

Are the GLAMs going to bring a steampunk/neo-Victorian sensibility and aesthetic to Linked Open Data?

Ingrid Mason
eResearch Analyst
Intersect Australia
ingrid.mason@intersect.org.au

Abstract

The premise underlying this paper is that the cultural collecting sector, that is, galleries, libraries, archives and museums (GLAMs,) has established description practices, communities of practice, and shared interest in collaboration and exploiting LOD (Linked Open Data) technologies to improve resource discovery of cultural collections online. The question posed is a means of unpacking: what LOD means in technical terms; what other technologies support resource discovery; what impact using LOD might have on collecting sector description practices and more widely in other sectors with increasing levels of openness in the provision of data and metadata, for example eResearch and government sectors.

The Digital Revolution

The Digital Revolution¹ is the phrase coined to encapsulate a phenomenon affecting all areas of work and life in the developed and developing worlds. This revolution is an accelerated, and at times, overwhelming social change driven through and by computing technology and programming. Computing technologies are embedded in the environment, daily habits and workplaces. Social mores are re-examined in the light of increasing use of mobile technologies and social media. Individual and community expectations, that is, political, social, economic and cultural expectations, are changing as a result of the introduction of new technologies. Societal structures and processes, and therefore social institutions, and their collection description practices, are necessarily therefore also undergoing change.

The cultural, research and government sectors² are all affected by, and affecting, this phenomenon and the rate of social change. This paper focuses on the change to be wrought in cultural institutions, with the production and reuse of linked open data. Many of the issues and challenges that arise in the cultural sector also arise in the research and government sectors. New computing technology and programming practices can socially impact in common ways, that is: changes in traditions of practice; take up of and collaboration around the use new technologies; and community awareness raising and capability building to enable use. Social institutions play a key role in this revolution and each type of social institution may provide a particular characteristic in this process of practice change. In the case of cultural institutions the question in the paper's title asserts the possibility that GLAMs (galleries, libraries, archives or museums) may bring a steampunk/neo-Victorian sensibility and aesthetic to that socio-technological change.

The metaphor of steampunk/neo-Victorian³⁴ sensibilities and aesthetic is a reference to the Enlightenment thinking and what seems to be an interesting parallel of the Digital Revolution with the Industrial Revolution. Steampunk and neo-Victorian sensibilities and aesthetics have been variously interpreted as new art forms and genres, with historical connotations from Edwardian and Victorian times overlaid and embedded in cultural expression and technology. Significant technological innovation, new industries and public infrastructure, including transport and communication, were established in the Edwardian and Victorian periods, through the Industrial Revolution. Social mores and status were put to the test by liberal thinking, 'new ways', the emerging nouveau riche, and ideas of nationalism. Steampunk and neo-Victorian sensibilities and aesthetics as a genre or cultural expression are used as a metaphor to explore whether there is a way forward for GLAMs to tap into the rich, historical and viscous nature of their practice. GLAMs may build these sensibilities and aesthetics into the practice of making collection descriptions available online, while employing linked open data in that approach.

Broadly speaking, the steampunk/neo-Victorian cultural movement encourages a reconnection with the ethos and values that emerged in Enlightenment thinking to examine, inform and shape intellectual engagement and cultural production in the present. This sensibility and aesthetic can be seen in: jewellery or designer goods made with mechanical parts, reflecting hand made crafts and mechanical parts; the production of eBooks, reflecting the concept of a book but the artefactual result is multimedia; and textual encoding of books, reproducing and reflecting the print production features, names and reading conventions of the codex in the eXtensible Markup Language (XML) coding; popular films like *Harry Potter* series, with a

narrative exploring social norms and individuality and featuring steam engines; or *Howl's Moving Castle*, with a narrative exploring feminism and free will and a steam-powered flying machine in the form of a castle. These cultural artefacts are all derived from, or set, in the past, and they exhibit the aesthetic of fashion, history, steam power, mechanisation, and explore power structures, social boundaries and mores. There are parallels with the Classical values that informed political, social, economic and cultural practices in the Renaissance. The steampunk/neo-Victoria cultural movement operates in the Digital Revolution as a renaissance of Enlightenment thinking that drove social change and the Industrial Revolution.

Characteristics of social change impacted by new technology are mass production, innovation, openness, and new ways to access resources. A steampunk aesthetic or sensibility seems to reflect ideas of throwing off the social strictures of the past, the punk, at the same time as embracing the technological past, the steam, in a manner of being unconventional and yet mannered in the embrace of computing technology. In counterbalance to these characteristics: individuation, special production, tradition and heritage, competition and privilege are revisited and at times re-embraced. A neo-Victorian aesthetic or sensibility seems to reflect a renewed enthusiasm for and clarity in purpose of, the neo. The stolid emancipatory embrace of the ideas of egalitarianism and social equity embedded in the past reflect the Victorian. The kind of socio-technological change occurring at this time has strong resonance with the Industrial Revolution. In Victorian times, the anticipation was that new technology would afford new opportunities to improve life and broaden the opportunity for cultural exchange, exploration and engagement, rather than continue to privilege it. The same anticipation underlies the kind of cultural shift and democratisation underway and is driving the changes in description practices in the cultural sector.

Cultural Institutions

Citizens and cultural professionals are implicated in this socio-technological change and the changes in cultural sector description practices. The focus of access to cultural collections in the Victoria era was the physical access to the monumental buildings and collections formed around scientific discovery or international exhibitions. The focus of access to cultural collections in the Digital era will be the online access to the digital representations of those collections and how that online interaction is linked, or not, to the onsite experience. This shift in focus of activity to resource discovery and online access is a significant challenge for cultural practitioners and institutions to grapple with. Some of those challenges are practical, such as what to stop doing or do more efficiently, or how to start doing something else. Other challenges are which curatorial decisions and methods for the integration of external linked open data sources with cultural collection data should be part of a revamped online service. Many cultural practitioners and citizens alike need to understand the whys and what-ifs before they can begin the process of practice change. Cultural institutions are heavily dominated in principle by public benefit and in practice by ideas of control, actual and intellectual, of cultural artefacts and information. New technologies and approaches that enable unconstrained access and use of digital cultural heritage, and multiple perspectives, are not readily embedded in professional practices and services as they currently operate in the cultural sector because of this tradition. Cultural institutions and professionals need to be active in this time of social-technological change so that as their practices shift, their social relevance is retained. The upshot of cultural institutions not making this

shift in collection description practice is to test that traditional approaches are viable and that their social relevance is maintained.

Cultural professionals have been accustomed to being the authorities of cultural information and social facilitators in the cultural sector. The role of cultural institutions will remain, but how that authority and social function is translated into new practice is a critical part of sustaining the value and relevance of cultural institutions in society. Cultural collecting institutions are repositories of cultural heritage and generators of cultural collection data. GLAMs have a history of diverse means of providing access to collections and diverse collection description practices. Collection description practices are being tested in the efforts to make cultural collection data available online. Where description traditions and technical capability have served the needs of the institutions in the push to move resource discovery services online, the broader goal to support search and reuse across diverse cultural collection data is yet to be fully realised. There is an inherent tension and conflict between the collection description practices in GLAMs and divergent approaches to providing collection data and the liberal ethos underlying online development and more specifically linked open data.

There is, however, hope that the situation will improve and cultural collection descriptions practices will enable more extensive cross-collection resource discovery and therefore access. There are signs that the divergence in approaches to collection description practices and the resulting disconnect in resource discovery is beginning to be resolved. This hope comes in the form of:

- A ten-year dialogue around sharing and mapping to common schemas such as Dublin Core.
- More recent interest in higher-level description, for instance RIF-CS (Registry Interchange Format – Collections and Services) and DC-CAP (Dublin Core – Collection Application Profile).
- Heightened interest in using common codification of people and organisations for example VIAF (Virtual International Authority File), FOAF (Friend of a Friend), and EAC-CPF (Encoded Archival Context – Corporate Bodies, Persons, and Families).
- Wide interest in the common codification of places and place names including GeoNames, ANZLIC-MAP (Australian and New Zealand Land Information Council – Metadata Application Profile) and Gazetteer of Australia).
- Active research and testing of the use of linked open data to build new datasets from diverse datasets or to align datasets in related domains.
- Very recent attempts to talk through issues and address challenges associated with realising a semantic web such as LOD-LAM (Linked Open Data – Libraries, Archives, Museums).

