



Looking Beyond the Archive: Utilizing Encoded Archival Context in a Broader Societal Context

Gavan McCarthy, Ailie Smith & Michael Jones

To cite this article: Gavan McCarthy, Ailie Smith & Michael Jones (2015) Looking Beyond the Archive: Utilizing Encoded Archival Context in a Broader Societal Context, Journal of Archival Organization, 12:1-2, 143-164, DOI: [10.1080/15332748.2015.1001206](https://doi.org/10.1080/15332748.2015.1001206)

To link to this article: <https://doi.org/10.1080/15332748.2015.1001206>



Published online: 03 Mar 2015.



Submit your article to this journal [↗](#)



Article views: 447



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Looking Beyond the Archive: Utilizing Encoded Archival Context in a Broader Societal Context

GAVAN MCCARTHY, AILIE SMITH, and MICHAEL JONES

University of Melbourne, Melbourne, Australia

The eScholarship Research Centre (ESRC) has produced EAC-CPF XML outputs from web resources for many years. Using the Find & Connect web resource project as a case study, this article explores the uses that the ESRC has developed for EAC-CPF data, including online presentation of descriptive contextual information, as the basis of online search services, as a system-independent and preservable copy of research datasets, and as a harvestable data source for interoperability with other datasets. Also covered are expanded uses of the EAC-CPF schema for describing a wider range of contextual entities beyond corporate bodies, persons and families.

KEYWORDS *EAC-CPF, XML, interoperability, archives, context, search, communities, preservation, online presentation*

“Contextual information is that extra, associated, related, assumed and perhaps *a priori* information or knowledge that is required to meaningfully interpret the content of any given information source.”¹

INTRODUCTION

The eScholarship Research Centre (ESRC) is a research center, staffed predominantly by archivists, located in the University Library at the University of Melbourne, Australia. The Centre and its predecessor organizations have spent the past 30 years developing innovative ways to document the

© Gavan McCarthy, Ailie Smith, and Michael Jones

Address correspondence to Ailie Smith, eScholarship Research Centre, Level 2, Thomas Cherry Building, The University of Melbourne, Victoria, 3010, Australia. E-mail: ailie.smith@unimelb.edu.au

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/wjao.

contextual information that brings meaning to and is evidenced by archives, as well as widely disseminating this information. The international archives community is continuing to grapple with the challenges facing the archival profession in a digital and networked age. The ESRC comes to this context with a long history of web publication of archivally based information resources and was an early adopter of the use of EAC as a means of exporting an XML version of this contextual information in a standardized format.

The Find & Connect web resource, a project of the ESRC begun in mid-2011, is an online gateway to discovering the history of Australian orphanages, children's homes, and related institutions, providing information about those institutions and, importantly, details of records held by many large and small repositories across the country. The primary goal of the project was to use the network of related authority records, or context entities, to document this landscape and to assist people who have experienced out of home "care" in their childhoods in finding and accessing the records that are evidence of that part of their lives. The resource was built on the well-established informatics framework developed by the ESRC for capturing and disseminating this type of information and taps into existing capabilities for producing EAC XML versions of the data. The Find & Connect project also posed some unique challenges to the existing informatics frameworks and afforded the ESRC with an opportunity to further develop and extend its use of EAC-CPF.

The following article explores the background of the ESRC's development, maintenance, and dissemination of archival context entities, as well as its implementation and use of EAC, and extension of EAC-CPF. Using the Find & Connect web resource as a case study, the article discusses recent developments and extensions and imagines the future of EAC-CPF as an evolving standard.

BACKGROUND

The eScholarship Research Centre (ESRC) and its predecessors at the University of Melbourne, Australia, have been in the business of contextualizing archives for almost 30 years. From its beginnings as the Australian Science Archives Project in 1985²—followed by the Australian Science and Technology Heritage Centre (1999–2007)—the Centre recognized the need to document not just individual archival collections, but to identify and link related records, wherever they were located. The resultant long-term research project identified records of Australian scientists in institutions and repositories across the country and led to the publication of a print volume—*Guide to the Archives of Science in Australia*.³ By grouping resources under biographical information about scientists, this publication took a contextual, rather than resource perspective to archival descriptions. The aim was to take

information about archives out of their siloed local information environments and allow records to be explored and understood in different ways. From the beginning, the descriptions of people, descriptions of collections, and descriptions of repositories were managed as separate but related information objects.⁴

The arrival of the Internet enabled the *Guide* to be redeveloped and published online as *Bright Sparcs* in May 1994.⁵ The content of the web resource came from the same relational database as the *Guide* but was exported as a stand-alone dataset that was transformed into HTML for web representation.⁶ This attempt at the disentanglement of content and representation (or data and information) proved to be significant conceptually, technically, and pragmatically over the subsequent years, enabling the evolution of both the conceptual framework and the systems that underpinned this work. Of note, it supported a firmly held principle, that to enable the long term intergenerational transfer of knowledge, technology dependency should be as low as practically possible. For example, one outcome was a set of web publishing protocols and processes that posited the creation of independent set of relatively referenced and therefore self-contained HTML pages. The separation of data curation and storage, and information output facilitated the application of new technologies, and the serialization of the core data in different forms. The most significant of these being the export to Extensible Markup Language (XML) document type definitions (DTDs) and schemas.

SHARING INFORMATION BEFORE THE ADVENT OF ENCODED ARCHIVAL CONTEXT

In the late 1990s, requests from other research groups to create public web resources based on the *Bright Sparcs* model for different subject areas led to the development of a generic system, the Online Heritage Resource Manager (OHRM). It was conceived as a tool for registering and mapping the historical actors or agents (context entities) that make up any particular world; for registering the evidential sources (record entities) that support the existence of those actors; and for documenting the interconnections and dependencies both within and between actors and evidence (relationships). The OHRM has subsequently been used to create, manage, and disseminate all of the contextually modeled information resources produced by the Centre since 2000. It became not just a database system with a web publication service, but a way of thinking about the world that attempted to deal with its ontological foundations and then, through defined purpose-driven relationships, establish an epistemological framework that assisted interpretation and the search for meaning.

