BUILDING OUR AUSTRALIAN CLOUD

Monika Szunejko

PAPER

Turning the first sod

On 2 November 1981, six libraries from three states across Australia took the first steps towards building the Australian Bibliographic Network (ABN). The National Library of Australia, Canberra Public Library Services, the University of Adelaide, the University of New South Wales, the State Library of South Australia and the Commonwealth Attorney-General's Department took the bold move to collaborate by sharing their bibliographic content and began contributing catalogue records to the ABN.

By 2014, the successor to the ABN, Libraries Australia, comprised over 1,200 member libraries, of which 737 contribute records and holdings, and 800 libraries use the Libraries Australia Document Delivery service for their resource sharing operations. The Australian National Bibliographic Database (ANBD) reached the milestone of more than 25 million bibliographic records and over 50 million holdings of Australian libraries. Since its beginnings as the database underpinning ABN, Kinetica, and now Libraries Australia, the ANBD lies at the heart of our national resource sharing and discovery services – Libraries Australia and Trove – and enunciates the desire of Australian libraries to collaborate. Building ever more sophisticated services on the data aggregated over 33 years, Australian libraries have created a national resource sharing infrastructure unparalleled in the world. The smallest of Australian libraries can participate in the description, resource sharing, and discovery infrastructure built within the Libraries Australia service in the same way as the largest of Australian libraries can. The service offers equal opportunities for all members, and when the data is aggregated, creates a national infrastructure and service larger than the sum of its parts.

This paper reflects on the big bibliographic data 'cloud' that was seeded with catalogue records in 1981 for the purpose of cataloguing, and which has evolved into an aggregated metadata database with uses beyond shared cataloguing, to Trove and the big data cloud we are now managing.

A vision for flourishing

The primary driver for initiating the ABN was the need to develop an Australian shared cataloguing network in order to eliminate the duplication of original cataloguing and enable the sharing and copying of catalogue records immediately with other participants.

"By early 1975 the National Library was anxious to emulate or exceed the success of the OCLC network in the United States by establishing an online system, under the auspices of the government, that could not be challenged as Australia's only shared cataloguing network" (1991. Cathro, 37).

From the outset, ABN contained the National Union Catalogue (NUC) and its major role was to support shared cataloguing. ABN also provided inquiry, bibliographic verification and library location services, and was based on cooperative participation. In time the participant base extended to every State and Territory across Australia, and from all library sectors.

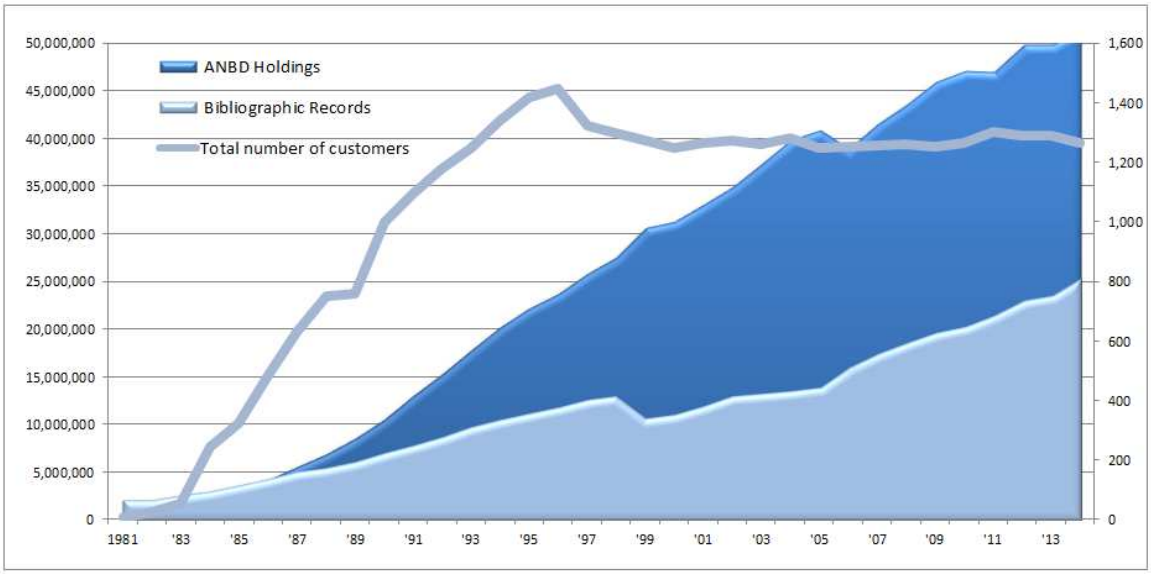
Underpinning the services of the national bibliographic network was the National Bibliographic Database (NBD), a unique combination of records drawn from a range of trusted/authoritative sources. In early 1980, before any Australian participant data was even added, the NBD had 1.5 million bibliographic records derived from the Library of Congress and the Washington Library Network (WLN) – the consortium and software system on which ABN was launched. Following the launch of ABN, the NBD included ABN participants' original cataloguing, the WLN database as it stood in January 1980, and national agency records from the Library of Congress (LC), U.S. Government Printing Office, British National Bibliography (BNB), Canadiana, Australian National Bibliography (ANB), and New Zealand National Bibliography (NZNB). New sources for bibliographic data from non-ABN participants were considered and proposed for addition to the NBD. Accompanying the bibliographic data, the NBD contained holdings data of participants and the serials holdings data of non-participants reported to the National Union Catalogue agencies. It also contained authority information and verified LC subject and name headings.

