Archives and Manuscripts, 2015
Vol. 43, No. 1, 42–60, http://dx.doi.org/10.1080/01576895.2014.959535

Routledge
Taylor & Francis Group

Towards Archive 2.0: issues in archival systems interoperability

Gregory Rolan*

Gregory Rolan is a PhD Candidate at the Monash University Centre for Organisational and Social Informatics investigating archival systems interoperability and community access to archives. He returned to study following a 30-year career in IT spanning enterprise systems management, software architecture and development, industry training, and a high-tech start-up. His research comprises theoretical as well as design-science/action-research investigations of systems interoperability; conceptual modelling in archival informatics; metadata standards-setting; and organisational/social factors in archival systems design and implementation.

Despite recent developments in archival systems, standards and practice, the delivery of web-based archival services continues to present significant barriers to access by members of the community at large. One cause of this is the lack of web-based archival systems interoperability that would otherwise facilitate discovery of, and access to, records by a broad constituency. Instead, monolithic archival control systems continue to position archives as jurisdictional resources that privilege a research-oriented audience. This paper describes a research project that explored the issues surrounding the development of web-based services for interoperability of archival control systems conformant with the Australian Series System. The study identified a number of significant interoperability challenges for archival systems and provided further evidence that requirements for interoperability must be 'designed in' and cannot be retrofitted with reliability or ease. It also identified areas in which conceptual and representational recordkeeping and archival standards could be improved. Current recordkeeping and archival standards appear to be insufficiently prescriptive to ensure interoperability, and do not model all of the required elements to facilitate discovery and access by the members of the wider community. From an organisational perspective, the study found structural barriers to progressing interoperability initiatives for community access.

Keywords: archives; discovery; interoperability; research; web

Introduction

The archive in society has become a heterogeneous mix of institutional, corporate and community repositories of memory.[1] However, this fragmented and contested Archival Multiverse[2] presents significant barriers to access by members of the community at large – whether they wish to locate and access personal records, or simply make their voices heard as a counterpoint to official narratives. These obstacles are a result of the interplay between: the historic provision of administrative services to a predominantly research-oriented clientele; a gatekeeper approach to custodial practice; and monolithic archival control systems that position archives as jurisdictional resources that serve privileged narratives.

Despite recent developments in systems, standards and practice, these obstacles, for the most part, are well entrenched. For example, a succession of Australian

*Email: greg.rolan@monash.edu

government-led enquiries has consistently recommended a consistent and integrated approach to the provision of archival services.[3] Similarly, academics and non-institutional researchers alike have long documented the importance of interoperable public domain archival information services, some going as far as to propose a deeply integrated and federated archival commons.[4] Yet, archival service providers continue to interpret their custodial role as one of jurisdictional protection; bolstering the walls of 'fortress archives' instead of opening up and embracing the commons.

Web 2.0 interoperability techniques are one mechanism by which distributed, heterogeneous systems may be integrated in the absence of centralised control. Such techniques, based on open standards and comparatively simple protocols, are fundamental to the Web as an emergent platform for both collaborative interaction and systems interoperability. Web 2.0 behaviours are implemented using Application Programming Interfaces (APIs) – sets of programming instructions and standards that determine how software components should interact with each other. Using Facebook 'like' buttons on news stories, posting blog links to a Twitter feed and employing Google Maps to display locations are all examples of employing interoperable, web-based APIs. In the archival context, Web 2.0 techniques may be a way to overcome organisational and cultural divides and achieve integrated discovery of, and access to, records, as well as facilitating memory narratives from a plurality of perspectives.

This paper describes a research project that explored the issues surrounding the development of a web-based API for archival systems interoperability. An API reference model was developed that comprised a conceptual model of recordkeeping and other metadata required for interoperability, together with definition of encoding standards and programming syntax. Additionally, a proof-of-concept implementation of the API service was developed and evaluated, leading to insights regarding barriers to transition from 'fortress archives' to a standards-based archival commons.

This paper is organised as follows. The following section provides a background to the evolution of, and requirements for, an archival commons. The third section explains the research approach taken for this study while the fourth describes the major research outcomes. The issues relating to the archival systems interoperability and the prognosis for an archival commons in the current milieu are discussed in section five. The paper then concludes with suggestions for further research.

Background

Archival practice has developed around a primary concern for the integrity of the record, that is, as 'defenders of the … record of social and organisational activity'.[5] As a result, archival institutions and the archival and recordkeeping profession exhibit an inward focus that is directed at ensuring the integrity of records in custody. Interestingly, this focus is often couched in terms originally described by Hilary Jenkinson, of ensuring the moral and physical defence of records.[6] In fact, Jenkinson prioritised 'defence against all kinds of dangers' ahead of 'generally making [records] available for use by *students*' (emphasis mine). Such archival practice is reflected today in the use of monolithic systems that are used to describe incoming records, arrange for their preservation and provide for orderly access by records-seekers.[7] The term *monolithic* is used in this context to refer to systems conceived and executed as stand-alone repositories without consideration of interoperability. Unfortunately, this inward focus has been at the expense of a view that considers how records could or should be accessed and used by those who are not members of an administrative or research audience.[8] For example,

restorative justice can be made possible by opening-up the archive for access by the broader community.[9]

To give an Australian example, a number of government inquiries over the last 15 years have addressed the dislocation of individuals that took place throughout the late nineteenth and most of the twentieth centuries.[10] These inquiries consistently highlighted fundamental problems in the conception and execution of archival services in Australia in relation to access to records by the broader community.[11] Such issues can be attributed to a failure of archival service providers to embrace community needs and contextualise their role in relation to other archival services.[12] This context comprises (co-)creators, custodians, stakeholders and seekers of records, and spans individuals, community groups, commercial organisations as well as state institutions. Interoperability in this context involves continua of users and uses – from creation to discovery, and from individuals to institutions as shown in Figure 1. Such interoperability within the Archival Multiverse has been identified as a core principle that needs to be upheld by public domain archival information services.[13]

Unfortunately, archival institutions, in general, do not see interoperability in terms of such continua. Institutional archival culture regards individuals primarily as record-seekers or, perhaps, in possession of knowledge that may be useful for augmenting official records. Similarly, eminent persons, community groups and other organisations are generally considered as potential donors. It is only other institutional archives that are recognised as evidentiary custodians of records and suitable peers for systemic interoperability.