The combination of sharing common schemas and vocabularies and providing and using linked open data services will radically change the way resource discovery and access is enabled. The cultural, research and government sectors are all partisans in this change in practice and process of enabling change. This partisanship around change in practice in the cultural sector is multi-dimensional. Key cultural institutions in these sectors need to: show leadership in practice change; provide infrastructure in the form of guidance and data services; and educate developers and searchers

alike to exploit the new resource discovery services provided and potentially explore and reuse the data in new ways. This practice change references core public access principles and reinvents process and thereby enables wider practice change by other data providers in diverse sectors. Opportunities to explore and recheck conventions should be tested with questions around privacy and ethics. This is to ascertain if the borders separating content can be open or if they need to remain controlled and therefore privileged.

Linked Open Data

The kind of resource discovery service that linked open data offers, in the context of GLAMs, is the equivalent to that which can currently only be provided by a human being, often a registrar, curator, librarian, or archivist with either knowledge of the area, or information literacy skills. Linked open data offers the ability to translate quickly for the new searcher multiple ways to define an entity.

“There are several ways to make sure data is open and linked: the data is freely accessible and verifiable; it can be accessed by other people and turned into other works; the data is connected to other databases. Knowledge is built into the database so more people can work with it.” (Williams, 2011)

The searcher can discover new information about the entity without restarting the search or seeking specialised advice. Space and place are good examples of the benefits of linked open data. If a point or area is defined geo-spatially, how many ways can that spot or area be referenced and by whom? The answer is that many parties provide multiple perspectives, in diverse languages and conventions. ‘Annandale’, a suburb in Sydney, was once known, in part, as ‘Johnston’s Bush’. Annandale is also known once to be the ‘land of the Eora people’, and as ‘2038’ through its postal code. This multiplicity of data types, information perspective and authority is a conundrum and is a challenge that linked open data can help to address. If a searcher is able to put the word ‘annandale’ into a search box and retrieve results that relate to all of these perspectives on that place and space, that is a significant feat of semantic connectivity across diverse datasets. This is a major goal for GLAMs to aspire to in this process of reinvention, to reduce the friction, in terms of time, and the gap between datasets, currently, to enable smarter and richer resource discovery.

The resource discovery service that linked open data enables, in the context of cultural institutions with collection data, requires a change in practice in the structuring and delivery of that collection data. Currently, the standard approach to providing information about cultural collections is to provide an online catalogue for searchers to seek cultural information from, and where possible through an aggregated metadata service. This constrains search in general to one source of information related to one cultural collection. Advances on this have been to aggregate cultural collection data to enable several pools of data to be interrogated at the same time. Unionisation as an approach to enabling resource discovery of cultural collection data requires considerable metadata standardisation and data alignment. Federated search is another means to enable cross-collection searching. In some instances collection data has been loosely coupled where it has not been feasible to undertake data alignment, or unionising, and share the same description standard. The overall benefit of the metadata aggregation across diverse datasets is a gain. However, loosely coupling cultural collection datasets across the GLAMs

without unionisation has had correspondingly variable results. Using linked open data to connect similar, but not unionised, cultural collection data will improve the information retrieval quality and improve resource discovery.

Linked open data, in the cultural context, does not mean discarding resource discovery and access services currently already provided; it is a means of augmenting them. Providing cultural collection data as linked open data is the repackaging of the data for search services that then enables it to be threaded into cross-collection or multi-resource searches. A linked open data approach enables the searcher to discover relevant contingent or contextual information to the cultural collection data being examined. The trick with linked open data is around the need to share description standards, vocabularies or schemas. Where unionisation of collection data relies heavily on all three methods of information standardisation to formalise collection data to be in place, linked open data can enable the description standard and schema to be in common or diverse. The primary value of linked open data is that it will enable linking through a common reference point to a shared vocabulary, whether controlled or not, and whether the same collection description schema is being used and in the same way.

Unpacking "linked open data" as a method of connecting datasets programmatically will inform how multi-resource and collection connectivity can be constructed, which will then enable broader and richer resource discovery. Open data means providing the collection data for exploitation by others on the web, sometimes constrained with specifically licensed terms of use. Linked data⁵ (LOD) means references to and from data in data services, by matching terms and using technical protocols. It is possible to enable a searcher to discover information from diverse collections or resources using a linked open data term common to them all. A good demonstration of LOD is the capacity to use a controlled vocabulary, such as a personal or organisational name authority, to retrieve information from multiple resources that have embedded that name into their metadata and provided it as linked open collection data. Having the name from the controlled vocabulary in the collection metadata in LOD format therefore enables the link to that name authority service. The beauty of this shared use of, and referencing to, a common vocabulary is that multiple perspectives and languages can be retrieved. Cultural authority can be asserted by cultural institutions, and that authority more firmly established and endorsed through consensus by the referencing or uptake in the wider cultural sector or in other sectors.

A term in a name authority such as 'Robert B Bruce 1947-', in the context of a library name authority file, is the formal identification of an individual that could be linked to from a data service and presented beside a term that refers to 'Dr Bob Bruce', in a scholarly context, who is in fact the same person. The link though needs to be validated as exact or expressed as a possible match by those that provide the data. In intellectual access terms, this is akin to presenting the linked open collection or vocabulary data in context and out of context. The concept of trust needs to stay in balance with open access to ensure that the value of cultural authority is not undermined through spurious term matching and therefore poor semantic linkage in resource discovery.

"The National Library has defined resource discovery as 'bibliographic and metadata' services, but these are to be provided to the Australian library system, not to all Australians using the Web to access resources. The term

'resource discovery' has thrived as part of the Web generation of services, in answer to concerns about what search engines provide. The term is underpinned by an expectation of the creation and dissemination of metadata. Structured services built on metadata have made digital resources freely available to everyone with some guarantee of authenticity and reliability. Similarly to search engines, structured services are provided with an expectation that no mediation is required to use them." (Campbell, 2002)

The matching process, in the case of a controlled vocabulary, enables a term in the context of the vocabulary to be matched in another context, and is supported by two critical questions. The first question is: Is the data in the datasets describing the same entity? The second question is: Is the data matching relevant and useful to searchers? If a controlled vocabulary is already used in collection description, that use can enable a straightforward task to establish the technical link, with that vocabulary delivered as a linked open data service. It is clear that the same entity is being described and in the same manner. This is an obvious case of knowing and trusting what data there is in common and programming in the link. If a controlled vocabulary is not used in the collection description practice, but the data in the collection records use common but not validated reference points then matching that data against a vocabulary data service can identify those links. This term matching can establish what data is in common and therefore the datasets can be linked. Depending on the level of accuracy of the matching algorithm, the example of "Dr Bob Bruce" being suggested as a possible match with "Robert B Bruce 1947-" may or may not be made. So being sure or conservative in the matching process is critical, or being clear about the match as a suggested one, means that the searcher understands the validity and integrity of the linkage between the terms being asserted and validated or suggested.

There are specific questions cultural institutions interested in providing or using linked open data need to ask. GLAMs that provide collection data as linked open data services need to ask questions to help inform what collection data services should be developed for others to exploit and what data services are exploited. These are questions such as: Who will be interested in this linked open data service? or What do these organisations need to be able to present this collection data in context with their own data? Cultural institutions need to understand some benefit and recognise common semantics as data characteristics to make the effort to integrate LOD sources to improve resource discovery. The National Library of Australia has partnered with the Powerhouse Museum and used the Museum's collection API (Application Programming Interface) to integrate the Museum's search results with its own (Chan, 2011). This is a great example of carefully curated collection data integration, because of an obvious alignment of subject matter reflected in the collection data. There is also necessarily mutual benefit through this collaboration in achieving wider exposure of cultural collection data. Notions of mutual gain, provenance and authority, and data curation questions, are involved to understand how well matched collection datasets are for this type of integration. In this instance the Museum collection API is well documented, and the collection data is structured. The Library ingests the Museum collection data, gives each collection record a local link, and uses the Persistent Uniform Resource Locator (PURL) to link back to the site of the Museum. The Library has used the PURL provided by the Powerhouse to link back to the collection record on the Museum's website. This is a federated approach to aggregating collection datasets that uses linked open data as

the technical means to do so. The cultural data provided by the Library and the Museum to describe their respective collections is not unionised but those datasets are now integrated and linked. The next areas of practice cultural institutions need to explore are search patterns and web statistics to ascertain how cultural data alignment and integration is affecting use of the collection descriptions online (Chan, 2011).