While the contributions from ABN participants dated from 1980, the data contained within the bibliographic and holdings contributions dated back to the late 1960s and early 1970s for some of the national contributing agencies (such as the British National Bibliography, New Zealand National Bibliography, Singapore National Bibliography and Vietnamese National Bibliography). (From its beginnings the NBD held a generous 'back catalogue' of bibliographic data at the disposal of ABN participants.

It is clear that the vision for the NBD was to create the richest aggregation of high quality bibliographic, holdings, and authority records that would support and streamline the cataloguing and document delivery operations of Australian libraries, and ABN was the 'platform' through which those services would be delivered.

Over time, the number of participants of ABN grew from six in 1981 to 1,100 users in 1991. Concurrently, the size of the ANBD grew from the initial 1.5 million records to 8 million records in 1991 (Cathro 1991, 40-41) and by 2013 the ANBD recorded over 25 million records (National Library of Australia 2014A). In order to enrich data and support evolving description and discovery requirements for electronic holdings, records are now loaded from a number of agencies such as OCLC, Blackwell's Table of Contents, and Serials Solutions. Authority data from the ANBD is sent to the Virtual International Authority File (VIAF) and then works its way into the International Standard Name Identifier (ISNI) database.

The graph below (see Figure 1) shows the growth of participants of ABN against the growth of the ANBD (bibliographic and holdings data). While the number of Australian library participants stabilised in the late 1990s, the ANBD has continued to grow.



(Figure 1) Growth of the ANBD and ABN customers 1981-2013
Prepared by Libraries Australia, National Library of Australia, 2014

In its early years participants saw ABN as a cost-effective and efficient shared cataloguing operation. However, responses to the ABN Forward Plan in 1983 indicated that many participants "understood that ABN has the potential to be far more than this and this it could ultimately provide the framework with which a large range of co-operative and resource-sharing activities can flourish" (National Library of Australia. 1985, 17). In hindsight, this vision augured the evolving suite of description, collection management, and analytical services available to libraries via Libraries Australia and the range of discovery and engagement services available via Trove.

<u>We came to catalogue, we stayed to do more</u>
At the outset ABN functioned as a large cooperative bibliographic utility. In 1981, building the NBD involved receiving data from the national agencies on magnetic tape which were then loaded by the National Library's ABN Office into the database via periodic batch mode updates. Retrospective files from ABN participants were also loaded in this way. After the introduction of reliable dial-up access in 1983, participating libraries contributed their records directly online.

