

**Making past industry  
information from R&D  
more accessible**

Leigh Walters  
SA Farmers Federation

Project Number: PT00027

## **PT00027**

This report is published by Horticulture Australia Ltd to pass on information concerning horticultural research and development undertaken for the potato industry.

The research contained in this report was funded by Horticulture Australia Ltd with the financial support of the potato industry.

All expressions of opinion are not to be regarded as expressing the opinion of Horticulture Australia Ltd or any authority of the Australian Government.

The Company and the Australian Government accept no responsibility for any of the opinions or the accuracy of the information contained in this report and readers should rely upon their own enquiries in making decisions concerning their own interests.

ISBN 0 7341 1394 3

Published and distributed by:

Horticultural Australia Ltd

Level 1

50 Carrington Street

Sydney NSW 2000

Telephone: (02) 8295 2300

Fax: (02) 8295 2399

E-Mail: [horticulture@horticulture.com.au](mailto:horticulture@horticulture.com.au)

© Copyright 2006



*Know-how for Horticulture™*

PT00027 (July 2006)

# **Making past industry information from R&D more accessible**

Leigh Walters et al  
South Australian Farmers Federation

<b>HAL Project Number</b>	PT00027 (July 2006)
<b>Project Leader</b>	Leigh Walters Technology Transfer Manager (Australian Potato Industry) South Australian Farmers Federation PO Box 6014 Halifax Street Adelaide SA 5000
<b>Other key personnel</b>	Trish Dempsey Alison Pitman Kieran Walters Wendy Fishers
<b>Purpose</b>	This work reports on the development of an electronic library system for the Australian potato industry consisting of HAL Final Reports, articles from Potato Australia and Eyes on Potatoes, and Conference Proceedings.
<b>Funding</b>	The project has been funded by the potato levy with matching funding from the Australian Government through HAL.
<b>Date</b>	July 2006



## Acknowledgements

The project has had a very dedicated local Management Committee (Wayne Cornish, Barry Philp, Neil Perry, Clinton Zerella, Paul Frost and past members Jim Kelly and Adam Gray) and a national Steering Committee - Potato Industry Advisory Committee (previously the Potato R&D Committee).

Thanks also to past staff (Sharon Baker, Jonathan Eccles, John Oakeshott) and present staff (Danika Houghton, Liz Romanos, Simon Drum, Phil Roeth, Karen Hellwig and Richard Stephens) of Horticulture Australia Limited (HAL) who provided so much support. Their contribution was critical to the success of the project.

The Archives would not exist without the authors who contributed to the publications for which we owe our thanks.

## Disclaimer

*Any recommendations contained in this publication do not necessarily represent current HAL Limited policy. No person should act on the basis of the contents of this publication, whether as to matters of fact or opinion or other content, without first obtaining specific, independent professional advice in respect of the matters set out in this publication.*

# Contents

Media Summary .....	2
Introduction.....	4
Strategy .....	6
Method.....	7
Legal aspects.....	7
Publications to be included .....	9
Producing the electronic files .....	9
Front end.....	15
Field testing.....	21
Marketing .....	22
Discussion .....	24
Recommendations .....	28
Acknowledgements .....	28
Bibliography.....	29
<b>Appendices</b>	
A: Publications included in the Potato Archives CDROM .....	30

## Media Summary

*Producing information of value to industry is one challenge, having a means of accessing it when it is needed is another.*

Potato Archives is the Australian potato industry's digital library. It provides easy access to articles in past editions of Potato Australia and Eyes on Potatoes, HAL Final Reports of potato research and development (R&D) projects and papers from selected conference proceedings.

By using Potato Archives, levy payers and other industry people can see the range of potato information generated from the potato R&D program from 1990 to 2005.

This will result in more informed decisions that have the potential to improve management on farms, enable better quality advice to be provided by service industries and improve the ability of researchers to properly appreciate the body of work that has been done when they formulate new project proposals.

Potato Archives is available on a CDROM for use on a personal computer with an internet version being developed by AUSVEG at [www.ausveg.com.au](http://www.ausveg.com.au).

Levy Payers have received the first version of Potato Archives free of charge while others will be able to buy it from AUSVEG.

### Background

The potato levy payers and the Australian Government invested about \$22 million in potato Research and Development (R&D) between 1989 and 2005. This was further bolstered by significant investment by state governments and other industry groups.

If industry was to gain maximum value from the investment, it had to be able to access information already generated from the R&D when it was needed. Before Potato Archives existed, there was no simple way of doing this.

A lot of information from the R&D program came out in Potato Australia, Eyes on Potatoes, Final Reports and Conference Proceedings. However, finding the relevant information when it was needed, assuming a person had all the publications, was not an easy task. This made it difficult for industry to capitalise on its investment in R&D.

Many of the industry friendly publications such as Potato Australia, were not always kept and before problems in the distribution system were sorted out, not necessarily received.

Final Reports had to be bought and Conference Proceedings either bought or received as part of attending a conference.

### Funding

Potato Archives is a potato levy project with matching funds from the Australian Government and managed through Horticulture Australia Limited.

### Recommendations

Potato Archives has the potential to expand considerably from its current form. Recommendations are:

1. Explore linking the CDROM version and the internet version of Potato Archives to capitalise on the benefits of each and maximise speed of access for most users.

2. Explore with state departments of agriculture and industry organisations the range of publications to be included in future Potato Archives, to produce a centralised electronic library
3. Hold an industry workshop with stakeholders and technical specialists in one to two years time to map the future evolution of Potato Archives.
4. Suggest HAL save all new Final Reports in a form suitable for inclusion in Potato Archives to reduce the amount of extra work.
5. Update the Potato Archives CDROM in 3-5 years time.
6. Suggest HAL explore a broader format for Final Reports that encourages use of new technologies to enhance the user's learning experience and the value gained from Final Reports.

## Introduction

The potato levy payers and the Australian Government invested about \$22 million in potato Research and Development (R&D) between 1989 and 2005. This was further bolstered by significant investment by state governments and other industry groups.

A lot of the research outcomes and progress were reported in Potato Australia, industry newsletters, workshop notes, information leaflets and HAL (Horticultural Research and Development Corporation/ Horticulture Australia) Final Reports. However, gaining access to the information when it was needed was often difficult.

Final Reports had to be bought from HAL and it was often difficult to know whether a purchase would be worthwhile. The decision to buy a report also needed to be made in advance if the information was to be available when it was needed. Many growers, processors and service industry people therefore did not buy reports even when they were good value.

Many of the publications that were more user friendly for industry, such as Potato Australia, were not always kept and prior to the problems in the distribution system being sorted out, not necessarily received.

Given the large number of these publications, it was often difficult to find what was needed, assuming the person still had the relevant publication. This did not help industry capitalise on its significant investment in R&D.

We needed a simpler way for people to gain access to information that had been generated by the research program - when they needed it.

The last four words were very important - when they needed it. People were not always receiving information when they needed it and often when they needed it, they couldn't find it or get it quickly.

If people could have quicker access to the information generated from the R&D program, it would have resulted in greater use of the information, which in turn had the potential to improve management on farms, the quality of advice provided by service industries and the ability of researchers to better appreciate the existing body of work when formulating new project proposals.

To provide rapid access to past articles and reports, information needed to be stored electronically on a computer and accessed by a simple to use search engine either locally on a personal computer or through the internet.

Why not produce the information in printed form? The answer was it would have made for an enormous and impractical document.

Even at the time the project was proposed, there were ten editions of Potato Australia, eight editions of Eyes on Potatoes and about 60 HRDC Final Reports which came to just under 3,300 pages of text, 1,466 tables and 1,512 pictures.

By project end, Potato Archives contained 16 editions of Potato Australia, 26 editions of Eyes on Potatoes, 160 Final Reports and two conference proceedings with a further eight conference and workshop proceedings prepared but not included in the CDROM version of Potato Archives. This represented just under 12,500 pages of information.

In printed form there was also the problem of how all this information could be indexed so people could find what they wanted quickly. Putting all this information into one document was therefore not practical due to the size of the resultant publication, the difficulty in accessing what was needed and the expense of producing and distributing it.



Producing the information electronically and storing it on a CDROM had several advantages. It was relatively cheap, CDROM players had become standard on most computers and they held an enormous amount of information. Once the information was installed on the computer, a search engine could be used to find items of interest simply and quickly. Ask for information on powdery scab and the documents with powdery scab would be listed. Click on a document in the list and the document would be displayed. If the item was an article in Eyes on Potatoes or Potato Australia, then the relevant page would be displayed.

Potato Archives was also to be available through a password protected site on the internet. It was being included on the new AUSVEG potato internet service being developed at the time this report was being written.

The internet was not chosen as the preferred delivery method in the first instance due to the slow download times many people were experiencing (PT96009, PT98037). For those on fast connections, the internet was a viable option and in the future as the quality of our telecommunications improves more people would be able to use the internet version of Potato Archives.

When the project was initially conceived, it was also envisaged that a hybrid system between the CDROM version and the new Potato Internet Service would provide the best of both worlds. Unfortunately the ability to use such an approach was hampered by the introduction of firewalls and other computer protection systems.

The idea was that people could obtain the CDROM version and have the bulk of the documents on their local computer but would carry out their searches using the new Potato Internet Service. As the internet service was password protected, an option within the users' profile would indicate whether they had the CDROM version of Potato Archives installed on their computer. If a user selected an item from the search results located on their computer, it would display the item using the local copy otherwise it would download a copy from the internet site. This meant recently released material could also be accessed.