Australian archivists have long been in the practice of capturing rich contextual information about records. From the work by people such as Peter Scott at the Commonwealth Archives Office (now the National Archives of Australia) on the development of the Australian “Series” system, to the development of the records continuum model, the inherent complexity of the context in which records were created, used, and kept, and the importance of records in capturing individual, institutional, and collective memories beyond the boundaries of a single collection or repository have long been recognized.⁷ Richard Szary, in his article on EAC, recognized Australia as being an exception to the “lack of recognition in previous systems ... of the interaction of bibliographical and contextual information.”⁸ The development of the OHRM coincided with initial development work on EAC and *Bright Sparcs* entries were used as a reference implementation of the conceptual thinking that underpins the schema.⁹

At the data level, the OHRM has evolved through five main versions and sixty incremental code updates (from version 1 in 1999 to version 5 in 2006), and since early 2003 has consisted of four informatic domains. The first enabled the systematic documentation of information regarding context entities, including, but not limited to, corporate bodies, people, families, places, and events. A context entity was regarded as anything that could be named and defined in terms of space and/or time that was useful in helping understand records or indeed other context entities. The remaining three domains captured information about record entities, including the archives, publications, and material objects that provide the evidential base for the existence of context entities. For pragmatic reasons, separate domains were created for the citation and annotation of published resources; the citation and documentation of archival and heritage resources; and the documentation and management of digital resources such as images, digital multimedia recordings, and other digital documents selected for utilization in the web output. Relationships emerged as a significant part of the conceptual thinking during this period but remained underdeveloped both technically and informatically. They were used to define the interconnections within each domain and between the domains but were not managed as primary or first-class entities in their own right. The appearance of RDF in the late 1990s also influenced the directions of this thinking.¹⁰

Although the publication of the *Guide to the Archives of Science in Australia* (1990) and the online publication of its successor, *Bright Sparcs* (1994), preceded the publication of *International Standard for Archival Authority Records—Corporate Bodies, Persons, and Families (ISAAR(CPF))*,¹¹ the structure of the information domains mapped well to the international standard. Investigations leading to the mapping of the data to *ISAAR(CPF)* resulted in the decision not to change the structure of the underlying data model to fit with the standard, but rather establish the ability to map to

and generate outputs that were compliant with *ISAAR(CPF)*.¹² This organizational practice established an environment that readily supported the export of data to XML, in particular EAC.¹³ The mechanisms (script and code) and protocols for mapping descriptive contextual data to external standards and schemas were in place in OHRM Version 1 for data exchange and the development of other services. Unfortunately, at that time there was nobody in an equivalent conceptual or technical position with whom to share data on the history and archives of Australian science.

In a 1999 paper presented at the Australian Society of Archivists' annual conference, the Director of the Australian Science and Technology Heritage Centre (now ESRC Director) Gavan McCarthy noted that:

The re-development of the systems supporting *Bright Sparcs*, *Australian Science at Work* and the *History of Australian Science and Technology Bibliography*, are focused on the expression of the encoded context entities and their relationships in XML. This will enable specific meaning to be given to data elements within an encoded entity and permit a much greater level of control over elements that define an entity in space and time. This in turn opens the door to the exciting analytical and access opportunities offered by data visualisation and graphic representation.¹⁴

During the subsequent 14 years this work and this vision have been realized with the generation of EAC, and later Encoded Archival Context—Corporate Bodies, Persons and Families (EAC-CPF) encoded data from the Centre's systems, and the harvesting of this data by the National Library of Australia's *Trove*¹⁵ service and the Humanities Networked Infrastructure (HuNI) service.¹⁶ In addition to embracing the harvesting and sharing capabilities of XML, the Centre has also developed further uses for this data in feeding indexing and application tools for online presentation and exploration of contextual information, and the generation of network visualization graphs to display data and the connections between related pieces of data as innovative visual presentations to support exploration and analysis. These developments will be explored further below, using the Find & Connect web resource project as a case study.

Using XML for documenting archival collections and their surrounding context has long been recognized as having "a low technological dependency."¹⁷ The ESRC, founded and run by archivists, has always endeavored to capture documentation and create information resources of enduring value that are sustainable long into the future. The creation of an EAC-CPF XML view of data was seen as a critical element of this strategy as it provided system-independent and highly preservable copies of the data.

UTILIZING ENCODED ARCHIVAL CONTEXT TO SHARE DATA FROM THE EARLY 2000s

When the National Library of Australia (NLA) explored ways to use their name authority files to support a searchable online interface, linking biographical information from a number of other resources, they found that EAC provided the best framework for achieving their goals.¹⁸ Started as the People Australia project, the goal was to link entries in the Australian Name Authority File, maintained by the National Library of Australia, with biographical information in other collections,¹⁹ creating a hub of “collective memory”²⁰ beyond the institutional understanding of the name authority files. *Bright Sparcs* was identified as a source of valuable authority information on Australian science and was included in early planning work for People Australia.²¹ Initially EAC was seen as overly complex and this led the NLA to shy away from using the schema. The informatic richness of the EAC schema, especially in its ability to document relationships both between EAC records and with external public domain resources, were identified as significant strengths and it came to be the metadata schema of choice for the representation of what the NLA referred to as “parties.”²² The ability of the NLA to harvest the content of *Bright Sparcs* and link entries with the relevant authority files would mean that the information about a person from a variety of sources could be gathered together in a centralized service, providing multiple views (or contexts) for that person in a single location. It would also mean that while the content of each resource could continue to be maintained independently, information in each individual resource could be greatly enriched by the ability to tap into a much broader pool of resources about a context entity with a single link.

EAC output from the *Encyclopedia of Australian Science*²³—the successor of *Bright Sparcs*—was amongst the first resources to be harvested by the National Library of Australia for this service, known by this time as *Trove*, in 2010. This provided the ESRC with a powerful working example of the utility and conceptual strength of EAC. Being able to harvest and share data and interoperate with other systems simply by generating data using a common schema meant that both *Trove* and the *Encyclopedia of Australian Science* could enrich their content and expand the network of contextual information related to any single entry.

Similarly, since 2012 the ESRC has been working with the Humanities Network Infrastructure project (HuNI).²⁴ At the time of writing there are sixteen EAC-CPF datasets from OHRM projects being harvested to create an Australian research resource that will allow users to search, analyze, synthesize, and interact with multiple humanities and cultural datasets through a single “virtual laboratory.”

