site authors to seek to “spam” the search engines.

There are, of course, a set of perfectly legitimate design principles which content providers should follow in their site with a view to enhancing both useability and presentation. These are discussed in section 3.5.1 below. It is when these degenerate into “spamming” that the problem arises. “Spamming” techniques are usually dubious, often unethical and sometimes illegal (or ought to be). Unfortunately, the Internet is a medium that brings out some of the worst, as well as some of the best, in human nature. “Spamming” techniques include⁶³

- multiple submission of the same pages, or multiple copies of what is essentially the same page but with trivial differences (such as the URL)
- multiple repeats of certain keywords in an attempt to fool search engines into rating a site more highly against those keywords than it deserves
- seeking to attract visits on the basis of false premises. It is possible to fool a search engine into thinking a site includes content that it doesn’t through text that is invisible to users but visible to search engines, such as meta-tags and/or through using the same colour for some text as for the background
- “bait and switch”, which takes advantage of the fact that search engine “spiders” identify themselves on arrival. This opens an opportunity to provide those spiders with pages tailor-made to the requirements of those particular search engines. A quite different page will be presented to users.⁶⁴

From a user’s point of view, the problem with “spamming” is that it throws up false, misleading and repeated “hits” in response to a particular query. As a result, a constant battle has developed between “spammers” and search engines - a battle in which other, honest, content providers can get caught. For example

- search engines have limited the number of repetitions of the same word that will have an influence on boosting the position of a page against those words. Such limits are inherently arbitrary, and could disadvantage the prominence given to some sites where repetition of some words would be appropriate
- search engines refuse to list sites known to be spamming them, such as from people reporting such sites and those frequently submitted. The problems with the latter are that⁶⁵
 - it has given rise to a particularly pernicious practice, viz, of “spammers” repeatedly submitting their competitors’ sites in an attempt to lower their prominence
 - a number of organisations provide free Web pages (eg, “GeoCities”). These are particularly popular with spammers because of the ease with which they can be used to flood a search engine with multiple submissions. If, however, search engines restrict submission from one domain in order to stop the spamming, this will affect other users of the domain
 - as noted above, some search engines limit the number of pages per site that can be submitted per day, which affects honest content providers as well as spammers
- search engine operators are reluctant to recognise meta-tags because of their potential abuse by spammers. As discussed in section 3.5.2 below, the development of meta-data schema holds out one hope for improving the accessibility of information on the Internet - but this is being retarded by search engine fears over spamming
-
- to prevent spamming via text hidden in the same colour as a page’s background, search engines have started rejecting pages with text in the same colour as the background. This can, however, have unintended consequences - an example is where a table is highlighted in white against a black background, but if black text is used in that table the page may be falsely rejected by the search engines⁶⁶
- it has been alleged that search engines are slow in applying new anti-spamming rules to existing sites in their index, giving those sites an unfair advantage compared with new sites.

It can safely be concluded that this war will never be “won”, but that the battlefield will constantly change. Content providers need to be aware of how these battles affect their site design and presentation.

3.2.3 OTHER PROBLEMS OF INFORMATION OVERLOAD

The difficulty users face as a result of information overload is discussed in Chapter 2 above, viz, the search engines return a very large number of “hits” in response to

searches, with many of the hits being irrelevant and many others being apparently repetitive.

Search engines do try to avoid indexing identical pages - not always successfully. Moreover, they can still be “spammed” through submission of very similar pages.

Quite aside from spamming problems, if one page on a site rates highly against certain keywords, it is quite possible that other pages will rate highly. In the example quoted above of a search for “What do Australian farmers want from the Internet?”, AltaVista returned Australian agricultural sites in the first ten “hits” - but several were from the Western Australian Farmers’ Federation site on the Farmwide server.⁶⁷ There is no suggestion that the Western Australian Farmers’ Federation had “spammed” the search engine - rather, the algorithm used by the search engine simply returned several pages from the one site.

Judging from the more eclectic set of “hits” shown by the other search engines in response to this same question, it would appear that their algorithms try to identify a range of responses in order to give the user more variety to choose from. There are strengths and weaknesses to each approach - if AltaVista gets it right, then it will return more relevant hits than the others; but if it gets it wrong, then it will return no relevant hits at all whereas the other search engines at least give themselves some chance.

The fundamental problem with irrelevant “hits” is that the search engines cannot know what is in the mind of the user when (s)he enters certain keywords. The subtleties of language escape them - they can’t account for variations in word meaning. The techniques of refining searches rely on Boolean theory through the use of “AND”, “NOT”, “NOR” etc between keywords. The bulk of users, who of course have little or no understanding of Boolean theory, are left frustrated by a deluge of hits when a single word is entered.⁶⁸

3.3 SEARCH ENGINE DEVELOPMENTS

In the competitive market in which search engines operate, the search engines have, of course, sought to respond to these problems and so improve their service to users

- with advertising their main source of revenue, search engine operators prime commercial incentive is to maximise the number of people using their search engine and so viewing the advertising. As a result, their focus is on users more than on content providers (other than advertisers) - much to the chagrin of many content providers.

In addition, other Internet users and service and content providers have sought to address these problems in a number of ways.

This sections reviews some of these developments. Adopted pro-actively some of these developments have the potential to go a long way to solving the problem of improving the accessibility Australian farmers have to information on the Internet.

3.3.1 REFINING SEARCHES

The major Internet search engines are making a number of attempts to help users refine their searches. Prominent “help” facilities and “tips” constitute one such mechanism. Field limits (see below) are another. In fact, one of the major forms of competition among the search engines is in giving users more options. Excite, for example, allows users to specify the depth searched within sites⁶⁹ - a useful way of reducing multiple hits from the same site.

AltaVista has introduced a system it originally called LiveTopics, but is now simply a “refine” button.⁷⁰ Following a search, the user can click on “refine”, and a number of keywords related to their initial keywords will be displayed. The user can then decide which of these related keywords (s)he considers relevant, with the list of “hits” refined accordingly

- as at early 1998, this facility was available at AltaVista’s main US site, but not at its Australian mirror site.

The new Northern Light search engine returns not only returns a list of documents, but also a list of related, browsable, topics.⁷¹

Other search engines have introduced a refinement system of a different nature, seeking to exploit the eclectic nature of the “hits” they return in response to a specific query. This is that users are given the option under each “hit” to click on a button which calls up more hits “like this one”. This allows searches to be narrowed quite quickly - more quickly than under AltaVista’s system. On the other hand, AltaVista’s system allows for searches to be widened as well as narrowed.

AltaVista claims that around 3 to 5% of its users take advantage of the “refine” function.⁷² It is not clear whether this indicates the limited value of the feature to users or to the fact that most people find what they are looking for without the need to refine a search.

Another development is that of “channels” in search engines covering general topics such as: arts and entertainment, sports, business and investing, computers and the Internet, careers and education, games, health and science, lifestyle, news, people and chat, politics and government, shopping, travel, and “my channel” customised by the individual.⁷³

In reality, the “channels” operate like a directory. This is one of the ways in which search engines are becoming “hybridised” directories. When a user clicks on a “channel” a number of Web sites relevant to that topic are presented from the search engine’s index. Additional features, such as news stories on that topic, can also be presented. It is expected that these “channels” will become more and more like a directory structure with sub-options within each category provided.

“Channels” are being touted as a way of improving accessibility to information. To date, they are primarily a way of attracting users to particular pages which can be selectively sold to advertisers. Given that most search engines are losing money, a situation which cannot continue indefinitely, revenue-raising gimmicks such as this can hardly be criticised. The reality at this stage is that “channels” do not assist in finding specific information. A search entered from within a “channel” searches the entire index used by the search engine - not just a sub-set of sites considered relevant to the topic of the “channel”.

3.3.2 “PUSH”

A development of the “channel” theme is the so-called “push” technology.

This allows users to specify topics they are interested in, allowing information on those topics to be drawn from specified sites and delivered to the user, including “news” items, site updates etc. The number of “channels” available is rising rapidly, although still well short of Netscape’s announced intention of building up to some 700 “channels”. From a user’s point of view, the “channel” is much like a Web site, but one which is constantly updated to the user without the user going looking for it.

By making information on topics selected by the user available automatically, including information on where further information can be found, this may assist substantially in improving users knowledge of where to find information and in actually providing some of the information they may want. It would not, however, of itself assist in finding a specific piece of information not included in the information received from the “push” channels.

One potential problem is that the number of “push” services around, and the number of channels offered by each, could threaten to recreate the very problem they are developed to solve. It has been suggested that “push has started to become a victim of

its own success”⁷⁴ - when most Internet users have yet to experience it at all! On the other hand, there is scepticism of the value of the technology, at least at this early stage.⁷⁵

There are no agricultural “channels”. Australian agricultural content providers could consider the development of Australian agricultural “push channels” where users can modify the service according to their particular interests and needs.⁷⁶ Given the similarity to a Web site, this is really a matter for content providers to decide whether they want to take their Web site this stage further.

3.3.3 SPECIALISED SEARCH ENGINES

A number of specialised features exist in various search engines. For example

- WebWombat searches Australian and New Zealand sites only, which is useful for people trying to avoid the great mass of overseas, particularly US, content. A similar result can, however, be achieved in, say, AltaVista by typing +host:au in a search query
- What’s New Too only provides indexes on sites that have been submitted in the last seven days
- a number of search engines allow separate searches of newspaper and magazine sites⁷⁷
- a number of search engines can search usenet newsgroups (see section 4.1 below)
- Forum One searches discussion forums
- Web Seer searches for graphics.

A particularly promising development is the development, still in embryonic stages, of search engines that specialise in particular topics. These are search engines along the lines of the general Internet search engines but which search only a restricted number of sites known to contain content relevant to a particular topic.

Argos is described as a “Limited Area Search Engine”, or LASE.⁷⁸ In this particular case, it is designed to search for sites relevant to the study of the ancient and medieval worlds. Argos searches only a small set of Associate Sites, and all the pages they link to (with the exception of pages it is told to avoid). An Editorial Board decides on which sites are included as Associate Sites, with the content of those sites and their links then the responsibility of the content providers concerned. As well as restricting the search area only to sites relevant to the topic concerned, this also serves a quality control function. Because of the limited number of Associate Sites, it is possible to keep the index fresh.

Argos is being developed at the University of Evansville. The Web site says that it’s software will be made available to others once development is completed. A mailing list on the issue of Limited Area Search Engines is available - although there has been few messages on it.