In this light, archival institutions have thus far attempted to address the issues of interoperability and community relevance in two ways, as shown in Figure 1. The first involves the exchange of description metadata between cooperating institutions. These batch-oriented transfer mechanisms have their roots in library automation systems[14] and provide for the aggregation of finding aids and linking of records among archival consortia. Examples of such arrangements include: the Humanities Networked Infrastructure (HUNI) project in Australia, Europeana and the Archives Portal Europe in Europe, the SNAC and ArchivesSpace projects in the United States and the Archives Canada network in Canada. The second mechanism has been termed the 'participatory archive', whereby the community is given the opportunity to contribute to institutional collections in various ways.[15] Often conceived as addenda to the main collection, such initiatives

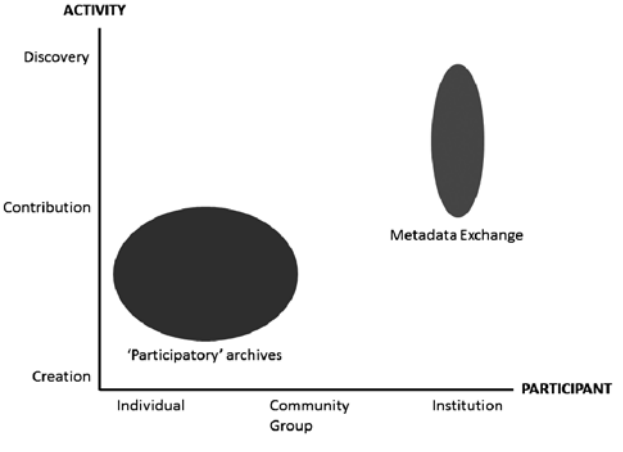

Figure 1. Interoperability continua and institutional perspectives.

allow the annotation or transcription of material, or the submission of additional records that complement existing holdings. Such participatory archive initiatives have met with varied success as 'simply building [participative] initiatives does not mean that users will come to use them'.[16] What is also apparent is that institutional archives do not consider interoperability in terms of the primary creation of material, nor in regards to discovery and access, save for the sharing of catalogue metadata with other institutions.

Interestingly, the term *participatory archive* is becoming rather overloaded. For example, research in the context of indigenous or post-conflict archives has surfaced a range of stakeholder roles, rights and archival principles for participatory archives that embrace both dimensions of the interoperability continua described above. Similarly, although it is not an archival control system (and operates at arms length from the holders of archival material), the Find & Connect Web Resource Project is an exemplar participatory project, enjoying a high degree of involvement from individuals through to institutions in the creation and maintenance of a highly curated and deeply linked register of archival material.[17]

Both of the institutional approaches shown in Figure 1 bring into sharp relief the boundaries of control, provenance and authority that pervade the archival 'gatekeeper' role and mindset.[18] They situate the individual/institutional contexts as discrete rather than as extremes of a continuum. It also appears that institutional archives are willing to accommodate a hospitality to 'otherness' 'as long as it does not challenge inherited ways of knowing'.[19] Thus, while there have been many calls for the democratisation of archives,[20] true 'hospitality' may not be achievable. This intractability and inaccessibility of institutional archive services has led some communities to create their own collections, providing services and narratives that complement, if not contrast with, institutional accounts.[21] In this post-custodial world, we appear to be left with a fractured archival landscape. How then can this fragmented landscape be connected into a coherent, navigable and usable whole? How may conflicting narratives be represented and broad community access to archives be provided?

The archival commons and web-based interoperability

These are exactly the sorts of issues that have been addressed by Web 2.0 approaches for interoperability in a wide range of information contexts.[22] For example, the interoperability catalogue 'programmableweb.com' currently has more than 11,000 registered web applications.[23] Indeed, the potential for interconnected archival systems has been recognised since the advent of the Web.[24] More recently, this concept has been refined and articulated as a distributed archival commons, based on 'links between objects using accepted Web standards' that would 'allow users to engage with archival materials as they pursue their own needs regardless of repository or institution'.[25] Such a post-custodial approach to archival systems interoperability not only sidesteps the political issues that would mire a consensual approach to contested management,[26] but would provide a structural basis for integrating access to 'other' voices within the Archival Multiverse.

There are many models of systems interoperability, all of which posit layered functionality that progresses in increasing levels of sophistication from simple technical conventions, through complex semantic mapping, to potentially labyrinthine inter-organisational negotiation of meaning. In order to interoperate successfully, heterogeneous entities require agreement at each model layer. For the purpose of this study, interoperability was defined using the classification of interoperability layers proposed by Ouksel and Sheth[27] and was primarily concerned with the technical layers of interoperability, involving

systemic (being able to comprehend transmissions), *syntactical* (understanding the exchange of information), *structural* (understanding collections of information) and *semantic* (understanding the face meaning of information) aspects of interoperability. The API reference model is intended to enable archival systems to expose an interoperable, web-based interface for discovery and retrieval. The semantic interoperability of such archival control systems is dependent on the mutual understanding of exchanged metadata and means that participating systems must be able to resolve the meaning of metadata structures and elements.

Throughout Europe and North America, archival description is tightly linked to the International Council on Archives (ICA) family of international standards,[28] which, in turn, were informed by conceptualisations contributed by Australian practitioners with experience of the Australian Series System model. These ICA standards address archival description (ISAD(G)), authority record(s) for corporate bodies, persons and families (ISAAR(CPF)), functions (ISDF) and institutions with archival holdings (ISDIAH).[29] Again, the path dependency of international archival practice has seen these standards interpreted in the context of traditional custodial description. Consequently, such use of these standards has been criticised for not directly supporting the documentation of the plurality of contexts under which records may have been created.[30] In contrast, the Australian Series System requires metadata documentation that explicitly declares all contextual entities and their relationships with those representing records.[31] Much work has been performed in the Australian context to develop suitable standards and have them adopted in broader recordkeeping practice. For example, archival practice at national, state and local levels in Australia and New Zealand has led to the development of International Standards Organisation (ISO) standard 23081 and Australia/New Zealand standard (AS/NZS) 5478 for records metadata.[32] These standards have also influenced international archival metadata initiatives such as the InterPARES2 metadata registry project[33] and regional archives in several Canadian jurisdictions.[34]

The relative maturity of Web 2.0 paradigms, together with developments in Australian recordkeeping and archival standards in recent years, have provided ample opportunity to marry a standards-based approach with Web 2.0 interoperability techniques. That such an approach has not been embraced by the Australian archival community (for example, only one Australian institutional archive has a public web API) suggests that there may be organisational, social or informatics problems with such an approach. The next section describes the research approach used to investigate these issues.

