**WALLACE — Recording**

<2 tables; 1 figure>

**Recording Historical Connections in the Dictionary of Sydney**

Wallace, S.

This paper will outline the approach taken to encoding historical information in the Dictionary of Sydney project and, in particular, will assess five years of use of one of its core data structures—known to the project as ‘Factoids’.

**The Dictionary of Sydney (http://dictionaryofsydney.org/)**

The Dictionary of Sydney project arose from an initiative at the City of Sydney Council in 2004–5. A trust was set up to support and manage the establishment and operation of the Dictionary, and it was subsequently the subject of an Australian Research Council linkage grant from 2005 to 2010 involving the University of Sydney, the City of Sydney, the State Library of NSW, the State Archives of NSW, and the University of Technology Sydney (UTS). The author was project manager of the ARC project from 2006 to 2010 and has worked intermittently with the Dictionary since.

The Dictionary was ‘born digital’ and aimed to create a flexible repository of information on the history of the area now encompassed by ‘Greater Sydney’. A website, intended to be only the first of various digital products, went live in 2009. The Dictionary continues to operate with the backing primarily of the City of Sydney Council and some other institutional and philanthropic support.

**Text and Data Structures**

The use of the term ‘Dictionary’ was debated and somewhat at odds with similar initiatives more often titled ‘Encyclopaedias’ (Te Ara [n.d.]; Encyclopedia of Chicago [2005]). The Dictionary was always intended to include narrative entries of varying length and was by no means expected to be confined to a list of defined terms or to be simply a gazetteer of names and labels (as ‘Dictionary’ might have implied). However, the obvious potential for a strong set of terms and labels to link and aggregate digital resources (text, multimedia, and maps) encouraged an approach that integrated narrative text and structured data.

This gave rise initially to two distinct sets of terms encompassing Entities (‘things’) and more abstract Subjects. The emphasis in all these deliberations was on simplicity and ease of use rather than the final word in academic rigour. So, in assessing useful categories of urban Entities, listings on popular directory websites such as Sensis (2014) and even those at the back of street directories were consulted along with more scholarly efforts at categorisation. Hierarchies were kept shallow, with generally no more than three levels.

The final list of Entity types was Artefacts, Buildings, Events, Natural Features, Organisations, People, Places, and Structures. Digital resources were identified as Entries, Maps, and Multimedia, and the Dictionary’s set of components was completed with Subjects, Contributors, and Roles.

By the time of writing, at least 50,000 of these elements have been created (alongside over a million words in Entries), and several years of experience and practice has been built up.

**Linking Entities and Defining Roles—The ‘Factoids’**

The Entities provided a solid framework to which could be attached various digital resources. This ensured that the Dictionary was not ‘entry-centric’ and could have multiple Entries attached to a single Entity—important in ensuring that the Dictionary was *not* confined to a single interpretative voice but instead did justice to the many opinions and discourses that a complex urban history demands.

There remained requirements that Entities alone did not satisfy:

• Linking entities.

• Consistently recording location spatial and temporal extent.

• Recognising that the ‘role’ of Entities can change (e.g., a building used for different purposes over time).

• Handling different names (Mark Foys aka The Downing Centre, Pyrmont Bridge aka Anzac Bridge).

• Finer-grained categorisation where the eight basic Entity types were not adequate.

There was reluctance to define different data structures for different Entities, thereby leading to a complex and inflexible relational ‘database of everything’. It was also recognised that the Dictionary could not (and should not) take over the role of every specialist information source; for example, the Dictionary might record the location, designer, builder, and perhaps function of a bridge and key dates, but it could be left to an engineering heritage resource to record—for the true aficionado or expert—the materials used, the number of rivets, the exact dimensions and load-bearing characteristics, etc.

The recognition of the temporal element and the need to record change also significantly influenced the data structure used to handle these requirements. The key element for encoding change was the Role. Different Entities could take on different Roles over time—a Building might be a gaol, then a cultural centre, and then a residential block; a Place might be a quarry and then a park; a Structure might be a railway viaduct and then an elevated park. Roles could be created flexibly as needed (with due attention to duplication and ambiguity).

Roles could similarly describe relationships, most obviously between People (spouse, child, mentor) and between People and Organisations (member, CEO). A Person could be assigned a Role to indicate a profession—a doctor—or a position—a doctor *at* a hospital. In this sense, some relationships were transitive (requiring a source and target on either side of the Role), others intransitive (only requiring a source). Some relations are clearly directional—e.g., child-of—and some are symmetrical—e.g., spouse-of.