Terms of use in providing collection or vocabulary data via API may be a critical tool in preventing use of data in ways that it was never intended. This kind of control begs the question whether this kind of caution is absolutely necessary. Aside from requiring attribution, what would be the value of constraining reuse for commercial gain, for example? What would happen if an ecommerce website pulled in and integrated the Powerhouse Museum's collection data in an attempt to increase interest in the wares for sale and the traffic on the website? Search engines have a role to play here in guiding searchers to the resource with the stronger ranking. Perhaps the consequence of the release of collection and vocabulary data needs to be more fully understood in terms of consequence. Is the answer to this concern 'so what'? Web savvy searchers know that search engine optimisation, spamming and link farms are a reality that search engines are continuously grappling with in an open web development environment. Potentially the benefits of data release in most cases far outweigh the downfalls, especially as LOD requires that at minimum a link must have at least two references. It may be important, as with web links, that links that work both ways from two separate data providers indicate a high level of semantic overlap and shared meaning.

This is uncharted territory, and it is worth knowing that a considerable amount of web traffic already goes from eBay to the Powerhouse Museum collection online (Chan, 2011). If eBay integrates the Powerhouse Museum collection in context with its own search results nothing changes, except the searchers on eBay perhaps do not click back to the Powerhouse record for more information. However, searchers would now have a new opportunity presented to them to do so, without having to restart their search, and they become aware of the alignment between the items bought and sold on eBay and those held in the Powerhouse collection. This is an example of the further democratisation of culture and public access to knowledge. The Powerhouse Museum makes images and descriptions about controversial objects in their collection, such as weapons and political posters, readily available for reuse in their online catalogue under a Creative Commons – By Attribution licence. The boundaries and relationships between commerce, culture, science, etcetera, will be tested, blur, get reinforced or reinvented, through new collection or vocabulary data alignment and integration.

Institutions that provide a controlled vocabulary as a linked open data service need to ask questions around how the data service will be used. These questions include how the data needs to be packaged, whether a download or a data service; how the data should be structured for reuse; and what guidance is best provided to enable the data to be exploited effectively. Providing, for instance, information on the history, development, review processes and structure of the vocabulary, that is, hierarchical, enumerative, or faceted, is important to enable developers to understand how to integrate the data into a resource discovery service. If a vocabulary is provided as a data service, developers need to know how the updates are handled, and, in particular, how term retirement is handled, so that links are not broken. There is a significant challenge for cultural institutions providing vocabularies

as data services in this environment in enabling community co-contribution. How do cultural institutions as providers of vocabularies as data services operate in an environment where term suggestion and updating could be community driven? How is authority maintained or spread into a community technically and socially so that change is well managed? Many of the questions around pushing practice into the digital arena are interchangeable, whether in the position of collection data provider or user or vocabulary data provider or user.

In all situations it is critical that those cultural institutions engaged and stepping into this territory think ahead. This is particularly so if a data service, that searchers and other cultural institutions or organisations rely on, is going to be established. The consequences of process and approach change around the exploitation of those services may have a wide impact. The quality, stability and assurance cultural institutions are expected to provide is a fundamental element in their approach to providing and using linked open data. Potentially this need for surety can be seen as the neo-Victorian sensibility of establishing robust infrastructure for public benefit. Impact works both ways in terms of invested time and effort and expectation setting for both linked open data service providers and users, whether it is collection or vocabulary data being made available. Always at the centre of this should be the searcher, or even the data miner⁶, the constituents who are supposed to get the benefit of this change in practice to making and using cultural collection or vocabulary data linked open data services.

Support for Resource Discovery Services

Linked open data simply defined is an approach to enabling the discovery and reuse of datasets and an understanding of the technical components that support that approach to data provision and use. A linked open data approach requires a community wide shared approach and use of technical standards, protocols, ontologies and formats, that is, URI (Uniform Resource Identifier), HTTP (Hypertext Transfer Protocol), RDF (Resource Description Framework)⁷ and XML. The need for an understanding of the potential interest in, use and alignment and relatedness of datasets, and a willingness to commonly encode and make available data, is less obvious. The willingness to commonly encode data is based on peer relationships and social conventions or domain overlap. Peer relationships are formed, around mutual need and shared goals, to serve a user community and potentially the interest in common approaches to collection resource discovery services. Linked open data is therefore one means of improving resource discovery and reuse of collection or vocabulary datasets should the GLAMs decide to encode their data in a common manner and in concert. There are some key design principles outlined and further refined by Tim Berners-Lee (Miller, 2010) that need to be considered in making collection data available online and many other resources (Sherratt, 2011) to support decision-making around providing linked open data. The technologies that support resource discovery services are reliant upon the quality and consistency of the data and how effectively it can be queried. In the library sector the cultural collection data is often of a high quality and consistently produced and there is a strong capacity for high quality resource discovery and aggregation of multiple datasets through unionisation. Where collection data across the GLAMs is of varying quality, inconsistent and less frequently aggregated, search algorithms⁸ can improve the quality of resource discovery. Using a linked open data approach can strengthen

the capacity for resource discovery, in either circumstance, and with either approach within each collecting community in the GLAMs and across the GLAMs.

The premise underlying this paper is that the cultural collecting sector, that is, GLAMs, has established description practices, communities of practice, and a shared interest in collaboration and exploiting linked open data technologies to improve resource discovery of cultural collections online. This is a viewpoint based on working in the collecting sector for approximately 20 years. That viewpoint is very much coloured by the bias of the approach that develops structured description schemas and that makes collection data available online, using whatever technologies are available and readily utilisable to that end. There are others who share this view: “Linked Data is really just a relatively simple, but best-practice way of publishing structured data on the web in a way that encourages discovery and re-use.” (Sherratt, 2011) There are also signs on the local horizon that linked open data is an approach that is increasingly being endorsed. This uptake and endorsement of a LOD approach is evident in Australia from key organisations in the cultural, research and government sectors. The National Library of Australia’s (NLA, 2011) extension of the People Australia services (funded by ANDS to support the Australian Research Data Commons (ARDC) Party Infrastructure), and GeoScience Australia’s release of the Gazetteer of Australia (GSA, 2011) (funded by ANDS to support the ARDC Location Infrastructure) are releasing these public information assets as data services. Governments, universities and cultural institutions from around the globe are now responding to the interest and support for LOD with pilot data services being run by Europeana (Europeana, 2011) and many other cultural institutions (W3C, 2011) involved in W3C initiatives. Workshops have been run at Stanford (Stanford, 2011) and Harvard (Harvard, 2011) universities and the LOD-LAM summits (LOD-LAM, 2011) started in 2011 are due to roll out for another two years.

While this action and activity demonstrates interest and support for a linked open data approach to providing cultural collection data, there are differing viewpoints about the role and function of linked open data in the semantic web. In discussions with several colleagues at Intersect and at the Australian National Data Service, all of whom have computer science backgrounds, it is apparent that there is a view that linked open data is not scalable, whereas search algorithm is. So, there is a question about where LOD fits in the technical ecology of the semantic web that is yet to be answered. The potential to exploit search algorithm exists, but at present it is an ineffective and inefficient solution to meet the goal of resource discovery across multiple cultural collection datasets nationally and internationally. By using a linked open data approach and undertaking data curation to enable the integration of well-aligned cultural collection data and vocabularies, GLAMs will go some way to achieving that goal. More crucially, the semantic foundation of structured and curated collection and vocabulary data, that serves either of these technologies, search algorithm and linked open data, to meet that end, maintains its value, by being made available as linked open data. That value lies in the careful, meaningful and consistent structuring of collection and vocabulary data through established and continuously improved description practice. Uptake of schemas, controlled vocabularies and the use of linked data techniques improve description practices and importantly provide a rich substrate for search technologies.

Discovery and Access with Linked Open Data

The kind of online discovery and access that linked open data will enable is both targeted and serendipitous information retrieval. Searchers will not have to compile multiple search strategies with complex search strings with alternate terms. Before the web this was the search approach used to interrogate complex and highly structured cultural datasets, one after the other. With the advent of the web a combination of simple searching strategies, improvements to search algorithms, and the exposure of metadata to search engines has meant less effort in constructing a search across cultural collection description on the web. Search on the web has not enabled the level of accuracy that search in a cultural collection database affords. The greater exposure of metadata and linked open data available online can only improve the quality of search on the web. The value of the online collection catalogue is retained, but it is likely to sit more firmly behind the search interface as a foundation for search through linked open data as the search interface evolves. Again, this shift in collection description practice, through using new technologies and approaches, could be construed as neo-Victorian in nature. Victorian, in that this approach aims for the development of stable infrastructure based on established cultural authority to enhance public access to cultural knowledge. Neo, in that the cultural authority and access are reinvented and re-established in new ways.

