

Bringing Australia's first parliamentary records into the digital age

The Digitisation and Preservation Access Program at the Parliament of New South Wales

By Tanja Zech, Committee Officer, Legislative Assembly, Parliament of New South Wales (formerly Project Coordinator, Digitisation and Preservation Access Program)

The Parliament of New South Wales is the oldest legislature in Australia. It holds records from 1824 onwards, when the First Legislative Council was formed. These documents cover all but the earliest period of Australian settlement. In 2014, NSW Parliament obtained Treasury funding to digitise all pre-Federation tabled papers and make them publicly available on the internet.

The documents span in time the expansion of white settlement across Australia, New Zealand and Norfolk Island, the progressive creation of independent State administrations, responsible government in New South Wales in 1856 with the instalment of a bi-cameral system, the depression of the 1890s and the lead-up to Australian Federation. The Parliament's records are unique and afford insights into social issues of the times as well as into performances on the political stage. They pertain to liquor licensing and pub opening times, train schedules on weekends, limits on immigration and many other socially volatile issues. Many of these topics are still or again relevant in today's society.

The digitisation of these documents will allow unprecedented access to previously relatively obscure material. The decision to place the documents on a publicly available part of the NSW Parliament's website will allow users to conduct independent research without having to engage library services. This in turn will enhance the Parliament's public profile and affirms the relevance of the institution in the wider community.

Digitisation of the records also contributes to their conservation. In their 150 years of existence, the records were not stored in optimal conditions, causing some to become illegible or lost. To prevent further deterioration, the records have been relocated to State Archives after digitisation and are now stored in a controlled environment. As a result, considerable storage space has been freed up on the premises of Parliament House.

This paper sets out the content of the Digital Archive in terms of what was digitised, the process of digitisation and metadata work, the conceptual work behind the online presentation, and describes the online Digital Archive itself.

The Digital Archive

Content

The Digital Archive comprises tabled papers, Votes and Proceedings or Minutes for every sitting day, and, for the earlier dates, correspondence of an administrative nature. It holds documents starting in 1824, when an advisory council to assist the Governor with legislation was established, and continues through to the Australian Federation in 1900. This material is internally divided into papers from the First Legislative Council up until 1855, and papers from the early bi-cameral period from 1856 onwards.

In the course of the project, the Parliament also digitised records of proceedings in the House from this period. This includes full sets of **Journals** for both Houses, consisting of Votes and Proceedings / Minutes and papers ordered to be printed for each session from 1824 to 1900, as well as **Hansard** from 1879 to 1979 and the precursor to Hansard, reports of debates which were published in the Sydney Morning Herald between 1843 and 1879. These records will be made available on different parts of the Parliament's website. The scanning of Hansard complements earlier digitisation efforts, so that a comprehensive set of all NSW Hansard transcripts is now available online.

Tabled papers

In total, the archive contains over 60,000 **tabled papers**. These can be subdivided into different document types, such as annual reports, bills (including bills that were not passed), petitions, returns to order, official correspondence, special reports of inquiry and others. Papers were tabled in both hand-written and printed versions. Some printed documents also contain hand-written annotations. Unfortunately, we do not have a full set of the hand-written papers as some suffered damage in storage. From 1856, papers were tabled in both Houses, which increases the chance that at least one version has survived. This also means that the Digital Archive may contain duplicates where the papers tabled in both Houses have survived. The Digital Archive represents the state of the tabled papers at the time of scanning in 2015, with only a few exceptions where the Houses decided not to include some very large petitions in the scanning process.

For a large part of the period covered by the Digital Archive, incoming **administrative correspondence** addressed to the Clerk was stored and registered together with the tabled papers. The content may range from Members' requests for copies of certain documents and changes to Library arrangements or to the smoking room to applicants seeking employment at Parliament House or even employees requesting the instalment of an oven in the messengers' room.