The redevelopment of ABN was impelled by the network's participants who were seeking more flexibility in data contribution options and further efficiencies in workflows. Developments in technological capabilities, search functionality, and the development of networked services also drove the re-development of ABN into the Kinetica service, and later into the Libraries Australia suite of services.

Today, the ANBD is built with a range of tools. Member libraries contribute their library's bibliographic and holdings records in a variety of ways: record-by-record via the online web cataloguing form, via the Cataloguing Client with direct access to the database, via the Record Import Service (a batch load ftp service), using a Harvester, or via their library cataloguing supplier. Data files from national agencies such as the Library of Congress and the British National Bibliography are loaded regularly via the Record Import Service. In addition the records and holdings of Australian libraries contributing records directly to OCLC's WorldCat are transmitted to the ANBD, thereby allowing greater flexibility for Australian libraries to participate in the national

infrastructure and collaborative arrangements available in our national resource description and sharing ecosystem.

The flow of bibliographic and holdings data does not move in one direction only. Records in the ANBD are transmitted to WorldCat every day via the SRU Update protocol. Furthermore, every minute of every day, records contained within the ANBD are transmitted to Australia's national discovery platform Trove. Through this combination of metadata aggregation and syndication to national and international services, Australian libraries collaborate and participate in forging a data footprint that extends beyond their local library catalogue to the World Wide Web. The Australian National Bibliographic Database plays a vital role in promulgating Australia's global data footprint through data synchronisation to WorldCat, data flow to Trove and subsequent discovery in both via Google. On a national scale, there is nothing else like Libraries Australia in the world where libraries from all sectors across an entire nation contribute their content and set the foundation for our Australian global data footprint.

The ANBD forms the foundation for ongoing strategic transformation and collaboration in the service of Australian libraries and their members. ABN and its following services Kinetica and Libraries Australia have built a range of additional applications onto the ANBD that build not only on the aggregated data but on the desire to collaborate. Libraries readily contribute their bibliographic and holdings data for others to re-use, provide information about their collections, and enable resource sharing. The ANBD aggregates the supply of bibliographic records and provides the core from which libraries can collaborate. As such it serves as a library form of a public good – no matter how many times data is used or viewed, its value is undiminished.

The impact on national and international collaboration

With aggregated supply, demand collects and more services are built into the national infrastructure. Over 33 years, Australian libraries have built a national database of records. In that time, the size and use of the data has evolved into an ecosystem of data and services that not only supports the daily operations of over 1,200 Australian libraries, but also reflects the ongoing professional practices, standards, and pragmatic decisions made by the thousands of dedicated library and information employees who have been responsible for keying, loading, and inscribing the database. Concealed within the database are the thousands of hours of human effort that have gone into constructing and crafting records, and the hours of peer-review and collaboration that lies behind the submission of each record to the database. Similarly, the work to ingest records into the ANBD and manage the quality and integrity of the database require hours of human labour to maximise the use and re-use of the content for Australian libraries. Inscribed in invisible ink is the investment of human intellectual capital that has created this unique national body of data.

The database inscribes and expresses the character of our national collection: declaring what libraries hold, their collecting practices and policies, descriptive practices and human devised notes, individual shelf-marks of unique items, large-scale movements of materials, library openings, mergers, and now, sadly, library closures. Data goes in, data goes out. The database is an ever expanding and changing body of data that describes the distributed national collection (DNC) of Australia. Cathro asserts that the DNC really describes an attempt to articulate a comprehensive National Collection Access Strategy that, among other goals, aimed to support resource discovery and resource sharing in the broadest sense. (Cathro. 1994). In this context, the ANBD contained the foundation data that supported the

ambitions of the distributed national collection, and enabled Australian libraries to declare their collection holdings in a common infrastructure for the common good.

