Time Period Directories: A Metadata Infrastructure for Placing Events in Temporal and Geographic Context

Vivien Petras
School of Information Management and Systems
University of California, Berkeley
Berkeley, California, USA, 94720-4600
vivienp@sims.berkeley.edu

Ray R. Larson
School of Information Management and Systems
University of California, Berkeley
Berkeley, California, USA, 94720-4600
ray@sims.berkeley.edu

Michael Buckland
School of Information Management and Systems
University of California, Berkeley
Berkeley, California, USA, 94720-4600
buckland@sims.berkeley.edu

ABSTRACT

Metadata is ordinarily used to describe documents, but it can also constitute a form of infrastructure for access to networked resources and for traversal of those resources. One problematic area for access to digital library resources has been the search for time periods or events. If there is a capability to search for time, it is usually a date search - a standardized and precise form but unfortunately rarely used in common chronological expressions. For example, a user interested in the "Vietnam war", "Clinton Administration" or the "Elizabethan Period" must either know the corresponding dates, or rely on simple keyword matching for those period names. We consider the ability to interpret user statements of periods or eras as ranges of dates and to associate them with particular locations an important feature of an information system. This paper describes the Time Period Directory, a metadata infrastructure for named time periods linking them with their geographic location as well as a canonical time period range.

Categories and Subject Descriptors

H.3.1 [**Information Systems**]: Content Analysis and Indexing—*Indexing Methods, Thesauruses*
; H.3.7 [**Information Systems**]: Digital Libraries—*Systems issues, Standards*

General Terms: Design, Standardization, Experimentation

Keywords: Time Period Directories, Metadata Infrastructure

1. INTRODUCTION

Historical research of any kind is commonly organized by chronological periods and events marking a point of change. Especially in the humanities it has been found that scholars search in three major categories: biography (person), chronology (period or event) and geography (place). In a 2-year study of humanities scholars' search in online databases at the Getty Art History Information Program Marcia Bates found that:

> "Humanities scholars searched for far more named individuals, geographical terms, chronological terms, and discipline terms than was the case in a comparative science sample." [2]

In her analysis of 150 queries, 49% were searches for an individual as subject, 25% for a geographical name, and 17% for a chronological term[2]. Donald Case, in his study about the search habits of twenty American history scholars writes:

> "Of particular interest in history is the dimension of time; were it not for the temporal dimension, most topics in history would be the same as those covered by other fields within the social sciences and humanities". [3]

Helen Tibbo, in her dissertation on abstracts and searching in the humanities, points out that historians put a particular emphasis on the chronological and geographical dimension. When she asked historians which kind of materials they most likely wanted to see in abstracts of historical materials, 100% of respondents wanted to see (a) specific dates and time span indicators and (b) names of geopolitical units. Slightly less important were (c) names of individuals and/or groups (96%) and interestingly, (d) the main topic or subject of work (92%) was only ranked fourth[11]. This evidence led Bates to conclude that time should be made available as a search aspect:

> "Dates should be considered another category of term used in thesauri, and thesauri should instruct indexers in how to represent chronological terms effectively. Only in that way will good date indexing make effective online retrieval possible." [2]

Rarely, however, does one find an organized way to search for time periods in standard information systems like library

JCDL '06 June 11–15, 2006, Chapel Hill, NC, USA.

catalogs and bibliographical databases. If one finds the capability to search for time, it is usually expressed in dates - a standardized and precise form of expression but unfortunately rather seldom used in common language.

It is much more common to speak of important historical time spans in terms of named time periods like the Renaissance, Fin de Siècle or the Cold War. These named time periods, however, are not "merely convenient collections of years"[3], they have social, political and cultural context and, as a category of research, they are contested:

> "As its simplest, we can say that a period is the outcome of the need to relate generic continuity to temporal discontinuity. [...] Period, as we have said, provides that context, but it can only do so according to some organizing principle which must be the work of the historian". [9]

In this paper, we will describe our research on constructing relational linkages between named time periods, date ranges and their associated localities in order to provide effective chronological access to information in a digital library search environment. We present a proposed content standard to describe records in these "Time Period Directories", which organize named time periods or events, their associated dates or date ranges and locations in one structure, similar to the way that gazetteers relate place names to latitude and longitude coordinates. We have implemented a prototype Time Period Directory with named time periods and events harvested from Library of Congress subject headings and constructed a web interface to demonstrate different access and search strategies for potential users.

This work stems from a vision of future online information resources where the user can easily, and effectively, explore the historical context and interconnections of people, topics, location, and events. We expect the results of interaction with such systems to yield a rich dynamic portal of interconnected resources with maps, biographies, timelines and chronologies, in place of flat lists of web pages or bibliographic records.

The remainder of this paper is organized as follows: in section 2, we provide a brief description of how the Library of Congress describes time in its subject headings and in section 3 we provide a short overview of two statistical studies analyzing the distribution of chronological subdivisions in subject headings in a sample of authority records and a sample of real catalog records. In section 4, we introduce the Time Period Directory Content Standard, which is an XML schema on how to structure Time Period Directory records. Section 5 describes the procedure of harvesting chronological subdivisions from LOC subject heading authority records and their transformation into Time Period Directory instances. A description of the web interface implementation of the prototype Time Period Directory is provided in section 6. Finally, in section 7 we present our perspectives on future research and discuss the improvements that might be made to the existing implementation.

2. LCSH CHRONOLOGICAL ELEMENTS

The Library of Congress Subject Cataloging Manual states under instruction sheet H620 (Chronological Headings and Subdivisions: Methods of expressing chronological focus in subject headings) that Library of Congress Subject Headings (LCSH) provide five ways to express time and time periods. These are:

1. General headings to express time periods not limited by subject (e.g. Renaissance).
2. Headings with inherent or implied chronological concepts limited to a specific time period due to nature of the subject expressed, such as historical movements, artistic styles, technological development (e.g. Post-communism).
3. Headings with an adjectival qualifier (e.g. Greek drama, Modern).
4. Headings with a date qualifier (e.g. Culpeper's Rebellion, 1677-1679).
5. Chronological subdivision ($y).