Over time, as new publications became available, more material would be downloaded through the internet slowing down response times so at some stage this would trigger an update of the CDROM. With the updated files on the local computer, response times would improve again and the cycle continue.

At some stage in the future, all files will be able to be downloaded when internet connection speeds increase. This may be many years away though given the isolated locations in which many people in the industry live.

Potato Archives also provided the opportunity for HAL to rationalise their stocks of Final Reports and manage distribution of hard copy versions more simply.

An important benefit of industry obtaining better access to R&D information would be a better appreciation of its value and the opportunities R&D can provide. It is hoped better understanding will lead to greater debate and ultimately better research outcomes for industry.

In summary the development of Potato Archives, the industry's digital library, should change the way information from the R&D program is produced, used and valued. This will ultimately lead to improved adoption of new technologies and better investment decisions by industry and government.

## Strategy

The strategy for developing Potato Archives was quite simple but the process of carrying it out proved much more difficult than expected.

The following steps summarised the proposed approach:

- (1) Obtain legal opinion on copyright and resolve any issues prior to starting.
- (2) Gain agreement as to what publications were to be included on the CDROM and obtain a copy of those publications (where possible in electronic form).
- (3) Contract a multimedia company to convert the documents into a suitable format, catalog the information, create the ancillary support information for the package and incorporate a search engine.
- (4) Field test and refine the package based on feedback and produce it as a CDROM.
- (5) Sort out distribution arrangements and market it through Eyes on Potatoes and Potato Australia.

The aim of the project was to produce an easy to use product that would be valued by industry and could be easily expanded to meet future needs.

## Method

### Legal aspects

A lawyer that specialised in copyright and intellectual property was identified.

To ensure the lawyer was clear about what we were doing, a mock-up was produced of Potato Archives. This made the process of identifying issues that needed addressing much simpler.

The following were key issues that came out of the process.

*(A note of caution: At the time of writing this report, Copyright law was still evolving, so it is important to take care how you use the following information and always obtain legal advice when in doubt. Not being legally trained, my wording may not be strictly correct in all instances.)*

#### (i) Ownership

It was important to establishment ownership of items so permission could be gained to use them in Potato Archives. Ownership may rest with an individual, a group of individuals or an organisation.

When reproducing the entire publication, often an organisation rather than individual needed to be approached for permission - HAL in the case of Final Reports, Potato Australia and Eyes on Potatoes.

If publications such as Potato Australia or Conference Proceedings were broken up, then permission would need to be obtained from authors or their organisation.

As there was a possibility of the Australian Potato Industry Council being dissolved, copyright for Potato Australia and Eyes on Potatoes was transferred to Horticulture Australia Limited in 2002 with the option for industry to take it back at a later date if they desired. This was done by a simple letter between the Australian Potato Industry Council who had current ownership seeking transfer of ownership of the publications to Horticulture Australia Limited, and their acceptance.

This meant copyright for all Final Reports, Eyes on Potatoes and Potato Australia was now with Horticulture Australia Limited.

The main problems with copyright lay with Conference Proceedings. Most early conferences had no clear publication ownership or had statements indicating copyright rested with the authors. This meant in some instances permission had to be obtained from authors of articles even though the whole conference proceeding was to be included.

Due to time constraints and difficulties with other parts of the project, only two conferences were included on the CDROM – Potatoes 2000 and Potato 2005. The former had clear ownership by the South Australian Research and Development Institute while the latter did not. This meant for Potato 2005, we attempted to get permission from all authors of material.

Eight other conferences or workshops were produced electronically and will require copyright permissions before they can be used.

To keep track of who owned the items, HAIDB Archive Manager (part of the potato industry's National Database System) had the facility to record copyright details for each item and where necessary link to a PDF copy of the copyright permission from the author or organisation. A copy of this database has been lodged with HAL at the conclusion of this project and forwarded to AUSVEG who will be carrying out future work.

#### (ii) Presentation of material

Presentation of the material was also important. Initially it was thought the documents would be converted to HTML or XML. When presenting material on a computer, a number of issues needed to be considered. Two points were particularly pertinent to our work.

All reasonable means needed to be taken to maintain the link between copyright owners and the material in the article were made so when an article was printed, there was no separation of ownership details from content. For example, if using frames in HTML, a person could easily print out only one part of the article, therefore separating the ownership details from the material itself. Whereas with a PDF, people would print out the entire item and would have to make a conscious decision to exclude the ownership details for them to be left off the printout.

Use of colour, particularly in warnings or important parts of the document such as ownership details needed to take into account colour blind users.

### **(iii) Opportunity to object**

The international nature of the internet had raised many interesting legal questions as the technology had evolved faster than the law. It was therefore worthwhile to give people the opportunity to object through an industry notice if they had concerns about the use of their material in Potato Archives.

An industry notice was put in the March 2002 edition of Eyes on Potatoes. No responses were received.

**Potato archives - Copyright notice**  
Potato archives are being developed to provide better access to past information generated from the levy funded research and development program.  
The archives will be available for sale on a CD ROM, or through a proposed national internet site for the Australian potato industry.  
This notice is to inform organisations and people who have supplied editorial material for Eyes on Potatoes, Potato Australia, potato conferences and Horticulture Australia Final Reports that if they have any concerns with their information being used in this way to contact:  
**Leigh Walters**  
☎ (08) 8232 5555  
Fax: (08) 8232 1311  
✉ [lwalters@saff.com.au](mailto:lwalters@saff.com.au)  
Where copyright does not reside with Horticulture Australia and the Australian Potato Industry Council for the publication, the author/organisation will be contacted directly to obtain permission to use the material.  
**Leigh Walters**  
**Technology Transfer Manager**  
**Australian Potato Industry**

Figure 1: Copyright notice from March 2002 Eyes on Potatoes

### **(iv) Finalising the product**

Once the product was finished, it was supplied to the lawyer with the ancillary information for checking. The wording for the 'Legal issues' in the program, 'Disclaimer' and 'License Agreement' on the CDROM cover and the license statement on the CDROM were produced by the lawyer or reworked from information supplied to the lawyer.

Other legal issues considered in production of the archives and not involving our lawyer were:

- (a) Fonts – If fonts were to be purchased and embedded it was important to read the conditions of use.
- (b) Runtime royalties – Some compiled programs could incur runtime royalties. In this instance it was not an issue.
- (c) Adobe command restrictions – Some commands for use with PDF documents were freely available for use, while others could only be used under a licensing agreement. The former applies for Potato Archives.

- (d) Inclusion of Adobe Acrobat Reader on CDROM - Inclusion of Adobe Acrobat Reader on the Potato Archives CDROM was considered but after discussions with HAL, it was decided not to include it due to the license conditions.

### **Publications to be included**

The question of what publications to include was relatively straight forward for Potato Australia and Eyes on Potatoes. We needed to include all if possible. As early editions had to be converted into electronic form, we were lucky to have a full set of both publications thanks to Jonathan Eccles from Horticulture Australia who had spare copies of the earlier editions of Potato Australia we were missing.

As the start of the project was delayed, the number of hard copy publications to include increased nearly three fold. Given that we thought many would have been digital, particularly in the case of Final Reports, we did not anticipate any problems with handling the extra workload. However, as it turned out, only half of the Final Reports had digital files, far less than we anticipated.

Some Final Reports were not included for the following reasons:

- They were confidential - as determined by HAL.
- For some projects (eg national potato publications, The Workboot Series - The story of potatoes in Australia, National Seed Potato Certification Standards) Final Reports provided no useful information for levy payers and the service industry as their outcomes were the publications themselves which everyone in the industry received or had access to.
- The decision was made not to include Cross Industry Projects (eg. AUSHORT) in the first edition of Potato Archives due to the extra workload and the likelihood we would not have the space to fit them on the CDROM..
- In the case of one project, PT96010, we could not establish the publication status of the Final Report before finalising the project so it was excluded. (I expect this be resolved before the next edition is released.) Unfortunately the problem did not surface until late in the project.
- Final Report has not yet been submitted.
- Final Report not submitted but may be available as some form of Business Plan or working document (eg. VX00023, PT04011). It is possible these will become available in future editions of Potato Archives.

We identified ten conference and workshop proceedings after approaching a number of people in the industry. We borrowed a number of the publications, scanned them and returned them after final proofing.

Initially we were going to look at other publications for inclusion but, given we had so much material to start with, we restricted it to Final Reports (excluding Cross Industry Project reports such as AUSHORT), Potato Australia, Eyes on Potatoes and selected conferences and workshops. Other material could be added once the archives were established.

### **Producing the electronic files**

It was at this stage we ran into difficulty with the project. The initial project aim was to sub-contract the scanning and interpretation of most of the material to an external specialist company. In preparing the material, especially the Final Reports, we found the quality of the originals in many cases quite poor and there were many problems such as missing pages, duplicate sections, upside down pages, etc. Many of the problems would have come back to us to resolve if we had outsourced the work.

Due to the number of issues with original copy quality, the decision was made to bring the job back 'in house' and do it ourselves. The challenge in doing that at the time was not fully appreciated.

The process we set up for publications without digital files included:

1. Checking all publications for problems and going back to HAL and other service providers to find missing material.
2. Developing HAIDB Archive Manager to track all work and be the repository for copyright information.