MACISS (Meteorological and Climatological Information Server Searcher) is a search facility based in the Atmospheric Science unit at the School of Earth Sciences at Macquarie University. It is based on the Harvest system. It collects information from servers thought to contain meteorological, climatological or atmospheric science content. In some cases only single documents are searched, in others an entire server. The database currently contains URLs from over 600 servers. URLs are collected both manually and automatically from a range of sources, such as mailing lists, newsgroups, links from sites included in the data base, and those suggested via a form.⁷⁹

The Harvest system is already used as a site-specific search engine by a number of organisations in Australia, including on the Commonwealth Department of Primary Industries and Energy's PIENet. It was developed at Colorado State University in the USA, and imported into Australia in 1996 in a pilot project by a number of universities. While most (indeed, almost all) users use it solely for searching their own server, it can be extended to search other servers. Indeed, it can be used as a general Internet search engine - with the advantage of being much less demanding of bandwidth.⁸⁰ Like other search engines, it follows all the hyper-text links it finds, subject to stops that can be imposed.⁸¹ More details on Harvest are available at its Web site,⁸² including examples of applications.

A specialised search engine along the lines of the above would provide substantial improvements in the accessibility of agricultural information on the Internet. Users would go to the site containing the search engine, and search for agricultural information in the same way as with a general Internet search engine. The specialised search engine would, however, search only those sites known to contain relevant and useful information. For information relevant to Australian farm businesses, probably one or two thousand sites world-wide would qualify. This would go a long way to eliminating the problem of thousands of irrelevant "hits". Desirably, the search engine should also be able to search the limited number of agriculture newsgroups and forums attached to Web sites.

Harvest is not the only option available - it is understood that most search engine software could be adapted in this way.⁸³

Some of the issues that need to be addressed include

- which organisation(s) would fund, create and maintain the facility. Argos has been built from the ground up - a repeat effort would obviously be expensive. Harvest requires a Unix system for its servers,⁸⁴ implying an organisation of some substance
 - a commercial organisation may be able to fund the necessary development and maintenance from advertising. Moreover, an organisation such as Rural Press would have a marketing advantage given its newspaper and magazine circulation and the prominence of its existing Web site, plus the advertising arrangements it already has to fund that Web site
 - however, the risks may be too great for a commercial site to consider - there is a strong "public good" element to the proposal and hence a strong

case for a public sector site. An obvious candidate is the Department of Primary Industries and Energy, or one (or a consortium) of the other Commonwealth authorities using PIENet (such as RIRDC), given that PIENet already operates Harvest as its site-specific search engine

- the cost of development and marketing the facility. It is claimed that a site with httpd already installed should be able to get Harvest installed and running in about 30 minutes,⁸⁵ but in starting from scratch a few days work seems to be required. Some effort would be involved in gathering site information, maintaining the site information and in monitoring usage and acting upon user feedback - but it appears that this would be much less than one full-time job
- some marketing effort would be required - the purpose of the exercise is to make agricultural information more accessible to users, so this purpose is defeated if users do not know about it. A marketing program would include
 - announcement in the usual way once the facility is up and running
 - seeking marketing support from sites in the index, which would be readily given in most cases.⁸⁶ If necessary, some support (eg, a link) could be made a condition of inclusion in the index
 - regular marketing through announcements in relevant newsgroups, forums and mailing lists. An e:mail address list could be maintained for announcing significant changes to the facility. These could be an ongoing responsibility of the person nominated to maintain the currency of the index of sites
 - submission to all the general Internet search engines and directories and Australian agricultural directories. The page containing the facility should be well designed from the point of view of search engine prominence (see section 3.5.1 below)
 - ongoing marketing by the organisation responsible for its maintenance - eg, through newsletters etc.

In all, this is simply a matter of an organisation with an interest in providing information going beyond its own area of responsibility, and with the wherewithal, to call for expressions of interest in the IT industry on the development of such a facility. It just needs someone to do it!

3.3.4 OTHER DEVELOPMENTS

Given the competitive environment in which search engines operate, and the well-recognised deficiencies as discussed above, it is not surprising that technological progress is continuing.

As noted in section 3.2.1 above, existing search engines cannot search data-bases that may lie behind Web pages in the more substantial sites. AltaVista has, however, announced a plan to develop a system that will be able to search such data-bases.⁸⁷

A search engine which searches only meta-tags has developed by the Distributed Systems Technology Centre. This could become significant if the meta-data initiatives described in section 3.5.2 below become widely accepted.

Another development is a Hyperlink Search Engine which indexes hyper-text links pointing to site rather than the sites themselves.⁸⁸

Another set of developments in search engine technology relate to attempts to overcome the deterministic approach of Boolean theory.⁸⁹ These systems go beyond basic text-string matching to use probability techniques and knowledge-based and natural language search methods. Examples of these techniques include

- automatic summarisation functions take each hit from a query and automatically generate a coherent summary
- clustering returns results of a general search in individual subcategories
- relevance ranking uses techniques such as word frequency, the distance between multiple query words, and position in a document where a word is found to assess the probability that the document concerned is relevant to the query
- co-occurrence looks in the general area close to where a query term is located and creates a list of additional words with a statistical relationship to the original query word
- Query-by-Example uses a document vector method to calculate similarity to a selected document or paragraph
- techniques to index and retrieve unstructured data such as video and audio files and digitised pictures.

Many of the products being developed using these techniques have the corporate Web site and Intranet in mind, and so are primarily site-specific search engines. Some of the techniques, however, have been applied by the general Internet search engines to produce the results discussed in section 3.3.1 above.

3.4 USER TRAINING

As noted in Chapter 2 above, many farm users of the Internet consider their problems in finding information to be due to their own lack of skill and experience. No doubt this is true in many cases. This section discusses some of the issues involved in improving user training.

3.4.1 ONLINE TUTORIALS

Web browsers include a number of features to get people started on searching the Internet. Netscape Communicator, for example, has buttons called “Search”, “Guide”, “Lookup”, and “New & Cool” which can get people started in navigating the Web.

The “Search” button calls up the Net search page where users can directly query one of several search engines. Four search facilities are listed at the top of the page, each on its own tab. Each of the four appear in rotation.⁹⁰ A fifth “customise” tab allows users to add a fifth engine from a list of five other search engines. The page can also be made to highlight the same search engine each time it is opened rather than rotating. Further down the page, are a short list of both these and other search facilities, including AltaVista, and hyper-text links to them. The search engines pay Netscape a considerable amount (on average it is believed to be about \$US5 million per year each for those that appear at the top of the page, according to a formula of the number of users who click on each) - after all, Netscape is used by an estimated 60% of Web users and so prominence on its Search page is of considerable value to the search engines.

Similarly in Internet Explorer, pressing the “Search” button brings up a page where search engines can be directly queried.⁹¹ Presumably this will change once Microsoft’s search engine becomes available.

In either case, users who want to use a different search engine, or one of the regional mirror sites such as the Australian mirror site for AltaVista, have to find it for themselves.

Many users would confine themselves to using the Search button - indeed, many incorrectly say they do a “Netscape search” when in fact there is no such thing.⁹² Many others would add their favorite search engine(s) to their “bookmarks” (“favorites” in Internet Explorer).

Netscape’s Search page provides a help facility on how to use the search engines. Each search engine also provides a help facility, and some (like AltaVista) provide regular and prominent tips on search techniques. The best advice anyone can give a user who is unfamiliar with searching or having problems is to read these help sections!

Most search engines use a simple Boolean schema for searches. Some use quotation marks for words that are required to appear together (eg, “army worms” in the example quoted in Chapter 2 above). Some search engines regard a space between words not included in quotes as the same as AND (ie, both words must appear somewhere in the text - not necessarily together), others regard it the same as OR (ie, either word can appear). Generally, the operator AND is used for words that must appear, NOT for words that must not appear, and so on. “NEAR” requires words to be close to each other in the text. Confusingly, AltaVista requires plus and minus signs for its simple query form, but Boolean AND and NOT operators for its advanced query form.

Another set of search techniques which many users are unaware of are the field limits that can be placed on a search. In AltaVista, for example, fields can be limited to title, URL, link, text, image, language, or country of origin. A limitation to the title field,

for example, is achieved simply by typing title: before the keywords. A search can be restricted to Australian sites by typing +host:au after the keywords.

Other relevant issues to searching include case sensitivity, use of parentheses, wildcards (eg, in AltaVista, agricultur* covers agriculture, agricultural and agriculturalist), strings (eg, farm should return farm, farmer and farmers), and multiple spellings (eg, organisation and organization).⁹³

When a useful site is found, users should then “bookmark” it (the Netscape term - it is called “favorites” in Microsoft Internet Explorer) so they can easily return to it (and easily find its URL in order to tell others!).

Just as there is a plethora of search facilities, so there is a plethora of tutorials and tips on searching for users. Some are included in the table at Attachment A.

A usenet newsgroup has recently been established on this subject - comp.infosystems.search - along with a Web and Internet Search Engine FAQ (frequently asked questions).

One interesting gimmick is WebCrawler’s “search ticker”. It shows in a continuous stream the searches being conducted on WebCrawler at any one time.⁹⁴

There are even courses available over the Internet on how to use the Internet, including search techniques. An example is LearnSkills’ courses, such as a “Hands on Internet” course, which costs \$US54 for a 6 week course delivered electronically.⁹⁵

3.4.2 OTHER TRAINING ISSUES⁹⁶

A number of organisations provide user training for farmers (and others) on using the Internet. Examples include the Queensland Department of Primary Industries, the Victorian Government, universities such as Central Queensland University as part of its project on information and communication systems for the meat industry and the University of New England, commercial providers, community groups, telecentres, and Farmwide. This is in addition, of course, to more formal training courses through vocational education and training institutions.

Many of these training sessions involve either

- awareness raising, providing participants with a brief taste of what’s available and readily access can be provided
- more detailed, highly specific, training on the mechanics of using particular Internet software.

The main gap in this “system” is a dearth of ongoing support. In most cases, this is left to informal arrangements between individuals, or between individual users and their Internet Service Providers. It can be argued that Internet Service Providers should do more in this area⁹⁷ - but this is a competitive industry so the greater the user support, the greater will be the cost to the user.

A more complete service is provided to participants in the Farmwide Online Services Pilot Project. In addition to initial training sessions, they have access to help lines and regular Monday evening training “chat” sessions, with searching for information a common topic. While these sessions are accessible to anyone, e:mail notification of forthcoming topics is provided only to project participants. Even this more complete service is regarded as deficient in that it has not, to date, provided follow-up sessions to the initial training session.