Except for the chronological subdivisions where events or time periods can be automatically detected by harvesting the $y subdivision in catalog or authority records, the first four cases of time expressions occur in the MARC tag 150 $a, which is the general entry for topical term headings, making these headings virtually indistinguishable from other topical terms from a machine-processing standpoint. Without manual inspection, it is not possible to extract headings that contain a chronological statement from the thousands of authority records describing topical headings. As a consequence, in our analysis and implementation we concentrate on the records containing a chronological subdivision, providing a starting point for automated processing.

For chronological subdivisions, instruction sheet H620 of the Subject Cataloging Manual specifies eight different forms of chronological subdivision statements:

1. Unspecified starting date (e.g.150 ## $a Libraries $x History $y To 400)
2. Unspecified ending date (e.g. 150 ## $a World politics $y 1989-)
3. Unspecified starting or ending date (only for geological periods) (e.g.150 ## $a Geology, Stratigraphic $y Devonian)
4. Specific century or centuries (e.g. 150 ## $a Wood-engraving $y 17th century)
5. Specific date spans with explanatory words: (e.g. 150 ## $a Christian art and symbolism $y Renaissance, 1450-1600)
6. Specific date spans without explanatory words (e.g. 150 ## $a Music $y 500-1400)
7. Single date with explanatory words (e.g. 151 ## $a Great Britain $x History $y Edward VIII, 1936)
8. Single date without explanatory words (e.g. 150 ## $a Solar eclipses $y 1854)

Of these eight forms of chronological subdivisions, only a subset could be termed an event or a named time period in our interpretation. In particular, these are geological periods (3) and chronological subdivisions that contain explanatory words (5 & 7). Looking at the examples given above,

it is entirely probable that the other five subdivision forms contain a number of records that describe events (e.g. Solar eclipses, 1854), especially when they are tied to a particular location with a geographic subdivision - however, most chronological subdivisions seem to subdivide a subject by time in order to split the number of records into a manageable size (e.g. Libraries, History, To 400 or Wood-engraving, 17th century) without denoting specific events.

One could describe the majority of records as subjects that lend themselves to be qualified or subdivided by a time span - in this study, however, we are most interested in those records that primarily describe a time span or event (where time is the primary aspect).

3. CHRONOLOGICAL SUBDIVISIONS

Frommeyer[7] summarizes a number of quantitative studies analyzing the distribution of LCSH subdivisions in samples of catalog records, of which we will mention the largest (in terms of the number of records analyzed)[10] and most detailed (in looking at chronological subdivisions)[5].

Form of Subdivision	Topical heading	Geographic heading
Abbreviated century (e.g. 15th century)	75.5%	7.6%
Dates or date ranges (e.g. 1933-1945)	10%	40%
Names of events + dates (e.g. Civil War, 1861-1865)	10%	45%
Dates following "To" (e.g. To 400)	4.3%	3.7%

Table 1: Distribution of chronological subdivisions in topical and geographic headings (Drabenstott, 1992)

O'Neill and Aluri[10] analyzed a sample of 33,455 bibliographic records of monographs that contained 47,036 LC subject headings. Of the 33,597 topical subject headings (650), 31% contained a geographic and 3% a chronological subdivision. Of the 6,826 geographic subject headings (651), 5% contained a geographic and 23% a chronological subdivision. These percentages are calculated over the number of all subject headings, not only the subdivided ones, which is the case for the other studies described here.

For Drabenstott's study[5], a 0.1% sample (2,903 unique topical headings and 857 geographic headings) of subdivided assigned subject headings for topical and geographic terms from the OCLC Online Union Catalog was analyzed. Of the 2,903 topical headings, 52.4% contained a geographic and only 1.6% a chronological subdivision. Of the 857 geographic headings, 3.1% contained a geographic and 9.4% a chronological subdivision. Drabenstott also lists the different forms of period subdivisions in her 2 samples (see Table 1). Geographic headings have a much higher percentage of event names than topical headings - as could be expected considering that events are commonly associated with a place.

We have undertaken two similar studies with a sample of LOC subject heading authority records from 2004 (280,000) and about five million records from the University of California's MELVYL catalog. Looking at authority records is similar to Drabenstott's study that looked at unique headings whereas looking at catalog records is similar to O'Neill's study.

Our sample of authority records contained about 59,000 subdivided topical headings (MARC code 150) and 13,000 subdivided geographical headings (MARC code 151), which we further analyzed. Of the 59,000 topical headings, 28.1% contained a geographical ($z) and 4.6% a chronological subdivision ($y). Of the 13,000 geographical headings, 0.6% contained a geographical subdivision and 62% a chronological subdivision.

Table 2 lists the different forms of chronological subdivisions in our subdivided samples:

Form of Subdivision	Topical heading	Geographic heading
Abbreviated century (e.g. 15th century)	43.8%	9.4%
Dates or date ranges (e.g. 1933-1945)	13.3%	47.4%
Names of events + dates (e.g. Civil War, 1861-1865)	17.5%	30.7%
Dates following "To" (e.g. To 400)	14%	7.2%

Table 2: Distribution of chronological subdivisions in topical and geographic subdivided authority records.

Form of Subdivision	Topical heading	Geographic heading
Abbreviated century (e.g. 15th century)	81.5%	14.8%
Dates or date ranges (e.g. 1933-1945)	6.0%	59.1%
Names of events + dates (e.g. Civil War, 1861-1865)	10.3%	21.7%
Dates following "To" (e.g. To 400)	2.2%	4.4%

Table 3: Distribution of chronological subdivisions in topical and geographic subject headings in catalog records.