EXTENSION OF EAC

While seeing EAC as looking towards a digital future, Sue McKemmish, Barbara Reed, and Michael Piggott noted in 2005 that the schema was “still closely tied in to custodial models and the records group.”²⁵ Working on upgrading the existing EAC output from the OHRM to EAC-CPF, limitations in the XML schema as strictly defined were found that precluded the inclusion of a range of context entities types. While corporate bodies, persons, and families—the CPF of EAC-CPF—are important elements in providing a contextual information framework, it was found that when working in the broader public domain beyond the narrow world of the archive (in the 4th, or pluralized, dimension of the records continuum)²⁶ further types of context entity were often required. These established a more complete understanding of the world in which archives exist, and thus an understanding of the archives themselves. Events, such as Australian federal and state elections proved important to understanding the *Irving Saulwick Archives*²⁷—the records of a major Australian pollster and social researcher. Key strikes, disputes, and lockouts were identified as important context entities in the *Australian Trade Union Archives*²⁸ resource. Legislation played an important role in contextualizing the records of Australian orphanages, children’s homes, and other institutions within the Find & Connect web resource project.²⁹ Place—not the physical location of archives, but locations key to the information within archival collections, such as locations of memorials, expeditions, and parks and gardens—can also be understood as entities in a contextual sense, triangulating dependencies or links between dispersed collections.

Creating EAC-CPF records just for corporate bodies, persons, and families risked excluding critical information that enhanced the understanding of the world being described. The choice was either to extend the capacity of EAC, or to build a new schema that met the requirements. The decision was taken to extend EAC-CPF records to include a broader range of types of context entities. This involved allowing the use of additional generic entity types—concept, cultural artefact, event, language, natural phenomenon, and place—within the `<entityType>` element, as well as the inclusion of a `<localControl>` element to detail the specific entity types used within a dataset (see Figure 1).³⁰

THE FIND & CONNECT WEB RESOURCE PROJECT

The Find & Connect web resource project provides a useful case study through which to examine the development of the ESRC’s use of extended EAC-CPF records. The ESRC’s work on Find & Connect evolved from a Linkage Project funded by the Australian Research Council, “Who Am I?:



```

- <control>
  <recordId>E000422</recordId>
  <maintenanceStatus>revised</maintenanceStatus>
  + <maintenanceAgency></maintenanceAgency>
  + <languageDeclaration></languageDeclaration>
  + <conventionDeclaration></conventionDeclaration>
  - <localControl localType="typeOfEntity">
    <term>Legislation</term>
  </localControl>
  + <maintenanceHistory></maintenanceHistory>
</control>
- <cpfDescription>
  - <identity>
    - <entityId>
      http://www.findandconnect.gov.au/vic/biogs/E000422b.htm
    </entityId>
    <entityType>concept</entityType>
  </identity>
</cpfDescription>

```

FIGURE 1 Example of the ESRC's Use of Entity Types in EAC-CPF.

The Archive as Central to Quality Practice for Current and Past Care Leavers (Forgotten Australians)” (2009–2012).³¹ A key project output was the *Pathways* web resource, an attempt to map the history of out-of-home “care” in the state of Victoria from the 1840s to the present as a contextual information network, including organizations, record collections, legislation, events, people, published resources, and digital images.³² Archivists and historians conducted the research and developed and entered the relevant entities and their relationships as part of a carefully structured informatic managed in the OHRM. The *Pathways* web resource was then generated from this data as a self-contained set of HTML pages for publication online. At this time, though EAC records could be generated from the OHRM, they were not used as a functional part of the project.

As conceived by the project team:

The broad aims of *Pathways* were twofold. Firstly, to provide multiple avenues through which individuals might discover information about records relating to their time in care, and secondly to provide information that would help them understand those records. The “Who Am I?” team had much experience in this area and believed that much of the

contextual information necessary to discover and understand records was already in the public domain, but was neither systemically accessible nor utilised.³³

Providing information in this way was viewed as vital to the community of “care” leavers and their families to assist them in piecing together their personal history and identity, through historical information and links to records and resources. Furthermore, *Pathways* was seen as an “intervention,” an attempt to change the way in which the community of organizations, governments, advocacy groups, and “care” leavers functioned through the provision of authoritative, well-structured, and publicly accessible knowledge.³⁴

During the first year of “Who Am I?” the needs of “care” leavers were highlighted when then Prime Minister of Australia Kevin Rudd issued a national apology to Forgotten Australians and Former Child Migrants, on November 16, 2009. This was followed by a scoping study by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA), which aimed to assess options for a national “Find and Connect” service which would help people locate, access, and understand relevant records and have recourse to counseling and support services.³⁵ The scoping study was released in September 2010 and identified *Pathways* as the preferred model for a national resource to meet these needs, stating “No other example of a specialised website designed to cater for care leavers, with the search functionality and archival integrity of *Pathways*, was identified in Australia or overseas during the scoping study.”³⁶

The study included reference to the ten underlying principles of the *Pathways* resource. These were provided by the “Who Am I?” project team, who affirmed that a public knowledge resource should be:

1. Standards based
2. Evidential
3. Persistent and meaningful through time
4. Resilient
5. Available through multiple access and reference points
6. Geared towards person-to-person knowledge sharing
7. Structured to enable computer-to-computer data sharing
8. Public knowledge focused
9. Based on the science of networks
10. Coherent and purposeful in its interface design³⁷

Following confirmation of funding from FaHCSIA, the 3-year Find & Connect web resource project commenced in May 2011, and the web resource was published for the first time on November 15, 2011. Though at a much larger scale than “Who Am I?” the early launch was made possible

because of the existing systems and the conceptual models underpinning them.

Prior to launch, nine OHRM databases were established—one for each Australian state and territory, and one national database as an entry point—populated with content from pre-existing state guides and other readily accessible public material. State Based Historians were employed by the Australian Catholic University to develop content for each OHRM, with archivists, a National Editor and technical staff based at the University of Melbourne's ESRC. As with "Who Am I?," the project would be overseen by three academics: Gavan McCarthy (archival science and cultural informatics), Shurlee Swain (history), and Cathy Humphreys (social work).