It should be noted that Internet software has reached a stage of user-friendliness that it is now possible for even computer novices to reach a stage of reasonable proficiency in remarkably short time. It can be argued that the importance of training may be overrated, as many users are succeeding with minimal formal assistance.⁹⁸

Nevertheless, there is an appropriate place for training programs to at least start people off with confidence. As suggested in Chapter 2 above, farmers clearly demonstrate a demand for appropriate training, including in search techniques.

There is no suggestion of any national program of training in the Internet as this would risk substituting an excessively expensive centralised approach for the substantial efforts being made by a number of organisations and individuals to meet more specific local needs. Support for such initiatives may, however, be an appropriate use of some of the new Rural Telecommunications Infrastructure Fund resources - but such support should only be provided in partnership with delivery organisations. It should be recognised that a number of major content providers (such as State Agriculture Departments) and commercial operators have an incentive for, and an existing role in, the provision of training without Commonwealth assistance.

RIRDC and Farmwide have co-operated in the publication of *The Australian Farmer's Guide to the Internet*, which will assist those farmers interested in a book on these issues. An online version has been created for purchasers of the book, which is being regularly updated.

As well as the myriad of books to assist users with the Internet generally, at least one Australian book specifically on search techniques has been published.⁹⁹ Even among frequent Internet users generally, some 3.3% use printed Internet directories to find Web sites.¹⁰⁰

One of the issues that needs to be addressed in training provision is on minimising time on-line,¹⁰¹ particularly for a greater focus more on search techniques. Even a small amount of training can go a long way in assisting users to find the information they require, and thereby helping them to make the best use of the technology.¹⁰²

3.5 THE ROLE OF CONTENT PROVIDERS

Content providers¹⁰³ do, of course, want to attract users - or at least users from among a target group - to their sites. Hence it is in their interests to make their sites as

navigable and informative as possible. For a number of reasons, however, they may fail to achieve this aim.

More broadly, organisations - particularly public sector organisations with a charter to provide information to farmers - have a responsibility to go beyond their organisation's immediate institutional interests to consider how to best make the Internet as a whole work from the point of view of their stakeholders.

3.5.1 DESIGN, PRESENTATION AND MARKETING

Web publishing has come a long way in a short time, and a number of design "principles" have developed to assist in enhancing the prominence and navigability of Web sites. The use of site-specific search engines has been discussed in section 3.1.2 above. This section looks at other design and marketing issues.

Not all Web sites adopt the various principles, however, for a variety of reasons. These reasons include deficiencies in Web authoring software tools and ignorance on the part of the Web designers that may be hired to design an organisation's Web site. A survey of content providers for *The Internet Organisation* showed that only half (52%) had had their site designed by someone with experience in communicating with Australian farmers. In particular, it can be argued that Web designers tend to over-emphasise attractive graphics and similar features, often at the expense of people with difficulties of access, such as poor quality telephone connections.

Some of these principles have been set out in "Universal Specifications Accessibility Standards for Web Design by the NSW Attorney General's Department (AUS Standards).¹⁰⁴ These standards are designed for people with disabilities, those lacking multi-media computer functions, people using non-current Web browsers, and people from non-English speaking backgrounds.

People, such as many farmers, with poor communication links are not explicitly mentioned - but many of the standards would be equally relevant to them. Indeed, around 20% of Web users are believed to use the "graphics off" option in their browser - this is likely to be the case for an even higher proportion of farmers given that so many have poor communication links. Of the 19 AUS Standards and recommendations, 8 are of particular relevance to people with poor communication links (or, for that matter, users generally). These are

- the provision of ALT tag (ie, alternative text) for all graphics on a website, and where a particularly lengthy description is required, a D-tag (ie, a link to a separate page which describes the graphic)
- including within link text a description of what is at the other end of the link. This is of obvious value to the user, and could become more important if the hyper-text search engine mentioned above becomes popular
- provision of an alternative text list of hyperlinks contained in an image map. This also avoids the problem with those search engines that cannot follow links contained in image maps

- in frames, the provision of an alternative non-framed version of the content. This is not only of value to people with browsers that cannot view frames - it is also important to allow access by those search engines that cannot support frames
- defining all colour schemes in the body tag. Otherwise, the colours shown could be a mix of the pre-defined colours and the defaults from the reader's browser, creating the possibility of a low, or even unreadable, contrast
- ensuring that background colours and images contrast well against foreground text
- supplying a transcript or a description of the sound for all audio sound files
- if using non-standard HTML formats or higher level features such as javascript or java, providing an alternative text-only page.

A large proportion - if not the majority - of sites would not fully comply with these standards, although most would largely comply. For example, in *What's on the web now for Australian farmers?* we found 5 sites with poor or marginal readability because of background colours or designs. And even the Australian Prime Minister's site has been deficient in respect of the requirement for the use of ALT tags for images!¹⁰⁵

There are a number of other design features that should be observed to the extent possible by content providers concerned about users with poor communication links. These were set out in *What's on the web now for Australian farmers?* They include

- page length. While this is a subjective judgement, an optimal page length appears to be about 4-5 screens (of an average PC) in length. Shorter page lengths may create too many links for people to "drill down" through before finding information; longer page lengths take too long to load
- minimising the use of graphics, and graphic size. "essential" graphics - ie, those necessary for navigation without text-based alternatives, are particularly undesirable
- minimising the use of plug-ins and other software required to be downloaded
- minimising the number of pages users are required to "drill down" through before reaching information. Some sites provide too many successive pages of menus.

Another option for content providers to consider is the use of customisation technology. This involves user's indicating to the Web site their main interests and, through the placement of a "cookie" in the user's PC, the Web site recognising that user on returning to the site and delivering content "on the fly" relevant to those interests. Examples of such technology in operation in agricultural sites today are in the Canadian CAFFEINE directory and the US Farm Journal site.¹⁰⁶

Not only must sites be designed in a way that is comprehensible and navigable, they must be promoted, and be designed to be promoted.¹⁰⁷ Some of the ways to effectively market a Web site to the general Internet audience include

- submitting the site, and important pages within it, to as many search engines and directories as possible. This can be time consuming, even with the assistance of services such as Submit-It.¹⁰⁸ Nevertheless, with an estimated 71% of frequent Web users using search engines to find sites,¹⁰⁹ this is the most important way of marketing an Internet site

- including design features in a site that assist search engines in locating it and rating it properly (without “spamming”). Such features include a descriptive title on each page, use of description and keyword meta-tags and including a comprehensive list of keywords in the meta-tags (including plurals,¹¹⁰ commas and mis-spellings¹¹¹), a well worded summary of each page near the top of the page, text at the top of a page,¹¹² content and links in HTML (as distinct from graphics, which cannot be read by search engines), and well written text with an appropriate usage of keywords¹¹³
 - there are software agents available to test how well a site is placed by various search engines against different keywords¹¹⁴
- trying to arrange links from other sites, on a mutual basis. Some 8.4% of frequent Web users find sites via such links, plus the 8.1% who find sites via “surfing”.¹¹⁵ At the basic level, this simply involves contacting other relevant sites seeking mutual linking - much like submitting to a directory. More developed versions of this are banner exchanges and “web rings”
 - a banner exchange is an arrangement whereby participating Web sites submit banner advertisements¹¹⁶ on a mutual basis. Banners rotate every time a user accesses a participating Web page, with each Web site’s banner appearing on other participating Web sites in proportion (50%) to the number of times other sites’ banners are accessed on its Web site. The remaining appearances are used for advertising purposes by the operator of the banner exchange. The service is provided for free. Examples of banner exchanges are Cowley’s Australian Banner Bank,¹¹⁷ the Link Exchange,¹¹⁸ and Free 4 All¹¹⁹
 - “Web rings” are groups of sites with a common interest. Numbers of participating Web sites in each ring ranges from a handful to several hundred. Users traverse the ring by clicking on one of a number of options (eg, a “next site” link) provided by the central ring database at each participating site. Thousands of web rings exist on the Internet. A search under “agriculture” on 5 August 1997, found 7 relevant rings - but even the largest, called The Green Earth, had only 19 participating sites¹²⁰
- regular participation in relevant newsgroups, forums and mailing lists, with the organisation’s Web site address (URL) included in the “signature” at the bottom of each message posted. As long as relevant and topical messages to the newsgroup - which can include announcements of significant changes to the Web site - are posted along with the signature, this is not regarded as “spamming”
 - the Queensland Department of Primary Industries markets itself as well as providing a service for users through the “aussie-xxxx-nets” it has created. The same can be said of Rural Press’ F@rming Online site, with its user forums. Some individuals from other organisations (eg, State agriculture departments) participate in newsgroups and mailing lists and so implicitly assist in enhancing the Internet profile of their organisations - although this appears to be an ad hoc process

- using non-Internet promotion opportunities, eg, including Web site addresses in all print materials produced by organisations. Even among frequent Internet users, 8.5% of them most often found Web sites from newspapers and magazines.¹²¹

A number of the above marketing devices (eg, banner exchanges) would be of more value to commercial sites than to public sector information sites. Indeed, it can be argued that it would be inappropriate for a site to attempt to link to all other relevant sites.¹²² Nevertheless, all sites containing valuable content should be marketed as well as possible in order to assist users find the information contained within them. The question for this exercise is how to encourage providers of the more useful content on the Internet to adopt those techniques relevant to reaching their target audience. One of the aims of *What's on the web now for Australian farmers?* Report was to raise some of the issues involved.

Options (not mutually exclusive - indeed potentially complementary) include

- an unofficial, polite, “audit” of sites
- development of a “how to” guide for farm sites.

3.5.2 META-DATA

The boffins responsible for the creation, development and maintenance of the Internet and the World Wide Web as an accessible open-architecture environment attuned to the needs of users have, of course, recognised the problem users have of finding information. Their proposed solution is the development and implementation of meta-data schema.

