Geography, Time, and the Representation of Cultural Change: Experience from a Large Collaboration: The Electronic Cultural Atlas Initiatitive

Michael Buckland
University of California, Berkeley.

INTRODUCTION

The Electronic Cultural Atlas Initiative." The Electronic Cultural Atlas Initiative (ECAI, <ecai.org>) is a very informal collaboration among many scholars in several different countries, with librarians, information technology specialists, and others. Administratively, it is based at the University of California at Berkeley. The mission of ECAI is to transform education and research through better understanding of place, time, and culture. We are not creating an atlas. We are developing methods, recommending "best practices," and encouraging collaborations so that everybody can make more effective use of cultural resources by making them more accessible and more interoperable. As better and more interoperable techniques are adopted internationally, a collective atlas will, in effect, begin to emerge.

Recent work has been based on five ideas:

1. Understanding requires knowing the context. Nothing has meaning in isolation.

2. Finding and using Internet resources should be like using reference books in a library and as easy and as reliable.

3. A design goal: Find the context of any museum object, document, or performance: What is related to it in what it is, where it came from, when it originated, and who is associated with it?

4. WHAT, WHERE, WHEN and WHO (the four Ws) provide a useful structure.

5. Make better use of *existing* descriptive metadata.

Resources are not available unless somebody collects them. Collected resources are not accessible unless they are cataloged and two problems arise: First, catalogs (and indexes, ontologies, thesauri, etc.) differ, so the vocabulary used in any of them is likely to be unfamiliar. Seaches in unfamiliar vocabularies tend to be less efficient and less reliable. Second, the facets WHAT, WHERE, WHEN and WHO are different from each other and need different methods, different vocabularies, and different forms of display.

WHAT, WHERE, WHEN and WHO

WHAT

If you are interested in KUNG FU MOVIES, in the Library of Congress Subject Headings you must ask for MARTIAL ARTS FILMS. Previously you had to ask for HAND-TO-HAND FIGHTING, ORIENTAL, IN MOTION PICTURES. Description is a language activity, and there are many different languages and dialects. Here are the descriptive terms for *automobile* in four different dialects used in U.S. federal government databases:

 -- PASS MOT VEH, SPARK IGN ENG in the U.S. Import/Export statistics is an abbreviated version of "Passenger motor vehicle, spark ignition engine."
 -- TL 205 (Library of Congress Classification).
 -- 180/190 (US Patent classification).
 -- 3711 (Standard Industrial Clasification).

Even when a natural language is used it is often difficult to imagine which word will have been used and the meaning can still be unclear even when the proper term has been found. Consider, for example, from the International Harmnized Commodity Classification System: "*HS 847120* Digital auto data proc mach contng in the same housing a CPU and input & output device." It does not say so, but it means *computer*.

People understand the need for cross-references *within* a vocabulary, but a network environment raises the far larger problem of cross-references *between* different vocabularies. There is enthusiasm for a "Semantic Web." These examples illustrate the reality of the semantics of the Web.

WHERE

Special, separate treatment is needed for WHERE. *Place* is a cultural construct and *space* is a physical construct. A place name gazetteer, a kind of bilingual dictionary between place and space, is needed. A place name gazetteer gives the name of a place, indicates what kind of place it is (geographical description code or "feature type"), latitude and longitude, and when that name was used. Latitude and longitude are very important. They relate place to space and, thereby, enable us to show the locations of places on maps and to see how places are related to each other spatially. WHAT and WHERE can be related to each other when a place name gazetteer and a catalog talk to each other, and, with latitude and longitude from the gazetteer, the catalog can have a map display.

WHAT is WHERE? As one example, a search in a library catalog for books on *Folklore* may retrieve many books. Books on *Folklore* are often about folklore in some place. The place names in the catalog records can be passed to a place name gazetteer which can supply the latitude and longitude, so that the interface can display a geographical analysis on a map. One should be able to click on any location to find which books are about folklore in that place (Buckland et al., 2007).

WHERE is WHAT? In the other direction, if you are interested in, say, mass transit in the capital cities in South America, one might use the cursor to draw a box on the map around the area of interest, then limit the gazetteer feature type to capital cities. The gazetteer should then list the names of the capital cities in South America, ready for searching in the catalog.