3. Scanning in material as PDF images
4. Used Optical Character Recognition to convert the PDF images to an intermediate format that separated images from text and converted text into a format that could be edited by a word processor.
5. Checked and corrected – 1<sup>st</sup> check.
6. Converted the intermediate format into XML.
7. Edited on screen, printed out and checked – 2<sup>nd</sup> check.
8. Converted the XML files into standard PDF.
9. This was printed out, checked and recompiled if changes needed to be made – 3<sup>rd</sup> check.
10. Navigation links were added to the PDF Contents page
11. Open settings and password security added.

In some instances pages had been written on or had blemishes, especially the front cover or opening page of conference proceedings. In these instances, if it was being recognised through OCR, the offending element was deleted or if saved as an image, edited using the cloning tools of Jasc Paint Shop Pro.

For publications with digital files, the process started at Step 9.

If the file was a PDF, it was checked and edited as required. Earlier PDFs had a range of problems, including:

- Duotone headings did not reproduce correctly, requiring some headings to be replaced.
- Graphics did not always reproduce properly and needed to be replaced by images scanned from the publications.
- Layout was corrupted requiring editing of the PDF.
- Final Reports did not have the HAL front section, which had to be produced separately from the hard copy and inserted.

If the file was in Word it was checked, edited if required by comparing it to the hard copy and then converted to PDF and checked again.

The following is a more detailed explanation of each of the steps in the process.

### **Step 1 - Checked all publications for problems and went back to HAL and other service providers to obtain missing material.**

This involved manually checking all documents and was quite slow as some of the problems were easily missed with a quick scan of the pages. An upside down page was obvious but a missing appendix may not have been if it was not reflected in page numbering. Pages printed on an angle were also a problem as the OCR process could correct for minor angles (ie. Up to about 5%) but started to have difficulties at greater angles. The types of problems experienced included missing pages, missing sections, missing appendices, missing paragraphs, missing pictures or diagrams, duplicate sections, angled pages, poorly reproduced pages, upside down pages, pages out of order, graphs and diagrams that had been corrupted, Final Reports that we had that had not been apparently published by HAL and missing reports.

### **Step 2 - Developed HAIDB Archive Manager to track all work and be the repository for copyright information.**

HAIDB Archive Manager was a purpose built Filemaker Pro database to track all work and be the repository for copyright information. Archive Manager was also used to produce the index of articles for Potato Australia and Eyes on Potatoes, and HAL Final Reports that was included in the 2005 edition of Potato Australia. The following screen shots from the database show the information collected and produced.

# Archive Manager

Find Find all Replace Duplicate Sort Print page

ResID  
1000244

New item PA Index About AM

Delete

---

Item

**Title** Industry priorities for Research and Development

**Publication** Potatoes 2000 Conference Proceedings **To be archived**  
Unknown

**Publication date** Jan 2000

**Publisher** South Australian Research and Development Institute

**Copyright holder** Organisation ISSN 0 7308 5259 8

**Topic** Industry management

**Volume** **Page** 1 **Item number** **Conf/Worksh**

**Project number**

---

Author/s

**Author/s** Milton Rodda

**Contact author name**

**Title**

**Organisation**

**Address**

Copy to copyright holder

**Town** **State** **P/icode**

**Country**

---

Copyright holder

**Name**

**Title**

**Organisation** South Australian Research and Development Institute

**Address** South Australian Research and Development Institute  
Plant Research Centre  
GPO Box 397

**Town** Adelaide **State** SA **P/icode** 5001

**Country** Australia

---

Notes

**Potatoes 2000 Conference Proceedings**

Adverts, Title Pages and Table not in .xml (Folder containing all missing pages = Potatoes 2000  
Pages missing from XML - adverts, titles, & table) all pages added as images to final pdf. All

Find

Browse

Archive Manager Archive Manager Archive Manager Archive Manager

Figure 2: Information about Final Reports was downloaded from HAIDB Project Manager (National Database System) and entered directly for the other publications. The Final Report information was then edited as required as the publication details were not always the same as the information from the HAL database. The final title may have been changed by the Chief Investigator and in some cases Chief Investigators for projects changed or more than one author was indicated.

**Progress** ◀

[Explanation of the process](#)

**Item**

**Title** Industry priorities for Research and Development **To be archived**  
Unknown  
**Publication** Potatoes 2000 Conference Proceedings **ISSN** 0 7308 5259 8  
**Publication date** Jan 2000  
**Publisher** South Australian Research and **Volume** **Page** 1 **Item number** \_\_\_\_\_  
**Copyright holder** Organisation **Project number** \_\_\_\_\_

**Situation** Except for HTML only enter information once for a publication - Enter for first article or paper

**No of pages \*** **Pictures done** No **Chapters for Final Reports** 1 Default 1 **Data page**  1st item

**Originals** Only enter information for the first article or paper

**Have**  **Word file**  **Yes**  **PDF**  **No**  **Hard copy**  **Yes** \* Processing that needs to be done

**PDF Image (Scanned)** Only enter information for the first article or paper for conferences, PA and EOP

**Done**  **No** **Size in KB** **or MB** **Filename**

**OPD /TIFF folder (Processed)** Only enter information for the first article or paper for conferences, PA and EOP

**Done**  **No** **Size in KB** **or MB** **Filename**

**XML (Processed)** Only enter information for the first article or paper for conferences, PA and EOP

**Done**  **Yes** **Size in KB** 2958 **or MB** **Filename** Conf 2000 Potatoes 2000.doc

**PDF - text + images (Processed)** Only enter information for the first article or paper for conferences, PA and EOP

**Done**  **Yes** **Size in KB** 7350 **or MB** **Filename** Conf 2000 Potatoes 2000.pdf

**PDF - CD ready (Final)** Only enter information for the first article or paper for conferences, PA and EOP

**Done**  **Yes** **Size in KB** 5768 **or MB** **Filename** Conf 2000 Potatoes 2000.pdf

**PDF - Internet ready (Final)** Fill in for all articles, papers and chapters.

**Done**  **No** **Size in KB** **or MB** **Filename**

Figure 3: Each stage of the process was tracked and details of file names recorded. This was an important function given the large number of files being dealt with and the need to be able to monitor progress quickly. The Archive Manager structure was also replicated in the folder system used to store the files. Working files would be stored on local computers and when the work was completed, the files would be saved to the server. Both local computers and the server were backed up nightly. One person managed the file system to minimise the possibility of mixing up files.



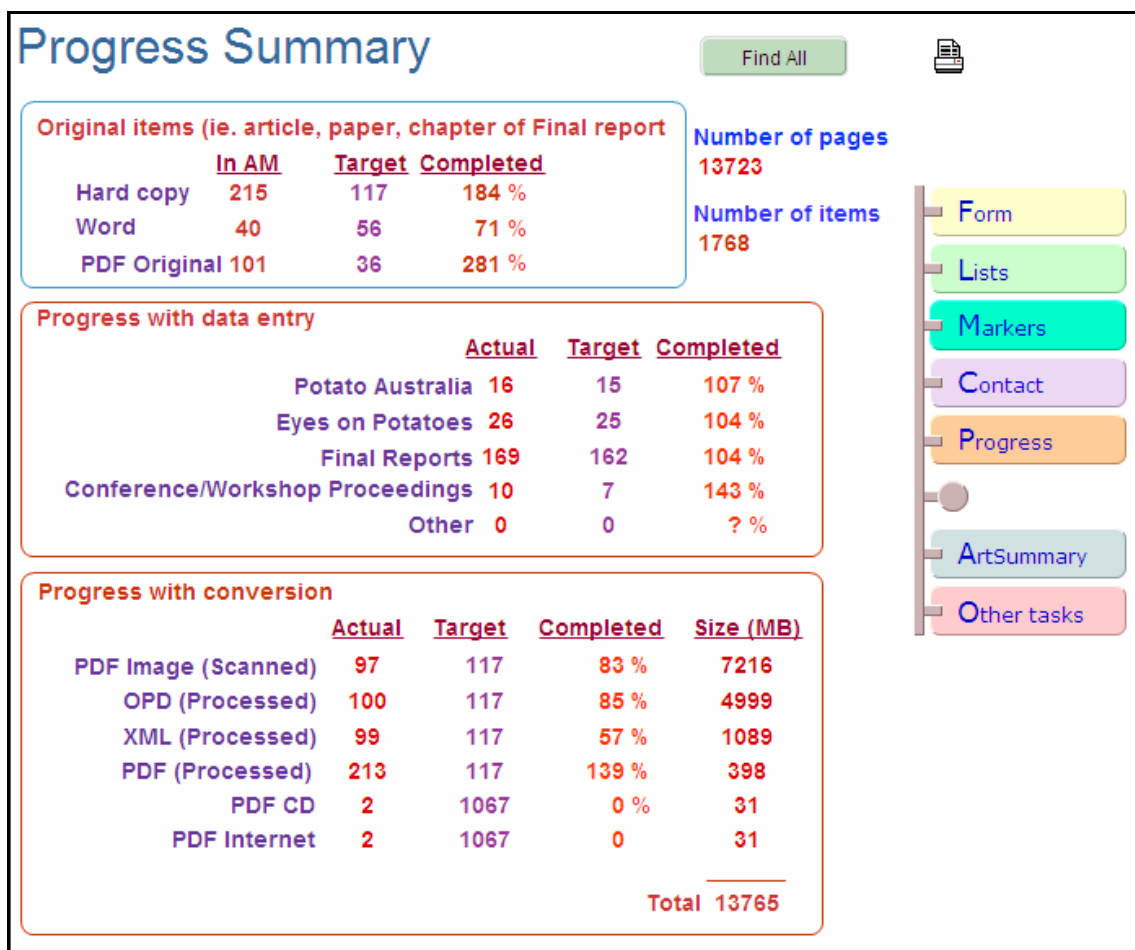


Figure 4: This screen summarised the current work situation at any time.