Investigating interoperability

The purpose of the research project was to explore the issues (for example, necessary conditions, impediments, success factors and implications) concerning post-custodial interoperability of web-based archival control systems for community discovery and access. The approach was to create a reference model for a web-based API that could be implemented by archival control systems conforming to the Australian Series System model. This reference model would specify concepts, standards and syntax for web-based archival services used to mesh archival metadata from two or more archival control systems and/or combine it with other web data. In addition, this reference model was to be instantiated as proof-of-concept implementations to verify the feasibility and usefulness of the approach. It was hoped that examining these issues in a research context could result in practical implementation guidelines and, perhaps, some technology or approaches reusable by interested organisations.

The objectives of this study were formally articulated as the following research question: *What design, implementation and evaluation issues will have an impact on the development of a standards-based API reference model for web-based archival systems interoperability?*

This primary question was predicated on a number of other, secondary, research questions. For example, despite there being a theoretical basis for such interoperability, it was necessary to investigate whether the reference model was a feasible and useful approach for solving records pluralisation issues in the archival space; and, if so, if further research was needed or warranted. It was also recognised that there may be complex organisational, social or technical impediments to archival systems interoperability and so this investigation, while primarily technical in nature, would have to touch on all of these areas.

There were major questions concerned with the conceptual modelling for interoperability. What metadata standards should form the basis of the model? What was the need for, or scope of, crosswalks between the metadata that underpins inplementations of the Australian Series System? More specifically, what would be appropriate entities, attributes and qualifiers? How should relationships and events be modelled? What schemes should be supported for values? On a technical level, investigation was needed to determine the functional and implementation approach for the API. For example, the access mechanisms that the API needed to support; the functions to be delivered through this interface; and, finally, the encoding and representation of information.

Systems development research methodology

Any attempt to answer these questions in the face of divisive social and organisational politics, divergent archival practices, and a multiplicity of overlapping standards and technological approaches, is to take on a complex problem – even when addressing the issues at the simplest of technological levels.[35] Rittel and Webber found that linear systems approaches were not suitable for tackling such *wicked* problems; problems that are ill- or un- definable have no end to the causal chains that link interacting systems of their components. They determined that an '[iterative] process in the course of which an image of a problem and of [a] solution emerges gradually' was more likely to yield useful results.[36]

Similarly, the rise of agile systems development techniques over the past 15 years has been a response to challenges in software development that exhibit wicked characteristics.[37] From a research perspective, design science methods based on a systems development approach may result in improved understanding of the requirements under consideration (that is, the problem definition).[38]

Design science research is usually considered to be a problem-solving process that is typically concerned with build and evaluation activities that lead to conceptual constructs and models, algorithmic methods and/or instantiations.[39] Originally, design science theorists framed the building of artefacts in terms of demonstrating feasibility, with evaluation activities determining whether progress has been achieved. More recently, there has been a growing recognition that information systems research is 'conducted in a multi-disciplinary and multi-cultural context' and inherently involves socio-technical systems thinking.[40] Similarly, McKemmish and Gilliland note that design science is emerging from its positivist origins and is embracing interpretivist approaches 'when the [research] focus widens to include the sociotechnical and human contexts of systems'.[41] In the case of socio-technical information systems, this approach

focuses on reflection of the design process, as much as any resulting artefact, to produce insights into the possibilities of technology to address these requirements.

Thus, a reflective design science research approach was selected as an appropriate methodology for investigating the complexities of archive interoperability. This approach encompasses three distinct activities: *concept building*; *systems building*; and *systems evaluation*, as shown in Figure 2.[42] While the research iterated between concept-building and implementation activities, there was only one pass through a formal system evaluation phase.

The concept-building stage involved the articulation of meaningful research questions together with the requirements for the artefact design. During this stage, a literature review was conducted to form an initial set of design requirements. Technical standards and extant systems were compared for similarities, differences and relative merit. Additionally, one-on-one interviews were conducted with expert practitioners to confirm and augment design requirements as well as to use the experience of others to avoid unfruitful research approaches.

Throughout the system-building stage, it was important to identify the set of transition paths from the start to the end states, and to follow a systematic process in an attempt to narrow these down to a solution transition path.[43] A *research journal* was maintained for the duration of the study to keep track of research activities, details of relevant resources and people, designs and implementation results. It also provided a record of thoughts, challenges, intermediate research questions, impressions and discrepancies

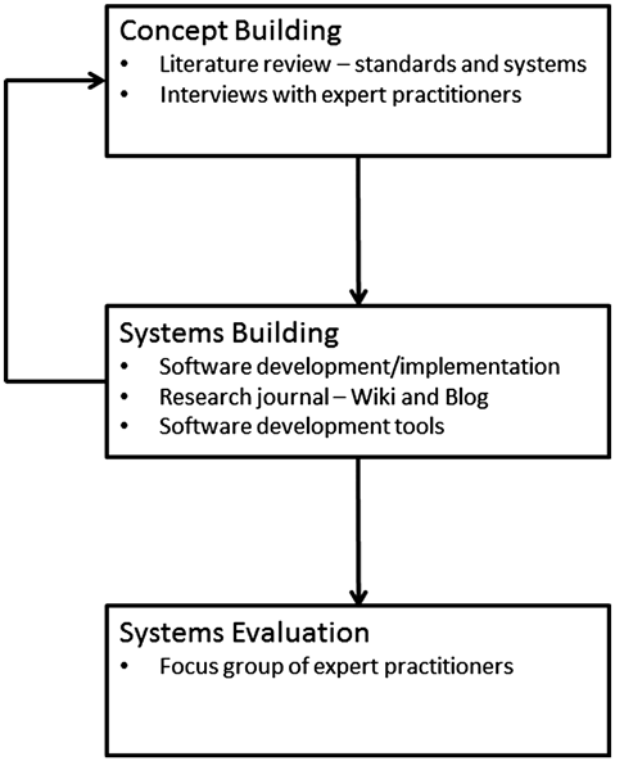

Figure 2. Systems development activities.

as well as problem (re)definitions and interim design solutions. The *development tools* used to build the instantiations also provided insights into the design process. For example, the source-control history of code check-ins afforded a means of charting the problem-solving process in terms of paths, milestones and development/implementation effort.

Finally, the reference model and instantiations were formally e*valuated* by a focus group of expert practitioners convened to review and evaluate the reference model and instantiations.

Results

The results of this study can be considered in terms of the initial interviews, the reference model itself and the proof-of-concept instantiations together with their impact on the research questions concerning design, implementation and evaluation.

Design issues

The purpose of the API reference model was to reconcile disparate approaches to the Australian Series System model and provide the basis for a concrete web-based API for interoperability. This reference model was primarily concerned with standardising recordkeeping and contextual entities together with relationships that inform archival provenance. It also introduced the concept of an *Archive Map* by which an archival control system can declare its entities and their interrelationships. In addition, some encoding standards and element schemes were mandated where they had implications for interoperability.