WHEN

When people speak or write about time they often do not use calendar dates. Instead, they mark time by mentioning events or named time periods, e.g. "In reign of Charlemagne…," "during the Weimar Republic," "after Vietnam," or "when I was a student." This practice is

culturally situated and it resembles the use of place names, so we designed a "Named Time Period Directory," like a place name gazetteer, and created an example using 2,000 records derived from the chronological subdivisions used in Library of Congress Subject Headings.

 -- A *place name gazetteer* has: a place name – type of place – latitude & longitude – and when the name was used.

 -- Our *named period directory* has a period name – type of period – calendar time – and where it happened.

Sample entries for the 13th century, organized by country, include:

  -- China: Yuan dynasty, 1260-1368.

  -- Czech Republic: Mongol invasion, 1242.

  -- Denmark: Erik Glipping, 1259-1286.

Who was Erik Glipping? What did he do? To find explanations and context our interface provides icons for links. Clicking on a small blue book generates a live search of the Library of Congress catalog, a "federated search" using the Z39.50 protocol. A live search retrieves an up-to-date set of references, as up-to-date as the library's cataloging. Another symbol sends a query to the Wikipedia. The Wikipedia uses structured URLs for biographical articles, so one can program the interface to go directly to the right page.

WHO

The Wikipedia biography of King Eric V (alias Erik Glipping and Eric Klipping) provides a concise text beginning:

Eric V "Klipping" (1249- November 22, 1286) was King of Denmark 1259-1286) and son of Christopher I. Until 1264 he ruled under the auspices of his mother, the competent Queen Dowager Margaret Sambiria. In this period he was for some time 1261-1262 a prisoner in Holstein after a military defeat . . .

This is helpful, but it may mean little if you do not know the context. Where is Holstein? Who was Margaret Sambiria and what did she do? Fortunately the Wikipedia has several links (where underlined) to additional related material in other Wikipedia pages. These are internal links. For "knowledge by networking" we need *external* links, dynamic links that will send searches for the latest versions of the best resources on the Internet.

As another example closer to the theme of this conference, consider what a biographical record for the designer of first desktop search engine would look like:

Emanuel Goldberg, b. Moscow, 1881; son of Grigorii Goldberg; Univ. of Moscow, 1900-04; Ph.D w. Robert Luther, Leipzig Univ., 1906; Assistant, Adolf Miethe, TU Charlottenburg, 1906-07; m. Sophie Posniak, 1907; Prof, Akad. f. graphische Künste, Leipzig, 1907-17; ICA, Zeiss Ikon, Dresden, 1917-1933; Kinamo cine camera, 1921; microdots, 1925; search engine, 1927; Contax 35 mm camera 1932; kidnapped by Nazi SA; refugee in Paris, 1933-37; Laboratory, Tel Aviv, Palestine, 1937; d. 1970.

Several people important in Goldberg's life are mentioned: Goldberg himself, his father, his dissertation advisor, Luther; his first employer, Miethe; and his wife. Who were they? What did they look like? What did they do? In the spirit of "knowledge by network" one would want links

for each name leading to a portrait and a biography wherever they could be found. Various places are named: Moscow, Leipzig, Charlottenburg, Dresden, Paris, and Tel Aviv. Where are these places? What were they like at the time when Goldberg was there? His "life-path" could be displayed as timed movements on a map. Achievements are mentioned: the Kinamo cine camera, microdots, the search engine, and the Contax 35 mm camera. Where are descriptions of these inventions? Do patents exist? Techniques for marking-up biographical texts to link significant names and words to external explanatory resourcers are under development in a new project, "Bringing lives to light: Biography in context," led by by Professor Ray Larson (*Bringing* 2007).