**Step 3 - Scanned in all material as PDF images for which we did not have a digital copy.**

The South Australian Farmers Federation (SAFF) Konica 7165 (black and white) and Konica 7920 (colour) office printer/scanners had capacity for high speed scanning. Most documents were dismantled and scanned internally and PDF Image files produced.

Conference proceedings were contract copied using the book copiers at the Waite Campus of the University of Adelaide which allowed copying right to the edge. These copiers allowed books to be opened 90 degrees and copied without breaking the spine. The photocopies were then scanned through the SAFF system to produce the PDF Image files.

**Step 4 - Used Optical Character Recognition to convert the PDF images to an intermediate format that separated images from text and converted text into a format that could be edited by a word processor.**

Initially Adobe Capture 3.0 was purchased for the OCR work but we found it difficult to use and the error rate was unacceptably high with the material we were processing.

The error rate in OCR was an important issue when converting thousands of pages. A 10% error rate when converting 10,000 pages was about 1,000 pages of corrections. These may be clustered or spread throughout the document. Quoted error rates were not always a good guide as different OCR programs responded differently to problems experienced.

Omnipage Scansoft Office 12 (later upgraded to 14) was then purchased which greatly reduced error rates and was much easier to use but very demanding on the computer processor.

In early 2005 we were processing a large quantity of PDF Images but it was taking a long time. Our main processing computer consisted of a Pentium 4 2GHz computer with 1GB RAM, 128MB video card, 17 inch high resolution (1280 x 1024) LCD monitor (with zero dead pixel warranty) and 80GB HDD (later upgraded to 160GB).

We investigated whether any other OCR packages had been released since our first assessment and found a package called Abbyy FineReader from a Russian company. An evaluation copy was downloaded from the company's internet site and tested.

It quickly became clear it processed our files more quickly, had better overall accuracy in zoning and recognition, was relatively easy to use and had a good range of features. A commercial copy was obtained and all OCR work carried out on Abbyy FineReader 7.0 Professional Edition from that point forward.

A number of issues confronted us in the OCR process:

- In some old publications, particularly conferences, old fonts were used that were not in our font library on the computer. These were purchased from MyFonts ([www.myfonts.com](http://www.myfonts.com)).
- Although layout features were recognised and zoned by the OCR program, the process was not perfect, requiring manual intervention. Advertisements in particular caused quite a few difficulties so most were zoned as images.
- Due to the poor quality of some original reports, the OCR had difficulty interpreting the print, resulting in errors. So as the quality of print declined, the error rate increased. This greatly slowed down processing time with some reports.
- It took a long time to process large documents. Work needed to be planned so other jobs could be done while the processing computer did its job.

#### **Step 5 – Checked and corrected – 1<sup>st</sup> check.**

Although a lot of checking was done on screen, it was sometimes easier to check documents printed out. At this stage the focus was on layout.

#### **Step 6 - Converted the intermediate format into XML.**

The 'Automatically increase page size' option was unchecked and the image (JPEG) quality set to 50-300. Documents were then saved as Microsoft Word XML to help in editing.

#### **Step 7 - Edited on screen, printed out and checked – 2<sup>nd</sup> check.**

It was relatively easy in Microsoft Word 2003 to address many of the textual and formatting problems.

#### **Step 8 - Converted the XML files into standard PDF.**

Using Adobe Acrobat Professional Version 7, the Word files were converted into standard Acrobat files. This was the most space efficient of the Acrobat formats. The following were the main settings used:

- Fast web view – Yes
- Tagged PDF – No (Due to size restrictions with CDROM)

#### **Step 9 - This was printed out, checked and the PDF recompiled if changes needed to be made – 3<sup>rd</sup> check.**

This was the third and final check unless more changes were needed. At this stage most documents had been checked by three different people. Sometimes the PDF process threw up some unusual results which had to be tracked back to the cause - usually some aberrant formatting commands.

### **Step 10 - Navigation links were added to the PDF Contents page**

Adding navigation links using Adobe Acrobat was straightforward but relied on true page numbering as opposed to actual page numbering. The former starts at the first page of the document as opposed to the latter which could start after the contents or before the contents page.

### **Step 11 - Open settings and password security added**

The PDF open settings were standardised and then using the batch option in Adobe Acrobat Professional encryption was placed on all files. The main aim was not to stop information being printed or copied but prevent the files being altered. (For documents supplied as PDFs, document settings may differ slightly.) The open settings and security settings were as follows:

- Document Properties - Initial View – Document Options – Show – Page only
- Document Properties - Initial View – Document Options – Page layout – Continuous
- Document Properties - Initial View – Document Options – Magnification – Fit Width
- Document Properties - Initial View – Document Options – Open to – Page number: 1
- Document Properties - Initial View – Windows Options – Show – Document Title
- Document Properties - Initial View – User Interface Options - Hide menu bar
- Security Method: Password Security
- Document Password: No
- Author Password: Yes
- Encryption Level: High (128-bit RC4) – Encrypts all document contents including metadata
- Accessibility: Allowed
- Fill and Sign: Not Allowed
- Extract Text and Graphics: Allowed
- Document Assembly: Not Allowed
- Changing the Document: Not Allowed
- Content copying or Extraction: Allowed
- Printing: Allowed
- Print High Quality: Allowed

### **Front end**

The front end or user interface of the package was produced using Filemaker Pro 8 Advanced. This allowed a simple user interface to be developed and compiled as a standalone program. The key components of the user interface consisted of:

- Search engine and listing
- Publication lists as an alternative to accessing PDFs
- Getting started guide to help users quickly learn how to use the program
- Background information on Potato Archives
- Legal issues which discuss copyright ownership
- Acknowledgements
- Settings which allow the type of PDF reader to be selected

### **Program map**

Opening page

Search (Home) page

- Search engine
  - Search list
    - Publication PDF*
- Eyes on Potatoes (EOP)
  - EOP PDF*
- Final Report
  - Final Report categories
    - Final Report list
      - Final Report PDF*
- Potato Australia (PA)
  - PA PDF*
- Conference proceedings (CP)
  - CP PDF*
- Getting started
  - the beginning
  - searching
  - Acrobat
- Background
- Legal issues
- Acknowledgements
- Settings
- Help
  - PDF will not open
  - Page size
  - Other

### Accessing the PDFs

The following screenshots show the main screens from which publications can be selected.

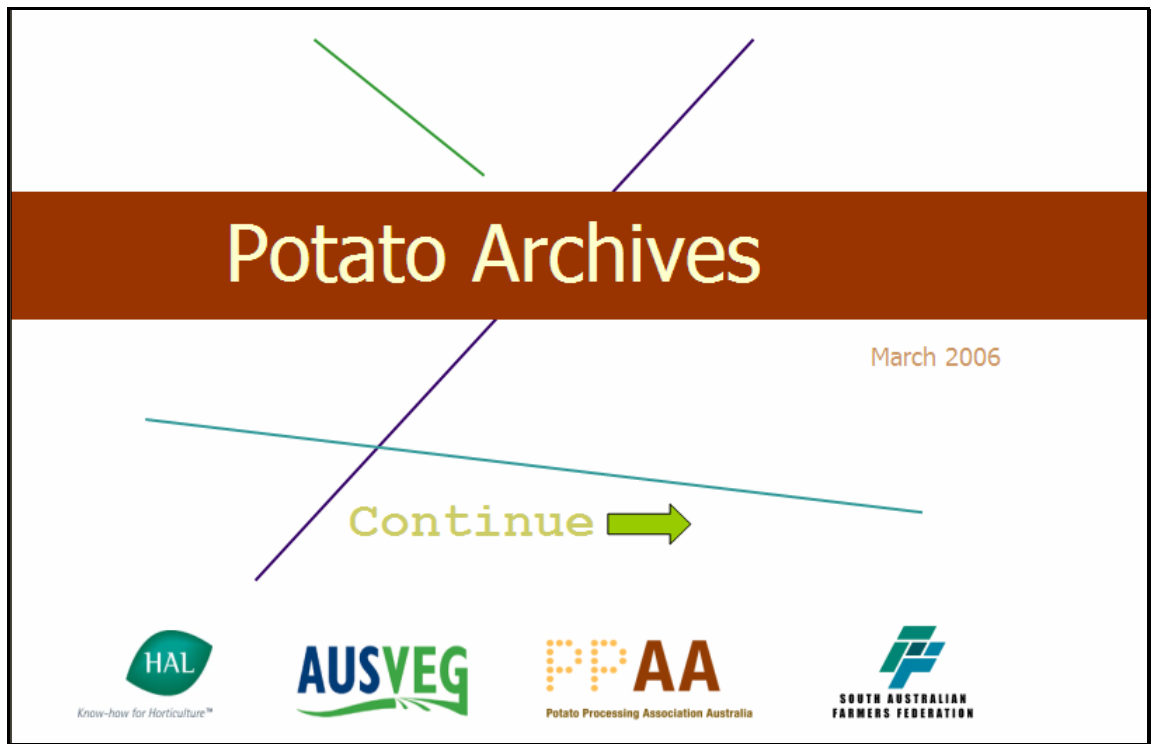


Figure 5: What users see when Potato Archives was first opened

# Potato Archives

[Eyes on Potatoes](#)

[Potato Australia](#)

[Final Reports](#)

[Conference proceedings](#)

**WARNING**

These historical archives contain recent and old documents produced between 1990 - 2005. Always seek professional advice before using the information in the archives, especially when it involves important management decisions.