Meta-data is data about data.¹²³ A library catalogue card is the classic example of meta-data. In the Internet context, meta-data is information about a Web site contained in meta-tags and invisible to a visitor to the page, but visible to an appropriately set up search engine. The aim of a standard metadata set is to ensure consistency in the description of Internet information, and so facilitate more efficient searching and retrieval of Internet resources. It would help overcome the inability of search engine crawlers to understand the context or genre of a Web site. It would also overcome the inability of search engine crawlers to understand the nature of graphics - which provide, after all, one of the major attractions of the World Wide Web.¹²⁴

Such a scheme needs to be interpretable by both the search engines and by human beings, and it must also be simple to create so that any Web page author may easily describe the contents of their page. This effectively rules out some of the more sophisticated library-type cataloguing schema from general applicability.¹²⁵ Nevertheless, it is important that such schemes be accessible to others, and hence that various meta-data schema relate to each other.¹²⁶

The most prominent meta-data schema under development is the “**Dublin Core**”. This is a system of 15 meta-data elements - such as subject, title author, publisher, date - each of which may be further extended by use of qualifiers. Description under each element would be undertaken by content providers themselves, (hopefully) using standard definitions. Documents would not have to use all elements. The system is still under development,¹²⁷ although some applications projects are underway, including by Education Network Australia (EDNA).¹²⁸ Moreover, as noted above, an experimental metadata search engine based on the Harvest system discussed above has been developed by the Distributed Systems Technology Centre.

There are two obvious problems with the use of meta-data such as the Dublin Core scheme from the point of view of making information more accessible

- meta-data will only be useful in respect of those sites which incorporate meta-data into their sites in an honest (ie, non-“spamming”) way
- while representatives of several search engine producing companies are involved in this work, no search engine is yet able to read all the various meta-tags involved in implementing the Dublin Core, for example, and some ignore meta-tags altogether. The best recognise only site description and keywords. The scope for abuse of meta-data via “spamming” is the single greatest difficulty in getting the search engines to adopt the arrangement, and without the search engines adding meta-data does not assist users to find the site (rating systems such as PICS are more likely to be adopted because of community pressure on filtering grounds).

Despite the optimism of the developers of meta-data systems, this problem of mutual co-operation among content providers and search engines could prove to be fatal. One expert believes that there is “very little prospect” of the search engines recognising meta-data schema - that their main use will be within Intranets.¹²⁹ Similar conclusions have been reached by other experts

it could “end up as another example of a great idea cooked up by technologists living in a vacuum.”¹³⁰

and

some organisation freaks invariably suggest that users of the Information Marketplace should organise information as they generate it, according to certain present rules. This is fruitless because people won't bother to comply.... This resistance to front-end taxonomies is a consequence of human nature. It is also a valued characteristic of the Internet and Web's unordered egalitarianism. Any attempt to enforce organisational rules will surely cause users to flee unless the front-end organisation provides great value to the users.¹³¹

Another schema is the Platform for Internet Content Selection, or PICS, under development by the World Wide Web Consortium (W3C), which is the central agency responsible for recommending World Wide Web standards. PICS was

originally designed as a series of technical specifications for labeling sites for content (or the absence thereof) such as violence and sexually explicit material in order to assist in the filtering of such content - a number of PICS-compatible filtering software packages are available. It is possible, however, to imagine other uses for labels, including for search purposes.¹³² Moreover, using the scheme for search purposes is regarded as a useful inducement to content providers to adopt the system for filtering purposes (the other inducement is that some filtering software provides users with a “block unrated sites” option). As with other meta-data, labels would be included in meta-tags within Web documents. Typically, content providers choose a self-labeling service, and have the site labeled by filling out an on-line questionnaire. Sites can be rated independently (ie, without the content provider’s co-operation), and the resulting labels sent to a label bureau, which is checked by the filtering software.

It is likely that this labeling issue will primarily be confined to issues of censorship - using the labels requires some effort on the part of users (ie, adoption of appropriate software and, if customised options are sought, the configuring thereof), which may well restrict its applicability outside of the censorship arena.

3.5.3 OTHER IMPROVEMENTS ON THE DIRECTORY IDEA

The above discussion on the role of content providers focuses on how content providers can improve the accessibility of the information on their own sites. However, both Australian agriculture and the Internet have strong traditions of public-spiritedness. Many organisations, most obviously public sector, community and industry organisations, consider that helping users to find information from other sources is part of their responsibility to users.

Most organisations discharge this responsibility by providing hyper-text links to other Web sites - to the point where quite a number of Web sites have become directories, in addition to those organisations whose quite explicit aim is to provide a directory of agricultural sites.

The discussion to date has raised a number of possible improvements to the basic idea of a directory of Internet sites, of which there are several serving Australian agriculture. These ideas include providing additional information on site content, rating sites, and the addition of a specialised search facility which searches only sites known to contain useful content for Australian farmers.

A further development of the concept of a directory would be for a directory that links to the search page on those sites with a site-specific facility. As well as assisting users to find information from those sites, this would also serve as a signal to users of which sites were of sufficient substance to contain a site-specific search engine.

This could readily be taken one step further by having the user enter the search query from the directory itself.¹³³

This could be further developed into a specialised meta-search facility. That is, that users could enter their search query at a single point, and the facility would search all the site-specific search engines automatically

- this is not the same as the specialised search engine proposed in section 3.3.3. The specialised search engine would search all specified sites, not just those with a site-specific search facility of their own. On the other hand, by using site-specific search facilities, the meta-search facility will be able to go beyond HTML text into the data bases lying behind Web sites. The two proposals are, therefore, alternative but not mutually exclusive options to achieve similar but not identical ends. They could, therefore, be developed and implemented by separate organisations or by the same organisation. The same issues of responsibility, cost, development etc would arise for this proposal as for the specialised search engine proposed in section 3.3.3.

3.5.4 REGIONAL INITIATIVES

Many of the information needs expressed by farmers' are for local information, much of which could only realistically be provided on a local basis. Moreover, regional initiatives have the potential of integrating the demand for local information, training, ongoing user support, and aggregation of demand required for the provision of services on an economic basis.

Many regional communities have taken the initiative to develop local Internet-based information systems. These have received government support from Commonwealth programs such as the Telecentres program and from State Government programs in States such as Victoria, Queensland and Western Australia. Further support is available from the newly created Regional Telecommunications Infrastructure Fund.

Many ISPs and other content providers find that the provision of useful local information is a good way to attract business. Indeed, current trends in commercial Web site developments are to provide "gateway" pages providing good information for free¹³⁴ and involvement in "virtual communities".¹³⁵ There is, therefore, scope for collaboration between commercial operations, and community and government organisations in this area.

The Centre for International Research on Communication and Information Technologies (CIRCIT) in Melbourne has, with RIRDC support, recently published a framework for developing regional communications initiatives, which aims to guide interested regions through the process of successfully developing such systems. The elements of the framework include building the local organisation, setting the vision, identifying the needs and developing a strategy for meeting them, choosing the infrastructure, arranging finance, assessing feasibility, implementation and evaluation.¹³⁶

One option for regional organisations would be the appointment of a regional co-ordinator for the development and presentation of local information and for assisting

users to find both this local information and to access national and international sources.¹³⁷

Marketing this framework to local communities should be a major objective, not only for assisting farmers to find agricultural information, but for meeting a wide variety of local information and communication needs.

A related exercise might be to investigate the various regional services provided now, with a view to providing an overview of available community information in non-metropolitan areas and hence encouraging more widespread adoption of some of the better ideas currently being implemented.

4 OTHER ASSISTANCE FOR FINDING INFORMATION ON THE INTERNET

Chapter 3 focuses on what are largely technology-based solutions to the problem of finding information on the Internet. This Chapter discusses some other, more people-based, options.

4.1 NEWSGROUPS, FORUMS AND MAILING LISTS

As well as looking for information contained on the Internet, users can also use the communication applications of the Internet to ask for information of, and exchange information with, others. Relevant applications include usenet newsgroups, forums, e:mail based mailing lists, and Internet Relay Chat sessions.

Of course, it cannot be guaranteed that an answer will be provided in a timely manner, or accurately, by any of these means. Indeed, by being asynchronous, newsgroups and mailing lists cannot provide answers instantaneously. Nevertheless, the level of co-operation in a number of these groups is such as to be the beginnings of the creation of “virtual communities”. Certainly answers are provided sooner or later to most relevant questions. The accuracy or usefulness of those answers? Well, this is a case of *caveat emptor*.¹³⁸

A number of **usenet newsgroups** are available on agriculture, including sci.agriculture, sci.agriculture.beekeeping, sci.agriculture.poultry, sci.agriculture.fruit, sci.agriculture.ratites, alt.agriculture.fruit, alt.agriculture.misc, and alt.sustainable.agriculture. These are reasonably active groups, which some Australian farmers have used to their advantage.

A number of the major Internet search engines, such as AltaVista, can also search usenet newsgroups. In addition, Deja News is a popular search facility which specialises in searching newsgroup discussions. Given the thousands of newsgroups in existence, many carrying quite heavy daily traffic possibly numbering as much as hundreds of messages per day, it is possible that sought-for information, or advice of a Web site that contains such information, will already have been provided to a newsgroup. Indeed, many newsgroups have “Frequently Asked Questions” and answers (FAQs) posted on a Web site. FAQs generally set out guidelines for using the newsgroup, but they also frequently provide answers to questions of substance that have been asked in the newsgroup.

Like all newsgroups, they are subject to the problem of spamming (ie, the posting of irrelevant messages - usually for purposes of advertising or dubious money-making) and the fact that any fool can post any rubbish, and does. Some groups are moderated - ie, a “moderator” vets all messages. This allows the problem of spamming to be eliminated; the moderator usually has also some, limited, responsibility for ensuring that messages seeking information do in fact get a response.

However, the only newsgroups specifically devoted to Australian agriculture are the groups operated by the Farmwide project. The groups listed above are - like most

generally-available things on the Internet - dominated by North Americans. The Farmwide groups are open to anyone, but users are required to log onto Telstra's BigPond news server to access them. While not difficult, this introduces a barrier to access compared with usenet groups more generally, which can be accessed from any news server (subject to them being carried by the server concerned).

There is an "aus.*" hierarchy within the usenet system, with several hundred Australian-oriented newsgroups within it. (They can be accessed from anywhere in the world - including by spammers - but by the nature of their titles and content their more serious users tend to be mainly Australians).

One respondent to the survey of farmers for the user needs project suggested the implementation of an aus.agriculture newsgroup. Usenet newsgroups are very simple to establish - all that is required is an announcement be made in aus.general and aus.net.news, for some support to be demonstrated, and for a news administrator to create the group.

However, a newsgroup would also require some marketing if it were to be successful, while a moderator would be required for a moderated newsgroup. A moderator could undertake the necessary marketing, including encouraging content providers to participate in responding to questions and to use the newsgroup for announcing developments in respect of their organisation, such as their own Web site.