Big (bibliographic) Data

The often overlooked value for Australian libraries, irrespective of size or status lies in the fact that their collections are described and aggregated into a single national bibliographic database. The ANBD is big data: 25 million bibliographic records and over 50 million holdings records. In the course of each day, the ANBD receives approximately 50,000 search hits from libraries and approximately 700 document delivery requests. Each day records are added, deleted, changed, matched, merged, enhanced, and compared to de-duplicate exact matches. The daily data quality routines that trawl the database create millions of transactions on the database each year – more than 18.3 million transactions were performed on the database in 2013-14 (2014, National Library of Australia). The ANBD transports vast amounts of bibliographic traffic in and out of the database, and within its boundaries. That is big (bibliographic and holdings) data. Leveraging this big data using the ANBD, Libraries Australia has built a range of services around aggregated data that enable libraries to manage and interpret their collections, promote their services, advocate their position, enunciate their uniqueness, differences and similarities with other libraries, and collaborate on projects to better understand and manage our national information context. The corpus of Australian library data has been used to support national projects such as the National and State Libraries of Australasia (NSLA) project on Collaborative Collections, Connecting and Discovering Content, Description and Cataloguing, and Open Borders.

If we were to map the movement and use/re-use of data from its origins in local libraries in each state and territory through to the National Bibliographic Database, to Trove and beyond, we could see how hard our national database works on behalf of the Australian library community. A search of the ANBD is embedded as the starting point of most library transactions in Australia – from cataloguing, collection development, acquisitions, new title services, document delivery, and broader research activities. While the tools used may obscure the origins of the data, for example by searching via z39.50 connection from within a local database, the ubiquity and value of connecting to this data has enabled libraries to move beyond the limitations of their local resources and take advantage of a national resource sharing utility.

National collaborative services such as the ANBD create information capital. Australian authority data from the ANBD makes its way into international frameworks and services such as VIAF, ISNI, and then Wikipedia. Active contribution to the Australian National Bibliography enables libraries to demonstrate their value in a competitive information environment, measure their impact, and expose their collections to a global audience through the analytical use of their data in the national database.

Through the services built around the ANBD, Australian libraries have been able to participate in collaborative ventures, build context and interactions around their collections, and leverage their collections to connect with other global discovery and delivery services. This has been facilitated through a national infrastructure established to build the capacity of all Australian libraries to contribute and export data from their local systems. The ANBD and its associated services provided via ABN, Kinetica, and now Libraries Australia have freed individual Australian libraries from the burden of having to find individual ways to collaborate, contribute and share data with the hundreds of other Australian libraries. It has provided a common database and infrastructure (available remotely) through which to connect,

contribute, download, leverage, and participate in national collaborations and over time has proven it is scalable through its capacity to grow and respond to the evolving information environment.

A key maxim of the cloud is that it releases organisations to focus on their local business and expedite local actions rather than on the development of infrastructure at every local instance. (OCLC. 2011, 16) Connectivity from the local database to the aggregated national database and re-use of data has been a hallmark of building our Australian cloud.

<u>If it looks like a cloud, and acts like a cloud…</u>
"*Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources*… that can be rapidly provisioned and released with minimal management effort or service provider interaction" [author's emphasis] (National Institute of Standards and Technology 2011, 2).

Considering ubiquitous cloud-based services such as Facebook, LinkedIn, Amazon, Flickr, Dropbox, and GoogleDocs we can see the characteristics of big collaboration, scalable infrastructure and context-building are features of cloud-based technologies. At the core of cloud computing, is the concept of shared services and infrastructure. Users are distributed, and access to systems is location and device independent, via a web browser. Cloud systems are typically multitenant, meaning software and resources are contributed and shared by a large pool of users. Embedded within 'the cloud' is the notion of accessible resources and infrastructure that can be connected to anywhere in the world via the internet. The ANBD displays all of these features and could be considered an early 'prototype' or 'pre-web' cloud. Applying the cloud paradigm and metaphor to describe the body of data and national services underpinning the ANBD, highlights the position of the database in the Australian library and information ecosystem and reframes its value to Australian libraries. If ABN and the ANBD had not been conceived of years ago, would Australian libraries have the opportunities to leverage large-scale description, collection management, and discovery services such as Libraries Australia and Trove?

