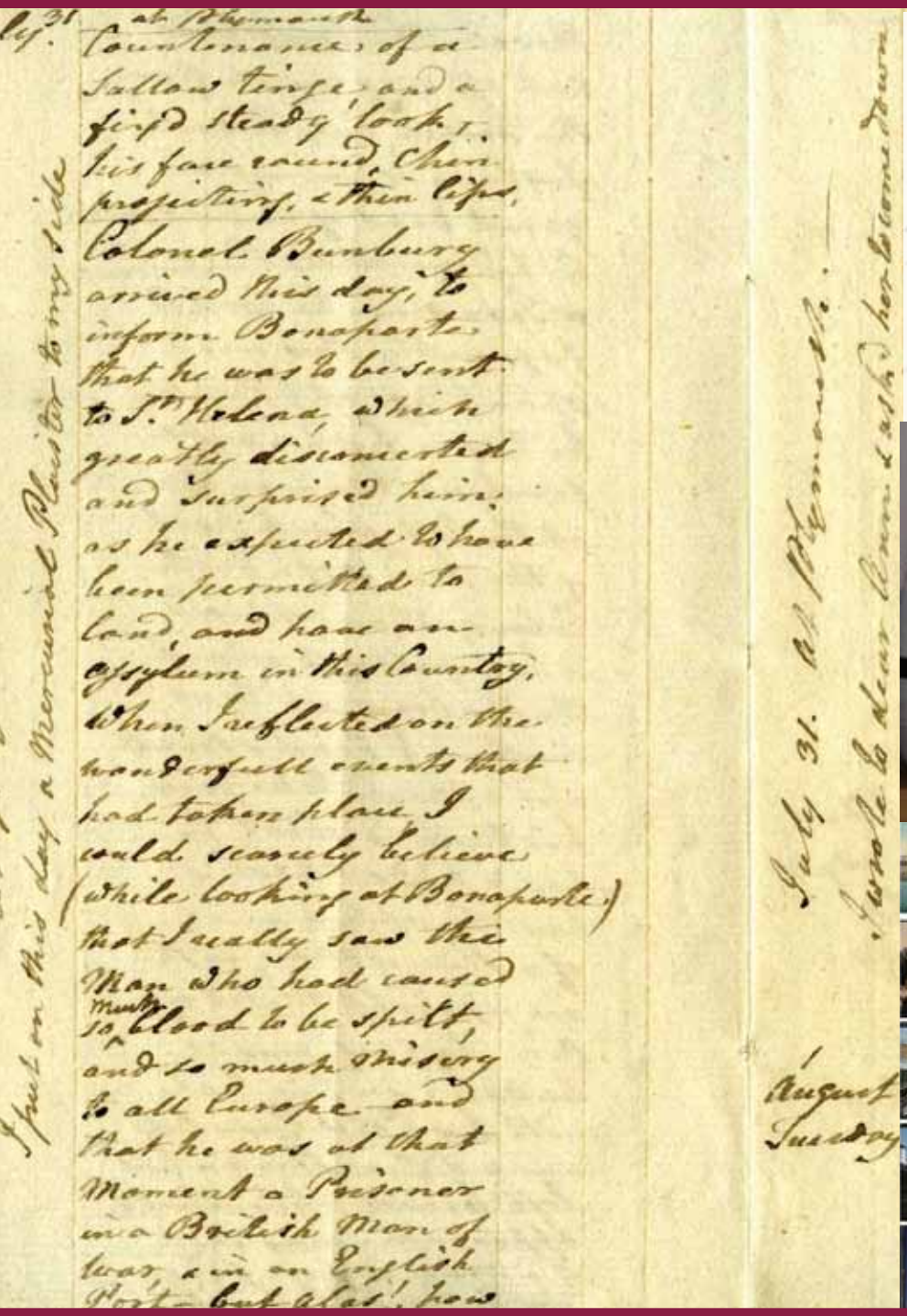
Archives & Records Association
June 2010
No: 250
ISSN: 1745-2120

ARC magazine

archives • records management • conservation



Data Standards Special Issue

Published by Society Media Sales Ltd

Linked Data and the Future of Archival Description

Much of the labour of archivists involves creating detailed descriptions of records and their creators that communicates context effectively. Archivists now have the opportunity to exploit archival description and expand the contextual ecosystem even further by applying a core component of the Semantic Web: Linked Data. Using a simple set of principles, archivists will be able to correlate and combine their descriptive data in new ways and with other data sources, as well as allow others to do the same. Furthermore, Linked Data will allow us to transform the highly implicit information contained in archival description into a highly structured, machine-readable representation of relationships.

The "Linked" Nature of Archives

The notion of "links," or cross-references, in archival description can be traced at least to the writings of Sir Hilary Jenkinson, if not further.[1] Nonetheless, it is the work of the Australians – specifically Peter Scott and Chris Hurley – that elucidated the highly relational aspects of archival description. In his 1966 article, "The Record Group Concept: A Case For Abandonment," Scott advocates for the separation of physical arrangement from administrative context.[2] Instead, he proposes recording relationships between "elements" (creators and records) as a set of "links" in sets of indexes. Hurley expanded upon the types of relationships in considerable detail to describe the interconnections between the diverse elements and entities involved in the records continuum.[3]

> "However, the true power of the Web is not contained solely in the resources available online, but rather in links between them."