It may be preferable, therefore, for Farmwide to make its newsgroups more widely available.

Forums are like usenet newsgroups but are accessed through Web sites rather than through the usenet directory called up by newsreader software. Forums are much less subject to spamming. Forums are usually moderated, either explicitly or implicitly, so assisting users to find answers to their questions. The disadvantage of forums is that participant numbers are limited to the number of people who know about, and are prepared to use, the Web site concerned. A specialised search engine, Forum One, searches over 100,000 forums.¹³⁹

An example of a set of very active forums is at the US @gricuture Online site.¹⁴⁰ A number of Australian agricultural organisations, such as Elders and the Victorian Farmers' Federation, have added forums to their sites, but they have not generated substantial usage - a "critical mass" of users is required for forums to be successful. The forums on the F@rming Online site,¹⁴¹ have, however, been more successful.

Mailing lists are e:mail based groups exchanging information, ideas, chit-chat etc. While participation involves "subscribing" to the list, this consists of sending a simple e:mail message to the list host - no money is charged. Typically, participants post a message to the list, which is then received by everyone on the list. Like newsgroups, some mailing lists are moderated

- in some groups (eg, DAIRY-L), traffic can become quite substantial - possibly too substantial for some Internet users. In this case, one option is for messages to be

sent to the moderator, who then compiles the more substantial messages into a single e:mail digest sent to participants on possibly a daily, or weekly, basis.

Mailing lists are a useful communication mechanism for both the exchange of information and for organisations to make announcements relevant to the group. Mailing list software contains a “who” function whereby any subscriber can obtain a list of the e:mail addresses of other subscribers (some lists, such as DAIRY-L, turn this function off, presumably on privacy and anti-spamming grounds).

Directories of mailing lists are available,¹⁴² and lists can also be accessed through Web sites of their originators. A number of US land grant Universities have created agricultural mailing lists as part of their extension responsibilities to farmers. Overseas agricultural lists include DAIRY-L, GRAZE-L, SHEEP-L, BEEF-L, and many others. Typically, they are dominated by North American contributors - but not overwhelmingly so.

Australia’s most comprehensive set of agricultural mailing lists are the 13 aussie-xxxx-nets operated by the Queensland Department of Primary Industries. Some, such as aussie-fieldcrops-net, are reasonably active; others are quite inactive. Not surprisingly, there is a Queensland focus to the discussions. Other Australian mailing lists include agwomen-l, and lists created for specific purposes such as welink-l created by the Queensland University of Technology project team on Enhancing Rural Women’s Access to Interactive Communication Technologies, and mecis-l created for Central Queensland University’s project on information and communication systems for the meat industry.

Mailing lists are a useful communication feature that are not used as widely as they could or should be. For example, even QDPI - which is to be commended for creating the aussie-xxxx-nets - do not make great use of them for announcements about the organisation’s activities.

The aussie-xxxx-nets give a good coverage of commodities. There is not, therefore, any need for any additional general mailing lists on commodity topics - what is required in that respect is for the Queensland lists to be promoted and used more broadly (eg, by extension authorities in other States).

There are, however, few Australian agricultural mailing lists focusing on cross-commodity topics - although there does not appear to be much demand for any either. There is plenty of scope for the development of more organisation-specific mailing lists (eg, for R&D Corporations to communicate interactively with their stakeholders). This could be a useful topic to include in the proposed polite “audit” of organisation Web sites suggested in section 3.5.1 above.

While on the subject of e:mail, mention should be made of the absence of a comprehensive “white pages” type of directory of e:mail addresses. Some directories do exist, such as WhoWhere, an Australian E:mail Directory, and Farmwide has created a directory of Australian agriculture e:mail addresses. All such directories rely on people voluntarily submitting entries - and only a tiny proportion of Internet users do. There are good anti-spamming reasons for refusing to divulge one’s e:mail

address, but it is more likely that the vast majority of users who do not list themselves with directories fail to do so simply through ignorance or because they couldn't be bothered. Australian agriculture users should be encouraged to use the Farmwide directory.

Internet Relay Chat (IRC) provides real-time interactive communication. Some chat facilities are available on Web sites.

Because of its synchronous nature, chat is even more subject to a "critical mass" of participants to be successful than are the asynchronous applications such as newsgroups and mailing lists. Put differently, there is an even higher probability of not getting an answer to a query on a chat channel than in newsgroups or mailing lists - although at least there is a possibility of an immediate answer. Just as live meetings tend to be dominated by the more verbally confident and articulate participants, so do chat sessions tend to be dominated by the more keyboard-confident and articulate participants

- chat sessions, of course, take time - leading many people to be wary of the connect time charges of their service provider. People with this concern should take advantage of the multi-tasking capacity of their PC, ie, they can search on their Web browser and, while waiting for sites to download to the browser, chat.¹⁴³

The main agricultural chat opportunity available to Australian farmers is the regular chats on the Farmwide chat channels. Technologically speaking, these are open to all comers, but announcements on discussion topics are only sent to project participants. Each Monday evening is devoted to training, and Wednesday evenings alternate between general chat sessions and guest speakers. Guest speakers have included a wide range of topics and people. With 1,000 participants, the Farmwide project has sufficient "critical mass" to make chat sessions work - indeed, when participant numbers climb above around two dozen, as is frequently the case, the conversation "threads" can become difficult to follow.

Opportunities for chat sessions can be expected to develop further as organisations use such sessions as a natural development of their Internet services for stakeholders. The Grains R&D Corporation, for example, has expressed interest in arranging discussions at specific times with scientific experts on grain issues.¹⁴⁴

4.2 INFORMATION SERVICES

Another form of human intervention in the task of finding information on the Internet would be for people, as well as technology, to be employed. Farmers in particular experience information overload not just on the Internet - they experience information overload from other sources as well. They want information filtered and structured so that it makes sense in a local context, and they want someone familiar with their

circumstances to do this for them.¹⁴⁵ As yet, no such agriculture-specific information service exists on the Internet.

There are, however, a number of general information services providing answers to questions available on the Internet at present. In this case, users enter their search query - because it will be read by a live human being, the search query can make better use of the language than the keyword searches of the search engines. The information service then finds the required information, sends it to the user along with a bill.

Examples of this in operation now include Answers.com, Findout, HumanSearch, Internet Research Services, and Sherlock@. The services are not expensive - and in some cases free. Answers.com charges \$US1.99, \$US7.99 and \$US19.99 for easy, medium and advanced queries respectively, and custom pricing for more involved projects. Internet Research Services charges \$A10 for a simple search, \$A50 for a search taking around one hour, and negotiates prices on longer searches. Findout is free, although its replies appear to be confined to a list of relevant Web links rather than the actual answer sought.¹⁴⁶ HumanSearch also does not charge for its services (other than for users to send their bookmark files to add to their database).

A related set of initiatives is where enthusiastic experts provide articles of interest, including in finding information on the Internet - usually including a directory of other sites. One business model has the authors being paid a very small amount, so it is largely a labour of love; another model gives the authors a share of the advertising on their part of the site. Service organisers are always looking for new authors come on board. The Mining Company, for example, boasts over 500 such experts, on a range of topics including agriculture.

Of course, charging for information services is not new to Australian agriculture - the Kondinin Group's FarmLine service has been in operation since April 1995. FarmLine receives calls through a 1800 telephone number. Readily answered inquiries are answered for free; prices of more complex responses are negotiated with callers. As at the time of its review in August 1996, FarmLine was receiving around 12 calls per day, with an average income of \$23 per inquiry.

The Kondinin Group has purchased the AgNet directory of agricultural Internet sites, and plans to develop it into an online facility along the lines of FarmLine. It is understood that FarmLine's database would be put online on a searchable basis. A bulletin board (ie, forum) would be available where users can seek information and other users can provide it if they wish, with FarmLine staff intervening to provide information where other users do not (and charging as appropriate). In addition, a "help" facility would be provided which would, in effect, be an e:mail alternative to the existing telephone service

- this raises the issue of pricing models. "Pay-per-view" pricing is not really an option in the absence of widespread availability and acceptance of micro-payment methods over the Internet.¹⁴⁷ Subscription services appear more practical, but as

noted above traditions of free information in both Australian agriculture and the Internet makes subscription services a very difficult proposition.

An even grander proposition along these lines was recommended to consultants for the South Australian agricultural sector in mid-1996.¹⁴⁸ This was for a “virtual venue” that would take information requests from farmers in whatever form (mail, phone, fax, Internet) and provide responses in whatever manner the farmer desired (mail, phone, fax, Internet). Information would be provided not only on agricultural topics but also on other topics relevant to farm households, including education and training and entertainment. This idea appears not to have progressed.

5 CONCLUSION - A STRATEGY TO IMPROVE THE ACCESSIBILITY OF AUSTRALIAN AGRICULTURAL INFORMATION ON THE INTERNET

The discussion in Chapters 3 and 4 above includes a number of suggestions for improving the accessibility of Australian agricultural information on the Internet. In this conclusion, some of these suggestions are presented as recommendations. They are presented in the order of the above discussion, along with some comments - most of the issues concerned with implementing the suggestions are discussed above.

RECOMMENDATION 1. Create a specialised search engine.

This proposal, discussed in section 3.3.3, is for a specialised search engine along the lines of the general Internet search engines, but restricted in its scope of search to those sites known to contain information relevant to Australian farm businesses.

With a substantial “public good” component to its operation, it may be necessary for this to be provided by one of the significant public sector content providers. One option would be for RIRDC, possibly as a lead agency for the other organisations using DPIE’s server, to call for expressions of interest from the industry for the development of such a facility. A decision to proceed could then depend on the adequacy and cost of the proposals received.

RECOMMENDATION 2. Provide training in search techniques.

Training initiatives need to include a search component.

RECOMMENDATION 3. Encourage significant content providers to adopt “best practice” in terms of the accessibility, navigability and marketing of their Web sites.

One means of providing this encouragement would be for a friendly “audit” of sites. This would rate the various sites against the various criteria set out in section 3.5.1, particularly the AUS Standards, and would make suggestions to the organisation concerned.

“Significance” of an organisation’s content could be assessed from the rating system developed for and applied in *The Australian Farmers’ Guide to the Internet*. Given the sensitivity of such an exercise from content providers’ point of view, it may be necessary for the comments of the “auditor” to be kept confidential to the organisation concerned.

A similar initiative would be a “how to” guide bringing together all the important elements that site authors should bear in mind when designing a site relevant to a farm audience.