Many older documents had to be converted from paper to electronic form. This process is quite accurate and the documents have been checked, but there may be occasional errors. Duplicate and blank pages have been omitted which is sometimes reflected in the page numbering. Blue links have been added in some situations to aid navigation.

[Getting started](#)

[Background](#)

[Legal issues](#)

[Acknowledgements](#)

[Settings](#)

© Horticulture Australia Limited 2006

Figure 6: Main page with an example search topic

# Search results



35 items found

1	Potato Australia Jun 1991	<a href="#">Common scab studies</a> ed: John Salvestrin, Jonathan Eccles
2	Potato 2005 Aug 2005	<a href="#">Detection and quantification of pathogenic streptomycetes SPP from soil that cause common scab of potatoes</a> Nigel Crump, Tonya Wiechel, Dolf de Boer
3	Final Report 2005	<a href="#">Evaluation and commercialisation of common scab resistant clones of commercial potato varieties</a> Calum Wilson
4	Potato Australia Sept 2005	<a href="#">How much is too much common scab on processing tuber seeds?</a> Hoong Pung
5	Potato 2005 Aug 2005	<a href="#">Identification of Streptomycetes species causing common scab of potato using molecular techniques</a> Tonya Wiechel, Nigel Crump, Dolf de Boer
6	Final Report 2005	<a href="#">International R&amp;D workshop and industry extension meetings on common scab disease</a> Calum Wilson
7	Potato Australia Sept 2005	<a href="#">Solution to common scab at hand</a> Calum Wilson
8	Potato Australia Sep 2004	<a href="#">Common scab threshold on tuber seeds for processing potato crops</a> Hoong Pung
9	Potato Australia Sep 2004	<a href="#">Finding common scab resistant clones of commercial potatoes</a> Calum Wilson
10	Potato Australia Sep 2004	<a href="#">International R&amp;D workshop and industry extension meetings on common scab disease</a> Calum Wilson
11	Potato Australia Sep 2003	<a href="#">Common scab threshold on tuber seeds for processing potato crops</a> Hoong Pung
12	Potato Australia Sep 2003	<a href="#">Development of extreme resistance to common scab</a> Calum Wilson, Alieta Eyles, Robert Tegg
13	Potato Australia Sep 2003	<a href="#">International R&amp;D workshop and industry extension meetings on common scab disease</a> Calum Wilson
14	Eyes on Potatoes Mar 2002	<a href="#">Common scab - lessons from Canada</a> Dolf de Boer, Nigel Crump
15	Potato Australia Sep 2002	<a href="#">Common scab threshold on tuber seeds for processing potato crops (New project)</a> Hoong Pung
16	Potato Australia	<a href="#">Development of extreme resistance to common scab</a>

Figure 7: The list from a search on common scab is in a simple to understand format with the highest priority articles first. (See below for more details.)

# Eyes on Potatoes



- |  |  |
|--|--|
| <a href="#">Volume 1, June 1997</a>      | <a href="#">Volume 15, March 2002</a>    |
| <a href="#">Volume 2, December 1997</a>  | <a href="#">Volume 16, June 2002</a>     |
| <a href="#">Volume 3, March 1998</a>     | <a href="#">Volume 17, December 2002</a> |
| <a href="#">Volume 4, June 1998</a>      | <a href="#">Volume 18, March 2003</a>    |
| <a href="#">Volume 5, December 1998</a>  | <a href="#">Volume 19, June 2003</a>     |
| <a href="#">Volume 6, March 1999</a>     | <a href="#">Volume 20, December 2003</a> |
| <a href="#">Volume 7, June 1999</a>      | <a href="#">Volume 21, March 2004</a>    |
| <a href="#">Volume 8, December 1999</a>  | <a href="#">Volume 22, June 2004</a>     |
| <a href="#">Volume 9, March 2000</a>     | <a href="#">Volume 23, December 2004</a> |
| <a href="#">Volume 10, June 2000</a>     | <a href="#">Volume 24, March 2005</a>    |
| <a href="#">Volume 11, December 2000</a> | <a href="#">Volume 25, June 2005</a>     |
| <a href="#">Volume 12, March 2001</a>    | <a href="#">Volume 26, December 2005</a> |
| <a href="#">Volume 13, June 2001</a>     |  |
| <a href="#">Volume 14, December 2001</a> |  |

Figure 8: The edition of Eyes on Potatoes can be selected from the list as an alternative to using the search engine.

## Potato Australia



<a href="#">Volume 1, June 1990</a>	<a href="#">Volume 9, September 1998</a>
<a href="#">Volume 2, June 1991</a>	<a href="#">Volume 10, September 1999</a>
<a href="#">Volume 3, June 1992</a>	<a href="#">Volume 11, September 2000</a>
<a href="#">Volume 4, June 1993</a>	<a href="#">Volume 12, September 2001</a>
<a href="#">Volume 5, November 1994</a>	<a href="#">Volume 13, September 2002</a>
<a href="#">Volume 6, August 1995</a>	<a href="#">Volume 14, September 2003</a>
<a href="#">Volume 7, August 1996</a>	<a href="#">Volume 15, September 2004</a>
<a href="#">Volume 8, September 1997</a>	<a href="#">Volume 16, September 2005</a>

Figure 9: The edition of Potato Australia can be selected from the list as an alternative to using the search engine.

## Final Reports



- [Crop management](#)
- [Breeding and evaluation](#)
- [Pest, disease and weed management](#)
- [Postharvest](#)
- [Market research and development](#)
- [Seed development](#)
- [Technology transfer](#)
- [Industry Management](#)

Figure 10: The category of Final Reports can be selected which will produce a list of reports. The categories are the same as that used in the Potato Australia HAL project tables.


Search results		
13 items found		
1	Final Report Jan 1997	<a href="#">After cooking darkening of potatoes</a> Adrian Dahlenburg
2	Final Report Feb 2000	<a href="#">Developing a product description language for potatoes</a> Andrew Henderson, Richard Bennett
3	Final Report Aug 2000	<a href="#">Development of a quality assured production and marketing system for the fresh potato industry</a> Eric Coleman
4	Final Report April 2004	<a href="#">Development of a universal grading system for ware potatoes in Western Australia</a> Deborah Pitter
5	Final Report May 2004	<a href="#">Evaluating a product for enhancing dormancy and storage qualities of potatoes</a> Ian Macleod
6	Final Report Jun 1995	<a href="#">Evaluation of the economic opportunities for potato by-product industries in Australia</a> Karen Freeman
7	Final Report Aug 2002	<a href="#">Increasing the opportunities for use of organic wastes in the Tasmanian vegetable industry</a> John McPhee
8	Final Report Sep 2001	<a href="#">Innovative Transport and Disease Control Systems: Potato Exports to Asia</a> Alister Sharp, Barbara Stephens, A.I. Bokshi, Stephen Morris
9	Final Report Mar 2003	<a href="#">Managing bacterial breakdown in washed potatoes</a> Trevor Wicks
10	Final Report Apr 2001	<a href="#">Potato Export Market Development</a> Russell Sully, Ralph Cadman
11	Final Report Jan 1996	<a href="#">Review of potato waste utilisation in Australia</a> Karen Freeman
12	Final Report July 2005	<a href="#">Supply chain handling systems for premium potatoes</a> Adrian Dahlenburg
13	Final Report Jan 1997	<a href="#">To improve postharvest handling, storage and processing quality of atlantic and cadima potatoes</a> Soon Chye Tan

Figure 11: List of Final Reports produced by selecting Postharvest

## How does it work

### (i) Generating the information and loading it into the Potato Archives front end

Information was collected and modified in the National Database System then loaded into the Potato Archives front end.

- HAIDB Project Manager – Contains information from HAL on all projects, which provides the basis for Final Report items.
- HAIDB Archive Manager – Information from HAIDB Project Manager was loaded for Final Reports. Information for other publications was entered manually. All details were checked against the publications.
- HAIDB Word Muncher – Information from HAIDB Archive Manager and HAIDB Project Manager (Industry summaries) was loaded into the database and articles from Potato Australia, Eyes on Potatoes and Conferences copied and pasted in. From this information, primary and secondary keywords were generated.
- The information was then exported to HAIDB Research Manager which was the repository for data required by Potato Archives and the Potato Internet Service. Author details were expanded and all details checked.

### (ii) Opening PDFs

The user interface program opened a selected PDF and in the case of articles within a publication, went to the appropriate page.

In Filemaker Pro the ‘Send Event’ command was used to carry out the action. The command line consisted of:

Acrobat Reader executable file name + Switch + Adobe Open parameter + File location

For example:



AcroRd32.exe /A "page=45=OpenActions" "C:/My Databases/Idb 2001/Potato Archives/Files/P12.pdf"

AcroRd32.exe – Adobe Acrobat Reader (Either the Reader or the full version could be selected from the Settings page.)