Sets of stand-alone HTML pages were generated from the OHRM databases, sharing a generic structural form but each with their own templates and cascading style sheet files to allow a different color scheme for each state and territory. System-generated browse lists were created to assist with content discovery, and search functionality was based on the indexing of HTML pages. Though EAC was generated, it was only used as part of a broader disaster recovery and technology dependency reduction strategy. With project data in the early stages of discovery and curation, the decision was taken not to share it with the National Library of Australia until the project had reached sufficient maturity.

USABILITY TESTING AND AUDIENCE

Though the *Pathways* model provided a starting point, there was an awareness from project commencement that technology design and development would be a key focus over the 3 years (2011–2014). The project team also knew that wholesale change was not feasible, instead focusing on incremental change over the course of the project based initially on user needs. With reference to the principles noted above, extensive usability testing was included in the project plan to determine whether the web resource effectively supported person-to-person knowledge sharing. And, though the OHRM's ability to export EAC-CPF would allow computer-to-computer data sharing, as noted above, this was not implemented when the Find & Connect web resource was launched. Overall, it was felt that the potential of standards-based data (including EAC-CPF output) was not being fully realized and that this would need to occur to meet the specific demands of the project and the Find & Connect audiences.

From inception the Find & Connect web resource project defined its primary audience as Forgotten Australians and Former Child Migrants, the approximately 500,000 people who experienced out of home "care" in Australia from the 1920s until 1989. Numerous secondary audiences were also identified, such as support services, advocacy groups, past providers,

record holders, family historians, researchers, and others with an interest in the history of out-of-home “care” in Australia. However, if the primary audience of the web resource had to rely on support from more experienced users it would not fully serve its purpose. Therefore, from March 1 to June 30, 2012, usability expert Roger Hudson conducted a review of the web resource and task-based heuristic evaluation of the site with a representative sample of potential users.³⁸

Usability testing was conducted with 27 participants in four states, with Forgotten Australians and Former Child Migrants comprising more than 80% of the test participants. The methodology was as follows:

An expert review of the usability or user-focus of a site involves assessing selected pages from different levels of the site with reference to predetermined usability and performance criteria. [...] Following the expert review, a task-based heuristic evaluation was used to assess the usability of the Find and Connect websites. This involved a representative sample of potential users of the site using it for a series of typical tasks. The process is based on the heuristic evaluation method developed by Jakob Nielsen and Rolf Molich in 1990 and later refined by Nielsen for use with the Web. The actions and utterances of each participant were observed and recorded along with any difficulties they encountered doing the tasks.³⁹

Among the many findings were the following key points:

- The most sought-after content was information about Homes (organizations that housed children), photographs, and information about who to contact when looking for records related to time spent in “care.”
- The division of the web resource by state and territory caused confusion.
- The overall structure and functionality of the site did not align with the main needs of the primary audience.
- Search functionality needed to include filters to allow users to target search requests more precisely.

Meeting the needs of Forgotten Australians and Former Child Migrants was a significant challenge whatever the site design. To generalize, because of the nature of a childhood spent in “care,” average levels of web literacy and general literacy were substantially lower than the general population. In addition, the knowledge domain being mapped—and the resulting informatic—was necessarily complex. Organizations changed names, merged, moved, or disappeared. Record collections had been moved, split, amalgamated, or lost. Records and resources were distributed across institutions, including governments, past and current providers, and state or national archives and libraries. And all this in a sector that included hundreds of

large and small organizations, record holders, and collections all impacted by changing legislation across multiple jurisdictions over more than 150 years.

Given this, the results of the first round of usability testing were perhaps inevitable (or at least not unexpected). Focusing on the 23 Forgotten Australians and Former Child Migrants who participated, they were able to complete only 40% of the tasks attempted without assistance from the facilitator. It was clear to the Find & Connect project team that **the necessarily complex data structures**, long lists of related entities, text-heavy pages, and limited search functionality **did not meet the needs of this audience**; but simplifying the complex data structures in the OHRM **would compromise the accuracy, authority, and sustainability of the contextual information network being developed**. And, though many users were struggling to find content, when shown relevant content they confirmed it was exactly the sort of information they were seeking.

One significant issue was an existing OHRM system limitation that implied a **one-to-one match** between entities or resources in the database and the HTML output. In other words, every entity (with relationships to other entities) was rendered by the OHRM as its own uniquely identified HTML page. Although useful as a complete record, a user **looking for record holdings, photographs, and contextual information** about a single Home might have to visit three or more web pages—sometimes many more—to locate this information. Success required the user to interpret and understand the complex network model used to map relationships between entities, and then to navigate a series of links on several long, textually rich pages, piecing together some details while ignoring or passing over others.

DEVELOPING A NEW DISSEMINATION MODEL BASED ON EXTENDED EAC-CPF

Addressing the usability issues required a substantial redesign of the way entities and their relationships were processed, rendered, and presented on the web. The project team began to explore the idea of creating **purpose-driven web pages**, created by aggregating information from multiple entities, based on the identified primary needs of users. The OHRM's inbuilt HTML generation functionality did not support this approach, and developing an active link to the underlying database was not a favorable solution. The latter was both technically problematic and conceptually undesirable, as **the effective use and preservation of the database and responsiveness of the web resource both required a clear separation of offline data and online dissemination output**.

Therefore, an online store of stable, structured data from which to render content as part of the dissemination layer was prepared. With EAC-CPF export capability already in place, this was the obvious candidate. From this

point the redesign of the Find & Connect web resource went through a number of stages, many of which were iterative and overlapping:

- Page mockups were designed based on user requirements.
- EAC-CPF output was analyzed and, where necessary, adjusted.
- Data structures were reviewed and, where necessary, adjusted.
- Usability testing was conducted based on the new page design.
- Specifications for a new EAC-CPF-based application layer were developed.
- The application was built, tested and deployed.

The first mockups were developed starting with user requirements. There was an inherent tension here. In moving away from the rendering of “whole” entities on individual pages it was important that previous assumptions and way of looking at the world (in part embodied by the OHRM informatic and EAC-CPF structures) did not result in simply rendering this complexity unquestioned. At the same time, it was important to maintain certain key principles in the way information was displayed. For example, although some information from related entities needed to be extracted and displayed on a page about a Home, the key identity metadata for that Home—including title, dates, and summary note or abstract—should remain visible and prominent.