RECOMMENDATION 4. Create a directory of the search pages on sites with site-specific search engines.

Options include (in ascending order of complexity)

- adding a list of site-specific search engines, with hyper-text links thereto, to an existing directory or “links” page
- a “submit-it” style service, whereby the user types the keywords sought once on the page of the directory, but then submits the query to each search page (s)he considers relevant
- a meta-crawler, that is, the user types the keywords in once and the software itself searches all the search facilities concerned to return a list of relevant pages to the user.

RIRDC could readily implement the first possibility on its own site, and it may be a simple matter to develop this into the second possibility. The third possibility could be explored with DPIE’s computer people with a view to implementing such a system on PIENet, if necessary through a contracting out arrangement with a private firm capable of implementing the system.

RECOMMENDATION 5. Support regional communication initiatives.

This is happening through RIRDC’s involvement in CIRCIT’s framework for developing regional communication initiatives and through the Regional Telecommunications Infrastructure Fund. Support for user training, including in search techniques, should be an important part of these endeavours.

A further initiative would be a study of existing community and local initiatives with a view to propagating ideas from each.

RECOMMENDATION 6. Create an aus.agriculture newsgroup.

As discussed in section 4.1, this may be best achieved through arranging for more ready access to Farmwide’s newsgroups.

ATTACHMENT A - LIST OF SEARCH FACILITIES

The following lists search facilities mentioned in this paper.

SEARCH FACILITY	URL
MAJOR GENERAL INTERNET SEARCH ENGINES	
AltaVista - main site	http://altavista.digital.com
- Australian mirror site	http://www.altavista.yellowpages.com.au/cgi-bin/telstra?
Excite	http://www.excite.com
HotBot	http://www.hotbot.com
InfoSeek	http://www.infoseek.com
Lycos	http://www.lycos.com
Northern Light	http://www.northernlight.com/
WebCrawler	http://www.webcrawler.com
OTHER GENERAL INTERNET SEARCH ENGINES	
Answers	http://www.answers.com.au
Metadata Search Engine	http://www.dstc.edu.au/RDU/Harvest/Meta/about.html
Open Text	http://index.opentext.net
Rankdex - hypertext search engine	http://rankdex.gari.com
World Wide Web Worm	http://wwwwww.cs.colorado.edu/wwwwww
GENERAL DIRECTORIES	
Magellan	http://www.mckinley.com
Matilda	http://www.aaa.com.au
WWW Virtual Library	http://www.w3.org/pub/DataSources/bySubject/Overview.html
Yahoo	http://www.yahoo.com
SEARCH ENGINE DIRECTORIES	
Linkaway	http://rio.atlantic.net/~bdarl/srchlnk1.html
Search.Com	http://www.search.com
WebTaxi	http://www.webtaxi.com
Worldwide Search Engines Annotated	http://www.lawresearch.com/cewwwwd.htm
Yahoo's Searching the Web	http://www.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web/
AUSTRALIAN AGRICULTURAL DIRECTORIES	
Agnet Australian Agriculture	http://agnet.com.au
Agriculture on the Web	http://www.comcirc.com.au/ag/
AgriWeb	http://www.ruralnet.net.au/AgriWeb/
Australian Agricultural Resource Centre	http://www.comu.net.au/gipsmirror/agri/
Australian Agricultural Sites	http://www.iconnect.net.au/farmers.html
Australian Agriculture & all things related	http://www.benet.net.au/farm.html
<i>The Australian Farmer's Guide to the Internet</i>	http://www.farmwide.com.au/farmersguide
Australian Farm Radio and Weather	http://www.aaa.com.au/Farm_Radio.html
Australian Indexes - Agriculture	http://www.moreinfo.com.au/ausindex/agriculture.html
Australian Web Servers - Agriculture	http://www.csu.edu.au/links/agriculture.html
Cowley's Farm Web	http://www.cowleys.com.au/public/farmweb.htm
Horsenet - Horse Information Service	http://www.horsenet.com.au/
Landcare Web	http://www.agfor.unimelb.edu.au/LCweb/LCweb.html
Silo Home Page	http://www.wagga.net.au/~silo
Weather Information	http://atmos.es.mq.edu.au/weather/
Websource	http://www.websource.com.au/science/agriculture/

List of Search Facilities (ctd)

OTHER AGRICULTURAL DIRECTORIES

Agfind	http://www.agfind.com
Aglinks	http://www.gennis.com/aglinks.html
Agrigator	http://gnv.ifas.ufl.edu/www/agator/htm/ag.htm
Agrisurf	http://www.agrisurf.com
Agscope Unified Search Interface	http://www.agscope.net/search.html
CAFFEINE	http://www.caffeine.ca
Not Just Cows	http://www.snymor.edu/~drewwe/njc
WWW Virtual Library for Agriculture	http://ipmwww.ncsu.edu/cernag/cern.html

E:MAIL ADDRESS DIRECTORIES

Australian E:mail Directory	http://www.sofcom.com.au/Directories/index.html
Bigfoot	http://www.bigfoot.com
Farmwide's agricultural e:mail directory	http://www.farmwide.com.au/community/directory/
Four 11	http://www.four11.com
WhoWhere	http://www.whowhere.com

WEB-BASED META-CRAWLERS

Ask Jeeves	http://www.askjeeves.com
Dogpile	http://www.dogpile.com
FireFly	http://www.firefly.com
Inference Find	http://m5.inference.com/ifind/
Internet Sleuth	http://www.isleuth.com
Kapow	http://www.bam.com.au/kapow/
Mamma	http://www.mamma.com
MetaCrawler	http://www.metacrawler.com
Metafind	http://www.metafind.com
Ultra-Search	http://virtual-mktg.com/jam237/UltraSearch/

PERSONALISED META-CRAWLERS

Autonomy	http://www.agentware.com
Echosearch	http://www.iconovex.com
Internet FastFind	http://www.symantec.com
WebBandit	http://www.jwsg.com
WebCompass	http://www.quarterdeck.com
WebEx	http://www.travsoft.com
WebSeeker	http://www.ffg.com/seeker

SPECIALISED SEARCH ENGINES

Argos Limited Area Search of the Ancient and Medieval Internet	http://argos.evansville.edu
Deja News - searches newsgroups	http://www.dejanews.com
Filez - searches for files on FTP sites	http://www.filez.com
Forum One - searches Web-based forums	http://www.forumone.com
MACISS - searches weather information	http://atmos.es.mq.edu.au/MACISS/
Web Seer - searches for graphics	http://webseer.cs.uchicago.edu
WebWombat - Australasian sites	http://www.webwombat.com.au/wombat/
What's New Too - new sites	http://newtoo.manifest.com/search.html

List of Search Facilities (ctd)

TIPS AND TUTORIALS ON SEARCH TECHNIQUES

Ask Scott	http://www.askscott.com/
Cnet.com Search Tips	http://www.cnet.com/Content/Features/Dlife/Search/ss06.html
How to Search the Web	http://issfw.palomar.edu/Library/TGSEARCH.HTM
Farmwide Tips Page	http://www.farmwide.com.au/faq/page3.htm
Finding Information on the Internet - A Tutorial	http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html
Understanding and Comparing Web Search Tools	http://www.hamline.edu/library/bush/handouts/comparisons.html
Web and Internet Search Engine FAQ	http://www.nauticom.net/www/krb/pageone.htm
Web Novice	http://www.webnovice.com/

PERSONALISED RESEARCH OPTIONS

Answers.com	http://www.answers.com
Findout	http://www.findout.com
HumanSearch	http://www.humansearch.com
Internet Research Services	http://www.rainbow.net.au/~ackirk/
Sherlock@	http://www.intermediacy.com/sherlock/index.phtml

EXPERT GUIDES

The Mining Company	http://home.miningco.com/
Suite 101	http://www.suite101.com

ATTACHMENT B - THE ISSUES PAPER

The following issues paper was circulated to a number of people interested in the provision of information for Australian farmers via the Internet. Responses are gratefully acknowledged.

OPTIONS FOR IMPROVING ACCESSIBILITY OF AGRICULTURAL INFORMATION ON THE INTERNET

I have been contracted by the Rural Industries Research and Development Corporation (RIRDC) to write a paper on improving the accessibility of agricultural information on the Internet.

The following represents a short issues paper on the project, with a set of questions at the conclusion of each section. Comments are requested - please send comments to me at the above address.

THE PROBLEM

The Internet contains a considerable amount of information relevant to Australian farm businesses. While there are also considerable gaps in this information, of the order of 200 Australian World Wide Web sites provide at least some information, in addition to many overseas sites.

This does not mean, however, that specific information is necessarily easy to find. There is, in short, a problem of information overload, which swamps even the substantial search and directories that are available.

SOME OPTIONS TO IMPROVE ACCESSIBILITY OF AGRICULTURAL INFORMATION

1. Comprehensive data base(s)

Some data bases relevant to Australian agriculture exist, but are by no means comprehensive. Some have suggested the creation of a comprehensive data base covering all information of relevance.

The difficulty, of course, would be the cost of establishing the data base and keeping it up-to-date. Given the diverse nature of farmers' information requirements, it is unlikely that any truly comprehensive data base would be achievable.

Nevertheless, a number of major content providers are moving in the direction of improving the quality and quantity of the data they make available on the Internet. Some of these are beginning to constitute partial data bases.

Are additional data bases feasible and cost-effective? If so, in what areas? Who is prepared to fund/develop such data bases?

2. Improved content provision

Content providers can - and many do - take a number of steps to make their information easier to find, including

- ensuring they take full advantage of the opportunities provided by search engines
- approaching other content providers to provide mutual links to their sites
- paying attention to the accessibility and navigability of their sites, including the provision of search facilities
- providing "help" or "tips" pages for users.

Do all major content providers for Australian farmers list their sites with as many search engines as possible? Do they take search engine prominence into account in designing their sites? Do they seek links from other sites? Is a "Web ring" or a "banner exchange" a viable option? Are sites designed to maximise accessibility and navigability? Do they provide search facilities? Do

they provide “help” or “tips” pages? What is “best practice” in these areas and how can content providers be encouraged to adopt it?

3. Assisting and training users

While a number of online directories of agricultural sites are available, none provide substantial information on what is contained within sites, and few provide ratings of the quality of sites. There is no paper-based compendium of Australian agricultural sites.

A compendium of sites for American farmers has recently been published - a similar project is under active consideration by RIRDC in conjunction with the National Farmers' Federation's Farmwide project. Ideally, such a compendium should be both online and paper-based, with regular updates (frequently in the case of the online version) and comprehensive reviews and objective ratings of sites.