Technical comment: In practice, descriptive metadata records are often complex. A library subject heading may have a geographic element and a chronological element, e.g. Architecture -- Japan -- Edo period, 1600-1868. Records in a place name gazetteer specify the type of feature as well as the name and location. This offers interesting possibilities for build links between the components of different kinds of records. Metadata serve two purposes. The first role of metadata is descriptive, to categorize a document; but, collectively, metadata form structures that can be navigated during search. In this second role descriptive metadata constitutes important infrastructure. These ideas are summarized in Figure 1.
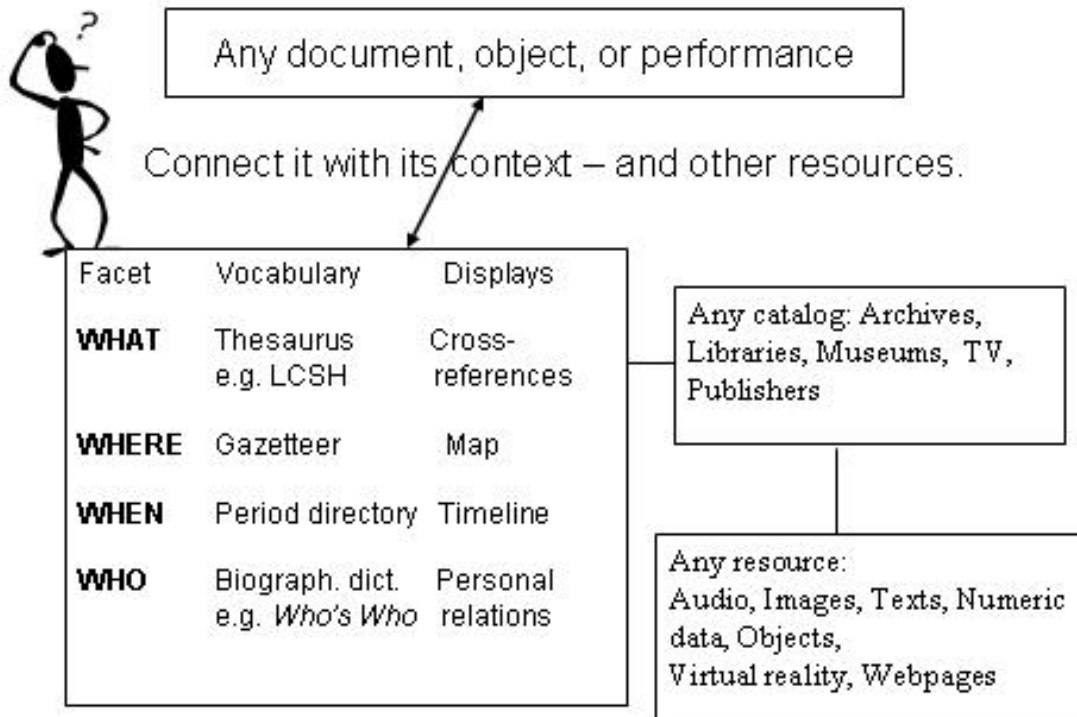

Figure 1: Connecting documents with their contexts.

AN EXPERIMENTAL 4W SEARCH INTERFACE

The needs reviewed above have been addressed in an experimental interface developed two years ago. (This description is based on the powerpoint slides presented at the conference <http://www.knowbynet.de/de/download/contibutions/buckland_kbn.ppt>. The interface has subsequently been redesigned but retains the same basic functionality at http://ecai.org/imls4w/.

To overcome the unfamiliarity of vocabulary, interfaces need search term recommender systems. Suppose we see the phrase DJENGIS KHAN and become curious: What is this? -- or who is this? We can ask the interface for a ranked list of Library of Congress Subject Headings related to *Djengis Khan*. This search term recommender system was derived by using statistical techniques to calculate the relative frequency with which words and phrases found in book titles are associated with subject headings assigned to those titles (Buckland et al. 2006)

The recommender service says that DJENGIS KHAN is not an established subject heading, but GENGHIS KHAN, a variant spelling, is. The recommender service lists ten subject headings that are statistically associated with DJENGIS KHAN. If you select one of the ten headings, the interface can send a search to several resources. In this example four of the recommended headings are for persons somehow associated with Genghis Khan and one is an apparently related period: The tenth heading says MOGUL EMPIRE. What is this? Is MOGUL the same as MONGOL or different? We can ask to see how MOGUL EMPIRE fits in the hierarchical structure of the *Library of Congress Subject Headings*, which shows that the MOGUL EMPIRE was part of the history of India in the 16th to 18th centuries, and that MOGUL is also spelled MUGHAL and MOGHAL, but not MONGOL. It is different. We could click on search links to find more. A live search of the Library of Congress catalog found 775 books about the *Mogul Empire*, the first written in Gujarati. Google lists a million links and the *Wikipedia* link leads to a simple, convenient introduction.