It was decided to include these letters in the digitisation project. From a practical point of view, it would have been very difficult to physically separate the letters from 'proper' tabled documents prior to scanning. The amount of work necessary and the potential for error precluded this course of action. Therefore, the letters were scanned and metadata was assigned to them. At this stage, the letters could have been removed from the data set more easily. However, keeping them from the public just because they did not fit into a category of official papers would have been a lost opportunity. As it stands, they present an interesting window into the work environment of Parliament and administrative processes a century and a half ago.

Proceedings in the House

The **Votes and Proceedings / Minutes** will be available separately for each day in the Digital Archive. The terminology is different for each House: while the Legislative Council refers to the record of proceedings in the House as Minutes, the Legislative Assembly term is Votes and Proceedings. During the First Legislative Council (1824-1855), they were also called Votes and Proceedings.

In the course of the digitisation project, all **Hansard transcripts** from 1879 onwards were digitised and are now accessible with full-text search on the Parliamentary website. They can be accessed as one record per House per day. This work complements earlier and current efforts to make Hansard available to the public in electronic format. Previously, they had only been digitised from 1979 to present.

For the period from 1843 to 1879, Sydney Morning Herald articles known as **Reports of Debates** will be added. These articles summarise the proceedings in Parliament without giving a near verbatim account as Hansard does. While they are also published on TROVE, the set included in the Digital Archive is more accessible as the articles have been indexed with key words (a digital version of the so-called Marsh index)ⁱ. This resource had previously been published on the Parliament's website but was taken down several years ago. It will be once again accessible via the Digital Archive.

Online presentation

The Digital Archive will be published in two parts on the NSW Parliament's website. Documents pertaining to the First Legislative Council (1824 – 1855) will sit in a separate dedicated location on the website, while later tabled papers and Votes/Minutes are being incorporated into the existing House Papers and Tabled Papers site sections to emphasise the NSW Parliament's structural continuity from 1856 to date. Keeping the First Legislative Council documents in a separate location acknowledges them and the First Legislative Council as separate from the current entity.

Both the current House Papers and Tabled Papers site sections and the First Legislative Council Digital Archive provide in principle the same functionality. Documents are accessible via a browse by date view (Browse), a filtering option allowing users to narrow lists down to certain pre-selected criteria (Filter), and a full-text search option (Search).

The **Browse** tab (see image 1) shows a calendar, displaying one year at a time. Dates appear only if a paper or debate exists for that date. The year can be selected from a slider at the top of the page. Clicking on a date displays all papers and debates for that day below the calendar. The Hansard and House Papers Browse view for 'current' (post 1855) documents displays the House Papers, Tabled Papers and Hansard transcripts for the selected day.

Browse
Filter
Search

17 | 1846 | 1845 | 1844 | 1843 | 1842 | 1841 | 1840 | 1839 | 1838 | 1837 | 1836 | 1835 | 1834 | 1833 | 1

January
Thu 5, Tue 10

February
Fri 10, Wed 22

March
-

April
Tue 11, Fri 28

May
Mon 8, Thu 11, Tue 16, Thu 25, Fri 26, Sat 27, Sun 28, Mon 29, Tue 30, Wed 31

June
Thu 1, Fri 2, Mon 5, Sat 10, Mon 12, Tue 13, Fri 16, Tue 20, Wed 21, Fri 23, Sat 24, Mon 26, Tue 27, Wed 28, Thu 29, Fri 30

July
Sat 1, Sun 2, Tue 4, Sat 8, Mon 10, Wed 12, Thu 13, Fri 14, Tue 18, Wed 19, Thu 20, Fri 21, Tue 25, Wed 26

August
Tue 1, Thu 3, Fri 4, Mon 7, Wed 9, Fri 11, Sat 12, Sun 13, Mon 14, Tue 15, Thu 17, Fri 18, Mon 21, Wed 23, Thu 24, Fri 25, Mon 28, Wed 30, Thu 31

September
Mon 4, Tue 5, Wed 6, Thu 7, Fri 8, Sat 9, Mon 11, Tue 12, Mon 18, Wed 20, Thu 21, Wed 27, Fri 29, Sat 30