Improved training of users in search techniques is also important. Despite extensive training efforts by some organisations and the ready availability of “help” screens and “tips” pages within search engines, it is clear that many farm users remain unaware of many quite basic search techniques.

What other measures can be taken to improve user training?

4. Provide Human Assistance

One solution to the problem of information overload is for the development of services assisting users find information or to provide the information for them. Newsgroups, mailing lists, forums and even Internet Relay Chat (IRC) can fulfill this function on a voluntary basis, and now commercial services are becoming available.

However, there are few newsgroups relevant to farming, and no Australian-based newsgroups. Moreover, with the possible exception of IRC, these services lack the immediacy that may be required by some users.

What developments are likely in the provision of human assistance? Are there ways to encourage such developments? Would an aus.agriculture Usenet newsgroup be worthwhile?

5. Specialised Search Engine

Part of the problem with search engines is that they search the entire Internet, and so locate a large amount of irrelevant information. Even with better user training etc, improved filters would be highly desirable.

From a layman's point of view, what seems to be required is a search engine that catalogues all information from only a defined set of sites, in this case, those known to contain information relevant to Australian farm businesses. World-wide, there would probably be less than 1,000 such sites - although all pages on a site would have to be accessed. Australian farmers can hardly be the only group to be facing the problem of “information overload”, but such a search engine does not appear to be in operation on the World Wide Web at present.

Does more specialised search engine technology exist? If so, can it be used by major content providers for Australian farmers? If not, is it possible to develop such a search engine? At what cost? Is any organisation prepared to bear this cost?

Jim Groves

5 March 1997

ENDNOTES

¹ Buy IT OnLine, *Electronic Commerce Newsletter*, 9 April 1997.

² *Ibid.* These results are from a survey of frequent Web users (defined as anyone who has used the Web in the 24 hours prior to the survey) on how they “most often” find Web sites. The answers were search engines 71%, recommendations from friends and relatives 9.8%, newspapers and magazines 8.5%, links from other Web sites 8.4%, television 3.6%, printed Internet directories 3.3%, and “surfing” 8.1%. The report does not explain how this adds to more than 100% - presumably some people nominated more than one search method.

³ “The Internet: Bringing Order from Chaos”, *Scientific American*, March 1997, <http://www.sciam.com/0397issue00397intro.html>.

⁴ J Groves and J Da Rin, *The Internet Organisation*, RIRDC, forthcoming.

⁵ L Monier, Chief Technical Officer of AltaVista, at http://www4.zdnet.com/anchordesk/talkback/talkback_13066.html.

⁶ A list of directory, search engine and data base titles and URLs is at Attachment A.

⁷ J Da Rin and J Groves, *What's on the web now for Australian farmers?* RIRDC, 1996.

⁸ See Table 3.1 below for the URLs.

⁹ This example was chosen because it was the example mentioned by a participant in the Farmwide Online Services Pilot project during a chat session on the subject in December 1996. The farmer concerned had needed urgent information on types of army worms, chemical applications and other specifics. He described his experience as having “bumbled around for so long”.

¹⁰ Yahoo is not a search engine - it is a directory. Its importance and prominence is such that it is included in the discussion on search engines for most practical purposes. When a relevant “hit” is not found within its directory, Yahoo pipes into a search engine - AltaVista by default.

¹¹ Table 2.1 should not be taken as a comparison test of the search engines themselves. Only one search was conducted, in an unsophisticated manner guaranteed to produce many irrelevant hits (eg, those referring just to the word “army”). InfoSeek’s algorithm was clearly best suited to this particular search - that would not necessarily be the case for a different search. “Relevance” of hits can also be in the eye of the beholder - relevancy in Table 2.1 is defined as any page that is clearly providing information on the subject - it may not be answering a specific query. The table is presented for illustrative purposes only.

¹² <http://etrng.findsvp.com/internet/overview.html>

¹³ Drawn from transcript at <http://www.farmwide.com.au/previouschats/970203-training.htm>.

¹⁴ *The Internet Organisation*, *op cit.*

¹⁵ Transcript at <http://www.farmwide.com.au/previouschats/9611%2D27a.htm>.

¹⁶ Electronic Trading Concepts Pty Ltd, *A Feasibility Study on the Most Cost Effective Strategies for Meeting the Electronic Information Needs of SA Primary Producers and Agribusiness*, July 1996, reports a survey result that over 80% of farmers are satisfied with existing sources for their various information needs.

¹⁷ *The “Third Revoution” and Rural Australia: An Exploration of the Economic and Social Impacts of Farm Internet Use.*

¹⁸ Aglinks - a directory.

¹⁹ With 38% nominating it as their favored search facility, compared with 19% for WebCrawler, 16% for AltaVista, 10% for InfoSeek, and the remainder spread among Excite, HotBot, Lycos and Search.com. Source: Calafia Consulting, *The Search Engine Report*, no 9, *op cit.*

²⁰ MA Hearst, “Interfaces for Searching the Web”, *Scientific American*, March 1997, <http://www.sciam.com/0397issue/0397hearst.html>.

²¹ N Ward, “Roadshow presentation”, 12 September 1997, <http://www.dstc.edu.au/events/roadshow/>.

²² Eg, Suite 101 - <http://www.suite101.com/>

²³ Response to issues paper from Ian Bailey of AgNet.

²⁴ AgNet, for example, receives 2,000 visitors (not “hits”) per week (Source: response to the issues paper from Ian Bailey of AgNet, *op cit.*).

²⁵ Calafia Consulting, *Search Engine Features Chart*, <http://searchenginewatch.com/features.htm>. Much of this section is drawn from Calafia Consulting’s site at <http://searchenginewatch.com>. Similar information can be found at NetGuide Magazine Online at

-
- <http://techweb.cmp.com/ng/may97/srchtick.htm> and at <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>.
- ²⁶ Calafia Consulting, *The Major Search Engines*, <http://searchenginewatch.com/major.htm>.
- ²⁷ Calafia Consulting, *The Search Engine Report*, no 6, May 1997.
- ²⁸ C/net, *Microsoft: Yukon ho?* 7 October 1997, <http://www.news.com/News/Item/0,4,15019,00.html>.
- ²⁹ This discussion of search engines is based on Calafia Consulting, *Search Engine Features Chart*, *op cit*. Some applications are discussed by S Pecor, *LinkExchange Digest*, no 93, June 1997.
- ³⁰ Danny Sullivan's *Search Engine Watch*, <http://searchenginewatch.com> and the Web and Internet Search Engine FAQ at <http://www.nauticom.net/www/krb/pageone.htm> are excellent sources of detailed comparative information.
- ³¹ Summaries of a number of such exercises are at Calafia Consulting, *Search Engines Strategic Alliances Chart*, <http://searchenginewatch.com/alliances.htm>. Examples include the Annual Search Engine Shoot-out, <http://www4.zdnet.com/pccomp/features/excl0997/sear/sear.html>; Search Engine Showdown at <http://imt.net/~notess/search/index.html>; The MICA Report at http://www.melee.com/mica/cr_yyyymmdd.html; and an academic study, HV Leighton and J Srivastava, *Precision among World Wide Web Search Services*, <http://www.winona.msus.edu/is-f/library-f/webind2/webind2.htm>.
- ³² Calafia Consulting, *Search Engine Report*, no 10, September 1997, quoting survey results at http://www.npd.com/c_online4.htm. Other such surveys can be found at Media Matrix, http://www.mediamatrix.com/c_pctopweb.htm and Relevant Knowledge, <http://www.relevantknowledge.com/>;
- ³³ With only 18% of respondents using all other search engines combined, there was insufficient data to provide a meaningful break-up of the satisfaction data for each search engine.
- ³⁴ Sums to more than 100 because some respondents ranked two search facilities equally first.
- ³⁵ <http://www.farmwide.com.au/previouschats/970203-training.htm>
- ³⁶ Information on the commercially-available (including freeware) search facilities available for this purpose is at <http://searchenginewatch.com/software.htm>.
- ³⁷ <http://www.dpie.gov.au>
- ³⁸ <http://www.dpie.gov.au/rirdc/>
- ³⁹ <http://www.dairy.com.au/>
- ⁴⁰ For examples, see Attachment A and <http://searchenginewatch.com/others.htm>.
- ⁴¹ <http://www.ariadne.ac.uk/issue7/search-engines/>
- ⁴² Some personalised meta-search engines are discussed and rated at <http://www1.zdnet.com/complife/fea/9708/findny56.html>.
- ⁴³ JWsg's announcement of its WebBandit product - <http://www.jwsg.com>.
- ⁴⁴ For the purposes of this paper, I downloaded WebBandit from <http://www.jwsg.com>. However, the effort required to set it up - and the absence of useful instructions in the demonstration version - was simply not worth it.
- ⁴⁵ <http://www.erin.gov.au>
- ⁴⁶ <http://www.dpi.qld.gov.au/longpdk/longpdk.htm>
- ⁴⁷ <http://www.sfe.com.au>
- ⁴⁸ <http://www.wesfarmersdalgety.com.au>
- ⁴⁹ <http://www.infoscan.com.au>
- ⁵⁰ <http://www.austlii.edu.au/austlii/>
- ⁵¹ Response to issues paper from Max Zeller of @griWeb.
- ⁵² Scott Newman of the Tropical Beef Centre.
- ⁵³ For example, State Agriculture Departments are willing to respond to requests from the general public for information on agricultural issues (eg, for school projects), but this is not their core business and so they would not want to encourage too many such requests (response to issues paper from Bruce Ward of NSW Agriculture).
- ⁵⁴ Eg, <http://www.submit-it.com>.
- ⁵⁵ Calafia Consulting, *The Search Engine Report*, no 9, *op cit*.
- ⁵⁶ Calafia Consulting, *The Search Engine Report*, no 5, April 1997.
- ⁵⁷ N Ward, *op cit*.
- ⁵⁸ Calafia Consulting, *The Search Engine Report*, no 9, *op cit*.
- ⁵⁹ Calafia Consulting, *Search Engine Features Chart*, *op cit*.
- ⁶⁰ *ibid*.
- ⁶¹ *ibid*.