Resource discovery and access to cultural information on the web relies significantly upon open standards with well-identified and defined use of data services, endpoints or interfaces and a collective approach in data provision including linked open data. Search can also be idiosyncratic and formalised with selected data sources to interrogate in a closed environment using a highly controlled schema and syntax to generate search results. Different outcomes of search can be better enabled through providing and using linked open data methods to expose collection and vocabulary data widely. This enables searchers to serendipitously move or browse through and across cultural collection datasets, rather than having to search in one dataset or then onto another, in a controlled and structured manner. There are definitely gains with well-defined and structured data for both technical methods of search and query, irrespective of context. In reading recent research, it appears that search techniques work well when applied on top of a rich substrate of well-structured data that is linked. There are efforts to provide that formalisation programmatically and to ascertain ways to reduce the human input to aligning information structures and standards. Bear in mind that all of this applied technology is to improve the potential to further search, retrieve or browse contextually relevant cultural information to that which an initial search has brought to light.

“Linked Data will be generated and consumed by those who seek its specific benefits. Having a structured data substrate makes the aforementioned pursuit exponentially easier.” (Idehen, 2011)

“Structured data is important, its [sic] actually more important than Linked Data at this point in time re. Web evolution.”(Idehen, 2011)

Library domain-specific methods of enabling library collection data aggregation (or unionisation) using Z39.50, pre-web search and retrieval protocol, and MARC (Machine Readable Cataloguing) schema were superseded by the use of SRU/SRW (Search Retrieval via URL/Search Retrieval via Web) web search and retrieval

protocol, query languages⁹ such as CQL (Contextual Query Language) and soon RDA (Resource Description and Access) schema. Providing and using linked open data, web standards and query languages such as, SPARQL (SPARQL Protocol and RDF Query Language) to build inference¹⁰ into search algorithms can be employed to effectively query data points, rank, filter and present specific resources in a result set. This result set can in turn contextualise a search with results from a closely aligned cultural collection dataset or content. Linked open data enables a wider group of collecting organisations to provide, and also parse, aligned and contextually meaningful and relevant external datasets, on the web, using common standards. The benefit of this ability to co-locate datasets is that it is possible to curate, that is, release and pull in contextual and related data, to augment a search within a single collection search interface. That external dataset may be more cultural collection data or another related dataset. With the alignment and integration of cultural datasets searchers can then move flexibly from a search, through a search engine, to a museum collection search interface, to information about a museum object, to a related publication about an exhibition, to a biographical entry about the object creator, to a gallery collection and so on.

In a sense, linked open data enables 'linked search'. Linked search is a search that retrieves results from multiple cultural collection datasets through their linkages. Linked search does not rely on overall collection description standardisation and unionisation or federation as a formalised means to support search across multiple resources; rather, it relies on LOD. The key question in here for cultural professionals, including developers, focused on providing resource discovery and access services in GLAMs is what related datasets are curated and co-located with the cultural institution's cultural collection data. The end result ideally enables searchers to begin to see the resources and information in their own collection, and, the relationships between that cultural content and content in other cultural collections or other information resources. Linked open data will operate quietly in the background enabling that incidental coursing through cultural collection data via the semantic seams and links established. What is perfect about linked open data is: the ability to retain a domain-specific schema and present cultural collection and vocabulary data in a way that meets the goals of the cultural institution as the provider; enables collaboration with diverse collecting agencies in a sector, or other agents in other sectors; and ultimately serves the searchers, as it has been designed to do. The steampunk/neo-Victorian metaphor is realised through the development of public digital infrastructure to support enhanced search and engagement with digital cultural heritage collections on the web.

Linked open data in this context permits similar, but different cultural collection datasets to work in concert with other relevant collection or vocabulary datasets, as a service for searchers. The value and importance of the heterogeneous nature of cultural collection and vocabulary data is maintained and reinforced with linked open data. The unique, rich and structured nature of cultural collection data is showing its potential as a great counterbalance to criticism of its idiosyncrasies in comparison with homogenous or highly unionised data. Homogeneity as a characteristic of data arises where the data generation is highly controlled. An example of this is where data is generated off instruments, and the schemas for the data are predefined as part of the instrument setting and the value of the highly consistent data through machine control, or where there are strict rules and controls for manual data input. In the realms of human intervention, whether it is to provide an interpretation and

cultural overlay to describe a cultural collection object or to record observations in the field or a hospital, there is less potential to control the data when it is hand notation or sketches, except where annotation standards have been established to capture and categorise that data consistently. Linked data provides a means to connect diverse datasets whether heterogeneous or homogeneous. The capacity to loosely couple diverse datasets, where there is alignment or convergence of data, is well served by using linked data approaches. There are opportunities to render cultural collection data more consistently open to search by using common data formats, in the GLAMs sector, and probably more widely, associated with culturally relevant data reflecting: people and organisations, date and time, and location. There are already common and standardised methods to encode this data being used, respectively, such as, EAC-CPF, UTC (Universal Time Coordinated), IAEC (International Address Element Code), and ANZLIC-MAP, used in cultural collection description practices.

Common use of standards to encode metadata is reflected in the Australian National Data Service eResearch infrastructure initiative to aggregate metadata in RIF-CS format in a search service, Research Data Australia, for research data discovery. RIF-CS is a schema focused around collection level description, to enable all research domains to provide high-level information about research data collections, in the context of information about researchers and research organisations, research activities and research services. There is the potential to aggregate collection data cross-walked (for two of the four of the entities in RIF-CS schema respectively) from the following standards: EAD (Encoded Archival Description) or DC-CAP to the Collection element; and EAC-CPF, FOAF, VIAF (Virtual International Authority File) or ORCID (Open Research and Contributor Identifier) to the Party element. Already ANDS Partners, that is, funded project participants, have either mapped to RIF-CS or cross-walked metadata in their own information systems to the RIF-CS schema, where the metadata is in MODS (Metadata Object Description Schema) or DC (Dublin Core) (McCulloch, 2011). There is a proliferation in many domains and sectors of information standards and methods being developed, and high interest in and capacity to translate those to enable data to be repurposed and more widely exposed. This is almost an explosion of information standards that has parallels with the kind of innovation and swift uptake of standardisation in practice in areas such as communications, manufacturing and transport during the Industrial Revolution.

Data services in Australia have been developed to support linked open data so that a searcher can use a search engine to shift from one information retrieval event to another easily. It will be possible soon to retrieve a record about a university research project in Australia around a historical period by searching on a research code from ANZSRC (Australian and New Zealand Standard Research Codes) and retrieve a research data collection with a strong connection to an Australian historical figure. If in the record of that research data collection the National Library of Australia Party Identifier has been included as a link, a searcher can then resolve to the National Library of Australia to their Trove service. The Trove resource discovery service uses the Powerhouse Museum collection data API, a data service, to extend that service. A searcher can then find publications by or about that historical figure held in the National Library collection or in the Powerhouse Museum collection. In this sequence of information retrieval events that constitute a 'linked search' there are elements of the linked open data approach highlighted. The ANZSRC is a controlled vocabulary used widely in Australia and New Zealand research sectors to

categorise research activities and outputs. This controlled vocabulary is used to enrich metadata and will be provided as a data service for organisations to build and thread into their collection description practices and resource discovery services. The National Library of Australia Party Identifier is a unique and resolvable identifier and can be used to enrich metadata and be exploited in resource discovery services. Using a linked open data approach to the provision of collection data increases the likelihood of targeted and serendipitous information retrieval to be achieved and more efficiently.

Collection Information Standards

The impact that embedding linked open data will have on the collecting sector description practices is manifold. Collection description practices will shift towards greater consistency overall and the way collection data is made available for reuse will diversify. Greater data consistency will not mean reducing or discarding the characteristics of individuation or heterogeneity; rather, it will mean a continuing increase of mappings, crosswalks, higher-level description, collection data sharing, and common use of controlled vocabularies and linked open data approaches to improve resource discovery via web search. The related and untested impact is the opportunity for community co-contribution to the development of controlled vocabularies and collection description thereby extending the labour and potentially shifting the role of collection description practice in cultural institutions towards moderation rather than complete control of cultural information. The increasing uptake, interest in and development of linked open data as an approach to making cultural collection data reusable by other agencies reinforces the need to continue a long tradition of formalised description of artefacts, whether human or machine generated. The ten years of expanding use of description standards across the GLAMs and shared use of common standards, such as DC, RDA or CIDOC-CRM (International Committee for Documentation – Conceptual Reference Model), either through adherence or through mapping has stood the community in great stead.