October
Wed 4, Thu 12, Mon 16, Tue 17, Fri 20

November
Wed 1, Fri 3, Tue 21, Thu 23

December
Mon 4, Fri 8, Thu 21, Wed 27, Thu 28, Sun 31

Monday August 28 1837

▲ Title	◆ Paper Type	◆ Document Type
Fire and Life Assurance Company: requesting the Fire and Life Assurance Bill to be withdrawn from Considerations of the Council for this session	Tabled Paper	Letter
Lamb, Chairman of the Australian Fire and Life Assurance Company: re the progress of the Fire and Life Assurance Bill through the Council, Deeds, By-Laws etc.	Tabled Paper	Letter

Image 1: Browse View – First Legislative Council (page under development)

The First Legislative Council archive displays the papers tabled on that day and the relevant minutes. From 1843 onwards, it also displays abstracts of parliamentary debates as published in the Sydney Morning Herald. The First Legislative Council documents also include administrative correspondence which was not tabled. These letters are associated with the date on which they were written because no tabling date exists. Therefore, the Browse view may also show Saturdays and Sundays.

In the **Filter** view (see image 2), both environments allow filtering by date range and by Paper Type. Selecting a Paper Type leads to further options depending on the selection. In addition, the Hansard and House Papers site section for ‘current’ (post 1855) documents allows filtering by House. The results are displayed in a list below the filter module and can be sorted by Date, Title, and Paper Type. Filters can be updated at any time to vary the list output.

Browse
Filter
Search

Session:

From Date:

To Date:

Paper Type:

Document Type:

Tabled By:

Results show Tabled Paper for session All and document type Bill

▼ Date	↕ Title	↕ Paper Type	↕ Document Type
31/12/1835	A Bill for lighting with Gas, the Town of Sydney in the colony of New South Wales. Manuscript bill.	Tabled Paper	Bill
30/09/1835	Flour Duty Exemption Act. Printed, numbered Act.	Tabled Paper	Bill
22/09/1835	Wheat Duty Exemption Act. Manuscript bill (two copies); printed, numbered Act.	Tabled Paper	Bill

Image 2: Filter view – First Legislative Council (page under development)

The **Search** is a powerful tool for finding relevant documents in the Digital Archive because all documents have been processed with optical character recognition. This means that the search covers not only titles but can include the content of documents as well, where they are printed. The hand-written parts of documents cannot be searched.

The Parliament's website utilises Microsoft SharePoint native search. At the time of writing, other than for Hansard, search is limited to a single search line which can process search requests with Boolean operators. An advanced search with more control options is being developed. The search operates across all parts of the Parliament's website and can be accessed from multiple locations. Both the Hansard and House Papers tab and the First Legislative Council archive will have access points to the search and can also be selected as separate areas of the site to be searched.

Once a search has been run (see image 3), the search bar with the search request remains displayed at the top, augmented by a drop-down filter to the left, which displays the section of the website that is currently being searched. Below, the search results are listed. Each document displayed in the search results shows a title which is hyperlinked to the actual document, followed by a three-line extract with search hits highlighted in bold text. Below, the location of the document including the file name displays in green.

House Home > Search

Legislative Assembly

Legislative Council

Document Type

Annual Report

Minutes

Questions and Answers

Votes and Proceedings

Notice Paper

SHOW MORE

Asking Member

Maguire, Daryl

Greenwich, Alex

Skinner, Jillian

Gay, Mr

Parker, Jamie

SHOW MORE

Responsible Minister

Skinner, Jillian Gell

Piccoli, Adrian

Hazzard, Bradley Ronald

Constance, Andrew James

House Papers health

Relevance

2807 - "CLOSE THE GAP"
 NSW Aboriginal Maternal and Infant **Health** Strategy (AMIHS ... In order to improve the **health** of Aboriginal mothers and their babies, the Report recommended a specific model of service provision which included a team approach to community maternity services (including midwifery, Aboriginal **health** workers, specialists and general practice), a flexible and non-judgmental approach, and ...
www.parliament.nsw.gov.au/la/papers/.../close-the-gap_214763.aspx

3345 - CLOSE THE GAP STATEMENT OF INTENT
 The Government's commitment to improving **health** and other outcomes for Aboriginal people ... for Aboriginal Children, Families and Communities Strategy and the Aboriginal Maternal and Infant **Health** Strategy, a range of other programs are aimed at improving access to **health** care by Aboriginal people, such as the Chronic Care Program for Aboriginal People and the Aboriginal Oral **Health** program ...
www.parliament.nsw.gov.au/.../close-the-gap-statement-of-intent_223...