The ANBD is egalitarian, offering equal opportunities for all libraries of all shapes and sizes to participate in a national infrastructure; is all-embracing, welcoming records describing all materials, from all libraries and collections; and is built on collaborative practices both of data management and collective management of the service – through strong representative governance structure. The data base is capability building: the infrastructure of ABN, Kinetica and then Libraries Australia has allowed library to participate in a platform that they may not on their own have been able to participate in. The ANBD has facilitated automation, collection management, and collection sharing, collaborative inter-lending, streamlining of workflows, and user group interaction. Professional development is a strong drawcard and focus of participating in the ANBD services. National infrastructure services include connections with other practitioners and accredited training programs. Furthermore, the ANBD is standards building. Supporting the development and implementation of national and international standards for the description and exchange of data, ABN and its successors Kinetica and Libraries Australia were grounded through the work of standards and network committees. The development of implementable international standards has always been a focus of the infrastructure the ANBD was founded on and was critical factor in the national push to establish RDA and move standards forward. These activities have ensured the ANBD remains a viable

resource for Australian libraries as they connect and manage their collections in a global context.

Embedded in operational and daily work the significance and value of the ANBD and Libraries Australia services may go unnoticed. Australian libraries can connect to the service remotely (most often from within their own local system), contribute, manipulate and store their data in the ANBD. Syndicating the ANBD to Trove and WorldCat (and by extension search engines such as Google) leverages the contextual power of the aggregated data using additional services. Participating in the ANBD enables libraries to build content, collaboration, enrich their capabilities, and extend the reach of Australian libraries through the aggregation, re-use, and syndication of bibliographic data. Contributing local data to the ANBD enables Australian libraries to maximise their global data footprint and place their collections alongside global collections and holdings.

The 'cloudiness' of our Australian cloud refers to the common infrastructure and shared services accessible remotely by libraries across Australia; it refers to the large body of aggregated data contributed by hundreds of Australian libraries since 1981; it refers to the scale of the aggregated data accreted over 33 years; and it refers to the emerging services made possible because of the scale of the service.

The ANBD continues to provide national infrastructure to underpin efficient and effective library services across the country, and supports Australian libraries in the twenty-first-century digital world.

Since 1981 we have been building our Australian cloud.

REFERENCES

Cathro, Warwick S. 1991. The Australian Bibliographic Network: A Survey of its First Decade *Australian Academic & Research Libraries* 22(4): 37-52. <http://dx.doi.org/10.1080/00048623.1991.10754736> (accessed 29 May 2014)

Cathro, Warwick S. 1994. Distributed National Collection *'Interlending and Document Delivery: Issues and Technology, A Seminar', Sydney, 17 January 1994*. <http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/977/1247> (accessed 29 May 2014)

Missingham, Roxanne and Boston, Tony. 2005. Libraries Australia: Creating a New National Resource Discovery Service *Online Information Review* 29 (3) 296-310 <http://www.emeraldinsight.com/journals.htm?articleid=1509088>(accessed 29 May 2014)

National Institute of Standards and Technology. 2011. *The NIST Definition of Cloud Computing* < http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf> (accessed 30 July 2014)

National Library of Australia. 1985. *ABN Handbook. Part 1*. Canberra: National Library of Australia.

National Library of Australia. 2014A. *Libraries Australia – About us* <http://www.nla.gov.au/librariesaustralia/about/> (accessed 07 June 2014)

National Library of Australia. 2014B. *Libraries Australia – ANBD* <http://www.nla.gov.au/librariesaustralia/about/anbd/> (accessed 30 July 2014)

OCLC. 2011. *Libraries at Webscale* < https://www.oclc.org/content/dam/oclc/reports/worldshare-management-services/libraries-at-webscale.pdf> (accessed 30 July 2014)

Wells, Andrew. 1999. Kinetica: Moving Australian Library Network Services Forward *ALIA Information Online & OnDisc 99 Proceedings* <http://conferences.alia.org.au/online1999/proceedings99/205b.html> (accessed 29 May 2014)