It is important to glean the distribution of subject headings in a real-world setting where headings might repeat and an uneven distribution of headings manifests itself. This is the reason why we also looked at catalog records. About four million records (of the five analyzed) contained one or more subject heading phrases (MARC tags 650 or 651). We found about 5 million subdivided topical subject headings and 1.27 million geographical subject headings. In our topical heading sample, 60.5% of the headings contained a geographic subdivision and 7.2% a chronological subdivision. In our geographic heading sample, 6.6% contained a geographic, and 27.3% a chronological subdivision.

We also looked at different forms of chronological subdivisions in the catalog records. The summary of this analysis in Table 3 shows the same trends in the catalog data as in the Drabenstott study and our authority records study.

An astonishing number of the chronological subdivisions in catalog records we looked at contained spelling or syntax errors. Table 4 shows an example of the chronological subdivision "Early church, ca 30-600", which occurred in 12 variants. An automatic checking and verification procedure for subject heading formulation and assignment as recommended by Drabenstott[5] would help consolidating these erroneous (and therefore difficult to retrieve) headings.

Early chuch, ca 30-600	1
Early church	4
Early church ca 30-600	3
Early church, ca 30-600	1
Early church, 30-600	3
Early church, 300-600	1
Early church, ca 3-600	1
Early church, ca 30 to 600	1
Early church, ca 30-60	1
Early church, ca 30-600	1692
Early church, ca 60-300	1
Early church, ca30-600	7

Table 4: Variants of chronological subdivisions in actual catalog records.

Most subdivisions with the structure of "Names of events with dates" are events, meaning that at least 16% of all records containing a chronological subdivision will reference an event. Many more subject headings (not containing a chronological subdivision) might also point to events.

For the Time Period Directory prototype introduced in this paper, we only look at those authority records (unique subject headings) that contain a phrase in their chronological heading possibly denoting a period or event name. Of the ca. 10,000 authority records in our sample that contain a chronological subdivision, around a third contains a phrase or potential event name. We have followed this approach for convenience reasons, because it was relatively easy for us to automatically extract those authority records that contained a phrase. Our application contains a little over 2,000 different period or event names. Although this shows the viability of the approach, we are certain that a closer inspection of LCSH authority records could find many more instances.

4. THE TIME PERIOD DIRECTORY CONTENT STANDARD

In 2003, Feinberg et al. prepared a report for the Electronic Cultural Atlas Initiative's IMLS project "Going places in the catalog" describing the design of a gazetteer-like standard to describe time periods:

> "Just as a gazetteer matches place names to coordinates, a time period directory could match time period terms to date ranges, location, and other information that characterizes the period." [6]

The motivation for the "Going places in the Catalog" project was to improve geographic access to library catalogs by providing a map interface and longitude/latitude data to provide a more stable reference system to unstable geographic names in the catalog records. We realized in doing

timePeriodEntry	* Time Period Directory Instance * Contains components described below
-periodID	Unique identifier
-periodName	* Period name, can be repeated for alternative names * Information about language, script, transliteration scheme * Source information and notes (where was the period name mentioned)
-descriptiveNotes	Description of time period
-dates	* Calendar and date format * Begin & end date (exact, earliest, latest, most-likely, advocated-by-source, ongoing) * Notes, sources
-periodClassification	* Period type, e.g. Period of Conflict, Art movement * Can plug in different classification schemes * Can be repeated for several classifications
-location	* Assocatiated places with time period * Contains both place name and entry to a gazetteer providing more specific place information like latitude / longitude coordinates * Can plug in different location indicators (e.g. ADL gazetteer, Getty Thesaurus of Geographic names)
-relatedPeriod	* Related time periods periodID of related periods * Information about relationship type (part-of, successor etc.) * Can plug in different relationship type schemes
-entryMetadata	* Notes about creator / creation of instance * Entry date * Modification date

Table 5: Main Elements of Time Period Directory Content Standard.

that work that time statements and time periods have the same problems. Even though there exists a precise system to specify time and time ranges (dates), we use ambiguous names in public discourse instead. As Feinberg notes:

> "The categorization of information by time period is a ubiquitous organizational device, especially for historical data. Yet instead of referring to specific years, we often use period terms to suggest calendar dates. Time period terms distinguish neolithic ruins, Elizabethan drama, and the Napoleonic wars. Indeed, these terms often carry a stronger meaning than calendar dates, because they implicate a subject, time, and place together. For example, the French term "le grand siecle" encapsulates a place (France), time (seventeenth century), and subject (a flowering of arts and culture, and the height of ab-

solute rule). In contrast, the term seventeenth century is markedly less informative."[6]

The result was a draft content schema description for a Time Period Directory Standard for describing named time periods and linking them to dates and locations similar to a gazetteer content standard describing a format for linking geographic names and their latitude and longitude coordinates (see, for example, the ADL Gazetteer Content Standard[1]).

They also provided a period type list categorizing the types of events that could be described in a time period directory (e.g. reigns, wars, revolutions, religious movements etc.). This is similar to a feature type thesaurus for a gazetteer describing the different geographic features (e.g. city, bridge, island).

Similar to the ADL Gazetteer Content Standard, we have created an XML schema describing the ECAI Time Period Directory Content Standard based on the content schema description put forth by the Feinberg report[6]. This XML schema describes the structure and elements of a Time Period Directory, including both the required and optional elements for Time Period Directory instances. A single time period directory instance contains information about a named time period or event, its associated dates and location as well as a period type categorization and information about related periods.

We have aspired to make our Time Period Directory Content Standard as flexible as possible: it is not only language- and script independent, therefore allowing multilingual and multi-script access to time period names, it also permits different location indicators and different time period categorization schemes to be "plugged in".

Table 5 describes the major components of our Time Period Directory Content Standard. The schema is very compact, trying to re-use elements in different components (e.g. notes, sources, scheme). The full XML schema can be found at our web site:
(http://ecai.org/imls2004/timeperiods.html).

5. HARVESTING THE LCSH AUTHORITY FILES

Although Time Period Directories could be manually populated with information from a variety of sources, we attempted to exploit automatic methods as much as possible for the development of our prototype Time Period Directory. In this section we discuss the extraction process and our use of Library of Congress Subject Heading authority records as basis for the Time Period Directory content. While LCSH has been mined for alternative representations of subject data (as in the OCLC FAST project[4]), these have considered dates alone without the corresponding events.

5.1 Record extraction

We predicted that the Library of Congress Subject authority headings containing a chronological term (MARC Authority heading 148), a chronological subdivision (MARC Authority heading 182) or other topical headings containing a chronological subdivision ($y) are very likely to reference event or time period names. We extracted these records from our sample of 280,000 authority records. The chronological term field (148) did not occur in our sample. The majority of the 10,731 records we extracted were geographic name authority records (151) and topical term authority records (150). We also extracted a few personal name (100), corporate name (110), uniform title (130), general subdivision (180) and chronological subdivision (182) records.

As discussed in section 3, the chronological subdivisions that contain a phrase are more likely to be an event, so we automatically extracted only those records that contained a phrase in the chronological subdivision. We ended up with 3,660 unique subject headings (authority records) in our preliminary sample, which we further processed. Of these 3,660 headings, 2806 (77%) were geographic name and 787 (22%) were topical term headings.

MARC authority records contain a lot of data for machine processing. Below we show in an example the fields of the authority records that we extracted for further processing:

```
<USMARC>
<Fld001>sh 00000613 </Fld001>
<Fld151><a>Magdeburg (Germany)</a><x>History</x>
    <y>Siege, 1550-1551</y></Fld151>
<Fld550><w>g</w><a>Sieges</a><z>Germany</z>
</Fld550>
<Fld670><a>Work cat.: 45053442: Besselmeier, S.
Warhafftige history vnd beschreibung des
Magdeburgischen Kriegs, 1552.</a></Fld670>
<Fld670><a>Cath. encyc.</a><b>(Magdeburg:
besieged (1550-51) by the Margrave Maurice of
Saxony)</b></Fld670>
<Fld670><a>Ox. encyc. reformation</a>
  <b>(Magdeburg: ... during the 1550-1551
  siege of Magdeburg ...)</b></Fld670>
</USMARC>
```

In this example, Field 001 is the control number, field 151 contains the main heading, field 550 refers to a see also from tracing and field 670 is a source data field.

Besides the control number and the main heading, we also extracted the see from and see also from tracing fields (4xx and 5xx) to mine them for alternate period names and the note fields (667-68x) to extract source information.

5.2 Record transformation

We devised a set of rules for each type of authority record (depending on whether they were geographic names or topical term records) in order to transform the MARC-like format to our Time Period Directory XML structure. We assigned a unique record ID (periodID) to each record, but also kept the control number in a source element to be able to track back to the original authority record. The main heading was then split up: the chronological subdivision ($y) was put in the periodName field (similarly we put the chronological subdivision of see from and see also from tracings in another periodName field), the geographic aspect ($a in geographic name headings, $z in topical headings) was transformed into the location field and form ($v) and other topical subdivisions ($x - mostly "History") were dropped. The notes fields were transformed into source elements. The date section of the chronological subdivision was then split into begin and end date fields. We also automatically entered entry information (e.g. creation date)to provide a administrative record for the directory entries. This rule-based transformation created a skeleton design for the final XML records that conforms to the Time Period Directory Content Standard. A number of elements were left empty and needed manual data entry (e.g. classification information, location indicator information).

5.3 Duplicate removal and problematic cases

We sorted the records by chronological subdivision and manually analyzed the 3,660 records to remove duplicates and records that weren't obvious events or named time periods. This was a highly subjective process and presented a number of problems, which we will try to enumerate here.

5.3.1 Obvious non-events and non place-able events

We first removed records that referred to historic language forms instead of events or could not be obviously associated to a place or did not denote events or named time periods at all. For some of these examples one could certainly contest whether they constitute a period or not, we did not spend too much time trying to force these records into our schema. Most records that were removed contained rather vague or broad date ranges. Examples include:

```
<Fld150><a>Church history</a><y>2nd
   century</y></Fld150>
<Fld150><a>Theology, Doctrinal</a>
   <x>History</x><y>Early church,
   ca. 30-600</y></Fld150>
<Fld150><a>German literature</a>
   <y>Old High German, 750-1050</y>
   </Fld150>
<Fld150><a>Egyptian language</a>
   <y>Demotic, ca. 650 B.C.-450 A.D.</y>
   </Fld150>
```

For other records that did not contain a geographic aspect, we looked up the associated location to place in the required location element. Examples include:

```
<Fld150><a>Crusades</a>
   <y>Eighth, 1270</y></Fld150>
<Fld150><a>Salzburgers</a>
   <y>Emigration, 1731-1735</y></Fld150>
<Fld150><a>Geology, Stratigraphic</a>
   <y>Eocene</y></Fld150>
```

For geologic periods, we generally entered earth as location. Geologic periods are one example where calendar and date format need to be flexibly defined. Our Time Period Directory Content Standard allows us to define the calendar format in "million of years ago" as is the common format in geologic time.

5.3.2 Duplicates

Artistic epochs and periods are often repeated in authority records. We collapsed these records by dropping the topical entry ($a) and just keeping the chronological subdivision and location information. For example, the Japanese "Meiji period, 1868-1912" occurs in 35 records. We collapsed them into one. Here are some examples for individual headings:

```
<Fld150><a>Architecture</a><z>Japan</z>
   <y>Meiji period, 1868-1912</y></Fld150>
<Fld150><a>Art, Japanese</a>
   <y>Meiji period, 1868-1912</y></Fld150>
<Fld150><a>Japanese drama</a>
   <y>Meiji period, 1868-1912</y></Fld150>
<Fld150><a>Japanese fiction</a>
   <y>Meiji period, 1868-1912</y></Fld150>
```

5.3.3 One event, several locations in the same geographic area

For geographic headings, we encountered a series of problems concerning the LOC practice of creating a new authority record for each new location encountered. For example, the "Revolution, 1917-1921" event in Russia occurs in 29 authority records - one for each place described in books at the Library of Congress. Because the creation of subject headings is based on literary warrant, it is indeed the case that the "Civil War, 1861-1865" event in the United States is described for 88 different locations, but not necessarily for each state!

Since we wanted to be able to display events on a map and initiate a catalog search to find records about these events, it was necessary to keep this location information. We could not simply search the LOC catalog for "Colonial period, ca. 1600-1775, United States" because that would only find the records indexed with this particular string but we also needed to expand the search with other locations recorded with the period "Colonial period" - especially if we wanted to find all records concerned with the geographic area of the United States. Keeping all the location information also enables us to provide a more geographically precise search, for example in the cases below:

```
<Fld151><a>Punjab (India)</a><x>History</x>
   <y>Partition, 1947</y></Fld151>
<Fld151><a>India</a><x>History</x>
   <y>Partition, 1947</y></Fld151>
<Fld151><a>Bengal (India)</a><x>History</x>
   <y>Partition, 1947</y></Fld151>
```

In these cases, however, we collapsed the records into one Time Period Directory instance with several location statements.

5.3.4 Erroneous data, several locations

Many records describing a monarch's reign occur twice or more under different locations but clearly describe the same king or period (geographical proximity). The most prominent examples are all the kings indexed under Germany and the Holy Roman Empire. This is one example:

```
<Fld151><a>Germany</a><x>History</x>
   <y>Charles IV, 1347-1378</y></Fld151>
<Fld151><a>Holy Roman Empire</a><x>History</x>
   <y>Charles IV, 1347-1378</y></Fld151>
```

Another case are those events or time periods that seem to be the same, occur in neighboring or overlapping locations but have varying dates. These may have been due to cataloging errors, but a more likely explanation is that they are due to different foci in the books used as authority sources by LOC. Here are a few examples:

```
<Fld151><a>Denmark</a><x>History</x>
   <y>Christian III, 1534-1559</y></Fld151>
<Fld151><a>Norway</a><x>History</x>
   <y>Christian III, 1537-1559</y></Fld151>

<Fld151><a>Rome</a><x>History</x>
   <y>Empire, 30 B.C.-284 A.D.</y></Fld151>
<Fld151><a>Rome</a><x>History</x>
   <y>Empire, 30 B.C.-476 A.D.</y></Fld151>

<Fld151><a>Austria</a><x>History</x>
   <y>Ferdinand I, 1521-1564</y></Fld151>
<Fld151><a>Germany</a><x>History</x>
   <y>Ferdinand I, 1556-1564</y></Fld151>

<Fld151><a>Holy Roman Empire</a><x>History</x>
   <y>Otto IV, 1198-1215</y></Fld151>
<Fld151><a>Germany</a><x>History</x>
   <y>Otto IV, 1208-1214</y></Fld151>
```

When we could ascertain that these reigns were in fact the same, we collapsed the records and input the most appropriate date and qualified this in the record with a "most-likely" or "earliest" or "latest" date qualifier.

After the removal of duplicates and general "cleaning" of records, we ended up with a total of 2,006 Time Period Directory records.

5.4 Manual entry of classification data

The next stage in the Time Period Directory creation was manual entry of time period type information. This involved looking at every record and determining the "type" of event described in this instance. We used the Time Period Type List suggested by Feinberg et al.[6]

The Time Period Type List was suggested after collecting and analyzing many different time periods and event names from a number of resources. This classification contains 6 top-level categories (Groups of People, Trends in thought and expression, Natural events, Acts of creation or discovery, Biography and Cycles and intervals) and 34 subcategories.

We have discovered that most events found in the chronological subdivisions analyzed for our Time Period Directory application fall into the Groups of People category. The overwhelming majority of instances were either categorized as Period of Conflict (e.g. revolution, war) (in 56% of the cases) or Period of Control (reign, occupation) (in 34% of the cases). We plan to expand the classification in order to provide more detailed structuring in these areas.

5.4.1 Manual entry of location data

We extracted place names from the LOC subject headings as far as possible but recognized the need for unambiguously identifying place names. Therefore we include more detailed location information, ideally latitude and longitude coordinates, (an approach we also suggested in the IMLS Going places in the catalog project) in the Time Period Directory records.

The Content Standard allows for various gazetteer or other location indication schemes to be applied. For our Time Period Directory, we used the Alexandria Digital Library Gazetteer
(http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp)
and the Getty Thesaurus of Geographic Names
(http://www.getty.edu/research/conducting_research/vocabularies).

We did not include the whole gazetteer record for a place or geographic area, which would be redundant, but just linked to the gazetteer record by id number. Using the gazetteer record id, one can easily retrieve longitude/latitude information (in case of the ADL gazetteer) or more information about the place linked to. We used the Getty Thesaurus of Geographic Names because it provides a record for historical place names that are not retrievable through the ADL gazetteer. Examples are Prussia, Holy Roman Empire, Galicia and Yugoslavia.