The set of *Entities* defined in ISO 23081 formed the basis of the reference model entity schema. This schema is shown in Table 1 and comprises four entity classes, with each class comprising a number of entity sub-types. Note that the relationship entity (from ISO 23081) is not explicitly modelled because it does not make sense for an API to return lists or instances of isolated relationships. Rather, relationships are always retrieved in the context of one of the entities that they link.

Table 1. Entity schema.

	Entity classes			
	Record	Agent	Business	Mandate
Entity sub-types	Archives Archive Series File Transaction Sequence Item	Institution Organisation Work Group Family Person Mechanism	Ambient Function Function Activity Transaction	Legislation Regulation Business Rule Stakeholder- Requirement Community-Expectation Standard Instrument Code of Conduct System Specification

There appeared to be greater diversity in *Relationship* modelling within standards and extant systems than for entities. The modelling of relationships in the various jurisdictional recordkeeping standards ranged from generalised statements of principle to highly constrained and detailed specifications. In order to remain semantically consistent with all conformant systems, various ontological approaches to relationship modelling were explored.[44] The status relationships category of the Purao and Storey multi-layered ontology was adopted as the basis of a relationship schema for the reference model, along with its relationship classes and sub-types shown in Table 2. All of the relationships defined in the examined standards and systems, bar the various extensive sets of 'recordkeeping event' relations, can be directly mapped to these relationship sub-types.

The minimum set of *Metadata Elements* specified by the reference model is shown in Table 3. Once identified and selected using these metadata elements, the original digital representation of the full set of metadata elements can be retrieved from the linked collection. Additionally, it was considered important to provide some guidelines for *Encoding* in the interests of promoting interoperability. The reference model therefore mandated that provider systems, at a minimum, encode entity representations as EAD (for Record class entities) or locally typed EAC-CPF (all others).

An ontological mechanism for semantic metadata resolution, termed an Archive Map, was conceived that documented, in a machine-readable way, departures from standard reference model entity, relationship and metadata elements. Specifically, an Archive Map serves a number of functions:

Table 2. Relationship schema.

Relationship class	Sub-type (outbound)	Sub-type (inbound)	Notes (synonyms)
Structural	Contains	Is part of	Totally contained
	Includes	Is member of	Joint membership; (Involves)
Influential	Creates	Created by	(Establishes)
	Destroys	Destroyed by	(Terminates, Abolishes)
	Owns	Owned by	
	Controls	Controlled by	
	Performs	Performed by	
	Concerns	Subject of	
	Regulates	Regulated by	
	(Familiar of)	(Familiar of)	Father, Mother, Daughter, Son, Brother, Sister, Spouse and son
	(Business Rel. of)	(Business Rel. of)	Partner, Advisor, Mentor
	Associated with	Associated with	Other general relationship
Temporal	Precedes	Succeeds	See note[1]
Spatial	Is location of	Is located at	
	Near	Near	
	Before	After	

[1]Temporal relationships can be derived from other (dated) relationships and therefore do not need to be explicitly modelled. Nevertheless, for the purposes of interoperability, the searching of (potentially large numbers of) relationship records necessary to determine these temporal relationships cannot be performed. Therefore this reference model needs to support explicit temporal relationships as 'convenience' relationships to facilitate interoperable processing. This is an example of the situation where data may not be explicitly modelled internally within a system, yet needs to be exposed as part of an interoperability API.

Table 3. Metadata elements.

Element	Interpretation
Entity/Relationship class	The entity or relationship class selected from the schema or, in the case of an entity, defined in an Archive Map.
Entity/Relationship sub-type	The entity or relationship sub-type selected from the schema or defined in an Archive Map.
Title	A short authoritative description of the entity or relationship.
Description	Free-form text of arbitrary length (but preferably limited to one paragraph) that provides a descriptive abstract of the entity or relation.
Existence Dates	The dates between which this entity or relationship existed. Qualified as *startDate* and *endDate*. Note that there are many such sets of candidate dates (for example, existence dates, content dates and so on). Not all archival control systems support multiple dates and so a single set was selected as the minimum requirement. For the purposes of this model, implementers should select the most appropriate date set.
Link	The web URL corresponding to the full item records retrievable from the system.

(1) It exposes access path details such as URIs for discovery and retrieval of different entity types.
(2) It maps constructs such as entities and relationships to those defined in this reference model.
(3) It defines additional entities and relationships in terms of the mapped constructs.
(4) It indicates the standard elements that are not in use.

Implementation issues

The interviews confirmed a lack of organisational capability relating to web-based interoperability of archival control systems. Many interviewees identified the lack of senior management support as being a core factor and indicated institutional dissatisfaction with current core and web systems. While several reported institutional fears of exposing malformed data to the community, others saw this as a net benefit, realising that exposing web-based interfaces actually drives requirements for core systems. The need for the ongoing management and support of participatory archive programs could also explain the lack of an institutional appetite for web-based initiatives.

As implementation of the API within actual archival control systems was impractical, the approach taken was to instantiate proxies for existing systems. Proxies were developed for two archival control systems: Public Record Office Victoria, and State Records NSW (SRNSW). Additionally a browser-based dynamic client was created that could be used to query the proxies and display results. Explanation about the proxies and the technical details of implementation may be found in Appendix 1.

An example of the client output is shown in Figure 3 and shows integration of each site's Archive Map in normalised hierarchical trees of entity sub-types used for searching and filtering. Results from queries of particular entity types were displayed in separate tabs, one for each service queried. Following links to particular entity instances resulted in a new window view of the selected entity, together with its related entities grouped by relationship sub-type.

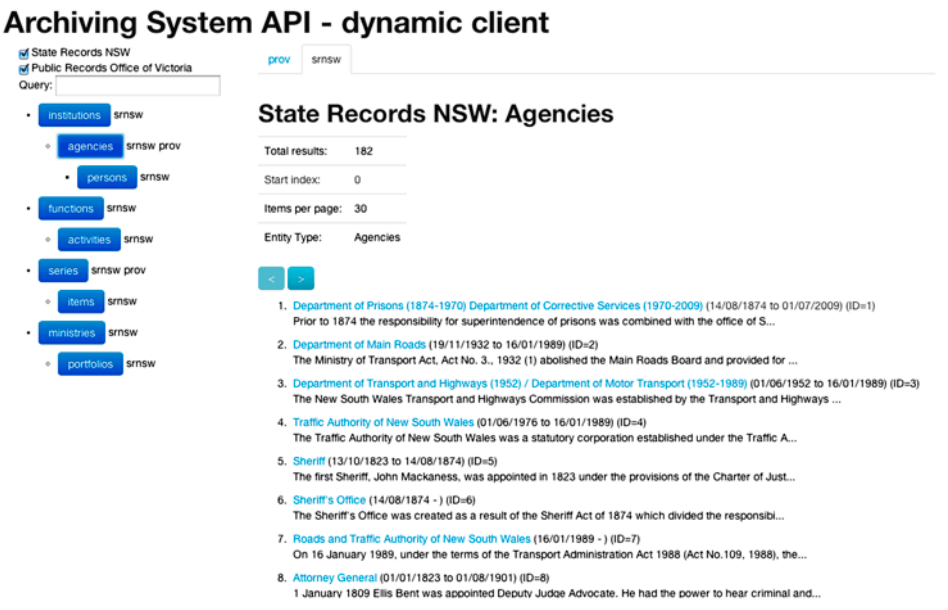

Figure 3. Dynamic client.

Evaluation issues

The focus group recognised the necessity of standardising for interoperability; however, it was suggested that the reference model needed to undergo further empirical testing. One of the focus group participants contrasted this post-custodial design with the HUNI project in Australia, which has taken a high-level conceptual approach in order to integrate disparate institutions, systems and ontological models. The all-embracing, Linked Data architecture of HUNI has resulted in integration problems stemming from the incompatibility of its constituent components, rendering the project goals far more difficult to achieve.

It was suggested that some effort be made to develop a suite of different sample visualisation tools to fully express the capabilities of the API design. Several of the focus group participants were struck by the ability to visualise deficiencies in the supplied metadata and differences in the quantity and quality of metadata provided by the two example institution proxies. Finally, the idea of a hackfest[45] was suggested as a way to verify assumptions of the reference model and develop innovative tools based upon the API implementation.

Discussion

The reported lack of progress in dynamic archival control system interoperability reported in the practitioner interviews validated the rationale for this study. The dearth of web-interoperable archival control systems conformant to the Australian Series System confirmed the justification for taking a systems-development research approach for this study.

Technical issues: the reference model and instantiations

Some aspects of the reference model metadata scheme require further work; in particular, the constructs related to relationship and event modelling. For example, the current modelling of binary, static relationships needs to give way to the nomenclature and representation of n-ary[46] relationships. While the use of relationship metadata for rich discovery is particularly important, the set of relationship classes needed for interoperability is relatively small. It is important that the set of relationships classes be strictly controlled to ensure interoperability is maintained.

Development of the API instantiations highlighted key areas of interoperability brittleness. For example, in the SRNSW data, the term *Persons* was used in some contexts and the term *People* in others. While a human would be able to resolve these discrepancies while querying the API, a machine may not do so as easily. These difficulties underscore the importance of the data model and API design for purposes of interoperability. They provide further evidence that requirements for interoperability must be designed into 'standards, processes, tools, and systems' and cannot easily or reliably be retrofitted.[47]

While the instantiated dynamic client employed a simple interface, there are, obviously, far richer ways of visualising archival metadata as well as providing faceted search interfaces.[48] In order for interface users to form appropriate mental models, interface visualisations need to expose description, relationship, time and provenance clues as well as hints about the existence of digital material. An unexpected insight from the focus group was the identification of interoperability visualisations as a useful tool with which archival professionals diagnose and maintain their metadata.

Discipline issues: standards and interoperability

There is no shortage of recordkeeping and archival standards, even within the context of the Australian Series System. They address identical recordkeeping endeavours (albeit within different jurisdictional contexts) but are inconsistent with each other, and often partially supported by purportedly compliant systems.[49] The proliferation of standards in the Australian context raises questions about their purpose(s). Are they backward-looking – designed for compliance by systems that have already been implemented? Are they able to support current successful contemporary practice subsequent to their publication? Or are they aspirational, guiding the next generation of development? It appears that the standards-setting culture within the Australian archival context is of this first empirical type, whereby standards follow systems. Such standards take the form of conceptual 'soft' standards (that is, containing broad guidelines) rather than 'hard' standards (that is, that comprise specific, prescriptive rules). This emphasis on compliance to a collective empirical norm is made at the expense of requiring and enforcing interoperability between conformant systems.

One characteristic of the empirical approach to standards-setting is that minor differences in nomenclature and schemes lead to the proliferation of jurisdictional standards. While local variation in systems requirements is understandable, should such variation and increased specificity be accommodated at the expense of interoperability? Certainly the experience from the Web is that emergent standards and interoperability can be a powerful driver of innovation.[50] Recordkeeping standards-setting needs to find a way to maintain systems interoperability while continuously incorporating variation that arises from jurisdictional pressures. In addition, there is still the question of when in the

innovation cycle standards should be set. Too early, and an immature standard can constrain innovation that would otherwise take place later in the cycle. An example of this was the early adoption of the library-oriented MARC format in archive automation systems.[51] Too late, and the result may be the creation of widely divergent implementations that cannot retroactively be made to converge.[52]

The Archive Map approach taken by this study attempts to address this dilemma and provide a mechanism for rules and exceptions. With the Archive Map, both conformance and interoperability are achieved through mapping to, and controlled extension of, a minimalist metadata subset. Thus the reference model takes the second approach to standardisation described above: that of reflecting successful contemporary practice. However, in doing so, it normalises current practice at the expense of guiding better recordkeeping modelling and stronger standards for interoperability. This is an example of the trade-offs that must take place in archival informatics. Flexible, interoperability standards such as this reference model have the potential to provide a circuit-breaker for the empirically based standards-setting that currently dominates recordkeeping and archives in the Australian context. Such a shift could see standards-setting moving to the third approach described above: that of providing guidance for future systems development. Ultimately, the standards should become blueprints for interoperability in their own right.

Interoperability also becomes an imperative for data free from inconsistencies, gaps and anomalies. For example, interoperability or visualisation mechanisms that depend on valid values, such as for relationships or date ranges, can be rendered inaccurate or inoperable by missing or invalid data. Rather than simply comply with a value in every slot, designers and users of archival systems need to understand that metadata is actually used purposefully for interoperability. Systems need to be designed to ensure that metadata is appropriately entered and applied at its point of capture. Where possible, data entry fields should be constrained to capture only meaningful values so that interoperational integrity can be maintained. In addition, data entry needs to be contextualised; feedback should be given regarding the consequences of missing or inappropriate information, perhaps through visualisation of the newly entered data in relation to existing networks of information. An additional approach may be to use community-based remediation and feedback programs to improve data quality.

Global issues: organisations and the archival commons

There appear to be structural organisational issues that impact capabilities for interoperability.

The study identified a catch-22 problem whereby appreciation of the benefits and implications of interoperability may be gained by implementing a public web API, but this awareness appears to be a prerequisite for such interoperability initiatives. Furthermore, archival institutions are often not in direct control of their operational environment, systems developers or website – particularly if integrated into a whole-of-government web presence. It is challenging for archival institutions to articulate requirements for better systems and operational environments, as such requirements are best derived from experience gained through the iterative deployment of interoperable systems in the first place.

Organisations may need to see concrete examples using their own data before they can appreciate the benefits of interoperability. Perhaps institutions could provide data for proxy instantiations that could be explored using tools similar to the dynamic client

instantiation – both in isolation as well as meshed with data from other collections. This could raise the institutional profile of interoperability and bootstrap the articulation of requirements, without requiring any infrastructural change or organisational commitment to systems development. Another approach could be for archival institutions to pool resources, sharing requirements and aligning archival systems at least at an operational level. A further suggestion by the focus group was that the reference model could be the catalyst for the formation of an archival systems development community that could inform requirements for development, staffing and operations. Furthermore, such a community could drive the development of open-source tools and systems that may better satisfy requirements and budgetary constraints.

Archival practice appears to be still largely centred on the *description* of records for use by archivists or institutional researchers. This means that much of metadata standards serves the recordkeeping business rather than the *documentation* of material for interoperability and discovery. As a result, archival metadata is oriented towards capturing the detail of records rather than the contexts of their origin. Members of the community at large seeking archival records 'should be able to enter the system with a knowledge of the world being documented, without knowing about the world of documentation'.[53] Archival practice needs to embrace this concept of documentation for interoperability. Only then can standards and systems change in order to ensure appropriate documentation of records that facilitates interoperable discovery and access.

Finally, consideration needs to be given to the issues that arise from meshing of archival metadata from two or more sources. Heterogeneous systems may make different, and possibly contradictory, assertions about descriptive elements, relationships and events. Sophisticated meshing may not consider all assertions with equal weight but, at its most simple, should document their origins. Similarly, archival discovery needs to be deterministic. As archival holdings and documentation change over time, an archival commons needs to ensure the consistency and evidentiality of discovery results through time and space.

Conclusion and further research

Despite comparatively recent moves towards the delivery of web-based archival services, there has been little progress in the web-based interoperability of archival systems in the Australian context. The primary aim of this study was to explore the issues concerning the technical interoperability of web-based archival control systems to achieve an archival commons. The core of the study was the development of a standards-based web API reference model that would be suitable for implementation with archival control systems conformant with the Australian Series System. A design science research approach was selected for this study that employed iterative and reflective software development methods to build and evaluate un-situated artefacts. The research encompassed interviews with archival control system practitioners, design of the reference model, creation of proof-of-concept implementations and evaluation.

The study identified a number of significant interoperability challenges for archival systems and provided further evidence that requirements for interoperability must be 'designed in' and cannot be retrofitted with reliability or ease. On a technical level, the research successfully proposed a web-based API for archival system interoperability in the context of the Australian Series System. It comprised a standard metadata schema of entities, relationships and elements, together with an ontological resolution mechanism termed an Archive Map. This research also identified areas in which conceptual

and representational recordkeeping and archival standards could be improved. Current recordkeeping and archival standards appear to be insufficiently prescriptive to ensure interoperability, and do not model all of the required elements to facilitate discovery and access by the members of the wider community. From an organisational perspective, the study found structural barriers to progressing interoperability initiatives for community access.

Possible avenues of further research could include investigation from a number of standpoints. From a conceptual perspective, the modelling of relationships and events; multi-source traceability and reproducibility; provenance of assertions; and authentication and authorisation all need further work in designing standards for next-generation, distributed, interoperable, archival systems. From a technical perspective, the reference model needs further verification against a broader set of real-world systems. Similarly, requirements need to be determined for interoperability, beyond that of a read-only API interface, in order to support truly participative archive activities. Investigation into user interfaces that facilitate appropriate mental models for community discovery is also warranted. And, finally, from an organisational perspective, what path(s) should an institution follow to participate fully in federated archives in terms of achieving technical, organisational and social levels of interoperability?

While the outcomes of this study have provided some approaches that may be used to unlock the records held in our 'fortress archives', interoperability remains a fertile area for research.

Acknowledgements

The author would like to acknowledge the guidance and support of Dr Joanne Evans, Professor Sue McKemmish and Associate Professor Gavan McCarthy in the research project that resulted in this article.

Endnotes

1. Terry Cook, 'Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms', *Archival Science*, vol. 13, nos 2–3, 2013, pp. 95–120.
2. Research Institute (AERI), Pluralizing the Archival Curriculum Group (PACG), 'Educating for the Archival Multiverse', *American Archivist*, vol. 74, no. 1, 2011, pp. 69–101.
3. Parliament of Australia, Senate, Community Affairs References Committee and Rosemary Crowley, 'Lost Innocents Righting the Record: Report on Child Migration', 2001, available at <*http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/Completed_inquiries/1999-02/child_migrat/report/index*>, accessed 16 August 2014. Parliament of Australia, Senate Community Affairs References Committee, (Stephen Patrick Hutchins and Jan McLucas), 'Forgotten Australians: A Report on Australians Who Experienced Institutional or Out-of-Home Care as Children', 2004, available at <*http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/Completed_inquiries/2004-07/inst_care/report/index*>, accessed 16 August 2014. Parliament of Auatralia. Senate. Community Affairs References Committee and Rachel Siewert, 'Lost Innocents and Forgotten Australians Revisited: Report on the Progress with the Implementation of the Recommendations of the Lost Innocents and Forgotten Australians Reports', 2009, available at <*http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/Completed_inquiries/2008-10/recs_lost_innocents_forgotten_aust_rpts/report/index*>, accessed 16 August 2014.
4. Barbara Reed, 'Beyond Perceived Boundaries: Imagining the Potential of Pluralised Recordkeeping', *Archives and Manuscripts*, vol. 33, no. 1, 2005, p. 176; Scott R Anderson and Robert B Allen, 'Envisioning the Archival Commons', *American Archivist*, vol. 72, no. 2, 2009, pp. 383–400.