-
- ⁶² One estimate is that there might be ten times more information in data bases than in HTML - and more valuable information at that (N Ward, *op cit*).
- ⁶³ A list and discussion is at <http://www.ariadne.ac.uk/issue10/search-engines/>.
- ⁶⁴ K Clementson, *LinkExchange Digest*, no 176, September 1997.
- ⁶⁵ Calafia Consulting, *The Search Engine Report*, no 9, *op cit*.
- ⁶⁶ Calafia Consulting, *The Search Engine Report*, no 6, May 1997.
- ⁶⁷ <http://www.farmwide.com.au/nff/wafarmers/wafarmers.htm>
- ⁶⁸ D Blacharski, *Your Search Has Ended - Improved search tools are helping Webmasters guide users through the Internet's data maze*, Communications Week, July 14 1997.
- ⁶⁹ Calafia Consulting, *The Search Engine Report*, no 10, *op cit*.
- ⁷⁰ Calafia Consulting, *The Search Engine Report*, no 9, *op cit*.
- ⁷¹ Calafia Consulting, *The Search Engine Report*, no 10, *op cit*.
- ⁷² Calafia Consulting, *The Search Engine Report*, no 9, *op cit*.
- ⁷³ J Peline, *Search engines turn on to TV*, <http://www.news.com/News/Item/0,4,9873,00.html>. The list given is Excite's.
- ⁷⁴ <http://www.zdnet.com.au/zdnn/content/zdnn/0929/zdnn0005.html> and <http://www.zdnet.com/zdnn/content/zdnn/0918/zdnn0014.html>.
- ⁷⁵ "Push comes to Shove," *The Australian*, October 21 1997.
- ⁷⁶ Response to issues paper from Simon Bryant of the Department of Communication and the Arts.
- ⁷⁷ Calafia Consulting, *The Search Engine Report*, no 4, March 1997, and no 7, June 1997.
- ⁷⁸ This information on Argos is drawn from its Web site at <http://argos.evansville.edu/about.htm>.
- ⁷⁹ This information on MACISS is drawn from the Web site at <http://atmos.es.mq.edu.au/MACISS/>.
- ⁸⁰ C Lynch, "Searching the Internet", *Scientific American*, March 1997, <http://www.sciam.com/0397issue/0397lynch.html>.
- ⁸¹ Frequently Asked Questions (and Answers) about Harvest, <http://harvest.transarc.com/afs/transarc.com/public/trg/Harvest/FAQ.html>.
- ⁸² <http://harvest.transarc.com/>
- ⁸³ Response to issues paper from Irene Goltz of DPIE.
- ⁸⁴ Frequently Asked Questions (and Answers) about Harvest, *op cit*.
- ⁸⁵ *ibid*.
- ⁸⁶ Responses to the issues paper from Ross Andrews of the Grains Research and Development Corporation (GRDS) and Ross Cutler of the Pig Research and Development Corporation (PRDC).
- ⁸⁷ *Business Wire*, May 1997.
- ⁸⁸ Calafia Consulting, *The Search Engine Report*, no 9, *op cit*.
- ⁸⁹ The following draws on D Blacharski, *op cit*.
- ⁹⁰ But not at random - Excite appears around one third of the time, InfoSeek and Lycos a quarter each, and Yahoo around 10% of the time (Source: Calafia Consulting, *Search Engines Strategic Alliances Chart*, *op cit*.)
- ⁹¹ This discussion is drawn from Calafia Consulting, *ibid*, and Calafia Consulting, *The Search Engine Report*, no 6, *op cit*.
- ⁹² *ibid*.
- ⁹³ TA Gray, *How to Search the Web*, <http://issfw.palomar.edu/Library/TGSEARCH.HTM>.
- ⁹⁴ <http://webcrawler.com/WebCrawler/SearchTicker.html>
- ⁹⁵ <http://www.learnskills.com>
- ⁹⁶ This section is drawn from J Groves and J Da Rin, *Farmers and the Internet*, *op cit*.
- ⁹⁷ R Simpson, *Farmwide OnLine*, paper to AIAS Learning Online Conference, Melbourne, November 1996.
- ⁹⁸ Response to issues paper from Bruce Ward of NSW Agriculture.
- ⁹⁹ M Henninger, *Don't Just Surf: Effective Research Strategies for the Net*, UNSW Press, 1997.
- ¹⁰⁰ Buy IT OnLine, *Electronic Commerce Newsletter*, *op cit*.
- ¹⁰¹ W Truelove, *Internet Use by Rural Residents and Farmers*, Report to RIRDC, 1997.
- ¹⁰² W Easdown, "Emerging Rural Expectations of the Internet and Perceived Training Needs", *Rural Society*, forthcoming.
- ¹⁰³ In this paper, "content providers" is defined to include any organisation with a Web site. It encompasses the ownership, information supply, authoring and publishing of the site - recognising that in practice these functions may be carried out by different people or organisations.
- ¹⁰⁴ <http://www.agd.nsw.gov.au/standards.html>.
- ¹⁰⁵ Net Cruiser, *The Australian*, 12 August 1997.
- ¹⁰⁶ <http://www.farmjournal.com/>. Source: W Truelove, *op cit*.

-
- ¹⁰⁷ Promotion tips are at <http://www.cnet.com/Content/Features/Howto/Promote/ss01.html> and in online newsletters such as the LinkExchange Digest and Gator's Byte Promotion Tips Newsletter.
- ¹⁰⁸ Submit-it, <http://www.submit-it.com>.
- ¹⁰⁹ Buy IT OnLine, *Electronic Commerce Newsletter*, *op cit*.
- ¹¹⁰ F Salvadori, *LinkExchange Digest*, no 92, June 1997.
- ¹¹¹ S Pecor, *LinkExchange Digest*, no 90, May 1997.
- ¹¹² B Holcombe, *LinkExchange Digest*, no 176, September 1997.
- ¹¹³ Gator's Byte Promotion Tips Newsletter, July 1 1997.
- ¹¹⁴ Eg, InterNetGain (<http://www.internetgain.com>), Position Agent (<http://www.positionagent.com>) and ScoreCheck (<http://www.scorecheck.com>).
- ¹¹⁵ Buy IT OnLine, *Electronic Commerce Newsletter*, *op cit*.
- ¹¹⁶ "Banners" are the 400x40 pixel strips advertising other sites - often saying "click here for ...". They include a hyper-text link to the sites concerned. Small graphics, including animated gifs, often appear in the banner.
- ¹¹⁷ <http://www.cowleys.com.au>
- ¹¹⁸ <http://www.linkexchange.com>
- ¹¹⁹ <http://www.free4all-exchange.com>
- ¹²⁰ The information contained in this paragraph, and the directory on which the search was conducted, was found at <http://www.webring.org>.
- ¹²¹ Buy IT OnLine, *Electronic Commerce Newsletter*, *op cit*.
- ¹²² Response to issues paper from Irene Goltz of DPIE.
- ¹²³ Much of the information for this section is drawn from P Miller, *Metadata for the masses*, <http://www.ukoln.ac.uk/ariadne/issue5/metadata-masses/> and R Iannella and A Waugh, *Metadata: Enabling the Internet*, <http://www.dstc.edu.au/RDU/reports/CAUSE97/>.
- ¹²⁴ In search engines such as AltaVista, images can be searched for through using a keyword appearing in the title of the image sought ("comet" is the example given on the AltaVista tips page). This remains, however, a text-based search - the image itself is not indexed, nor understood by the search engine. Web Seer is a search engine that can go a little further because it can distinguish between photographs and other graphics, and applies other filters such as file size and the number of people in a photograph.
- ¹²⁵ Such as the US Content Standards for Digital Geospatial Metadata used by the ERIN Web site (source: *ERIN DSDL - Metadata Files*, <http://www.erin.gov.au/technical/dsdl/metadata.html>).
- ¹²⁶ Iannella and Waugh, *op cit*.
- ¹²⁷ Such as through events like The 4th Dublin Core Metadata Workshop held in Canberra in March 1997. A detailed report of the workshop is at <http://www.dlib.org/dlib/june97/metadata/06weibel.html>.
- ¹²⁸ <http://www.edna.edu.au/edna/owa/info.getpage?sp=auto&pagecode=5210>
- ¹²⁹ N Ward, *op cit*.
- ¹³⁰ D Shafer, *XML and <META> tags: speaking in tongues*, <http://www.cnet.com/Content/Voices/Shaffer/072897/?dd>.
- ¹³¹ M Dertouzos, *What Will Be: How the new world of information will change our lives*, Harper Edge, 1997.
- ¹³² The information on PICS is drawn from <http://www.w3.org/PICS/> and P Resnick, "Filtering Information on the Internet", *Scientific American*, March 1997, <http://www.sciam.com/0397issue/0397resnick.html>.
- ¹³³ A general Internet version of such a system, falling short of being a meta-crawler, can be seen at <http://webusers.anet-dfw.com/~heinzmnl/link.html>.
- ¹³⁴ LinkExchange Digest, <http://www.ledigest.com>.
- ¹³⁵ J Hagel and AG Armstrong, *net.gain: expanding markets through virtual communities*, Harvard, 1997.
- ¹³⁶ Centre for International Research on Communication and Information Technologies, *A Framework for Developing Regional Communications Initiatives*, June 1997.
- ¹³⁷ Suggested by Marina Cavill of CIRCIT in her response to the issues paper.
- ¹³⁸ *Caveat emptor* also applies to information from Web sites, but with less force given the significance and reputation of some of the content providers.
- ¹³⁹ Calafia Consulting, *The Search Engine Report*, no 11, October 1997.
- ¹⁴⁰ <http://www.agriculture.com>
- ¹⁴¹ <http://www.rpl.com.au/farming/>

¹⁴² Eg, Liszt (<http://www.liszt.com>) and InterLinks (<http://www.nova.edu/Inter-Links/listserv.html>). Australian lists are at Ozlist (http://www.gu.edu.au/gint/ozlists_home.html), and agricultural lists worldwide are at <http://www.agr.com/consulting/maillist.html>.

¹⁴³ Response to issues paper from James Cameron, *op cit*.

¹⁴⁴ Response to issues paper from Ross Andrews of the GRDC, *op cit*.

¹⁴⁵ Response to issues paper from Marina Cavill of CIRCIT, *op cit*.

¹⁴⁶ R Basch, "Customised Research Options: Personalised Service", *Computer Life Online*, <http://www1.zdnet.com/complife/fea/9708/findny65.html>.

¹⁴⁷ While the technology exists (source: response to issues paper from James Cameron of Digital Equipment Corporation), its widespread adoption and acceptance has not yet happened.

¹⁴⁸ Electronic Trading Concepts Pty Ltd, *op cit*.