0488 - PEAT ISLAND HEALTH FACILITY
 Planning, Minister for Redfern Waterloo, Minister for Science and Medical Research, and Minister Assisting the Minister for **Health** (Cancer) representing the Minister for **Health** ... In relation to the Peat Island **Health** facility ... Are there plans to close the centre within 5 ... The Minister for **Health** has provided the following advice in relation to the Honourable Member's question ...
www.parliament.nsw.gov.au/.../peat-island-health-facility_174103.as...

Institute of Psychiatry - annual report 2012-13.pdf
 Our Mission: The NSWIOP aims to provide a quality learning and research environment to the mental **health** workforce and the wider community ... Our Goals/Strategic Priorities 1. To provide education that promotes mental **health** 2. To provide broad access to mental **health** education 3. To deliver high quality mental **health** education 4. To support development of a sustainable mental **health** workforce ...
www.parliament.nsw.gov.au/.../Institute of Psychiatry - annual repo...

Image 3: Search results view – House Papers

For the First Legislative Council search, the results will also display relevant metadata, such as date tabled, tabled by and document type, where available. This will permit easier identification of relevant content and offsets the lack of available context text to display where the document is hand-written.

To the left of the search results, facets allow the search to be narrowed further. They may include document type, Minister, decade, or House, among others, depending on the metadata associated with the full range of search results. For example, if none of the documents brought back as a result has an entry for the field 'Responsible Minister', that facet will not appear.

Building the Digital Archive

Organisational framework – oversight

Having secured Treasury funding for the Digitisation and Preservation Access Project, NSW Parliament set up a Digitisation Working Group to oversee the project's implementation. The Working Group comprises staff from the two Houses of Parliament which own the material, as well as from the Department of Parliamentary Services, which includes the Parliament's IT Services, Library, and Records and Archives teams. The Working Group in turn reports to a Steering Committee which oversees the transition of Parliament's web publishing systems from Lotus Notes to SharePoint and other systems.

The Working Group has met every second week for the duration of the project. The first meeting was held in January 2015, and the Group will continue to meet until the archive has been fully uploaded to the website.

In addition, a Project Coordinator worked on the project full-time for 8 months, from February to October 2015. During this time, the bulk of the digitisation work was completed, including scanning of the documents and books and compilation of metadata. The conceptual work for the digital presentation of the archive was also finalised.

Digitisation

The digitisation proceeded in several stages: the physical records were identified and retrieved, sent to a contractor's premises to be scanned, and returned to be prepared for transfer to State Archives. The resulting PDFs then underwent a series of digital enhancements such as compression and OCRing to enable better online presentation. At the same time, metadata was transcribed from registers and other sources. In a further step, the metadata was enhanced to include document type, House, and other necessary or useful data. Then, the images and metadata had to be combined. After this, several rounds of quality control were carried out to ensure the images were associated with the correct metadata as well as spell checks.

Identification of documents included in the project

At the outset, the first task consisted of identifying the documents included in the digitisation project. Clearly, the **tabled papers** from 1824 to 1900 were to be scanned. These papers were stored in three sets of archive boxes – 83 boxes for the First Legislative Council, 165 for the Legislative Council and 370 for the Legislative Assembly: 618 boxes in total. After completion of the scanning and quality control, these boxes were transferred to State Archives, where they are stored in a controlled environment. This will limit future deterioration.

Unfortunately, the documents have not always been stored in an ideal environment, which has resulted in some of them being damaged by water, pests or rodents. The Working Group decided to send all boxes to the scanning contractor so that they could provide expert advice on a case-by-case basis as to which documents could be scanned and which were not capable of being digitised.

The Working Group further decided to have A3-sized **registers** recording the tabling of papers scanned and transcribed. These ledgers provided the metadata framework for the scanned files, as they record the register number, title, date tabled and the Member who tabled the document. The transcription took place through a contractor off-site and the resulting data was proof-read for typographical errors and misunderstandings.

The **Journals** containing Votes and Proceedings / Minutes and tabled papers ordered to be printed were also included in the digitisation project as they are intrinsically linked to the tabled papers. In total, 380 volumes of books were scanned, relating to 99 sessions of Parliament.

In addition, it was decided to complete the online presentation of **NSW Hansard**. Prior to the digitisation project, only the Hansard transcripts from 1980 onwards were available online. 414 volumes of books were scanned, starting with the earliest record on 28 October 1879.

Scanning process

It was decided to outsource the scanning work rather than engage in-house staff in this activity. The age and fragility of the documents was part of the consideration, as was the aim to achieve a professional outcome. Quotes were invited from companies on the NSW Government's ICT Services Scheme panel of prequalified service providers. In evaluating the quotes, each company's experience with the digitisation of old and delicate records and estimated timeframes for the completion of the work were taken into account as well as cost.

The scanning contractor was instructed to provide the results in PDF and TIFF format and scan to a resolution of 300dpi. All tabled papers were scanned in colour to retain as much information as possible, including differing paper textures and pen colours, while the Journals were mostly scanned in black and white. The contractor was also required to provide basic metadata concerning the file names and their location in folders.

The tabled papers were to be scanned as one file per document. As the documents consist mostly of loose leaves, it was not always easy or possible to identify when one document ends and another begins. Staff at the scanning company therefore separated the documents to the best of their knowledge but were instructed not to spend too much time trying to provide an exact result. Therefore, many of the resulting PDF files contained more than one document and had to be separated into individual documents in a quality-assurance step by parliamentary staff with greater subject experience.

The Journals and Hansards as well as the A3-sized registers fall into the category of bound books. These were assessed as to how to best scan them, balancing the requirements of preservation with cost. Two options presented themselves: Firstly, the pages can be guillotined out of the cover, scanned and then re-bound in the original cover. This process results in the loss of one centimetre on the inside of the pages and the books are harder to open because they are bound closer. In an alternative approach, the books are laid open and scanned under and overhead scanner. This second process is easier on the books, but also takes longer and is more expensive. In addition, the resulting image can be harder to read because a shadow remains in the centre from where the page curves back to the binding.

The Parliament decided to overhead-scan all leather bound books as they would be difficult to rebind, and to guillotine and re-bind all others. This has brought about an optimum result in terms of balancing both preservation and economy.

For the Hansard transcripts, the scanning company was instructed to provide one PDF per sitting day. As the files were to be uploaded onto the existing website and to be presented alongside the current Hansard transcripts, the House, date and page number of the entry were required as metadata and the file names followed a specific convention allowing for this metadata to be associated. The scanning company was successful in providing all required metadata.

Digital enhancement of PDFs

The PDFs provided by the scanning contractor underwent a range of enhancements, including optical character recognition, compression, and for Journals bookmarking and separation. This work was carried out by another contractor specialising in this kind of work.

Optical Character Recognition (OCR)

All files were processed with optical character recognition software so that any printed text is now searchable. This means that in the Digital Archive, any search will include not only the title, but also return results within the document.

Unfortunately, the hand-written parts of the documents had to be included in this process, which can create problems for the displaying of search results. When an abstract is provided in the search result, the hand-written parts will display as what appears to be a random selection of characters. There was no practical way to separate hand-written from printed documents as many of the hand-written documents include pages in print and printed papers include hand-written annotations. To remedy this, additional metadata is displayed for First Legislative Council search results.

Compression

Many of the PDFs files as scanned were larger than 50 MB in size and thus less than ideal for displaying on a website or download. Therefore, all files underwent a compression process. The file size was reduced on average by 50 to up to 90 per cent, while retaining the quality of the images.