It is encouraging to see that many of the DCMI (Dublin Core Metadata Initiative) papers at the Dublin Core conference this year (DCMI, 2011) refer to linked data, linked open data, and improving understanding crosswalks and mapping across diverse schema. However, the following statement from the National Library of Australia about the change from using AACR2 (Anglo-American Cataloguing Rules Second Edition) to RDA (Resource Description and Access) says it all:

“RDA is a particularly exciting development for libraries as it will help them pursue their web-based resource sharing strategies by creating records that are more flexible in meeting users’ needs, and shareable with other metadata communities.” (Argus, 2008)

Both the clearly articulated goal from the National Library of Australia, to be responsive to needs and open record sharing work across the different collecting communities in the GLAMs, and the LOD-LAM summits, are encouraging signs on the horizon for those that understand how the web paradigm of search has changed expectations (Coyle, 2006). The LOD-LAM summits have afforded awareness-raising, peer networking and provided an opportunity to talk about the practical and professional issues with taking up linked open data as an approach to provide cultural collection data.

It is possible to describe the advent of RDA and the take up linked open data as a contentious and disruptive approach to cultural collection description practices. This is another place the steampunk metaphor can be applied, because the approach seems anarchic and requires traditions to be interrogated and, at times, thwarted. Dominant paradigms and practices in the cultural sector are challenged in this process of reinvention and use of new technologies and approaches. A reactionary response to increasing public access to cultural data seems ultimately rather ironic. This change in cultural collection description practices is about reaching for a broader social goal that serves tradition, that is, to enable richer and more flexible access to cultural collections, in this instance via changes to practice and use of linked open data and web search. In terms of impact, collection description as well as web search practices are changing and will continue to need to change. Fortunately the value demonstrated by early adopters provides a large incentive. By, for example, searching via Google it is possible to undertake an information quest using specific search terms and resolve through to National Library of Australia's Trove, the Museum Metadata Exchange, DigitalNZ and Europeana, because the cultural collection data is exposed to the web. This makes it clear that single, adjunct, like Trove, and multi-collection, like the Museum Metadata Exchange, DigitalNZ and Europeana, web search of collection data is here to stay. The Museum Metadata Exchange has been set up to enable collection-level search across Australian museum collections, and will be extended to archival collections soon. This metadata is mapped to the RIF-CS schema maintained by ANDS and extends it. DigitalNZ has been set up to support the exposure of collection metadata and the development of data services. Collection data from the Powerhouse Museum has been integrated into DigitalNZ's federated search and can be seen in context with the Te Papa Museum collection search (Chan, 2011). Europeana has been set up to provide access to digitised and multimedia European cultural heritage collection items. The Europeana uses a schema based on Dublin Core, extended for local implementation and includes the use of controlled vocabularies. The portal employs language translation technologies in support of multilingual information retrieval. These metadata aggregation portal developments require collaboration and the commitment of collection data providers to change and enhance collection description practice. Practice change and the use of new technologies support the initiative to increase access to cultural collection data through improved search.

The ability to support browsing from a single collection service to an adjunct collection, then onto an aggregation of collections service and onto another, by coursing through linked open data, just means the searcher does not necessarily have to start their information quest afresh at each service. Instead the searcher can use this layer of linked open data around standardised references to people, organisations, date/time and location to move in and out of cultural collection datasets. If those common points of shared meaning are standardised and proliferate in schemas, the GLAMs may also benefit from shared use of contextually relevant collections, vocabularies and metadata element sets (Isaac et al, 2011). Collection data can be extended and can be released for others to leverage in their own search interfaces. What collection data curation and value-adding in single search interfaces might do to web search is to provide a similar platform for reference ranking by search engine algorithms and exploit that shared value, alignment and linking of datasets. It may also assist with improving ranking in search results.

Open Data

The linked open data movement is about openness and the potential to widen value generation and remove barriers to access resources, especially publicly funded resources. Where there is a need to provide advanced online access to data and metadata for public benefit, there are large drivers to bring about those changes in other domains, where community service is a core and strategic outcome. The impact on all sectors with increasing levels of openness in the provision of data and materials, which includes cultural research and government sectors, of using linked open data, is entirely in keeping with their social mandate. The reference to neo-Victorian sensibilities is apt in an environment of reinvention or creation anew of public infrastructure with a large social values quotient.

“Linked Data makes sense if your data is inherently meant for global consumption. It does not make sense if your data is only meant to be consumed locally (within a document or graph, for example). The Web is filled with both types of data. If we make a spec that only deals with the first case and not the second, we will have created a useless spec.”(Sporny, 2011)

Where matters get interesting is in what publicly accessible data is exploited and what structured data, vocabularies and metadata element sets are contextually relevant and can be used. The ability to augment the value of providing single collection search with further resource discovery and enable multi-collection search, has already been demonstrated by the Powerhouse Museum, in partnership with the National Library of Australia and DigitalNZ. The ability to augment single collection search with further traffic and access, with adjunct datasets and also linked open data, is yet to be fully realised in the cultural sector. An approach to content provision, sharing and reuse using linked open data is an approach that should work for the research and government sectors too.

“The success of Web APIs has shown the power of applications that can be created by mashing up content from different Web data sources. However, mashup developers face the challenge of scaling their development approach beyond fixed, predefined data silos, to encompass large numbers of data sets with heterogeneous data models and access methods. In contrast, Linked Data realizes the vision of evolving the Web into a global data commons, allowing applications to operate on top of an unbounded set of data sources, via standardised access mechanisms. If the research challenges highlighted above can be adequately addressed, we expect that Linked Data will enable a significant evolutionary step in leading the Web to its full potential.”(Bizer, Heath, Berners-Lee, 2009)

It is clear that this approach works for Amazon; it is also working right now for the BBC (British Broadcasting Corporation).¹¹ The impact on these sectors is no different to that of the cultural sector. A change in the information practices and some measured thinking about what data and ‘whose’ to link to and why, and what flow-on effect that will have to information consumers nationally and internationally need to occur to enable constructive change. Questions around authority, authenticity and trust, given the trusted public role these public agencies have, continue to arise in this practice change. A change process that includes checks and balances is a good way of ensuring that public institutions make effective and well-considered change to improve resource discovery and access to information.

Conclusion

Two broad questions were posed in this paper's title and in the first paragraph. The first question was around the particular qualities of the collecting sector: Are the GLAMs going to bring a steampunk/neo-Victorian sensibility and aesthetic to Linked Open Data? The cultural collecting world has long been acquainted with cataloguing and classification as well as the natural science and medicine domains and developing formal systems for identification, that is, ontologies. Collection description practices in the cultural sector have in the past focused heavily on identification and reflecting the diversity of the content. This is entirely in keeping with the ethos underlying collection building, that is, to form a unique collection, to reflect the cultural heritage of the community and thereby meet their cultural needs. Collection description practices are moving onto a heightened phase of content alignment online, with data aggregation and sharing across domains and sectors. This heightened activity reflects the commonality and relatedness of the content across the GLAMs. Data alignment is an act of curation and will often highlight expected overlaps of data, and therefore collections, of related cultural institutions, or bring to light unexpected overlaps. This shift in practice breaks down many of the artificial intellectual barriers between collections and content held in cultural institutions and social institutions and more broadly in the research and government sectors. The semantic threads connecting cultural datasets in GLAMs are through the wide development and use of controlled vocabularies and, more recently, the interest in cross-walking metadata to commons schema in their collection description practices.

It seems entirely possible that a neo-Victorian sensibility and aesthetic, perhaps while more obviously associated with cultural heritage, because of a sense of vintage, is much more wide spread across all domains, but in particular social institutions. This is because the neo-Victorian sensibility and aesthetic underlying this change in collection description practices, and the linked open data approach to providing data, is the support of open access to culture and knowledge. Method, practice and outputs in cultural, research and government sectors are being codified more and more as the developed world moves concertedly into embedding technologies into is commercial work, education and recreation activities. Intensified codification is reflected in the digital infrastructure and artefacts being made online. One could say that the desire to codify and formalise are, at their simplest, the founding blocks of knowledge. If those founding blocks form an open infrastructure then the ability to access and exploit knowledge is increased. The development of linked open data is one small pieces of the epistemological puzzle, made real through a multi-cultural and collaboratively built artefact also known as the World Wide Web. The World Wide Web is a very different socially generated and digital artefact to that of the Dark Web, where digital content and services are privileged, hidden behind security layers, and not publicly accessible. The ethos of openness and public accessibility needs to be seen in perspective and in context. This shift in practice and use of new technologies occurs where there is a need to draw together ideas and data. The use of linked open data and the alignment of datasets can meet public, social, cultural and economic needs and it can meet private needs too. The critical point is that the linked open data needs to be open and standardised to enable the approach to be widely and flexibly used. These encoded cultural artefacts are presented mostly on the web and the cultural data does not change rapidly, so the information stream generated by cultural institutions, while weighted with historical data, can be readily reached via a search online. Therein lies the

steampunk element to the metaphor; that is, old and new social values embedded in old and new digital artefacts.

An instructive and robust debate occurred on a listserv thread (Idehen, 2011) recently between technical people around RDFa (Resource Description Framework – in – Attributes) and microformats¹² to determine how JSON (JavaScript Object Notation) will be defined so that it can reflect whether the JSON provided is a linked data version or not. One of those thread contributors made the point well that technological approaches that require openness, collaboration and shared understanding are necessarily ones that require enlightened and liberal thinking. The social ideology underlying this thinking is designed to support the diverse needs and the developments of a data Commons. Where there is value perceived, in coherence and connectivity across domains of activity and intellectual output, there necessarily is collaboration. Where there is not value perceived in alignment, there is necessarily a need to distinguish and not enable connectivity. Enlightenment thinking drove innovation, and barriers, actual, social or intellectual, were broken down as new concepts and approaches were worked through and taken up in the Industrial Revolution. Social values in Enlightenment thinking enable innovation and value generation in all domains: public, private, old, new, social, cultural and economic.

There is plenty of evidence already demonstrating the merit in opening up and sharing information standards to enable society to function more effectively and to enable cultural connection (Mason, 2006). This public value can be generated in any sector that has public benefit as a core responsibility irrespective of the nature of the outcome, social or economic. The core activity of the Victorian era was to open up public access to knowledge. While the GLAMs seem to have obvious connections to the past, the sensibilities of virtue inherent in the steampunk/neo-Victorian metaphor presented, so too do the institutions in the business of supporting public: health, education, welfare and citizenry. It is the aesthetic of new technologies presenting information where potentially the difference lies between the cultural and the research and government sectors in the use of this metaphor. The GLAMs will provide and share information about old and new cultural heritage artefacts and present that old and new information, using new practices of collection description. Cultural heritage information though does not change much. Cultural collection data slowly accretes in volume and value over time. The research and government institutions are most likely to provide and share information about new discoveries and social processes and present that using new practices of description. So, the aesthetic of steampunk/neo-Victorian refers to the characteristics of the GLAMs cultural collection data, the heritage content being provided and shared. It is the role of GLAMs to provide cultural heritage information, the legacy material and right up to the minute information, about natural history finds or material culture artefacts.

This leads to the second question: Is there a way forward for GLAMs to tap into the rich, historical and viscous nature of its practice and build this into the approach to making collection data available online and employing linked open data in that approach? Stronger and wider dialogue is needed to enable the technical discussion to filter across to the domain experts and vice versa. A key focus is needed on shared interest and goals for wider access, to enable allied domains to settle on information standards and find alignments, through use of common ontologies and open technical frameworks. The geo-spatial community is a positive exemplar of a community of practice built up around information standards. There seem to be few other domains or communities of practice to have managed to cooperate and

interoperate and share metadata openly on the web to improve resource discovery.¹³

“The upshot of this is that there is no one single destination for users involved in a search for information or in some other information-related activity. A survey conducted by OCLC on user perceptions of libraries shows that the vast majority of those interviewed use search engines to begin an information search (84 percent). Only one percent of users begin an information search on a library Web site. (p. 6-3) Which doesn't mean that they are not interested in library services, but they may not discover those services if the library does not find a way to go to the user, rather than waiting for the user to come to the library.” (Coyle, 2006)

It would be gratifying to think the GLAM community approach to using linked open data may provide another useful example to other areas of specialisation as a community of practice. This would make data interoperability and linking a common practice for related domains of data. A wider variety of cultural resources and information would be accessible online and reusable on the web and enable better search and browsing services. The way forward for the GLAMs community is getting clearer. As with any new movement or technology being explored, some initial testing has proved inspiring and exciting and more questions have been posed as interest in practice change deepens. The innovators in various fields have been busy debating and thinking through the technical challenges, running workshops, communicating, networking and piloting and are onto the next steps.¹⁴

“However much of these data are in non-linked formats. The data models are difficult to understand and re-use, and closed to web-scale integration. Publishing these data in linked data format would make it easier for them to be re-useable and interlinked. Resolving data and schema heterogeneity is a heuristic semi-automatic process. In future work, we aim to explore the application of data mining techniques to reduce the time it takes a human expert to align instances and/or schema. We have built a backlinking service to the Linking Open Data cloud. We aim to further integrate the backlinking service to our datasets. In addition, we aim to provide an efficient scalable user interface able to visualise and search multiple datasets.” (Omitola et al, 2010)

The acolytes¹⁵ and early adopters have established value; now the innovators need to be followed by the early and late majority, according to the theory of diffusion of innovation.¹⁶ The next and bigger wave of linked open data activity and the changes in description practices is upon us. The GLAMs seem to be on the cusp of building a community of practice across galleries, libraries, archives and museums, to realise a shared goal of improved resource discovery and access.

Acknowledgements

Acknowledgement and thanks go to Dr Tim Sherratt for prompting me to apply to go to the LOD-LAM summit in San Francisco (2011) and for providing useful advice¹⁷ and information to enable a case to be presented to advocate for exploiting linked open data methods in the provision of cultural collection data online. This support has meant that I could briefly reference the technology debate and, more importantly, the nature and value of collection data as highly structured, meaningful domain level metadata could be highlighted. Also, acknowledgements go to my colleagues at the Australian National Data Service, Dr Ming Fang Wu; and Intersect Australia Limited Drs Joe Thurbon, Cameron Maxwell, Luc Small and Anne Cregan,, and Jamie Norrish and Conal Tuohy (ex New Zealand Electronic Text Centre colleagues) for offering their expertise and informal insights to linked open data and search algorithm from computer science and semantic web perspectives and understandings in conversation.

Note

The examples of information standards used in this paper are not intended to be extensive, but illustrative. There are numerous standards for describing cultural collections and shared schemas for sharing data that are not represented in this paper. The same approach has been taken to highlighting examples of advanced access for searching cultural collection data online. There are numerous examples of innovators in the GLAM domain and in other domains internationally that are not represented in this paper.

Bibliography

Argus, Catherine. (2008) RDA is Coming! Gateways, No. 94, August 2008. National Library of Australia. <http://www.nla.gov.au/pub/gateways/issues/94/story06.html> (Last accessed: 30/09/2011)

Australian National Data Service. (2012) Registry Interchange Format – Collections and Services. <http://ands.org.au/resource/rif-cs.html> (Last accessed 13/01/2012)

Australian National Data Service. (2012) Research Data Australia. <http://services.ands.org.au/home/orca/rda/> (Last accessed 13/01/2012)

Bechhofer, Sean et al. (2010) Why Linked Data is Not Enough For Scientists. Preprint. <http://eprints.ecs.soton.ac.uk/21587/>In: *Sixth IEEE e-Science conference* (e-Science 2010), December 2010, Brisbane, Australia. (Last accessed 29/09/2011)

Berners-Lee, Tim. (2009) Linked Data – Design Issues, June 18th, 2009. W3C Semantic Web. <http://www.w3.org/DesignIssues/LinkedData.html> (Last accessed 29/09/2011)

Bizer, Christian, Tom Heath and Tim Berners-Lee. (2009) Linked Data – the Story So Far, Tom Heath. Preprint. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> In: *International Journal on Semantic Web and Information Systems*, Vol. 5(3), Pages 1-22. DOI: 10.4018/jswis.2009081901 (Last accessed 29/09/2011)

Bradley, Aaron. (2011) What Schema.org Means for SEO and Beyond, 3 June 2011. SEO Skeptic. <http://www.seoskeptic.com/what-schema-org-means-for-seo-and-beyond/> (Last accessed 29/09/2011)