/A – Switch for passing parameters from the command shell to the Adobe open parameter command.

“Page=45=OpenActions” – Adobe Open Parameter command which provides the ability to go to a selected page in the PDF after it was opened. In this case go to page 45.

"C:/My Databases/Idb 2001/Potato Archives/Files/P12.pdf" – Location of the PDF file. This is determined by Filemaker Pro working out the folder containing the Filemaker program and then assuming that the PDFs are in the Files folder contained within this folder.

### **(iii) Search**

To facilitate the search process, each database item contained information imported from HAIDB Research Manager. This consisted of the following:

- The four list items as text fields (ie. Publication, title, date, authors)
- Publication date (ie. as a date field)
- Topic (eg. Crop management – See Figure 10)
- Open parameter (provides page number)
- Primary and secondary keywords
- File name

### **(iv) The search list**

The search list consists of the number of items found (count of found records), list item (record number) and list elements. The list elements consist of:

- Top left – Publication
- Top right – Title
- Bottom left – Date
- Bottom right – Authors

The list elements were all exported from HAIDB Research Manager as text. The layout used was developed after viewing a large number of list formats and selecting a simple and easy to read approach.

### **Field testing**

No formal field testing was carried out as we ran out of time. The draft Potato Archives was used extensively by project staff and feedback used to refine the user interface. We also showed it to a number of individuals at the Potato 2005 conference.

Feedback can best be summed up by the Chairman of Board for the South Australian Farmers Federation who at the end of a demonstration said farmers would greatly appreciate how easy the product was to use.

Most comments were very supportive of the approach used and offered no suggestions for improvement. The most common comment was - when would it be available.

Those offering suggestions focused on cosmetic aspects of the user interface or why the Enter button had to be pressed twice to initiate a search. The latter was a limitation of the software being used.

Some people viewing the draft version thought the user interface was not as modern as it could be or could be organised differently, particularly the search list which was seen by some as not conforming to normal industry practice. Most people did not comment on the design at all so the people who did

comment were seen to be reflecting a personal style preference rather than a user interface usability problem.

## **Marketing**

The marketing of Potato Archives began in the project PT96009 when the concept was developed. Marketing can therefore be broken up as follows:

1. Awareness
2. User interface design
3. Production
4. Distribution
5. Pricing
6. Promotion

### **(1) Awareness**

Development of the Potato Archives started in 1997 in project PT96009 from a desire for people to be able to access information more easily when they needed it. This early phase was also an iterative process involving developing a concept to address peoples' needs then testing the concept with people as I moved around the country. By the time I had finished my group sessions and interviews as part of the project PT96009, I had the basis for a project proposal to submit to Horticulture Research and Development Corporation.

This early work also became the first stage of marketing the idea of Potato Archives as testing aroused interest and a desire to see such a product produced.

Once the project was funded, regular reporting in *Eyes on Potatoes* and *Potato Australia* resulted in a growing awareness of Potato Archives. A special insert as part of the project PT98037 in *Eyes on Potatoes – Making life easier with information technology* – also discussed the concept of Potato Archives and how it might be used by growers.

At the Potato 2005 national potato conference at Phillip Island, Victoria Potato Archives was demonstrated and received a good reception from conference attendees.

### **(2) User interface design**

The initial concept for Potato Archives was borne from industry needs and the user interface designed to address those needs.

Discussions carried out in projects PT96009, PT00001, PT98037 and VX00023 clearly indicated the need to be able to access information easily and quickly. Many potential users were not interested in the technology *per se*, only what it could do for them. Computer jargon and common computer practice also had no importance if it did not help them access the information.

The computer interface was therefore kept very simple and used terms, as much as possible, that most people using the product would be familiar with. The Final Report categorisation was identical to that used in *Potato Australia*.

The list format was designed in conjunction with PT04002 so a common, simple listing approach could be used for a range of industry services. Author names were expanded to make it easier for users to recognise the authors. Information was kept to a minimum to reduce clutter. Font size was kept large to make reading easier.

The Open Parameters for the PDFs focused on simplicity. Most features not likely to be needed were initially hidden. Links were added to the contents page to help movement around the document. Support information was included to help people get started (particularly understanding the navigation features) and if they ran into problems.

### **(3) Production**

The master files consisted of a folder of PDF files and a runtime copy of Potato Archives user interface produced with Filemaker Pro 8 Advanced with the following settings:

- Create Runtime solution application(s)
- Remove admin access from files permanently
- Custom image (displayed with the Filemaker details at the close of the program. This included the wording: A potato levy project – The Australian Government proudly supports research and development by matching the industry levy.

Surplus language and dictionary files were removed to reduce the size of the runtime application.

The development of the installation program and production of the final product was managed by ARRIS. This involved developing and testing the installation program for Windows 2000 and Windows XP operating systems including an option for custom install, registering program in Windows Registry, shortcut in Start menu and an option for a shortcut on the desktop; design work (based on user interface and powerpoint prototype), printing, duplication and sending out the required number of copies to state distributors and selected people.

The specifications to run the program were based on minimum specifications for Filemaker Pro 8. These were:

- Pentium III 500 MHz or higher
- Windows 2000 (Service Pack 4), Windows XP (Service Pack 2)
- 256MB RAM (more is better)
- CD-ROM drive and hard disk drive
- SVGA (800x600) or higher resolution video adapter and display
- At least 500MB of hard disk space for storing the program and files. (Potato Archives needs to be installed on a hard disk drive)

#### **(4) Distribution**

The plan was to offer a copy to Levy Payers free of charge while other people could purchase it through AUSVEG. Copies were sent out through the same distribution system used for Eyes on Potatoes and Potato Australia.

**Table 1: Potato Archive’s CDROM distribution**

<b>Copies</b>	<b>Distributors</b>
425	Ag-Challenge, Vic
235	AUSVEG (Commercial distribution for all saleable copies)
245	Department of Primary Industries, NSW (Bathurst)
285	Growcom, Qld
200	Potato Growers of Western Australia
300	South Australian Farmers Federation
395	Tasmanian Farmers and Graziers Association
15	Individually posted
<b>2100</b>	<b>Total</b>

#### **(5) Pricing**

The base price was set by the Potato Industry Advisory Committee. Levy Payers would receive the first version free and others could buy a copy from AUSVEG; in Australia for \$55 including GST and distribution and \$65 for overseas buyers (as per discussions with AUSVEG). HAL decided that Potato Archives be sold through AUSVEG rather than HAL, as originally proposed.

#### **(6) Promotion**

For non-Levy Payers, a brochure advertising that Potato Archives was now available was sent out with an industry booklet (not related to Potato Archives) while Levy Payers received a copy of the industry booklet and Potato Archives.

Given the Chief Investigator resigned before product release, all future promotion would be handled by AUSVEG.

## Discussion

Although the end result was very worthwhile, the process to achieve it was incredibly laborious. Optical Character Recognition technology, despite the marketing hype, still had a long way to go before tasks such as this one became acceptable as normal business practice.

### Current situation

The following summarises the process of putting together an archival system such as Potato Archives where a lot of material is not in digital form.

**Preparation** – Slow but unavoidable.

**Scanning** – Quick with today's high speed printer scanners that can produce PDF images. It is important to have access to a book scanner for publications that cannot be dismantled.

**OCR** – Slow for old publications with poorly defined layout. Newer publications developed with word processors or page layout programs generally process more quickly. OCR packages such as Abbyy FineReader and Scansoft Omnipage are improving all the time so this part of the process should become simpler and faster. Even so, OCR is a computer intensive process so a fast computer with plenty of memory is important. Key features of a good OCR software package are accuracy, speed and ease of use.

**Editing** – No shortcuts here. If OCR is accurate, this process becomes easier and quicker.

**Creating PDFs** – Quick and simple. Missing fonts can be bought (need to watch royalty issues).

**Setting open parameters and security** – Quick with batch processing

**Contents page links** – Simple but a slow and boring task.

**Checking** – Slow with few options for improvement if quality is not going to be compromised.

### Future processing

In the future there should be no need for conversion from hard copy as electronic files should be available for all publications. This will greatly simplify the process of including publications in the archives.

The challenge then becomes one of managing an efficient archival system which is not a simple process and requires careful planning. This would involve:

- Identifying publications to be included.
- Establishing a simple process for including new publications.
- Having clear guidelines about how files are to be supplied and to who they are supplied to.
- Clear policy on how the files are used.

All processes would need to be transparent and independent. Electronic files have raised many issues with regard to copyright and concern over how the documents are used. Some concerns may be unfounded and simply part of the learning process people need to go through as we enter the Information Age where information is much more widely available.

However, people's concerns are real to them whether unfounded or not. Some companies have approached this by having clear policies that reflect practices that are currently acceptable to their customers. Others have ignored their customer and suffered community and regulator backlash, particularly on privacy issues.

When working with new technologies, most of these problems can be easily overcome by forward planning and being sensitive to customer concerns.

### **Storage format**

Initially when the project was conceived, it was envisaged that HTML or XML would be used for a storage format.

Converting documents to HTML was problematic with software at the time and was abandoned after initial testing.

XML offered a real option but the size of files excluded it as a possibility with internet transmission speeds available at the time to many users. In future it may replace PDFs as the format of choice. To ensure minimal reworking in the future, most publications were also saved as XML documents.