5. Sue McKemmish, 'Are Records Ever Actual', in Sue McKemmish and Michael Piggott (eds), *The Records Continuum: Ian Maclean and Australian Archives: First Fifty Years*, Ancora Press, Clayton, 1994, pp. 187–203.
6. Hilary Jenkinson, *A Manual of Archive Administration Including the Problems of War Archives and Archive Making*, Clarendon Press, Oxford, 1922, p. 44.
7. Chris Hurley, 'Documenting Archives and Other Records', 2008, available at <*http://www.infotech.monash.edu.au/research/groups/rcrg/publications/ch-documenting-archives.pdf*>, accessed 16 August 2014.
8. Andrew Flinn, 'An Attack on Professionalism and Scholarship? Democratising Archives and the Production of Knowledge', *Ariadne: Web Magazine for Information Professionals*, no. 62, 2010, available at <*http://www.ariadne.ac.uk/issue62/flinn*>, accessed 16 August 2014.
9. Eric Ketelaar, 'Archival Temples, Archival Prisons: Modes of Power and Protection', *Archival Science*, vol. 2, nos 3–4, 2002; pp. 221–38; Sue McKemmish et al., 'Resetting Relationships: Archives and Indigenous Human Rights in Australia', *Archives and Manuscripts*, vol. 39, no. 1, 2011, p. 107.
10. Parliament of Australia, Senate, 'Lost Innocents Righting the Record'; Parliament of Australia, Senate, 'Forgotten Australians'; Parliament of Australia, 'Lost Innocents and Forgotten Australians'.
11. Cate O'Neill, Vlad Selakovic and Rachel Tropea, 'Access to Records for People Who Were in Out-of-Home Care: Moving Beyond "Third Dimension" Archival Practice', *Archives and Manuscripts*, vol. 40, no. 1, 2012, pp. 29–41.
12. Cate O'Neill, 'Accessing the Records of the Forgotten Australians: Learning from the Human Rights Context to Improve Archival Practices and Restorative Justice', presented at the International Council on Archives Congress, 'A Climate of Change', Brisbane, 2012, available at <*http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00354.pdf*>, accessed 16 August 2014.
13. Gavan J McCarthy and Joanne Evans, 'Principles for Archival Information Services in the Public Domain', *Archives and Manuscripts*, vol. 40, no. 1, 2012, pp. 54–67.
14. Daniel V Pitti, 'Encoded Archival Description: An Introduction and Overview', *New Review of Information Networking*, vol. 5, no. 1, 1999, pp. 61–9.
15. Isto Huvila, 'Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management', *Archival Science*, vol. 8, no. 1, 2008, pp. 15–36.
16. Flinn.
17. Find & Connect Web Resource Project, 'Home – Find & Connect Australia', 2012, available at <*http://www.findandconnect.gov.au/*>, accessed 16 August 2014.
18. Alexandra Eveleigh, 'Welcoming the World: An Exploration of Participatory Archives', presented at the International Council on Archives Congress, 'A Climate of Change', Brisbane, 2012, available at <*http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00128.pdf*>, accessed 16 August 2014.
19. Verne Harris, 'Seeing (in) Blindness: South Africa, Archives and Passion for Justice', *Archifacts*, October 2001, pp. 1–13.
20. Flinn; Harris; McKemmish et al.
21. Flinn; Sue McKemmish, Anne Gilliland-Swetland and Eric Ketelaar, '"Communities of Memory": Pluralising Archival Research and Education Agendas', *Archives and Manuscripts*, vol. 33, no. 1, 2005, pp. 146–174.
22. T O'Reilly, 'What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software', 2005, available at <*http://oreilly.com/web2/archive/what-is-web-20.html*>, accessed 16 August 2014.
23. Programmable Web, 'Mashups, APIs, and the Web as Platform', 2013, available at <*http://www.programmableweb.com/*>, accessed 16 August 2014.
24. Tim Sherratt, 'Pathways to Memory: Accessing Archives on the WWW', 1996, available at <*http://ausweb.scu.edu.au/aw96/cultural/sherratt/paper.htm*>, accessed 16 August 2014.
25. Anderson and Allen.
26. Chris Hurley, 'Parallel Provenance (If These Are Your Records, Where Are Your Stories?)', 2005, available at <*http://www.infotech.monash.edu.au/research/groups/rcrg/publications/parallel-provenance-combined.pdf*>, accessed 16 August 2014.

27. Aris M Ouksel and Amit Sheth, 'Semantic Interoperability in Global Information Systems', *ACM Sigmod Record*, vol. 28, no. 1, 1999, pp. 5–12.
28. Though it should be recognised that much description follows media-specific description standards in lieu of archival standards; see Marcel Ruhl, 'Do We Need Metadata? An On-Line Survey in German Archives', in *Proceedings of the 2nd International Workshop on Semantic Digital Archives (SDA 2012)*, available at <*http://ceur-ws.org/Vol-912/proceedings.pdf#page=30*>, accessed 16 August 2014.
29. International Council on Archives, 'ICArchives: Standards: Standards List', 2012, available at <*http://www.ica.org/10206/standards/standards-list.html*>, accessed 16 August 2014.
30. McKemmish; Hurley, 'Parallel Provenance'.
31. Australian Society of Archivists Committee on Descriptive Standards, *Describing Archives in Context: A Guide to Australasian Practice*, Australian Society of Archivists, Dickson, ACT, 2007.
32. Standards Association of Australia, *Australian Standard^{TM} Information and Documentation – Records Management Processes – Metadata for Records Part 1: Principles*, vol. 1, Standards Australia, Sydney, 2006.
33. Anne Gilliland et al., 'Towards a 21st Century Metadata Infrastructure Supporting the Creation, Preservation and Use of Trustworthy Records: Developing the InterPARES 2 Metadata Schema Registry', *Archival Science,* vol. 5, no. 1, 2006, pp. 43–78.
34. Steve Billinton, 'The Role of Archival Authority Records in the Finding Aid System of the Archives of Ontario', *Journal of Archival Organization*, vol. 5, nos 1–2, 2008, pp. 75–93.
35. Joanne Evans, 'Capacities and Complexities: A Reflection on Design Methodologies for Archival and Recordkeeping Research', personal communication, 2013.
36. Horst WJ Rittel and Melvin M Webber, 'Dilemmas in a General Theory of Planning', *Policy Sciences*, vol. 4, no. 2, 1973, pp. 155–69.
37. Kent Beck et al., 'Manifesto for Agile Software Development', 2001, available at <*http://academic.brooklyn.cuny.edu/cis/sfleisher/Chapter_03_sim.pdf*>, accessed 16 August 2014.
38. Joanne Evans and Nadav Rouche, 'Utilizing Systems Development Methods in Archival Systems Research: Building a Metadata Schema Registry', *Archival Science*, vol. 4, nos 3–4, 2006, pp. 315–34.
39. Salvatore T March and Gerald F Smith, 'Design and Natural Science Research on Information Technology', *Decision Support Systems*, vol. 15, no. 4, 1995, pp. 251–66.
40. Bjoern Niehaves, 'On Episemological Diversity in Design Science: New Vistas for a Design-Oriented IS Research?', presented at the Twenty Eighth International Conference on Information Systems, Montreal, 2007, p. 133, available at <*http://udoo.uni-muenster.de/downloads/publications/1992.pdf*>, accessed 16 August 2014.
41. Sue McKemmish and Anne Gilliland, 'Archival and Recordkeeping Research. Past Present and Future', in Kirsty Williamson and Graeme Johanson (eds), *Research Methods: Information, Systems and Contexts*, Tilde University Press, Prahran, 2013, p. 93.
42. Frada Burstein, 'System Development in Information Systems Research', in Kirsty Williamson (ed.), *Research Methods for Students, Academics and Professionals: Information Management and Systems*, Centre for Information Studies, Charles Sturt University, Wagga Wagga, NSW, 2002, pp. 147–58.
43. Ron Weber, 'Design-Science Research', in Williamson and Johanson (eds), *Research Methods*, pp. 245–66.
44. Yair Wand, Veda C Storey and Ron Weber, 'An Ontological Analysis of the Relationship Construct in Conceptual Modeling', *ACM Transactions on Database Systems (TODS)*, vol. 24, no. 4, 1999, pp. 494–528; Sandeep Purao and Veda C Storey, 'A Multi-Layered Ontology for Comparing Relationship Semantics in Conceptual Models of Databases', *Applied Ontology*, vol. 1, no. 1, 2005, pp. 117–39; Giancarlo Guizzardi and Gerd Wagner, 'What's in a Relationship: An Ontological Analysis', in *Proceedings of 27th International Conference on Conceptual Modeling (ER 2008),* Barcelona, 20–24 October 2008, pp. 83–97.
45. Steven Leckart, 'The Hackathon is on: Pitching and Programming the Next Killer App', *Wired Magazine*, 17 February 2012, available at <*http://www.wired.com/magazine/2012/02/ff_hackathons/all/1*>, accessed 16 August 2014.
46. The term n-ary refers to relationships that relate more than two elements. This is in contrast to most jurisdictional standards and systems that conceptualise only binary relationships - that relate two (and only two) elements. For example "Agency A controls series S" is a