British Broadcasting Corporation (BBC). 2011. Our Approach to External Linking. <http://www.bbc.co.uk/help/web/links/> (Last accessed 30/09/2011)

British Broadcasting Corporation (BBC). 2011. Editorial Guidelines, Guidance, Links and Feeds, Summary. <http://www.bbc.co.uk/guidelines/editorialguidelines/page/guidance-links-feeds-summary> (Last accessed 30/09/2011)

Campbell, Debbie. (2002) Feasibility Study for a Digital Australiana Resource Discovery Service, National Resource Discovery Metadata Initiatives, National Library of Australia. <http://www.nla.gov.au/initiatives/feasibility.html> (Last accessed 29/09/2011)

Chan, Seb. (2011) Actual use data from integrating collection objects into DigitalNZ. <http://www.freshandnew.org/2010/10/31/actual-use-data-from-integrating-collection-objects-into-digital-nz/> (Last accessed 13/01/2012)

Chan, Seb. (2011) Museum collection meets library catalogue. Fresh & New(er). <http://www.freshandnew.org/2011/11/08/museum-collection-meets-library-catalogue-powerhouse-collection-integrated-trove/> (Last accessed 11/01/2012)

Clifton, Christopher. (2011) Data Mining. In: *Encyclopædia Britannica*. <http://www.britannica.com/EBchecked/topic/1056150/data-mining> (Last accessed 29/09/2011)

Coyle, Karen. (2006) The Library Catalogue in a 2.0 World. Preprint. In *Journal of Academic Librarianship*, Vol. 33, No. 2. http://www.kcoyle.net/jal_33_2.html (Last accessed 30/09/2011)

Dublin Core Metadata Initiative (DCMI) (2011). DC-2011 Proceedings Abstracts. DCMI International Conference on Dublin Core and Metadata Applications. <http://dcevents.dublincore.org/index.php/IntConf/index/pages/view/abstracts-2011> (Last accessed 30/09/2011)

Europeana. (2011) Europeana Linked Open Data (LOD). <http://pro.europeana.eu/linked-open-data> (Last accessed 11/01/2012)

Geoscience Australia. (2010) Gazetteer of Australia. <http://www.ga.gov.au/place-name/> (Last accessed 9/01/2012)

Hartig, Olaf and Andreas Langegger. (2010) A Database Perspective on Consuming Linked Data on the Web, Olaf Hartig. Preprint. http://www2.informatik.hu-berlin.de/~hartig/files/Hartig_QueryingLD_DBSpektrum_Preprint.pdf
In: *Datenbankspektrum*, 10(2) Oct. 2010. (Last accessed 29/09/2011)

Harvard University. (2011) Global Interoperability and Linked Open Data Workshop Notes – Digital Library of America Project. http://cyber.law.harvard.edu/dpla/May_2011_Global_Interoperability_and_Linked_Open_Data_Workshop_Notes (Last accessed 10/01/2012)

Heery, Rachel. (1998) What is... RDF? Ariadne, Issue 14, 1998. <http://www.ariadne.ac.uk/issue14/what-is/> (Last accessed 29/09/2011)

Idehen, Kingsley. (2011) Re: Is Linked Data Useless? 28 July 2011. Public-Linked-JSON@w3.org Mail Archives. <http://lists.w3.org/Archives/Public/public-linked-json/2011Jul/0158.html> (Last accessed 29/09/2011)

Isaac, Antoine et al. (2011) Vocabulary and Dataset. W3C. http://www.w3.org/2005/Incubator/lld/wiki/Vocabulary_and_Dataset (Last accessed 30/09/2011)

Kobilarov, Georgi. (2010) Linked Data and Enterprise Data Integration, Georgi Kobilarov. <http://blog.georgikobilarov.com/2010/03/linked-data-and-enterprise-data-integration/> (Last accessed 29/09/2011)

Linked Open Data – Libraries, Archives, Museums (LOD-LAM). (2012) International Linked Open Data in Libraries, Archives and Museums Summit, June 2-3, 2011, San Francisco, CA, USA. <http://lod-lam.net/summit/> (Last accessed 11/01/2012)

Linkeddata.org (2011) Linked Data. <http://linkeddata.org/> (Last accessed 29/09/2011)

McCulloch, Julie. (2011) Metadata Mapping: InfoPath Form XML, MODS, DC and RIF-CS. Monash University. <http://www.researchdata.monash.edu.au/collections-project/toolkit/mods-dc-rif-cs-final-mar-2011.xls> (Last accessed 30/09/2011)

Miller, Paul. (2009) The Cloud of Data - Blog. Does Linked Data Need RDF? July 19th, 2009. <http://cloudfodata.com/2009/07/does-linked-data-need-rdf/> (Last accessed 29/09/2011)

Miller, Paul. (2010) Linked Data Horizon Scan. Tim Berners-Lee's Linked Data Principles. Cloud of Data and JISC. <http://linkeddata.jiscpress.org/tim-berners-lee%E2%80%99s-linked-data-principles/> (Last accessed 29/09/2011)

National Library of Australia (NLA). (2011) ARDC Party Infrastructure Project. <https://wiki.nla.gov.au/display/ARDCPIP/ARDC+Party+Infrastructure+Project+Home> (Last accessed 9/01/2012)

Omitola, Tope et al. (2010) Integrating Public Datasets Using Linked Data: Challenges and Design Principles. *CEUR Workshop Proceedings*, Volume 700. Proceedings of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly, Ghent, Belgium, December 16-17th, 2010, edited by Soren Auer et al. <http://ceur-ws.org/Vol-700/Paper8.pdf> (Last accessed 29/09/2011)

Pedrinaci, Carlos and John Domingue. (2010) Linked Services and the Future Internet. *CEUR Workshop Proceedings*, Volume 700. Proceedings of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly, Ghent, Belgium, December 16-17th, 2010, edited by Soren Auer et al. <http://ceur-ws.org/Vol-700/Paper11.pdf> (Last accessed 29/09/2011)

Sack, Harald. (2010) Linked Open Data Universe. Linked Data Universe – Large Scale Computing Tasks for the HPI FutureSOC-Lab Presentation for: 5th Annual Symposium on Future Trends in Service-Oriented Computing, June 16th, 2010. Hasso-Plattner-Institute for IT Systems Engineering, Potsdam. Slideshare – lysander07. <http://www.slideshare.net/lysander07/linked-data-universe-large-scale-computing-tasks-for-the-hpi-futuresoclab> (Last accessed 29/09/2011)

Sequeda, Juan and Olaf Hartig. (2010) Open Research Problems in Linked Data. Consuming Linked Data Tutorial Presentation for: World Wide Web Conference. Slideshare – juansequeda. <http://www.slideshare.net/juansequeda/07-openresearchproblems> (Last accessed 29/09/2011)

Sherratt, Tim. (2011) Personal email from Tim Sherratt to Ingrid Mason, 16 June 2011.

Sporny, Manu. (2011) Is Linked Data Useless? 27 July 2011. Public-Linked-JSON@w3.org Mail Archives.

<http://lists.w3.org/Archives/Public/public-linked-json/2011Jul/0153.html>

(Last accessed 29/09/2011)

Stanford University. (2011) Report of the Stanford Linked Data Workshop, 27 June – 1 July 2011.

http://library.stanford.edu/about_sulair/news_and_events/Stanford_Linked_Data_Workshop_Report_FINAL.pdf (Last accessed 10/01/2012)

Sun SITE Central Europe (CEUR). (2011) Proceedings of the 1st International Workshop on Semantic Digital Archives, Berlin, Germany, September 29, 2011.

<http://ceur-ws.org/Vol-801/> (Last accessed 10/01/2012)

Wikipedia. (2011) Diffusion of innovations. 23 September 2011.

http://en.wikipedia.org/wiki/Diffusion_of_innovations (Last accessed 30/09/2011)

Wikipedia. (2012) Digital revolution. 15 December 2011

http://en.wikipedia.org/wiki/Digital_Revolution (Last accessed 5/01/2012)

Wikipedia. (2011) Inference Engine. 25 January 2011.

http://en.wikipedia.org/wiki/Inference_engine (Last accessed 29/09/2011)

Wikipedia. (2011) Information Retrieval. 23 September 2011.

http://en.wikipedia.org/wiki/Information_retrieval (Last accessed 29/09/2011)

Wikipedia. (2011) Information Retrieval Query Language. 3 June 2011.

http://en.wikipedia.org/wiki/Information_retrieval_query_language (Last accessed 29/09/2011)

Wikipedia. (2011) Neo-Victorian. 11 September 2011.

<http://en.wikipedia.org/wiki/Neo-Victorian> (Last accessed 30/09/2011)

Wikipedia. (2011) Query Language. 29 September 2011.

http://en.wikipedia.org/wiki/Query_language (Last accessed 29/09/2011)

Wikipedia. (2011) Search Algorithm. 10 September 2011.

http://en.wikipedia.org/wiki/Search_algorithm (Last accessed 29/09/2011)

Wikipedia. (2011) Steampunk. 20 September 2011.

<http://en.wikipedia.org/wiki/Steampunk> (Last accessed 30/09/2011)

Williams, Alex. (2011) The Value of Linked Open Data [Infographic], ReadWriteCloud, February 1st, 2011.

<http://www.readwriteweb.com/cloud/2011/02/the-value-of-linked-open-data.php> (Last accessed 29/09/2011)

World Wide Web Consortium (W3C). (2011) Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets. W3C Incubator Group Report 25 October 2011.

<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>

(Last accessed 11/01/2012)

World Wide Web Consortium (W3C). (2009) Semantic Web Case Studies and Use Cases, W3C Semantic Web, 2009.

<http://www.w3.org/2001/sw/sweo/public/UseCases/> (Last accessed 29/09/2011)

Endnotes

¹"The Digital Revolution is the change from analog mechanical and electronic technology to digital technology that has taken place since c. 1980 and continues to the present day... Implicitly, the term also refers to the sweeping changes brought about by digital computing and communication technology during the latter half of the 20th century. Analogous to the Agricultural Revolution and Industrial Revolution, the Digital Revolution marked the beginning of the Information Age. Central to this revolution is the mass production and widespread use of digital logic circuits, and its derived technologies, including the computer, digital cellular phone, and fax machine. "(Wikipedia, 2011)

²The change associated with digital technology is represented by the use of terms: digital cultural heritage, eResearch, and eGovernment.

³"Works of steampunk often feature anachronistic technology or futuristic innovations as Victorians may have envisioned them, based on a Victorian perspective on fashion, culture, architectural style, art, etc. This technology may include such fictional machines as those found in the works of H. G. Wells and Jules Verne." (Wikipedia, 2011)

⁴ "Neo-Victorian is an aesthetic movement which amalgamates Victorian and Edwardian aesthetic sensibilities with modern principles and technologies. A large number of magazines and websites are devoted to Neo-Victorian ideas in dress, family life, interior decoration, morals, and other topics. (Wikipedia, 2011)"

⁵"Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information and knowledge on the Semantic Web using URIs and RDF."" (linkeddata.org, 2011)

⁶"Data mining, also called knowledge discovery in databases, in computer science, is the process of discovering interesting and useful patterns and relationships in large volumes of data. The field combines tools from statistics and artificial intelligence (such as neural networks and machine learning) with database management to analyze large digital collections, known as data sets. Data mining is widely used in business (insurance, banking, retail), science research (astronomy, medicine), and government security (detection of criminals and terrorists)." (Clifton, 2011)

⁷"RDF is distinctive in that it is designed specifically with the Web in mind, so it takes account of the features of Web resources. It is a syntax based on a data model, and this model influences the way properties are described in that it makes the structure of descriptions explicit. This means RDF has a good fit for describing Web resources, but, on the downside, it might cause problems within environments where there is a need to re-use or interoperate with 'legacy metadata' which may well

contain logical inconsistencies. (An extreme example might be the MARC format, the design of which was influenced originally by the needs of tape storage.)” (Heery, 1998)

⁸“In computer science, a search algorithm, broadly speaking, is an algorithm for finding an item with specified properties among a collection of items. The items may be stored individually as records in a database; or may be elements of a search space defined by a mathematical formula or procedure, such as the roots of an equation with integer variables; or a combination of the two, such as the Hamiltonian circuits of a graph.” (Wikipedia, 2011)

⁹Query languages are computer languages used to make queries into databases and information systems. Broadly, query languages can be classified according to whether they are database query languages or information retrieval query languages. The difference is that a database query language attempts to give factual answers to factual questions, while an information retrieval query language attempts to find documents containing information that is relevant to an area of inquiry.” (Wikipedia, 2011)

¹⁰“In computer science, and specifically the branches of knowledge engineering and artificial intelligence, an inference engine is a computer program that tries to derive answers from a knowledge base. It is the "brain" that expert systems use to reason about the information in the knowledge base for the ultimate purpose of formulating new conclusions. Inference engines are considered to be a special case of reasoning engines, which can use more general methods of reasoning.” (Wikipedia, 2011)

11

- “Pulling third party feeds onto BBC Online pages may be more editorially sensitive than linking externally, because users are more likely to expect the feeds to have been through a rigorous editorial examination and to comply with BBC editorial standards.
- Feeds and embeds require the approval of the relevant Interactive Executive.
- The person responsible for the page on which a feed or embed appears will be responsible for the content from that feed or embed.
- They will have to decide what level of editorial monitoring is appropriate and whether they have the necessary resources to carry it out.
- When we take in content from an external supplier, we should be transparent about this with our users.
- If one supplier is known to favour or advocate a particular view or position, we should normally take care to offer a reasonable range of views from different suppliers at the same time.
- We should be clear about what our users can do and where they should go if they see a mistake or find something offensive.
- It should be clear to a casual user that an inbound embed (for example, of video from a social networking site) is not under the editorial control of the BBC and is not primarily the responsibility of the BBC. But we are responsible for deciding what non-BBC content to embed on BBC pages and we should

be alert to the possibility of the original content being changed or removed and to rights issues.” (BBC, 2011)

•
¹²“A microformat (sometimes abbreviated μ F) is a web-based approach to semantic markup which seeks to re-use existing HTML/XHTML tags to convey metadata and other attributes in web pages and other contexts that support (X)HTML, such as RSS. This approach allows software to process information intended for end-users (such as contact information, geographic coordinates, calendar events, and the like) automatically.” (Wikipedia, 2011)

¹³

“The upshot of this is that there is no one single destination for users involved in a search for information or in some other information-related activity. A survey conducted by OCLC on user perceptions of libraries shows that the vast majority of those interviewed use search engines to begin an information search (84 percent). Only one percent begin an information search on a library Web site. (p. 6-3) Which doesn't mean that they are not interested in library services, but they may not discover those services if the library does not find a way to go to the user, rather than waiting for the user to come to the library.” (Coyle, 2006)

¹⁴

“In this way, Linked Data application developers would have access to an ever growing body of reusable components ready to be combined and exploited. Integrating services with the Web of Data as depicted before would give birth to a services ecosystem on top of Linked Data, whereby people would be able to collaboratively and incrementally construct complex systems by reusing the results of others, gradually supporting the development of increasingly complex applications for the Future Internet.” (Pedrinaci and Dominigue, 2010)

¹⁵

“Talk of Linked Data, Open Data, the Web of Data and related concepts in recent years have led to a quite remarkable shift in attitude amongst individuals, public bodies and private corporations. Almost everywhere my work takes me, clever people are seriously grappling with the implications of *consuming* from or *contributing* to these emerging ecosystems. Not all of their questions have good answers, and not all of the technological, strategic and business implications have necessarily been fully worked through. But these people are *asking* the questions, and they are asking them in all seriousness. That is a dramatic and welcome shift. ... What are we after? More Linked Data, or more RDF? I sincerely hope it's the former. ... So let's see loads more Linked Data, and plenty of evangelism as to why RDF could be the *best* way to do it. But let's not ostracise the vast majority of potential participants, contributors and beneficiaries in the world of Linked Data, just because they haven't wholeheartedly embraced RDF yet.”(Miller, 2009)

¹⁶“Diffusion of Innovations is a theory that seeks to explain how, why, and at what rate new ideas and technology spread through cultures.” (Wikipedia, 2011)

¹⁷“Useful references:

- Linked Data design issues - <http://www.w3.org/DesignIssues/LinkedData.html>
- Cool URIs for the semantic web - <http://www.w3.org/TR/cooluris/>
- How to publish Linked Data on the web - <http://www4.wiwiiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>
- Publishing and consuming Linked Data with RDFa - <http://ld2sd.deri.org/lod-ng-tutorial/>
- Also as an example of the process in designing and publishing Linked Data Jeni Tennison's blog is excellent. Particularly the series of posts that starts here: <http://www.jenitennison.com/blog/node/135y>“ (Sherratt, 2011)