The PDF format was the most flexible and efficient format at the time. The tools to generate final files were also well developed and easy to use. The only real limitation was it was a proprietary format which limited what could be done without entering into licensing arrangements.

### **Search engine**

The most important tool for an electronic library is the search engine. Its prime role is to find what you request. Not always an easy feat!

Search engine technology has evolved very quickly. Think of the type of results that could be achieved with Google in 2005 compared to other search engines five years earlier. A lot of these changes have been subtle but have resulted in a more satisfying experience for the user. Even with recent advances, today's search engines are still in their infancy.

The search engine in Potato Archives used the search facilities of Filemaker Pro 8 rather than a conventional web search engine. The approach provided a faster and more satisfying result. It was also designed to be compatible with the new Potato Internet Service.

The search approach reflects the long term goal of having a secure, scaleable and fully searchable extranet. Web search companies 'crawl' sites indexing information about the site for later use in searches. Current 'crawlers' cannot gain full access to the information on a site especially if much of it resides in databases.

To get around this problem and provide a more satisfying search experience, the Potato Internet Service, which the Potato Archives would ultimately reside on, would combine a dual indexing system based on conventional database indexing approaches combined with data from the 'crawled' contents on the site.

Indexing is also an important part of the search process. A major advance that will occur over time is development of index information that can result in more intuitive searching.

As a stopgap until these and other more advanced tools became available, Potato Archives used keywords generated from a purpose built keyword generator and dual keyword fields. The advantage of this approach was that the search outcome better emulated what people wanted compared to a simple keyword search.

Looking further forward, the potential exists, through relationship building with State departments of agriculture as proposed in the original concept of the Potato Internet Service, for external publications to also be included using a form of domain searching.

What this would involve is the selective crawling of specified sites to produce a potato index. In essence, the potato publication resources of the state departments could become part of the Potato Archives system without them residing on the industry's server. The advantage over conventional

search engines would be the information would come from clearly defined sources. This creates a powerful alternative to using the broad based search engines such as Google or Yahoo.

The potato industry has taken an important first step in digitising its information resources. Over time, the ability to access information will become accepted as normal practice resulting in users wanting information that is currently not in the Archives to be included. As more information is included locally, the opportunity to establish a distributed library system may evolve where different groups pool their potato information resources by linking computer databases using domain searching or a variant of this approach.

As raised in this report, the possibility exists of the CDROM version of Potato Archives to link with the internet version to overcome transmission speed issues downloading information.

The other possibility is to index the potato information on your computer and have it included in the search. Your index would only be accessed by you (unless you want others to be able to use it) but it would become a local part of the information resource the Potato Archives search engine could access.

Taking the next step, a group of people who work together could have sections of their hard disk indexed and shared. Linked in with the Potato Archives search engine, the user would have access to the industry resource as well as their group resource.

In overview, any form of electronic information can be used. The forms we currently use consist of text, pictures, video and music. The reality is that this is the 'tip of the iceberg' of what could be included.

In future any electronic information could be made accessible for indexing. This could be information stored on the computer about the farm, financial records, paddock records or local information databases.

External information could be crop statistics, crop monitoring data from automatic weather stations, supermarket information or location of bins in the supply chain (via RFID tags).

Searching therefore becomes an all inclusive exercise (Google through Potato Archives), an industry exercise (Potato Archives), group exercise (using a personalised form of Potato Archives) or a personal exercise (again using a personalised form of Potato Archives).

Some of this may sound fanciful but the reality is that all that has been discussed can be done now with existing technology.

## Recommendations

Potato Archives has the potential to expand considerably from its current form.

The following recommendations are therefore based on the vision for the future evolution of the product:

1. The linking of the CDROM version and the internet version of Potato Archives be explored to capitalise on the benefits each of the technologies have to offer and most importantly to maximise speed of access for most users.
2. That the range of publications to be included in the future on Potato Archives be explored with state departments of agriculture and industry organisations. The potential benefits for industry are enormous if R&D information is consolidated into a centralised electronic library that can be easily accessed. This may be a physical local library or a virtual library using such techniques as domain searching.
3. An industry workshop be held in one to two years time to map the future evolution of Potato Archives. The forum should include representatives from each industry stakeholder group to ensure the breadth of possibilities is explored. Technical specialists also need to be included so the full range of technological possibilities are considered.
4. HAL save all new Final Reports in a form suitable for inclusion in Potato Archives to reduce the amount of extra work.
5. The Potato Archives CDROM be updated in 3-5 years time.
6. Our present mindset for reporting has been constrained by the tools that we use. In the not to distant future there is no reason why Final Reports couldn't contain video and interactive features with the option of accessing different levels of detail based on user needs. Potato Archives need not be limited to only textual content.
7. HAL explore a broader format for Final Reports that encourages use of new technologies to enhance the user's learning experience and the value gained from the Final Reports.

## Acknowledgements

Potato Archives would not exist without the authors who contributed to the publications for which we owe our thanks. Many authors are from state departments of agriculture, CSIRO, Universities, Cooperative Research Centres, industry organisations, farms, processors, wholesale markets, service companies, and various government groups.

Thanks also to the authors who assisted in resolving problems with Final Reports. There were many small issues for which people responded promptly too to resolve which was appreciated.

A number of conference and workshop proceedings were obtained and processed but did not go into the CDROM version of Potato Archives. These will require copyright permission to be obtained from the authors before they can be used in Potato Archives. We would like to thank Chris Williams, John Fennell and Ken Jackson for providing copies of the conference proceedings for us to copy and process.

Thanks also go to past staff (Sharon Baker, Jonathan Eccles, John Oakeshott) and present staff (Danika Houghton, Liz Romanos, Simon Drum, Phil Roeth, Karen Hellwig and Richard Stephens) of Horticulture Australia Limited who provided so much support. Their contribution was critical to the success of the project.

In most cases copyright for the publications is owned by Horticulture Australia Limited. Where this has not been the case, permission has been kindly granted by the relevant organisations or individuals for including the publications or articles in Potato Archives.

Thanks also go to the management of SAFF for the ongoing support especially as our team and space requirements expanded rapidly and every effort was made to accommodate our needs with minimal disruption to our work.

Finally and very importantly thanks to those industry people who clearly articulated their needs which provided the idea for developing Potato Archives.

The project has been funded by the potato levy and matching contributions from the Australian Government through Horticulture Australia Limited.



## **Bibliography**

- Adobe Solution Alliance (2004) Adobe Acrobat 7.0 – PDF Open Parameters, July 11, 2005
- Jarosz N (Unpublished) Facilitating the introduction of electronic information products and services to the Australian Potato Industry, Horticulture Australia Final Report PT98037
- Walters LJ (2006) Coordinating technology transfer in the Australian potato industry, HAL Final Report PT96009 – in press
- Walters LJ (2006) Implementing the Potato Industry's communication plan, HAL Final Report PT00001 – in press
- Walters LJ and Feutrill C (2002) National internet service for the Australian vegetable industries, Horticulture Australia Final Report VX00023

## Appendix A: Publications included on the Potato Archives CDROM

Eyes on Potatoes (Vol 1 – 26)  
Potato Australia (Vol 1 – 16)  
Potatoes 2000 conference proceedings  
Potato 2005 conference proceedings

### HAL Final Reports

<b>Project number</b>	<b>Title</b>
HG97007	National strategy for the management of western flower thrips and tomato spotted wilt virus
HG97038	Fungal identification course at CABI Bioscience, UK 1998
HG98036	Commercial development of biofumigation technology across the horticultural industries
PT00010	Potato varietal evaluation for Western Australia's fresh and export markets
PT00019	Management of tomato spotted wilt virus in potatoes
PT00033	A comparative evaluation of different materials used for cut potato seed treatments
PT00034	Communication of management strategies for potato virus diseases in Western Australian potato crops
PT00035	Market research for potatoes in the food service sector
PT003	After cooking darkening of potatoes
PT004	Potato irrigation - development of irrigation scheduling guidelines
PT006	Rapid identification of <i>Streptomyces</i> spp. on potato, the key to integrated management of common scab
PT008	The control of volunteer potato plants in subsequent vegetable and poppy crops
PT009	Development of a test for potato leafroll virus (PLRV) & determination of PLRV strains in South Australia
PT01001	Control of black dot in potatoes
PT01008	Monitoring and developing management strategies for soil insect pests of potatoes
PT01019	Prediction and molecular detection of soil-borne pathogens of potato
PT01020	Evaluation and commercialisation of common scab resistant clones of commercial potato varieties
PT01030	Seed potato handling and storage - implementing best practice
PT01031	Enhanced detection of potato cyst nematode and bacterial wilt to improve market access for the Australia and New Zealand Potato Industries
PT01032	Breeding Australia's potato germplasm: the resource for varietal development
PT01033	Potato variety evaluation, commercialisation and adoption: Interim project
PT01038	Evaluating a product for enhancing dormancy and storage qualities of potatoes
PT01040	Pilot commercial crop monitoring for pests and diseases in WA seed potato crops
PT01042	Potato pink rot control in the south east of South Australia
PT01044	Evaluation of internationally bred potatoes
PT01048	Virus testing of early generation certified seed potato crops in Western Australia
PT01050	Putting the steam back into the potato market from commodity to product - a consumer and market study
PT01051	Factors affecting specific gravity loss in crisping potato crops in Koo Wee Rup, Victoria