Links and the World Wide Web

The World Wide Web is not new to archivists, and we certainly rely on the multitude of resources available on the World Wide Web. However, the true power of the Web is not contained solely in the resources available online, but rather in links between them. Sergey Brin and Larry Page recognized this early on, and incorporated this into the design of their prototype search engine that became Google.[4] As with the Web in general, archivists tend to take links for granted. In an average Web-based finding aid or register, users will usually find explicit links to further information about the repository that holds the records. In addition, archivists may in some cases link to descriptions of other related or complementary fonds, or to digitized material from the fonds itself.

Beyond these explicit kinds of links, we might find various kinds of "implicit" links as well. Implicit links can be subdivided into two categories. First, there are implicit hyperlinks, wherein the link in question leads to another location on the Web but may not represent an individual resource. An example of an implicit hyperlink would be a link to a "canned" search result in an online catalog system. On the other hand, there are implicit conceptual links, which occur without a corresponding hyperlink. These conceptual links are exemplified by the use of unlinked controlled vocabularies and authority files, or by references to related collections that do not link to other resources. Despite their presence, these implicit links are not crawlable or machine-actionable in the same way that explicit links are. Since archival description relies heavily on implicit conceptual links, the complex relationships between records and context cannot be readily processed or reused by computers.

Introducing Linked Data

One mechanism by which we can transform these implicit links to explicit ones is to implement the principles of Linked Data. As Dan Chudnov has said, "Linked Data is a way to link better."[5] Tim Berners-Lee, widely regarded as the inventor of the Web, has identified four key design principles for Linked Data.[6] Instead of relying on Web links alone, Linked Data requires that we provide

“it allows us to represent the relationships between those things rather than the documents that describe them.”

unambiguous, single names for real world “things” – people, organizations, fonds, and so forth – using Web addresses. Linked Data also requires that we describe these relationships in a structured way that can be understood by computers. This goes beyond regular Web links by providing a computer-parseable description of the relationship between two resources. For example, we may link often to an extended description of an organization from an online finding aid. Linked Data would allow us to represent the relationship between the organization and the records, be it as creator or collector. In addition, since Linked Data uses Web addresses to identify real world things, it allows us to represent the relationships between those things rather than the documents that describe them.

Future Directions

In 1994, Dan Cantrall argued that the “uniqueness” of archives was sufficient motivation to investigate delivering archival description via the Web. In his words, “[archival] information is too valuable and interesting to be left only to MARC-based electronic finding aids.”[7] Since we often spend a significant amount of time and energy on description, it makes sense that we should consider making it available in as many ways as possible. Archivists have several options to consider how to apply the principles of Linked Data. The easiest way is to make explicit links to other data sources. Linking to other data makes the work of archivists more valuable and visible. Finding aids and registers could link to controlled vocabularies available as Linked Data, such as the UK Archival Thesaurus[8] or Library of Congress Subject Headings[9], or to reference sources such as the Linked Data versions of Wikipedia[10] or Ordnance Survey data.[11] The number of available open data sources is steadily growing thanks to the work of the Linking Open Data project.[12]

Beyond linking to existing data sources, archivists can also provide descriptions of records and creators as Linked Data. Of course, there will be challenges in doing so, the largest of which appear to be technical. Archivists will have to familiarize themselves with supplemental metadata standards and models as well as the wide range of vocabularies and ontologies (i.e., specifications of relationships of entities). Overcoming these hurdles, however, will allow archivists as well as others to repurpose this data more easily, enrich the contextual relationships around records, and lead to future innovations in access systems involving visualization. Our description is much too valuable not to consider making it available as Linked Data.

Mark A. Matienzo

1. See Manual of Archive Administration (Oxford: Clarendon Press, 1922), p. 86.
2. “The Record Group Concept: A Case For Abandonment,” American Archivist 29, no. 4 (1966): 493-504.
3. See “Relationships in Records,” originally published in New Zealand Archivist, 2001-2004, http://www.infotech.monash.edu.au/research/groups/rcrg/publications/relationships-in-records-rev-3b.rtf.
4. See “The Anatomy of a Large-Scale Hypertextual Web Search Engine,” http://infolab.stanford.edu/~backrub/google.html.
5. “TCDL 2009 Talk: Better Living Through Linking,” http://onebiglibrary.net/story/tcdl-2009-talk-better-living-through-linking.
6. “Linked Data - Design Issues,” June 18, 2009, http://www.w3.org/DesignIssues/LinkedData.html.
7. “From MARC to Mosaic: Progressing towards Data Interchangeability at the Oregon State Archives,” Archives and Museum Informatics 8, no. 1 (1994), p. 4.
8. http://www.ukat.org.uk/
9. http://id.loc.gov/authorities/search/
10. http://dbpedia.org/About
11. http://data.ordnancesurvey.co.uk/
12. http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData