

# Events as a Structuring Device in Biographical Mark-up and Metadata

by Michael Buckland and Michele Renee Ramos

## Foundations of Information Science

There is little structure or best practice in the concise biographical texts found in biographical dictionaries and *Who's Who* volumes. This paper is a progress report on an investigation of using events as a structuring device for mark-up and metadata structures in biographical texts as part of a project entitled *Bringing Lives to Light: Biography in Context* [1]. The idea is that anyone's life can be usefully decomposed into events at any desired level of granularity and that each event could be described as a 4-tuple of the four facets what, where, when and who.

### Purpose

The difference between seeing and understanding lies in knowing the context, and it should be emphasized that we approached this problem area from a particular perspective: helping readers to understand. This paper reports on one part of a series of studies of how learning can be facilitated by making it easier to find relatively trustworthy explanatory resources, suitable both for a text being read and for the reader. During 2004-2006 a project entitled "Support for the Learner: What, Where, When and Who" explored this area in general terms [2]. A four-facet "4W" approach was adopted – *what, where, when* and *who* – because each has distinctive characteristics leading to different genres of search aid and different display requirements.

*Where* involves a duality of place (a cultural construct) and space (a physical construct) and, for this, place name gazetteers and map displays are well-developed genres [3][4]. *When* similarly involves a duality of events and calendar time. Historical events are calibrated by calendar time, and

calendar time is calibrated by events such as solar years and cesium radiation cycles. In practice, people tend to mark time by mentioning personal and historical events (for example, "after I graduated," "during [the] Vietnam [war]") more than by calendar dates, so an approach similar to place name gazetteers using named time period directories and timelines can be adopted [5]. *What* tends to be a residual category when other specialized concepts have been removed. Here thesauri, subject indexes, library classifications and other tools are used, and ways to express and display relationships and cross-references are well developed.

*Who*, however, emerged as a relatively underdeveloped area. The disambiguation of personal names – associating multiple names for the same person and distinguishing different persons with the same name – is a well-understood area with its own standards. Also, genealogists understand family relationships and how to represent them in family trees, but other kinds of interpersonal relationships have not received the same careful attention, although a variety of limited examples can be found. Further, there seemed to be a distinct lack of accepted standards or best practices for structuring the very concentrated, stylized biographical texts found in biographical dictionaries. Our impression of the situation was validated and detailed in a report issued by the Text Encoding Initiative [6]. Fortunately we have been able to explore these issues as part of a project entitled "Bringing Lives to Light: Biography in Context" [1].

If order is to be brought to disorder, appropriate organizing principles are needed. Time is one obvious principle for biographical arrangement, but, by itself, it provides little beyond a useful sequencing. What additional approaches would be more useful for organizing the details of people's lives? Events shape our lives, and we engage in actions and activities, so combining actions and activities with other kinds of events seemed a promising analytical device for organizing the untidiness of peoples' lives.

Michael Buckland is co-director of the Electronic Cultural Atlas Initiative, University of California, Berkeley. He can be reached by email at [buckland@at.ischool.berkeley.edu](mailto:buckland@at.ischool.berkeley.edu). Michele Ramos is a research assistant at the Initiative. She can be reached by email at [michelerenee@berkeley.edu](mailto:michelerenee@berkeley.edu).

### About Events

We use *event* in an everyday sense. The most basic characteristic of an event is that something happens. There is activity, some change; otherwise we would say it was a “non-event.” Change happens through time. An event may be very brief, seeming instantaneous, or it might be very prolonged, but all events occur in time and require at least some time to happen. Because an event is presumed to involve some change, events cannot be preserved, only described, represented or re-enacted. For this reason and also because people see and define events differently, we treat events as narrative constructs.

Events happen in some place. It could be everywhere (like the Big Bang), the location might be unknown or it could be an imaginary place, but an event that happens nowhere would be a non-event.

When considering events, one quickly recognizes a duality between state and action. For example, one could say that one spent the year 2008 *studying* (an activity), or one could say that one was a *student* during 2008, a status. These are alternative, comparable autobiographical statements that appear to be different but are equivalent for most purposes. (Strictly speaking, of course, being officially a student does not in itself guarantee that much studying is being done and, vice versa, one can study without officially being a student, although in a broader sense anyone who studies is a student.)

There is a corresponding duality with respect to defining what the event is. It could be the prolonged activity of studying, or it could be the transition to (or from) student status. That these alternative perspectives are different does not matter much because they are complementary, mirror images of the same life experience, and one can choose either approach to suit one’s need. We have preferred to adopt the activity as the event, rather than the change of state, believing it to be more useful unit for description and analysis for our purpose.