Conversion to PDF/A

After completion of necessary quality assurance work, such as combining or separating of files to ensure that every file contains only one document, all PDF files were converted to the PDF/A format. PDF/A is readable just as PDF, but provides a better guarantee that the file will always be displayed in exactly the same way. In addition, future PDF displaying programs are required to be able to read current PDF/A files so that the format will not be redundant in the near future.ⁱⁱ

Bookmarking and extraction

The Journals underwent additional processing. The Working Group decided to present the Journals as one file per book on a separate section of the webpage, which results in PDFs of 1000 pages and more. To make them more accessible, bookmarks were inserted into those files.

In a further step, all Votes and Proceedings / Minutes are being proposed for extraction as individual files so that they can be uploaded as individual entries into the Digital Archive and associated with specific dates. To facilitate this, a minimal amount of metadata has to be provided, namely title and date. These two sets of data can mostly be derived automatically as the individual files are extracted.

Metadata and PDF assignation

For most of Parliament's operation, tabled papers were recorded in large registers, noting the title of the document, date of the document, date tabled, tabled by whom, and the tabled paper number. For most of the period in question, incoming letters of an administrative nature were also registered in these ledgers and thus have a tabled paper number, even though they were never tabled.

As the Digital Archive was designed to mirror the current Tabled Papers site section, Excel spreadsheets were created incorporating all fields currently used. Then the registers were transcribed into these spreadsheets. The document type was assigned as far as possible in correlation with modern practice and modified where documents did not correspond to modern document types.

For the First Legislative Council, no registers could be found. The metadata for this period stems from contents lists contained in the archive boxes together with the tabled papers themselves.

Once the metadata spreadsheets had been created, the PDF documents were assigned to their corresponding metadata entry. This means that the location of every PDF image was recorded at the relevant entry in the spreadsheet, so that the metadata would be assigned to the correct PDF during the upload onto the webpage. In most cases, this assignment was largely based on the paper number, as this is noted on the actual document on the first page, thus indicating a new document. It was much easier to identify the number than reading the first page of the document to correlate the content of the page with the descriptive title contained in the register. During this stage, the PDFs were also separated where necessary to contain only one document per file.

A number of register entries do not have PDFs assigned to them as the corresponding documents may have been too damaged to scan or simply lost. It was decided to retain those entries and to incorporate them into the Digital Archive as the document titles in the registers are very descriptive and may have some value in themselves. Should the documents be found at a later date, it will be possible to assign them to the relevant entry in the archive.

Quality assurance

The metadata, PDF files and the correlation between the two underwent numerous rounds of quality checking. The metadata was spell-checked and cross-checked with the original register entries to identify misread entries. The PDFs were checked against the requirement to contain only one document per file and separated or combined where necessary. The assignment of images to metadata was also cross-checked to a large extent. Most of this work was carried out by one person to achieve consistency across the archive.

However, it is not possible to exclude all errors, especially considering the size of the archive, the limited human resources and the difficulty of reading faded century-old hand-written records. It is therefore known that the archive contains errors which will have to be corrected progressively through future projects.

Making the Archive accessible

The presentation of the scanned documents on a publicly accessible website was always an inherent aim of the digitisation project. The Digital Archive is in the process of being integrated into the NSW Parliament's website. This is a desirable outcome as it allows for all Parliamentary records to be presented with some consistency and to be searched comprehensively in one place.

The Parliament's website has been undergoing a major transition over the last 12 months, both in relation to the content and structure of the site as well as the underlying technology platform. The timing was fortuitous in that the needs and aims of the Digital Archive could be considered in the overall design of the new platform. As a result, the Digital Archive and the modern tabled papers will be presented in a consistent and comprehensive manner. It will also be possible for parliamentary officers to easily augment or correct the metadata in the archive in the future. Additional material will also be able to be added as required.

In the early stages of the project, the Working Group considered several other options for making the content available on the Internet, including via Trove or OpenGov.