Our practice of linking to the Alexandria Digital Library Gazetteer (ADL) deserves further consideration. The ADL gazetteer provides contemporary location information and not historical data. This ignores the fact that geographic places are unstable and their borders change over time. Consequently, if we link to Russia in the record for the Russian revolution of 1917, we are not linking to location information about Russia in the borders of 1917, but rather to Russia of today. It is obvious that this is a clear limitation of our data presentation at the moment. Although we can present the locations on a map thanks to the latitude and longitude data from the ADL gazetteer, these do not necessarily represent the actual borders of the place at the time of the event but rather a first approximation. (Note that the ADL standard does support application of date ranges to placenames and spatial locations, but these are very rarely available in the existing database).

To our knowledge, no comprehensive historical gazetteer of the scale of the ADL exists, that also provides spatial data that we could use to present our instances on a map. We hope to have provided enough context information using two gazetteers to be able to fill in the information of a more precise historical gazetteer.

The Time Period Directory Content Standard also provides an element for linking related periods. Our intent is to use this for successive events or events that are part of other events. Due to time constraints, we were unable to fully fill these elements to our satisfaction. Part of this information could be automatically constructed (looking at date ranges in the same countries) but some require human overview.

An example of a complete Time Period Directory instance can be found on our web site at:
http://metadata.sims.berkeley.edu/tpd/TPD-instance.pdf.

6. MAKING IT ACCESSIBLE: A PROTOTYPE WEB INTERFACE

The Time Period Directory Content Standard was created with specific access points to the data in mind. The XML structure is intended to enable a variety of different presentations of the data depending on a user's preferences.

Four major entry points for a search can be envisioned:

Search by location: A search by location identifies a place and for that place all available time periods (events) for that location. This could be represented by dots on a map and an accompanying table showing the time period records for this location. This feature could also be used for the creation of national timelines, for example. A search by location could be limited (specified) by time and category (time period type classification).

Search by time: A search by time identifies all time periods (events) within a given time range. This could be used for orienting oneself about a particular time period. A search by time could be limited (specified) by location and category (time period classification).

Search by time period type: A search by category will show events associated with a specific category from a time period type classification. This could be used for showing trends or related events across space (i.e. civil revolutions in the 19th century). A search by category could be limited by location and time.

Search for a specific time period or event: A direct search for a named time period or event will be enabled by having period names and alternative names enumerated in the Time Period Directory. This could be used for specific known-item searches. A search for specific time period or event could be limited by category, location or time.

For our 2,006 records of LCSH events, we have created a web interface to show a variety of search possibilities and to make this Time Period Directory available to users. We have also connected our Time Period Directory to the Library of Congress catalog to demonstrate the feasibility of other search entry points (specifically time!) into the catalog. It is now accessible at:
http://ecai.org/imls2004/TimePeriodDirectory/tpdindex.html.

Three principal entry points are provided on the starting page:

- Country Browse, which allows to search for events by country, US state or major city;
- Map Browse, which presents events as clickable dots on a map;
- Time Line Browse, which allows to search for events by time.

We have not yet implemented a search by category because our time period type classification (as explained in section 5) classified most of our 2000 records in 2 categories and therefore did not provide a good (valuable) entry into the data.

6.1 Country Browse

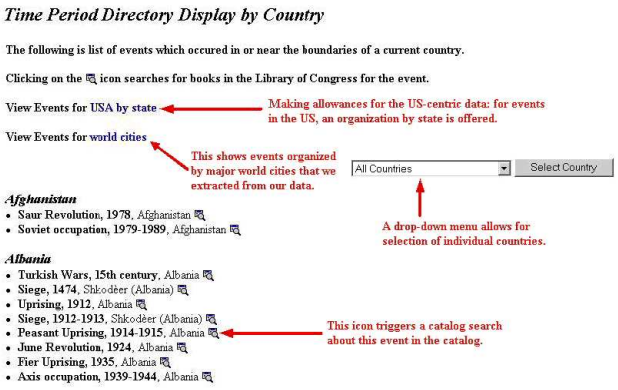


Figure 1: Sample view of the Country Browse

The Country Browse interface provides an overview of events or time periods by country (first alphabetically organized) or by US state or world city. Every event has an icon associated with it that will trigger a catalog search in the Library of Congress catalog. Since our Time Period Directory records are created from LC subject headings, we can use these subject headings (from which we extracted our instances) to trigger a precise subject search in the catalog. The drop-down menu allows selecting a specific country and to view events associated with it.

The catalog search looks up the Time Period Directory record and finds the LCSH id in the source field, it then transforms this LCSH id to the original subject heading (by searching in a local copy of LCSH Authority file) and submits a Z39.50 search for that heading to the Library of Congress catalog server. It shows the original heading, the number of found items and the items itself with all subject headings linked. Clicking on one of the subject headings triggers a new Z39.50 search to the LC catalog retrieving

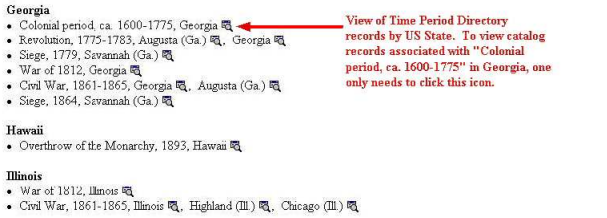
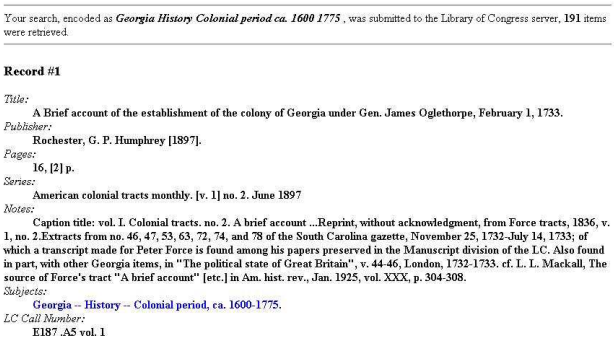


Figure 2: View of Time Period Directory by US State.

catalog records associated with that heading. Figure 3 shows an example for the search "Colonial period, ca. 1600-1775" in Georgia.


Your search, encoded as ***Georgia History Colonial period ca. 1600 1775*** , was submitted to the Library of Congress server, **191** items were retrieved.

Record #1

Title:
A Brief account of the establishment of the colony of Georgia under Gen. James Oglethorpe, February 1, 1733.
Publisher:
Rochester, G. P. Humphrey [1897].
Pages:
16, [2] p.
Series:
American colonial tracts monthly. [v. 1] no. 2. June 1897
Notes:
Caption title: vol. I. Colonial tracts. no. 2. A brief account ...Reprint, without acknowledgment, from Force tracts, 1836, v. 1, no. 2.Extracts from no. 46, 47, 53, 63, 72, 74, and 78 of the South Carolina gazette, November 25, 1732-July 14, 1733; of which a transcript made for Peter Force is found among his papers preserved in the Manuscript division of the LC. Also found in part, with other Georgia items, in "The political state of Great Britain", v. 44-46, London, 1732-1733. cf. L. L. Mackall, The source of Force's tract "A brief account" [etc.] in Am. hist. rev., Jan. 1925, vol. XXX, p. 304-308.
Subjects:
Georgia -- History -- Colonial period, ca. 1600-1775.
LC Call Number:
E187 .A5 vol. 1

Figure 3: Catalog search from a Time Period Directory instance.

All catalog searches are Z39.50 searches of the Library of Congress catalog using the Cheshire II search engine[8]. In the current version of the catalog search, personal name subject headings from LC records are also converted into searches of Wikipedia. Because standard protocols for search are used, many other compatible resources are available.

6.2 Map Browse

The Map Browse function provides the same fundamental search functionality as the Country Browse interface, but it presents events plotted on a map using the location information in the Time Period Directory records with the numbers of events in a country (or U.S. state) indicated by darker colors.

The map interface allows the user to zoom in to different countries or U.S. states. Clicking on a country or state triggers a search to the Time Period Directory showing the events associated with that country. All records are once again linked to a catalog search.

Eventually the "time bar" in the lower part of the map display will be used to restrict the events by date ranges. We also plan to color-code the different types of events shown on the map. Due to the limited set of classes in our classification this is not at present very useful. The map search interface is implemented using TimeMap (http://www.timemap.net/) with data stored in the ECAI Clearinghouse.

6.3 Time Line Browse

The Time Line Browse provides another view of the data

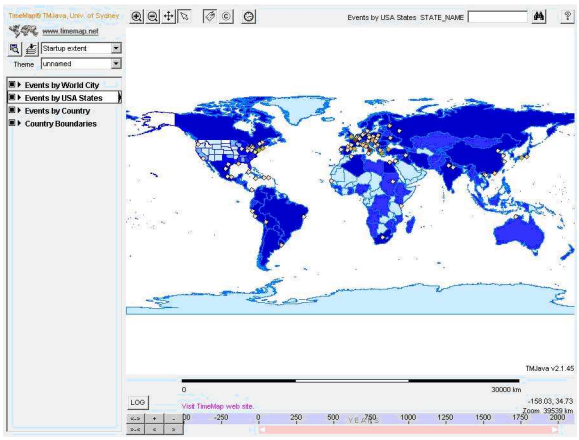


Figure 4: Map Browse interface.

Figure 5: Zooming in to Europe and clicking on the dot for Paris opens a menu allowing searching for events for the city of Paris or the country of France.

- organized by date. This entry point might be especially interesting to a general searcher, trying to orient him- or herself about a period of time. This prototype allows viewing all events and periods of our Time Period Directory in a given century, obviously finer distinctions would be needed for larger directories.

The events in the time line display are sorted by date and linked directly to a Library of Congress catalog search. Once we have implemented the time line slider in the map interface, we will also be able to provide a chronological entry point from the map.

For further implementation, we envision a time line feature that will - given a country, date range or time period type - create customized time lines on the fly. Our Time Period Directory certainly provides the data and structure for this application.

Especially for this interface, we are looking into more sophisticated visualization techniques to make the presentation visually more appealing.

The Time Line Browse is based on a Javascript visualization and database querying of the underlying Time Period Directory.

7. CONCLUSIONS

For the near future, we are considering several areas of improvement - both in our implementation of the prototype

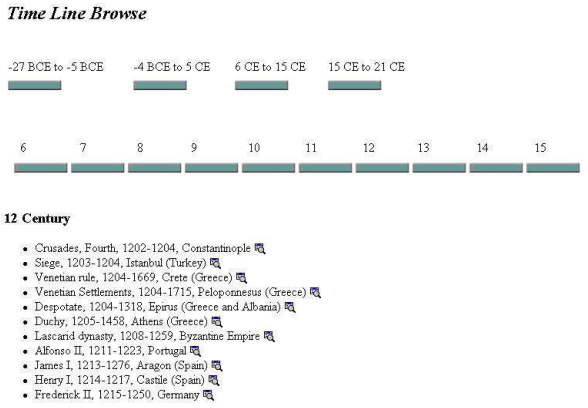


Figure 6: Time Line Browse. Showing results for the 13th century.

Time Period Directory and the fitting of Time Period Directories in a greater framework of a "metadata infrastructure" for better access to information systems. We are also adding support for search and access to the Time Period Directory data as a web service that will return entries in the XML schema used internally for the representation of time periods.

It became obvious from the implementation of our 2,006 Time Period Directory records that scale is a significant issue for the visual and textual representation of the data. For the Country Browse and Time Line Browse interfaces, alphabetical or chronological ordering alone is not scaling up - too many records in this view are confusing and exhausting to look at. We have several ideas how this problem can be alleviated.