PT011	Increasing potato productivity in Queensland by cultivar development and improved management
PT012	Soil fertility management in potatoes on the Atherton tablelands
PT015	Improved productivity of the french fry industry in Victoria
PT016	Yield and quality effects of post planting applications of nitrogen & potassium to potatoes
PT017	Potato variety evaluation for local, export and processing marketing Western Australia
PT018	Seed potato industry development (aphid monitoring)
PT02001	Biodegradable plastics: The potential for Australian potato as an input for biodegradable polymers
PT02008	Potato industry workshop and Pre-tour to South Africa and First Southern Hemisphere Potato Workshop, South Africa Study Tour and Paper
PT02009	Evaluation and development of new potato genotypes in South Australia
PT02013	International R&D workshop and industry extension meetings on common scab disease
PT02015	Supply chain handling systems for premium potatoes
PT02017	Effects of potato seed characteristics on seed-piece breakdown and poor emergence
PT02018	A partnership to build crisping potato capacity of West Java and Australian seed potato sales
PT02033	Review of potato research & development program
PT02036	Disease management of potatoes on Kangaroo Island
PT02039	Study tour for Kangaroo Island seed potato growers
PT02041	Fresh Product Retail Service - Potatoes
PT02042	Development of an industry plan for potato growers in South Australia
PT02045	The monitoring of potato crops for insect movement on a district scale
PT02047	Minimising virus infection in early generation seed potato crops in Western Australia
PT02048	Developing a pests and disease crop monitoring program for Western Australian seed potato crops
PT02051	Development of a universal grading system for ware potatoes in Western Australia
PT021	Soil Insect Pests of Potatoes
PT022	Improving cosmetic quality of ware potatoes
PT023	An investigation of black dot disease of potatoes and its control
PT024	To improve postharvest handling, storage and processing quality of atlantic and cadima potatoes
PT03026	Potato evaluation trials - Victoria
PT03028	Potato evaluation trials - McCain foods (Aust) Pty Ltd
PT03043	Ballarat potato drip irrigation study tour
PT03055	Market research for potato nutrition software
PT03057	Study tour to the UK and Netherlands to investigate value adding opportunities for potatoes, September 2003
PT03058	Seed potato workshop, Portland, Victoria, August 18 & 19, 2003
PT03061	Improving virus control in seed schemes by combining aphid monitoring and virus testing
PT03067	China study tour & attendance at the World Potato Congress, Kunming, April 2004
PT03070	Variety development for the fresh potato market in Western Australia
PT04013	Seed Potato Certification Workshop 2004
PT04018	Potato evaluation trials - Simplot
PT101	Rapid detection, epidemiology and control of tomato spotted wilt virus (TSWV) in seed and processing tomatoes
PT105	Control of black leg, black scurf and other postharvest storage rots of seed potatoes

PT107	Development of crop management strategies for improved productivity and quality of potatoes grown on highly acid soils
PT108	Decision support software for the nutrient management of potato crops
PT115	Utilising potato microtubers for field production of seed potatoes
PT201	Fresh Potato Marketing Research
PT204	Introduction of French Fry Potato Cultivars to Australia
PT205	Integrated management of potato common scab
PT209	Determination of factors causing stem end browning in russet burbank potatoes
PT212	A national survey of cadmium in potato tubers and soils
PT213	Phosphate, nitrogen and irrigation management in potatoes
PT216	Development of a commercial assessment to detect parasitoids of the potato moth
PT220	Non frozen fresh potato products
PT221	Survey for the incidence of black dot of potatoes
PT223	Potato breeding and cultivar trials in Victoria, Australia
PT230	The development & application of training programs for IPM techniques in SA/integrated crop management program for crisping potatoes in SA
PT300	Attendance at Potato Conference, Madison, USA, 1993 New developments in the control of potato diseases
PT303	Epidemiology and control of powdery scab of potatoes
PT304	Review of potato waste utilisation in Australia
PT314	Evaluation of round potato seed in Queensland
PT315	Rhizoctonia control on fresh market potatoes
PT319	Development and extension of potato hygiene strategies
PT320	The Trace Element Requirements of Vegetables and Poppies in Tasmania
PT336	Implementation of IPM in Northern Australian potato production
PT337	Sustainable potato production in highland areas of Australia
PT338	Production of virus resistant potato plants to enable reduced use of insecticides on potatoes
PT340	Improving international competitiveness of the French fry potato industry in SE South Australia
PT341	Integrated management of early and late potato blights in Australia
PT346	Identification of potato cyst nematode pathotypes
PT354	Use of natural sprouting inhibitors for potato storage
PT402	Sustainable Crop Management for Potato Farms on the Atherton Tableland
PT406	Development of biological control of potato wireworm
PT409	Evaluation of the economic opportunities for potato by-product industries in Australia
PT410	PCR protocols for the detection of chrysanthemum stunt and potato spindle tuber viroids
PT412	Potato early dying in Australia
PT423	A national strategy to reduce cadmium in potatoes
PT424	Integrated crop management & efficient irrigation for crisping potatoes in South Australia 1/7/94 - 30/6/96
PT432	Improved productivity of inland potato production, 1994-1997
PT436	Characterisation and detection of potato cyst nematode
PT437	Development of IPM strategies for potato moth
PT439	Attendance at Potato Harvest 1994 demonstration, United Kingdom
PT447	Integrated management with biofumigation to control soil pests and diseases in potatoes
PT448	An evaluation of the viability of the manufacture of potable spirits (vodka and gin) by the fermentation and distillation of potato biomass
PT452	Fresh potato industry technology transfer: Needs assessment and strategic recommendations

PT509	Irrigation management for crisping potato growers in southern Queensland
PT517	International Plant Protection Conference and organic amendments for potato pest and disease control: A report of a conference and study tour to the Netherlands and USA
PT519	National potato Improvement & Evaluation Scheme 1992-1995
PT520	Molecular markers for PCN resistance
PT538	Comparative benefit cost of IPM and conventional pest management in potatoes
PT603	Refining potato pest management practices in Australia -
PT625	Potato breeding study tour November 1996
PT653	Early production of french fry potatoes on the Swan coastal plain
PT656	National IPM program for potato pests
PT94028	Information packages and decision support software for improved nutrient management of potato crops
PT96005	Potato cultivar accession and testing in Tasmania
PT96014	Development of a quality assured production and marketing system for the fresh potato industry
PT96015	Selection and evaluation of potato cultivars in Queensland
PT96017	Potato breeding & cultivar trials in Australia - Western Australia component
PT96032	Influence of rotation and biofumigation on soil-borne diseases of potatoes
PT96047	Control of black nightshade ( <i>Solanum nigrum</i> ) and other weeds in potatoes
PT97003	More economically and environmentally responsible use of phosphorus fertiliser in potato cropping on krasnozem soils in Australia
PT97004	Potato pink rot control in field and storage
PT97004TR	Overseas Travel Report
PT97010	Sustainable potato production in highland NSW - Stage III
PT97011	DNA fingerprints and cryopreservation of potato cultivars for improved quality assurance
PT97013	Production and assessment of virus resistant potato cultivars
PT97015	New chemical treatments for fungal diseases of seed potatoes
PT97025	SpudNet, an electronic information system for Australian potato growers
PT97026	Developing Soil and Water Management Systems for Potato Production on Sandy Soils in Australia
PT97031	Innovative Transport and Disease Control Systems: Potato Exports to Asia
PT98007	Managing bacterial breakdown in washed potatoes
PT98008	Improving seed potato production
PT98009	Characterisation of Australian isolates of <i>Phytophthora infestans</i> and planning to manage new and more aggressive strains of the fungus
PT98011	Effect of calcium nutrition on decay of summer sown seed-potatoes
PT98018	Cleaning and disinfection practices for potato farms
PT98022	Potato Export Market Development
PT98028	Developing a product description language for potatoes (Final product)
PT98030	Reduced Chemical Usage in Seed Potatoes
PT98036	Biological and chemical control of rhizoctonia
PT98039	Australian Potato Research and Technology Transfer Conference, Adelaide: 30 July - 3 August 2000
PT98042	Evaluation of potato publications
PT99022	An agronomic and economic blue print for using whole, round seed for processing potatoes
PT99052	Potato tuber quality management in relation to environmental and nutritional stress

PT99057	Nitrogen dynamics in commercial seed potato crops and its effect on seed yield, quality, storage and subsequent commercial crop performance
VG006	Effects of soil conditions and fertilizers on cadmium in vegetables - a national approach
VG011	The reduction of cadmium contamination in Tasmanian vegetables and poppies
VG407	Improved irrigation management in potatoes and onions
VG97011	Remote sensing as an aid to horticultural crop recording and husbandry
VG97050	Biofumigation - bioactive brassica rotations for IPM of soil borne pests and diseases
VG98076	Screening potato and vegetable soil borne diseases that may be controlled by eucalyptus leaf mulch - pilot study
VX00012	Enhanced metalaxyl breakdown and its implication in Australian horticulture
VX01026	Building strategic alliances with young Australian and New Zealand vegetable and potato industry representatives
VX99002	Increasing the opportunities for use of organic wastes in the Tasmanian vegetable industry
VX99011	Improved herbicide management for increased yield and quality
VX99040	Coordination Meeting for Reviewing Cadmium Issues in Potatoes & Vegetables