The OHRM informatic contained data not supported by the EAC-CPF schema, for example the elevation of publications and other resources to first class objects. Therefore, the stand-alone HTML information pages generated by the OHRM (containing this complete data) needed to be retained as “reference” pages. “Super users” would therefore still have access to the full complexity of the information from the complete dataset as and when required.

Next, the OHRM’s current EAC-CPF output was analyzed to determine if all the required information was included. To this point the primary purpose of exporting EAC-CPF from an OHRM was for harvesting by the National Library of Australia’s *Trove* discovery service via OAI-PMH.⁴⁰ This meant EAC-CPF content was geared to meet the NLA-*Trove* needs (within the scope of the EAC-CPF schema itself). A review of the current data structures happened concurrently, with a short iterative cycle of EAC-CPF and data structure adjustment to ensure the required information was available. The overall strength of both the OHRM informatic and EAC-CPF schema were clearly evident during this process, with few adjustments required to meet the project’s needs.

Once the feasibility of the approach had been confirmed, a second round of usability testing was conducted with active, online mockups of the new web design. Based on the idea that the primary audience were likely to start by searching for a particular institution or Home, these pages were a



FIGURE 2 Example of a Summary Page in the Updated Find & Connect Web Resource.

particular focus. The design included the following internal page navigation tabs (see Figure 2⁴¹):

- **summary**—containing key identifying information about the Home, a key-stone image, and links to immediate predecessor and successor organisations (where applicable);
- **records**—drawing information from related archival entities, and the contact details of the organizations holding those records;
- **photos**—drawing information from related digital objects;
- **location**—drawing information from the events section of the entity information; and
- **full page**—linking to the full HTML entity page generated by the OHRM.

Testing was conducted using the same approach as the first round of testing, with most of the same tasks (some adjusted to reflect the new navigation options and functionality planned for the new site). Participants were able to complete all tasks **96.3% of the time**—a substantial improvement from the first round, and firm evidence the new design would go a long way to meeting the audience's needs. With the fundamental design all but

confirmed, the project team turned to the practicalities of implementing the new “application layer” of the web resource.

IMPLEMENTING A HYBRID SOLUTION

Retention of the full, OHRM-generated HTML pages meant that the new design was in essence a hybrid approach, moving toward a new way of presenting and rendering data rather than attempting to implement a completely new model in a single step.

The first step was to examine the informatics pathways between EAC-CPF XML files and the HTML layer of the site. In October 2012, following discussions with technical staff, an XML reference was added to the OHRM so that stand-alone pages—including those produced for Find & Connect—would include a specific link to the relevant EAC-CPF file without the need to infer this link from the page URL. Each EAC-CPF file also contained a link to its corresponding HTML page as its <entityID> (see Figures 1 and 3⁴²) and href links to the HTML files for related entities and resource relations. When it came to mapping out the application specifications, this made traversal between the HTML layer and XML layer relatively straightforward.

Moving through multiple XML files required an additional step. All the references in the EAC-CPF output were URLs for HTML files, with none directly to the XML files for related entities. Each URL does contain a unique identifier matching the identifier used to name each EAC-CPF file; however, after some discussion it was agreed that parsing every URL for the identifier then using this to construct a path for the corresponding XML record could be problematic and would be difficult to maintain. In contrast, the paths from XML to HTML and vice versa were complete. Therefore, the decision was taken to follow the full links already embedded in the HTML and XML layers. For example, starting at Entity A and retrieving the XML for a related

```
<meta name="DC.Title" lang="en" content="Ballarat Orphanage - Organisation - Find & Connect Victoria" />
<meta name="DC.Creator" lang="en" content="Find & Connect Project" />
<meta name="DC.Subject" lang="en" content="Ballarat Orphanage - Organisation - Find & Connect Victoria, ar
<meta name="DC.Description" lang="en" content="Ballarat Orphanage - Organisation - Find & Connect Victoria
<meta name="DC.Publisher" lang="en" content="eScholarship Research Centre, The University of Melbourne" />
<meta name="DC.Date.Created" scheme="ISO8601" lang="en" content="2009-02-13" />
<meta name="DC.Date.LastModified" scheme="ISO8601" lang="en" content="2013-10-22" />
<meta name="DC.Type" lang="en" content="Document" />
<meta name="DC.Format" scheme="IMI" lang="en" content="text/html" />
<meta name="DC.Identifier" scheme="URL" lang="en" content="http://www.findandconnect.gov.au/vic/biogs/E000031b
<meta name="DC.Language" scheme="ISO639" lang="en" content="en-gb" />
<meta name="DC.Rights" lang="en" content="Creative Commons Attribution-Non-Commercial-Share Alike 3.0 Australi
<meta name="Author" lang="en" content="Find & Connect Project" />
<meta name="Description" lang="en" content="Ballarat Orphanage - Organisation - Find & Connect Victoria, F
<meta name="Keywords" lang="en" content="Ballarat Orphanage - Organisation - Find & Connect Victoria, arch
<meta name="Generator" content="Online Heritage Resource Manager (OHRM) V Alpha 2.16+++" />

<meta name="EAC" lang="en" content="http://www.findandconnect.gov.au/vic/eac/E000031.xml" />
```

FIGURE 3 Example of the Metadata on an HTML Page for the Find & Connect Web Resource Showing the Reference to the Related EAC-CPF Rendering of the Same Data.

entity (Entity B) would require following the link from the entity relation in Entity A's EAC-CPF to the static HTML page for Entity B, finding the link to Entity B's EAC-CPF record, and following that link back into the XML layer.

It was critical that all the data required by the application layer was readily accessible. As described earlier, the OHRM could be used to generate EAC-CPF for all entities, regardless of type (not only for corporate bodies, persons, and families). The Find & Connect web resource treats archival collections, series, and items as context entities in their own right (rather than recording them in the OHRM archival resources informatic domain), because the existing OHRM informatics, based on a model of hierarchical relationships between levels of archival units, did not allow for the complexity of relationships required to accurately map how these resources were situated in the overall network of entities and resources. This outmoded informatic reflected what is now seen as an overly limited provenencial view of how records are contextualized.

The only required information that was less easy to access was the contact details of organizations. Previously this had been stored (with other information) in an open text field. During the planning stages a new "Contact Details" entity was added to the informatic to manage this information, with relationships to the relevant organization and to the archival entities to which the contact details applied. The application layer could then retrieve the required block of text from a single element (<biogHist>).