binary relationship linking A and S. "Legislation L mandates that agency A controls series S" is a more complex, n-ary relationship, linking L, A and S.

47. Joanne Evans, Sue McKemmish and Barbara Reed, 'Making Metadata Matter: Outcomes from the Clever Recordkeeping Metadata Project', *Archives & Manuscripts*, vol. 37, no. 1, 2009, pp. 28–56.
48. Mitchell Whitelaw, 'Visualising Cultural Collections at TedXCanberra', 2010, available at <*http://visiblearchive.blogspot.com.au/2010/11/visualising-cultural-collections-at.html*>, accessed 16 August 2014; DigitalNZ, 'DigitalNZ', available at <*http://www.digitalnz.org*>, accessed 19 October 2013.
49. Evans, McKemmish and Reed.
50. John Palfrey and Urs Gasser, 'Mashups Interoperability and eInnovation', *Berkman Publication Series*, vol. 11, no. 7, 2007.
51. Daniel V Pitti, 'Encoded Archival Description: The Development of an Encoding Standard for Archival Finding Aids', *American Archivist*, vol. 60, no. 3, 1997, pp. 268–83.
52. Joanne Evans, 'Building Capacities for Sustainable Recordkeeping Metadata Interoperability', Monash University, 2007.
53. David A Bearman, 'Documenting Documentation', *Archivaria*, vol, 1, no. 34, Summer 1992, pp. 33–49, available at <*http://journals.sfu.ca/archivar/index.php/archivaria/article/viewArticle/11839*>, accessed 19 October 2013.
54. Roy Thomas Fielding, 'Architectural Styles and the Design of Network-Based Software Architectures', PhD thesis, University of California, 2000, available at <*http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm*>, accessed 16 August 2014.

Appendix 1. Instantiations

The API was implemented using a ReSTful approach, whereby paths to related elements and actions were represented by URIs contained in response documents.[54] These URIs, together with the semantic mapping provided by the Archive Maps, enabled loosely coupled, yet intelligent, traversal of the providers' data space by API consumers. The API supported presentation in XML, JSON and HTML formats.

The Archive Maps were expressed as Linked Data documents, needing only to contain those entries necessary to define deviations from the reference model. For example, a fully compliant site needed only to supply the Archive Map header information. Alternatively, an Archive Map could also define changes to standard entity sub-types, additional entity classes and entity sub-types, as well as changes to standard relationship sub-types or additional relationship sub-types.

As implementation of the API within actual archival control systems was impractical, the approach taken was to instantiate proxies for existing systems. Proxies were developed for two archival control systems: Public Record Office Victoria (PROV) and State Records NSW (SRNSW), as shown in Figure 4. The SRNSW proxy communicated directly with the SRNSW archival control system to retrieve various entity sub-types. The PROV proxy was based on an extract of metadata obtained from PROV that comprised agency and series entities and their relations.

The API client was instantiated as a browser-side component to ensure clean separation from the server-side code as well as to demonstrate that it could be lightweight and easily implemented. This client dynamically queried the Archive Maps and API results from conformant archival control systems.

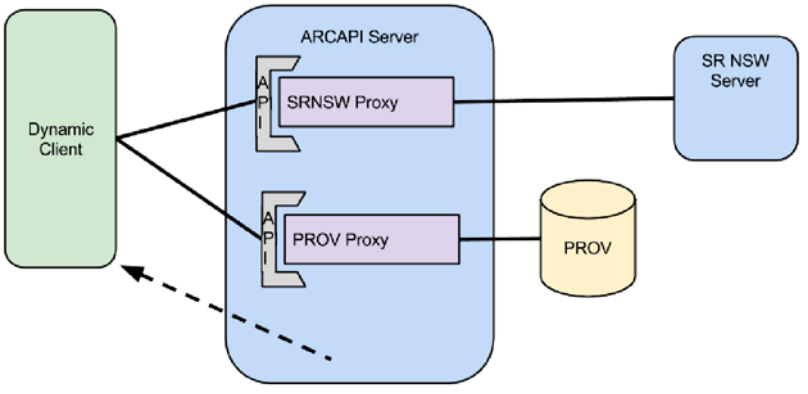

Figure 4. Instantiation architecture.