Trove (<http://trove.nla.gov.au/>) is a freely accessible online archive hosted by the National Library of Australia which provides access to a multitude of Australian books, magazines and newspapers,

reaching back to early colonial times. It is well known as a valuable research tool and harnesses the resources of the 'crowd', most notably with the OCR correction tool, which allows users to make corrections to the automatically produced OCR text. This tool is valuable as the OCR especially of old documents can be unreliable.

OpenGov was also considered as a potential host for the Digital Archive. The OpenGov website is the online location where government agencies publish their annual reports and other publicly available information under the *Government Information (Public Access) Act 2009*.

It became clear after a relatively short time that neither Trove nor OpenGov would be able to provide a completely suitable platform for the Parliament's historical documents. The OpenGov website was not at that time set up for hosting the types of documents included in the Digital Archive. Trove advised that its website was undergoing a rebuild and that for that reason new material of the size of the Parliament's Digital Archive could not be added in the foreseeable future.

Process of website development

NSW Parliament engaged the services of a specialist consultant to advise on how to best structure the web presentation of the Digital Archive. The consultant held a number of sessions with parliamentary staff, focussing on what data is contained in the archive and how researchers could best be aided to retrieve it.

It became clear that the archive should be accessible in multiple ways. In addition to a search of all the titles and OCR'd full text of the documents, they should also be accessible by date and by paper types. The system would make extensive use of filters on the side of search results to allow further narrowing down.

Testing of the resulting interface triggered a raft of reviews as the strengths and weaknesses of the imagined archive became clear 'in the flesh'. One of those changes concerned the display of additional metadata for the search results. Many of the hand written documents have been OCR'd and display as random characters in the preview of the document. To aid the searcher, the Working Group decided to display, in addition to the title, also the document type, date tabled, and who tabled it, where this metadata is available. As the archive continues to be improved, so the metadata will continue to grow.

Back end

The 'working part' of the website, the back end, is where the files are stored and changes can be made. Since May 2016, the NSW Parliament's webpage is run through SharePoint.

All files in the Digital Archive have been uploaded as PDFs along with associated metadata. Capacity is being allowed for ongoing changes to the individual files and metadata as well as uploads of new material to be made through the same interface through which current House Papers and Tabled Papers are managed.

As in the public presentation, the files digitised during the digitisation project are distributed between several locations. The First Legislative Council documents are kept separate from other documents on the site. Only authorised staff can access each site section to amend or add information.

Future work

While publication of the Digital Archive will be a significant milestone, much work will still be required to make it as good as it can be. Most importantly, the metadata which has been taken from registers and other lists will have to be manually compared to information stored in the Votes/Minutes and corrected if necessary. This work can potentially be completed by parliamentary staff.

In addition, there is potential for the Digital Archive to be expanded. A future project could see all tabled / printed papers from the Journals be put up as individual files and classified by date, so that they would also appear in the browse view alongside the equivalent hand-written tabled papers. Gaps in the tabled papers could thus be reduced if not fully closed.

Beyond that, the tabled papers from the time between 1900 and 1999 are not yet available individually online. Only future digitisation projects can remedy this. However, the experiences during this digitisation project will inform the processes of any future endeavours and the current website has the structures in place to accommodate such additions.

ⁱ Ian Marsh directed a project from 1987 – 1990 with the aim to provide a subject index to the SMH articles summarising the proceedings in the NSW Parliament from 1843 – 1879. The index is structured chronologically by Parliament, with each Parliament subdivided in two parts, 'by subject' and 'by speaker'. The latter lists all the speeches a Member made according to subjects.

The index was published in print in two volumes in 1990. It was subsequently made available digitally on the NSW Parliament's website, structured by date of publication in the SMH. The metadata identified in the printed index was linked to PDFs of the original articles to enable search and filter functions. The incorporation of those files into the Digital Archive is based on this previous work.

ⁱⁱ Further information on the PDF/A file format is available on the PDF Association's website: www.pdfa.org, specifically under the link to the PDF/A Competence Center.