With these pieces in place, the project team started mapping out the various pathways and pieces of content required to assemble the elements making up each page. This was accompanied by rendering information, screenshots, and descriptions of the intended functionality. The developer then used these pathways to write a Python application that would retrieve the necessary content from the EAC-CPF records, convert this information to JSON and render it as a web page in Find & Connect templates.

For example, looking at the organizational entity for Melbourne Orphan Asylum, a request to render the page starts at the EAC-CPF record:

<http://www.findandconnect.gov.au/ref/vic/eac/E000180>

The application then retrieves information from nine related EAC-CPF records:

Organization—St James' Orphan Asylum and Visiting Society:

<http://www.findandconnect.gov.au/ref/vic/eac/E000181.xml>

Organization—Melbourne Orphanage,;

<http://www.findandconnect.gov.au/ref/vic/eac/E000275.xml>

Glossary Term—Orphanage:

<http://www.findandconnect.gov.au/ref/vic/eac/E000116.xml>

Archival Collection—OzChild Records:

<http://www.findandconnect.gov.au/ref/vic/eac/E000734.xml>

Contact Details—OzChild—Contact Details:

<http://www.findand-connect.gov.au/ref/vic/eac/E000881.xml>

Archival Collection—Miscellaneous papers, 1857-1867, State Library of Victoria:

<http://www.findandconnect.gov.au/ref/vic/eac/E000687.xml>

Contact Details—State Library of Victoria—Contact Details:

<http://www.findandconnect.gov.au/ref/vic/eac/E000855.xml>

Archival Item—Anne's Story, State Library of Victoria:

<http://www.findandconnect.gov.au/ref/vic/eac/E000686.xml>

Archival Collection—Melbourne Orphan Asylum Records, State Library of Victoria:

<http://www.findandconnect.gov.au/ref/vic/eac/E000685.xml>

The content is converted to JSON, which is publicly accessible as part of the web resource: <http://www.findandconnect.gov.au/guide/vic/E000180/json>

Finally, the JSON is rendered to create the page seen by users: <http://www.findandconnect.gov.au/guide/vic/E000180>

OTHER FIND & CONNECT WEB RESOURCE DEVELOPMENTS

EAC-CPF output was also used as the basis for other developments. There was an identified need to improve search functionality. The previous version of the Find & Connect web resource relied on the indexing of HTML pages.

The screenshot displays the Find & Connect web resource interface. At the top, the logo features a stylized kangaroo and the text "FIND & CONNECT" with the tagline "History & information about Australian orphanages, children's Homes & other institutions". To the right, contact information for "Find & Connect Support Services" is provided: "1800 16 11 09" and "Freecall: Monday–Friday 9am–5pm".

A navigation bar contains links: "HOME", "ABOUT Find & Connect", "Look for HOMES", "Look for PHOTOS", "Look for RECORDS", "BROWSE Find & Connect", "HELP with this site", and "CONTACT support/counseling".

On the left, a "State" facet lists Australian states and territories with counts: New South Wales (1770), Western Australia (1739), Victoria (1274), South Australia (1247), Tasmania (901), Queensland (812), Northern Territory (413), and Australian Capital Territory (127). Below this, a "Function" facet lists categories: Home (1758), Care Provider (1163), Government-run (576), Children's Home (557), Catholic (383), Non-denominational (293), and Anglican (290).

The main content area shows search results. A warning banner states: "Some people may find content on this website distressing. [Read more](#)". Below is a search bar with a "Search" button. The first result is "The Salvation Army (Western Australia) Property Trust Act 1931", dated 1931–1931, Western Australia. It describes a digital copy of a statute held by the State Law Publisher in Western Australia, which was passed to regulate the financial affairs of the Salvation Army in Western Australia. A link to the document is provided: <http://www.findandconnect.gov.au/wa/objects/WD0000274.htm>. The second result is "Nedlands Library", Western Australia, which has a local studies collection including information about the Salvation Army and children's homes in the area. The third result is "Longmore Remand and Assessment Centre", 1965–2004, Western Australia, which provided short-term residential care for arrested children awaiting Court appearances, those on remand, and those committed to the Department as wards, who are assessed and placed.

FIGURE 4 Example of Search Results From the Find & Connect Web Resource, Showing Facets for Further Filtering of Results.