### A 4W Structure for Representing Events

We have identified activity, time and location as salient characteristics of events. Agents, human or other, are also commonly involved. In biographical narratives the biographee is ordinarily the primary agent, or at least a participant, but others may also be involved, starting with the mother at birth. It is reasonable to expect there to be values for each of what, where, when

and who when describing or representing any biographical event even though our knowledge of the details of each might sometimes be lacking. What emerges is a case for using the 4W facets – what, where, when and who.

The *what* facet initially proved to be problematic. There is a tendency to equate “what” with the grammatical object of any narrative sentence, but so long as we are concerned with activities, *what* should be more associated with verbs, and the solution is to reserve *what* for *what kind of activity*, or happening or change. If from 1940 through 1944 he grew potatoes, potatoes were what was produced, an outcome, but the activity was a kind of agriculture.

### The Granularity of Events

If we adopt “event” as the unit of analysis and representation for biographical narratives, how detailed an approach is desirable? This question takes two forms. First, should one be concerned with only large or also small events? Our answer is that so long as an event-based approach proves effective, it should be effective for any granularity – any level of detail – so the choice of how far to extend the treatment to smaller events should depend on the purpose of the analysis. Second, as with any form of mark-up, there is a tendency to want to be complete in the sense of including more and more detail. But for our purposes, this temptation should be to strongly resist because the more detailed the mark-up and the more encyclopedic the description, the greater the cost and the less the likelihood that the work would ever be completed. The mark-up itself (and its structure) can be treated as a convenient abstraction of the narrative text (and its structure), but that quality should not imply that it is a replacement. For detailed questions, reference should be made to the text itself.

### Collective Biography and Prosopography

A biography is ordinarily a narrative of a single person’s life, although other people will be mentioned. An account of the lives of several persons is usually called a *collective biography* and a volume of biographical descriptions a *biographical dictionary*. The word *prosopography* originally meant the description of the features of a single individual, but its meaning has been extended to the description of a defined group of people sharing some

significant cultural characteristic and, especially, the relationships among them and the attributes they have in common. Nowadays, a prosopography is ordinarily a database describing the same attributes of each of a set of people in relatively standardized terms. In this way, prosopographies are in the mainstream of humanities computing: the creation of a corpus and of methods to look into it to analyze in various ways the content of the corpus.

Our own interests are significantly different. We are primarily interested in making texts, and especially biographical dictionaries, easier to understand. Since the difference between *seeing* and *understanding* lies in knowing the background and context of whatever it is we see, our concern is in facilitating convenient access to relatively trustworthy explanatory resources. So instead of the inward-looking perspective characteristic of prosopographical studies, our interest is in looking outwards from any fragment of any description of any person to any accessible trustworthy explanatory resource outside of the text.

### Terminological Issues

Any description needs vocabulary. The nature of our interests has two consequences for choices of vocabulary. First, of course, the terminology has to be suitably descriptive of the 4W facets. As already noted, the naming of persons, places and time has received attention, so our interest has mainly been on *what*, what kind of event.

Our interest in linking outwards to external explanatory resources suggests that we should draw on a widely used external resource such as the *Library of Congress Subject Headings (LCSH)* [7] or at least a vocabulary that will map to it. Our rather limited initial examination of biographical records suggests that *LCSH* could be used. The attraction, of course, is that *LCSH*, used in library catalogs and some other bibliographies, takes one immediately to the available scholarly literature on whatever the activity is. Also *LCSH* is more likely to have (or acquire) interoperable mappings to other widely used vocabularies than a locally created one. Another possibility is to use article names from *Wikipedia* [8], which are organized in a thesaurus-like structure, are more comprehensive and up-to-date than the *LCSH* and are also being mapped to a variety of other vocabularies.

Locally developed vocabulary may be necessary for specialized texts

where fine distinctions and technical terms matter. Specialists use and need specialized terminology. With specialized vocabularies and, indeed, with more general ones, the need arises to map the terms used locally to the corresponding terms used in catalogs and other scholarly resources. This problem resembles the need to map from the geographical description codes (aka feature types like castle, lake, inhabited place, airport) used in place name gazetteers to *LCSH* in order to connect objects on the ground to literature concerning that kind of object. In this case we have found that comparison of the National Geo-Intelligence Agency's Geographical Description Codes [9] with *Library of Congress Subject Headings* reveals differences in style and in emphasis, as well as scope and scale, with some 600 NGA GDC codes to over 150,000 Library of Congress subject headings. Sometimes *LCSH* has greater detail, especially for kinds of historic sites; sometimes NGA has more detail (for example, in submarine geomorphology), but, in general, they match quite well [4, pp 380-381]. There is a need here for the adoption of search-term recommender services when moving to or between vocabularies.

Whatever vocabulary is used for the mark-up of biographical texts, if, as in our case, the intention is to connect fragments of the text to external explanatory resources, it is increasingly important that the vocabulary chosen lends itself to easy interoperability with the proliferation of naming services associated with the Semantic Web.