The Wikipedia article on the *Mughal Empire* explains that it was in Hindustan which included most of India. To look for resources on India in this period we go back to the interface home page where we have four options: (1) We can browse by region, looking for Asia and then South Asia; (2) We can choose a location from a list of states, a list of cities, or a list of countries; (3) We can browse by time period; or (4) We can use a map.

On the map, we zoom to south Asia using the TimeMap software developed by colleagues at the University of Sydney, Australia (*Time*M*ap* 2000), which also allows us to set the time period that is of interest. We select the desired period on the time bar, zoom to south Asia, click on India, and the database underlying the interface creates a menu for 15th century India in two parts. The top part has links for India generally, not limited to the 15th century, and currently offers general information about India from the *Wikipedia*, the CIA's World Factbook, the BBC News Country Profiles, and the Ethnologue Languages of the World website, where we could find out about Gujarati. Also a natural history museum website offers details of the animals and plants of India. Note that in each case the interface generates the searches that go directly to the section on India, not just to the home page of website, but the India page. The rest of the page provides links to resources on India in the 15th century. For example, a significant person was Babur, for which the Wikipedia has a biography.

The Metropolitan Museum in New York has a wonderful website at www.metmuseum.org. This website uses structured URLs in its Timeline of Art History ("toah") section at http://www.metmuseum.org/toah/. The structured URLs use a simple code so an interface can know and link to the correct URL in the timeline for any given time and place. It is so simple, so easy, so hospitable, and so effective. It is a wonderful design. All museum websites

should use this technique! By using a simple table mapping our interface categories to metmuseum toah codes and inserting the appropriate codes into the Met's URL template, the interface can send the searcher directly to the best page in the Metropolitan website, in this case the *South Asia, 1400–1600 A.D.* portion of their timeline, which shows the choices, with MUGHAL DYNASTY 1526-1858 in the top on the right. Our interface page also links an architecture website which can be searched by place and time.

Also listed are relevant cultural atlases created with or by ECAI affiliates. One is about the ivories of Begram (in present-day Afghanistan, but pre-dating the Moguls); another includes a video of the changing boundaries of the Mogul Empire; a third is about the Silk Road; and a fourth includes a video of the dramatically changing boundaries of the Mongol Empire from 110 – 1400 A.D., created ten years ago by colleagues at the University of Sydney and available at <http://ecai.org/Area/images/mongol.avi>. The red and white lines on it show the different routes of the Silk Road and the yellow line is the changing boundary of the Mongol Empire. We can observe the geo-political changes as Genghis Khan unites the Mongols, the Mongols conquer the Tartars, then attack and destroy the Chin Empire, conquer Russia and Korea, strike into Europe, and so on. When technology has advanced, one should, in the future, be able to stop the video at any year, choose any point, and say, "Find knowledge on the Internet about *this place* at *this time*!"

Understanding means knowing context. For this we need to develop geo-*temporal* systems, since geo-spatial is not enough, we need search term recommender systems to help deal with unfamiliar vocabularies, and dynamic links to well-maintained resources to find the most up-to-date information.

References
The Electronic Cultural Atlas Initiative website <ecai.org> provides extensive resources on ECAI activities. For more on the work described above see <ecai.org/imls2004>.

Bringing lives to light: Biography in context. [Project website] http://ecai.org/imls2006 visited
        Aug 29, 2007.
Buckland, M., A. Chen, F. C. Gey & R. R. Larson. 2006. Search Across Different Media:
        Numeric Data Sets and Text Files. *Information Technology and Libraries* 25, no 4 (Dec
        2006): 181-189. Also at
        http://www.lita.org/ala/lita/litapublications/ital/252006/number4decembera/buckland.pdf
Buckland, M. and others. 2007. Geographic search: catalogs, gazetteers, and maps. *College &
        Research Libraries*. Forthcoming Sept 2007, also at
        http://www.ala.org/ala/acrl/acrlpubs/crljournal/collegeresearch.cfm.
TimeMap: Time-based Interactive Mapping. [Website]. 2000. http://www.timemap.net/ Visited
        Aug 29, 2007.