First of all, using a more appropriate time period type classification will help us to provide a faceted view into the data by grouping events by similar types and provide a more cohesive view of different political, social and cultural developments for a given area or country. As a beneficial side effect, the new facets will also add other access points to the Time Period Directory instances - allowing to search for similar events across space and time.

Secondly, we believe that the visual impact of a time line is familiar to many users of our system. Being able to create a time line on the fly - customized by country, date, or event type - should prove to be helpful for educational and presentation purposes for our users. The structure of Time Period Directories permits the use of the data elements in this way. We are working on making this application available.

The Time Period Directory Content Standard also contains a component linking to related events and time periods. Because of time and labor reasons, we did not attempt to populate these elements in our current implementation of the LOC chronological subdivision data. If these elements were filled, we foresee more and more interesting ways of connecting and visualizing events in time lines and causal relationships that provide insight in a particular historical development. This feature should be especially interesting to domain specialists and teachers.

Additionally, we are conscious of the fact that in order to really prove the value of Time Period Directories and the Content Standard, it is necessary to try out the schema and implementation on other and more varied data sets than

the relatively homogenous and standardized LOC authority records. We are looking into mining other interesting sources as well. For general historical events, any encyclopedia or domain chronology would lend itself to being harvested. Other classification systems and thesauri can be mined as well. For artistic periods, for example, the Getty Arts and Architecture Thesaurus provides a "Styles and Periods Facet", which seems very suitable for incorporation into a more structured and searchable schema as the Time Period Directory Content Standard provides.

In the greater context of the ECAI/IMLS "Support for the Learner: What, Where, When and Who?" project, the Time Period Directory is just one part of the general framework that we plan to develop. Although our implementation nicely connects the "When" and "Where" parts and provides a direct access to the "What" aspect through the Library of Congress catalog search, we have yet to incorporate the biographical aspect into our project.

Chronological, geographical and biographical data lend themselves naturally to being connected: an event is associated with a place, a time and potentially with particular people; places are associated with different events and people; and individual people are also associated (in a variety of ways) with different places and events. One can foresee a plethora of relationships and possible search questions that a truly interconnected system should be able to answer. We have demonstrated with our Time Period Directory implementation that many different views and perspectives on the same data are possible and desirable.

We are now at work on adding a biographical aspect to our framework. The Library of Congress Name Authority records are an obvious place to look - not only do they provide already structured data on persons and other corporate entities, they also inherently connect to a topical search applications (the library catalog), potentially easing our task of connecting the different informational aspects. Additionally, we are researching other sources of biographical data to analyze. Wikipedia's people category (http://en.wikipedia.org/wiki/Category:People) appears to be an opportunity and a challenge and is our next step forward.

We are developing prototype web services to support the dynamic interaction of Time Period Directories, digital gazetteers, biographical data and ontological structures like thesauri and classification schemes, in combination with a variety of network-accessible digital library resources ranging from library catalogs to archival collections and digitized version of historical primary resources. Our vision, as noted in the introduction, is that in response to users' expressions of interest, their interaction with this system will construct a rich dynamic portal of interconnected resources with maps, biographies, timelines and chronologies, and primary research materials.

8. ACKNOWLEDGMENTS

The work presented draws on three projects partially supported by Institute of Museum and Library Services National Leadership Grants: Seamless Searching of Numeric and Textual Resources (1999-2002; http://metadata.sims.berkeley.edu/GrantSupported/seamless.html), Going Places in the Catalog: Improved Geographical Access (2002-2004; http://ecai.org/imls2002/), and Support for the Learner: What , Where, When, and Who (2004- ; http://ecai.org/imls2004/). PIs: Michael Buckland, Fredric C. Gey, and Ray R. Larson.

9. REFERENCES

[1] Alexandria Digital Library Project. ADL Gazetteer Content Standard Version 3.2. Technical report, University of California, Santa Barbara, 2004.

[2] M. J. Bates and D. N. Wilde. An analysis of search terminology used by humanities scholars: the Getty Online Searching Project Report Number 1. *Library Quarterly*, 63(1):1–39, 1993.

[3] D. O. Case. The collection and use of information by some American historians: a study of motives and methods. *Library Quarterly*, 61(1):61–82, 1991.

[4] R. J. Dean. FAST: Development of simplified headings for metadata. In *Authority Control: Definition and International Experiences conference, Florence, Italy, Feb 10-12, 2003*, 2003.

[5] K. M. Drabenstott. Period subdivisions in the Library of Congress Subject Headings system - some thoughts and recommendations for the future. *Cataloging and Classification Quarterly*, 15(4):19–45, 1992.

[6] M. Feinberg, R. Mostern, S. Stone, and M. Buckland. Application of geographical gazetteer standards to named time periods. Technical report, Electronic Cultural Atlas Initiative, Berkeley, 2003.

[7] J. Frommeyer. *Zeitbegriffe und Zeitcodierungen in allgemeinbibliographischen Datenbanken : eine Analyse von RSWK, LCSH und RAMEAU sowie Entwicklung eines theoretischen Modells fur ein Zeitretrieval.* Logos-Verlag, Berlin, 2003.

[8] R. R. Larson, J. McDonough, P. O'Leary, L. Kuntz, and R. Moon. Cheshire II: Designing a next-generation online catalog. *Journal of the American Society for Information Science*, 47(7):555–567, July 1996.

[9] G. Leff. *Models inherent in history. The rules of the game: cross-disicplinary essays on models in scholarly thought.* T. Shanin, London, Tavistock, 1972.

[10] E. T. O'Neill and R. Aluri. Research report on subject heading patterns in OCLC monographic records. Technical report, OCLC, Research Dept., Columbus, Ohio, 1979. Research and Development Div. Report No. OCLC/RDD/RR-79/1.

[11] H. R. Tibbo. *Abstracts, online searching, and the humanities: an analysis of the structure and content of abstracts of historical discourse.* PhD thesis, University of Maryland, 1989.