### Conclusion and Unresolved Issues

This paper is a progress report. So far we have experimented primarily with event-based mark-up of concise biographical texts created by archivists and by editors of scholarly texts. The former have been <bioghist> fields found in Electronic Archival Description-compliant descriptions in the UK *Archives Hub*, a repository of records for British archival collections [10]. The latter have been brief biographical records created as part of *The Emma Goldman Papers* project at the University of California, Berkeley [11].

We assume that RDF (Resource Description Framework) and OWL (Web Ontology Language) specifications should be used in order to achieve maximum interoperability in the emerging Semantic Web environment.

Interoperability is especially important since we are trying to relate words, names and phrases in biographical texts to external explanatory resources. A report on our preliminary vocabulary development will be published separately.

A number of questions and unresolved issues remain:

1. Institutions play a large role in archival and biographical texts, so the “biography” of institutions needs comparable attention.
2. We have found a need to distinguish between personal biographic events and contextual events. The Great Depression and the Second World War were major events that affected the environment of personal lives, but they were not major personal events in the same way as, for example, getting married is. These differences regarding life events and cultural-historic events are generally addressed in terms of mereological (part-to-whole) relationships in most event ontologies. Although in some instances biographic events can have a part-to-whole relation to larger contextual events, in many cases they may not. For instance, a person can get married during the Second World War, but it does not follow that the marriage was a part of the war. We see a need for a vocabulary of biographic event types that is distinguishable from the event vocabularies specified in other ontologies.
3. Events are often related to each other. To make these links explicit would imply explicit enumeration of events within or external to the mark-up.
4. The terminology used in the mark-up of biographical texts is likely to vary from the wording in the text being marked-up. Retaining or associating fragments of the original text along with the mark-up terminology could provide a training set for programs to provide computer-aided mark-up.

More experience is needed, especially in mapping between different vocabularies and in the varieties of uses of event-based mark-up, but the provisional conclusion is that an event-based 4W form of mark-up will work as intended. ■

*ACKNOWLEDGMENTS* – We are grateful for support from the Institute of Museum and Library Services for Award LG-06-06-0037-06 “Bringing Lives to Light: Biography in Context” and from the Advancing Knowledge grant PK-50027-07 “Context and Relationships: Ireland and Irish Studies” jointly funded by the National Endowment for the Humanities and the Institute of Museum and Library Services. We benefited from initial analyses by Ms. Aurélie Bénard of Paris.

### Resources Cited in the Article

- [1] *Bringing Lives to Light: Biography in Context*. [Project website]: <http://ecai.org/ims2006>.
- [2] *Support for the learner: What, where, when, and who* [Project website]: <http://ecai.org/ims2004>.
- [3] Hill, L. L. (2006). *Georeferencing: The geographic associations of information*. Cambridge, MA: MIT Press.
- [4] Buckland, M. Chen, A. Fredric C. G., Larson, R. R., Mostern, R., & Petras, V. (2007). Geographic search: Catalogs, gazetteers, and maps. *College & Research Libraries*, 68(5), 376-387. Retrieved October 19, 2009, from [www.ala.org/ala/mgrps/divs/acrl/publications/crljournal/2007/sep/Buckland07.pdf](http://www.ala.org/ala/mgrps/divs/acrl/publications/crljournal/2007/sep/Buckland07.pdf)
- [5] Petras, V., Larson, R. & Buckland, M. (2006). Time period directories: A metadata infrastructure for placing events in temporal and geographic context. In *Opening Information Horizons: Joint Conference on Digital Libraries (JCDL)*, Chapel Hill, NC, June 11-15, 2006. Retrieved October 19, 2009, from <http://metadata.sims.berkeley.edu/tpdJCDL06.pdf>
- [6] Wedervang-Jensen, E., & Driscoll, M. (2006, February 16). *Report on XML mark-up of biographical and prosopographical data [PERSW02]*. TEI. Retrieved October 19, 2009, from [www.tei-c.org.uk/Activities/PERS/persw02.xml?style=printable](http://www.tei-c.org.uk/Activities/PERS/persw02.xml?style=printable)
- [7] Library of Congress. *Library of Congress Subject Headings*. Information about editions and formats available at [www.loc.gov/cds/lcsh.html](http://www.loc.gov/cds/lcsh.html). Online search of LC subject headings available at <http://authorities.loc.gov/>
- [8] *Wikipedia*: [www.wikipedia.org](http://www.wikipedia.org)
- [9] *NGA GEOnet Names Server (GNS)*: <http://earth-info.nga.mil/gns/html/index.html>
- [10] *Archives Hub*: [www.archiveshub.ac.uk/](http://www.archiveshub.ac.uk/)
- [11] *The Emma Goldman Papers*: <http://sunsite.berkeley.edu/Goldman/>