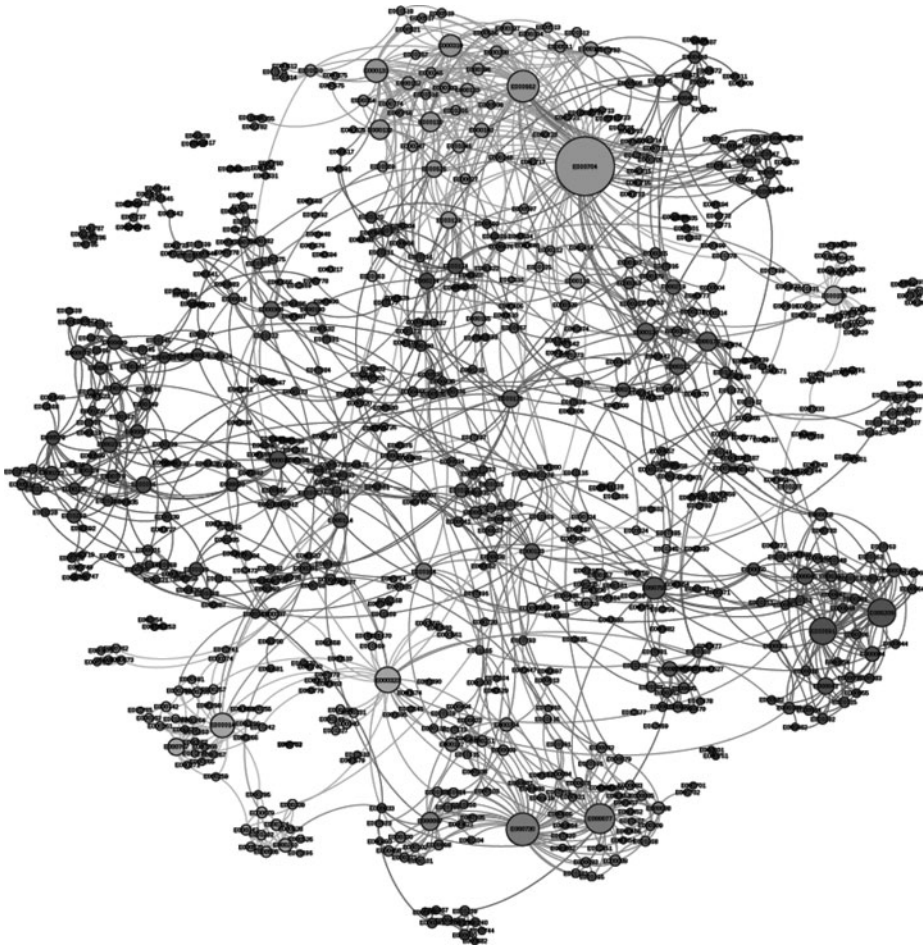


FIGURE 5 Example of a Network Visualization of the Find & Connect Data for Victoria, With Circles Representing Entities and Relationships Represented as Lines Between Related Entities. This Provides an Overview of All Entities and Relationships, Enabling the Easy Identification of Clusters.

Although these pages were well structured (including an embedded header of Dublin Core metadata for each individual page), it was not possible to effectively implement filters and facets, which proved increasingly problematic from a usability perspective as the dataset continued to grow.

A new Solr indexer was written using Apache Lucene to index the EAC-CPF records. The structured, well-fielded XML data allowed a more targeted approach to search, including pre-applied filters for particular types of entity (underlying the “Look for Homes,” “Look for Photos,” and “Look for Records” options on the Find & Connect web resource) and a general search page with user-controlled facets to filter by state, function, entity type, and date range, as seen in Figure 4.⁴³

ESRC staff also continued to explore the possibility of using extended EAC-CPF records as data for network visualization tools for analysis, exploration, and quality checking. Figure 5 shows a visualization of the Find & Connect Victoria dataset from June 2012.

And, though not publicly accessible at the time of writing, a geographic mapping tool has been developed which uses location information in the events (<chronItem>) section of EAC-CPF files to infer latitudes and longitudes and present entities on a map of Australia.

CONCLUSION: ONGOING DEVELOPMENT—2014 AND BEYOND

The enhancements for the Find & Connect web resource are being planned within the context of the Centre more broadly.

Current developments to the OHRM system have enabled not only all context entities to be exported as extended EAC-CPF XML files, but also a complete bibliography of published resources to be exported in MODS 3.5 XML format, and references to archival resources and digital items to be exported as a set of EAD XML files. All of the files will include references and links to related files both the same schema and across the set of XML schemas. This means that all data from the OHRM can be exported as a complete set of XML records for the world being described. It will also mean that it will be possible to navigate through the complete, standards-based XML output from any entry point to discover related information. Being able to render all of data from an OHRM in XML formats will enable the navigation through the XML datasets and will eliminate the current need to reference the stand-alone HTML renderings of data where there are gaps in the XML.

The complete set of OHRM XML records will be human-readable, independent from the system that created them and indeed independent of any specific technological platform. This means that resources are not locked into any single technology or system and an enduring preservation copy of complete datasets that can be exported, shared, and preserved long into the future is produced. OHRM XML records will also allow not only for the maintenance of the compliance of data with XML schemas such as EAC-CPF and capacity to collaborate with the international EAC-CPF community, but also to extend these schemas to ensure that all of the data captured for an entity can be fully expressed.

In late 2013, work was well advanced on developing the next generation of the OHRM system. The new version will move from using a desktop application to a web-based system, allowing for a multiuser environment and a more geographically distributed use of the system. Although many of the eventual outputs from the new system are still under discussion and in development, it is planned to maintain the use of a completely separate,

independent set of stand-alone HTML pages and XML data generated from the central dataset.

It is planned to build on the work done to use the EAC-CPF version of the data for web presentation in the Find & Connect web resource and it is likely that in the future most, if not all, of the online display version of datasets will be fed by EAC-CPF, EAD and MODS 3.5 data. Developments in search functionality, indexing, and other tools built for the exploration, access and discovery of all OHRM datasets will be built to interact with the XML feeds from the datasets.

2014 represents the twentieth anniversary of the first publication of *Bright Sparcs*, and the fifteenth anniversary of the OHRM. While many of the principles initiated in 1994 remain central to the ESRC's conceptualization of the archival mission, it recognizes that the immediate future is one of significant development and change. Version 6 of the OHRM is in active development and seeks to exploit the many advances in web service technology and the work of the International Council on Archives Expert Group on Archival Description in revisiting the conceptual modeling that underpins the work of the archival community. But, though technological change is inevitable, the informatic and conceptual principles—embodied by the OHRM informatic and expressed via schema such as EAC-CPF—continue to lie at the core of the ESRC's approach to documenting and describing human activity, its evidence and its consequences.

ACKNOWLEDGMENTS

Collaboration lies at the heart of the ESRC's work and the developments outlined above. Particular thanks are due to Dr. Joanne Evans, Daniel Pitti, Professor Cathy Humphreys, Professor Shurlee Swain, Helen Morgan, Rachel Tropea, Dr. Marco La Rosa, Davis Marques, Dr. Tim Sherratt, Rosanne Walker, Basil Dewhurst, Dr. Steve Melnikoff, all the staff of the ESRC and its predecessors, and the staff of the Find & Connect web resource project team.

NOTES

1. Gavan McCarthy, "Heritage and the Internet—Encoding Context Objects: Using Knowledge to Reduce Risks," Australian Society of Archivists 1999 Conference (1999), via the Internet Archive, Wayback Machine October 30, 2013, at <http://web.archive.org/web/20081006105606/http://www.archivists.org.au/events/conf99/mccarthy.html>.

2. "Our History," eScholarship Research Centre, accessed November 28, 2013, <http://www.esrc.unimelb.edu.au/about-us/our-history/>.

3. Gavan McCarthy, *Guide to the Archives of Science in Australia: Records of Individuals*, (Port Melbourne: D.W. Thorpe in Association with Australian Science Archives Project & National Centre for Australian Studies, 1991).

4. Gavan McCarthy, "Review of National Library of Australia's Reference Services" (1998), via the Internet Archive, Wayback Machine, March 13, 2014, <http://web.archive.org/web/20081123034658/http://www.library.yale.edu/~rszary/Authority/gavan1.html>.

5. Gavan McCarthy and Joanne Evans, "Mapping the Socio-technical Complexity of Australian Science: from Archival Authorities to Networks of Contextual Information," *Journal of Archival Organization* 5, no. 1–2 (2008): 164.
6. Ibid.
7. Sue McKemmish, Barbara Reed, and Michael Piggott, "Chapter 7: The Archives." In *Archives: Recordkeeping in Society*, eds. Sue McKemmish, Michael Piggott, Barbara Reed and Frank Upward (Wagga Wagga: Centre for Information Studies, Charles Sturt University, 2005): 159–195; Peter Scott, "The Record Group Concept: A Case for Abandonment," *The American Archivist* 29, no. 4 (1966): 493–504.
8. Richard V. Szary, "Encoded Archival Context (EAC) and Archival Description: Rationale and Background," *Journal of Archival Organization* 3, no. 2–3 (2006): 220.
9. McCarthy, "Heritage and the Internet."
10. Eric Miller, "An Introduction to the Resource Description Framework," *D-Lib Magazine*, (May 1998), accessed March 14, 2014, <http://www.dlib.org/dlib/may98/miller/05miller.html>.
11. Gavan McCarthy, "Engineering Utility: A Visionary Role for Encoded Archival Authority Information in Managing Virtual and Physical Resources," AusWeb99, the Fifth Australian World Wide Web Conference (1999), accessed October 30, 2013, <http://ausweb.scu.edu.au/aw99/papers/mccarthy/>.
12. McCarthy and Evans, "Mapping the Socio-Technical Complexity of Australian Science," 166.
13. Daniel V. Pitti, "Creator Description—Encoded Archival Context," Co-published simultaneously in *Cataloging & Classification Quarterly* 38, no. 3–4 (2004): 201–226; and: *Authority Control in Organizing and Accessing Information: Definition and International Experience*, eds. Arlene G. Taylor and Barbara B. Tillett, (The Haworth Information Press, an imprint of The Haworth Press, Inc.: 2004), 201–226, accessed December 20, 2013, <http://www.columbia.edu/cu/libraries/inside/units/bibcontrol/osmc/pittieac.pdf>.
14. McCarthy, "Heritage and the Internet."
15. Trove, National Library of Australia, accessed November 25, 2013, <http://trove.nla.gov.au/>.
16. Humanities Networked Infrastructure website, accessed November 25, 2013, available at <http://huni.net.au/>.
17. McCarthy, "Heritage and the Internet."
18. Basil Dewhurst, "People Australia: A Topic-based Approach to Resource Discovery", VALA2008 14th Biennial Conference and Exhibition (2008), available at http://www.valaconf.org.au/vala2008/papers2008/116_Dewhurst_Final.pdf.
19. Ibid., 1–2.
20. McKemmish, Reed, Piggott, "Chapter 7: The Archives," 159.
21. Dewhurst, "People Australia," 2.
22. Dewhurst, "People Australia."
23. *Encyclopedia of Australian Science*, eScholarship Research Centre, accessed November 25, 2013, <http://www.eoas.info/>.
24. Humanities Networked Infrastructure website.
25. McKemmish, Reed, Piggott, "Chapter 7: The Archives," 168.
26. Sue McKemmish, "Yesterday, Today and Tomorrow: A Continuum of Responsibility," first published in *Proceedings of the Records Management Association of Australia 14th National Convention*, 15–17 September 1997, accessed December 3, 2013, <http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-smckp2.html>.
27. "Saulwick Polls and Social Research," The University of Melbourne, accessed November 25, 2013, <http://www.saulwick.info/>.
28. "Australian Trade Union Archives," eScholarship Research Centre, accessed November 25, 2013, <http://www.atua.org.au/>.
29. "Find & Connect," Find & Connect Web Resource Project for the Commonwealth of Australia, accessed November 25, 2013, <http://www.findandconnect.gov.au>.
30. "Adoption Act 1984," Find & Connect web resource, accessed December 20, 2013, <http://www.findandconnect.gov.au/ref/vic/eac/E000422.xml>.
31. Cate O'Neill and Nell Musgrove, "The "Who Am I?" Project (2009–2012)," Find & Connect (2013), accessed November 28, 2013, <http://www.findandconnect.gov.au/ref/vic/biogs/E000123b.htm>.
32. "Pathways," Who Am I? Project, eScholarship Research Centre, via the Internet Archive, Wayback Machine, accessed December 20, 2013, <http://web.archive.org/web/20091227051326/http://www.pathwaysvictoria.info/>.
33. Gavan J. McCarthy and Joanne Evans, "Principles for Archival Information Services in the Public Domain," *Archives and Manuscripts* 40, no. 1 (2012): 55.

34. Gavan J. McCarthy, Shurlee Swain, and Cate O'Neill, "Archives, Identity and Survivors of Out-of-Home Care," *Archives and Manuscripts* 40, no. 1 (2012): 2.
35. McCarthy and Evans, "Principles for Archival Information Services," 55.
36. A. Elliott and T. Smith, "Find and Connect Service Scoping Study," Department of Families, Housing, Community Services and Indigenous Affairs, Commonwealth of Australia (September 2010): 32, accessed November 28, 2013, http://www.dss.gov.au/sites/default/files/documents/find_connect_scoping_study.pdf.
37. McCarthy and Evans, "Principles for Archival Information Services," 57.
38. Internal project report by Roger Hudson, "Find and Connect—Usability Evaluation Report. V1," Web Usability (June 30, 2012), 1.
39. Ibid., 8–11.
40. "Open Archives Initiative Protocol for Metadata Harvesting," accessed December 20, 2013, <http://www.openarchives.org/pmh/>.
41. "Ballarat Orphanage," Find & Connect web resource, accessed December 20, 2013, <http://www.findandconnect.gov.au/guide/vic/E000031>.
42. "Ballarat Orphanage," Find & Connect web resource, accessed December 20, 2013, <http://www.findandconnect.gov.au/ref/vic/blogs/E000031b.htm>.
43. "Browse Find & Connect—Search Results," Find & Connect web resource, accessed December 20, 2013, http://www.findandconnect.gov.au/browse/search-results/#/q=*&rows=10&fl=*&json.wrf=JSON_CALLBACK&wt=json.