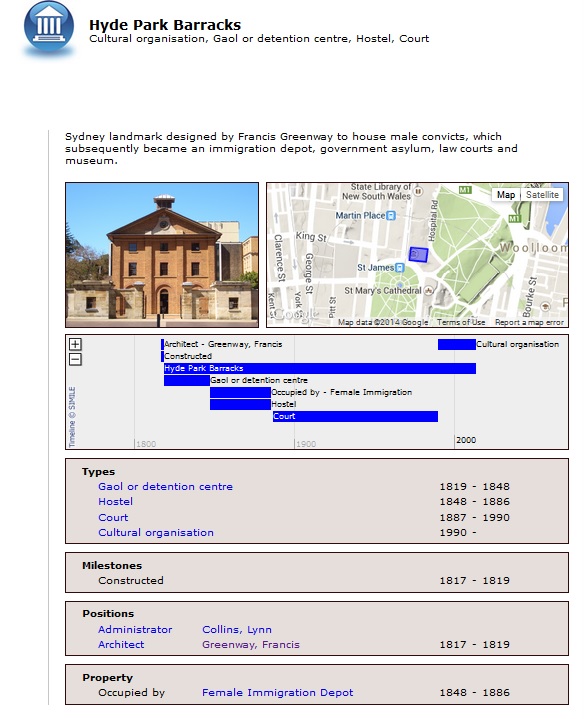
A single data structure called a ‘Factoid’ (an unsatisfactory term and only used internally) was fashioned to handle all these requirements. The structure was as follows:

|  |  |  |
| --- | --- | --- |
| *Field* | *Purpose* | *Example* |
| Source | An Entity | Francis Greenway |
| Role | Describes the relation or adds information | Architect |
| Target | A ‘target’ Entity | Hyde Park Barracks |
| Type | The type of Entity (see below) | Position |
| Target-text | A text substitute for the target Entity where it is not appropriate to create it as a full record |  |
| Start date | Date on which the Role commenced | 1819 |
| End date | Date on which the Role ended | 1822 |
| Location | A geo-spatial encoding (point, line, or polygon) |  |

The Type field was used to distinguish Factoids created for different purposes and to group them meaningfully. The types were as follows:

|  |  |
| --- | --- |
| *Type* | *Purpose* |
| Name | for recording different names that Entities had over time |
| Sub-type | used to subtype the main Entity types, e.g., Hostel as a sub-type of Building |
| Milestones | key events for an Entity—birth, death, construction, demolition, opening |
| Occupation | normally held by People, e.g., architect |
| Position | an occupation applied to a specific Entity, e.g., architect of Hyde Park Barracks |
| Relationships | most commonly between People—brother, spouse, mentor, etc. |
| Property | recording occupation or ownership |
| Time-Space | a special type of Factoid used only to record ‘existence’ and location, commonly used for Natural Features and Places |

Factoids can be seen on most Entity pages on the Dictionary website:

Figure 1. The Entity page for Hyde Park Barracks with four groups of Factoids.

**Factoids in Use**

The Factoid is one of many attempts to encode this sort of historical information (HEML [n.d.]; Vicodi [2002–4]; CIDOC-CRM), and it is reflected in other initiatives such as the Humanities Networked Infrastructure (Huni, 2014).

The Factoids have been in use now for over five years. This provides a valuable opportunity to assess their utility in a real-world, growing resource. This paper will report the results of an assessment undertaken with the Dictionary of Sydney staff in the first quarter of 2015, and an analysis of the use of the Factoids in the Dictionary and through a comparison with recent developments in historical encoding.

Areas to be assessed will include the following:

• Limits the expressivity of the Factoid structure—At what point does the information being encoded become too complex and need to revert to narrative description?

• Has the original structure of the Factoid been amended or new types added?

• Has custom and practice developed around the use of the Factoid that was not originally anticipated?

• Has the Factoid (as defined in the Dictionary of Sydney) been deployed in any other projects?

• Has the Factoid had advantages in processes not originally envisaged (e.g., fact/error-checking)?

• How has the divide between useability and rigour been navigated—Has the Dictionary’s position as a public (rather than academic) history project influenced this process?

• The Dictionary Factoids are not referenced or attributed—Has this affected their credibility or reliability?

• Who has been able to contribute Factoids effectively, and how has the process been managed and quality-controlled? Have amateurs contributed or only professional staff? (Reference may be made here to emerging models for managing ‘cyber-infrastructure’ such as that proposed by Lynch [2014]).

• What challenges and/or opportunities do the Factoids present for sustainability and preservation of the data they contain?

• Has it been possible to link the Factoids to other data standards?

• ‘Factoid’ was a term of convenience for the Dictionary but it has been used in other areas, such as prosopography—Has its use in other disciplines influenced its use in the Dictionary?

While it may not be possible to cover all these questions in depth, an assessment of the Factoids in practice may provide useful insights into the utility of data structures of this sort for digital humanities projects.

**References**

**Encyclopedia of Chicago.** (2005). http://www.encyclopedia.chicagohistory.org/.

**HEML.** (n.d.). http://heml.mta.ca/.

**Huni.** (2014). https://huni.net.au.

**Lynch, T. J.** (2014). Social Networks and Archival Context Project: A Case Study of Emerging Cyberinfrastructure. *Digital Humanities Quarterly,* **8**(3).

**Sensis.** (2014). http://www.sensis.com.au/.

**Te Ara—The Encyclopedia of New Zealand.** (n.d.). http://www.teara.govt.nz/en.

**Vicodi.** (2002–4). www.vicodi.org/.