0. Place, Period, and Setting for Linked Data Gazetteers

Karl Grossner, Krzysztof Janowicz, and Carsten Keßler

Introduction

There is a clear need in many domains of scholarship, and in society at large, for a comprehensive digital historical gazetteer. Or rather, we need to stitch together enough gazetteers to cover all places and historical periods, with unifying interfaces permitting human and programmatic access to them. The digitization of historical texts and maps is proceeding at a fast pace, and as algorithms for extracting place names from them continually improve, the enormous potential for indexing the world's data and knowledge on the many dimensions of place becomes more and more feasible. Expressions of this need have appeared with increasing frequency in various symposia and publications (throughout this volume; see also Bol 2011, and Southall, Mostern & Berman 2011). Information science researchers have been working for several years on some of the inherent challenges, in concert with historians and cultural heritage specialists (Janowicz 2009; Keßler, Janowicz & Bishr 2009; Kaupinnen, et al 2010; Mostern and Johnson 2008). An important early exemplar, the Pleiades project, is now reaching a critical threshold of adoption thanks in no small measure to the Pelagios project.[1] That important and ongoing work is reported in the next chapter of this volume.

Tasks involving organizing access to the world's knowledge should not fall to a few Silicon Valley companies alone. The motivations of commercial companies are not necessarily commensurate with those of academics, archivists and librarians employed at research universities, museums, and national libraries (or, for that matter, with the public at large). It seems intuitively the case that such institutions should be deeply involved.

Repositories of geographical data have grown in number recently in the form of *geoportals* and *national historical GIS systems*, gradually realizing a distributed global S*patial Data Infrastructure* (SDI) driven by governmental and academic interests. But a vastly broader realm of geo-referenced information awaits us, envisioned variously as *distributed geolibraries* (National Research Council 1999; Goodchild 1998), a spatially aware Wikipedia, and even a Digital Earth (Craglia, et al 2012; Grossner, Goodchild & Clarke 2008). If all place references found within the holdings of the world's libraries, museums and archives can be discovered and exposed effectively, the value of those works would be enhanced immeasurably. A federated global digital historical gazetteer is a lofty vision, but the technological impediments to achieving it are being identified and one by one falling by the wayside.

But apart from being large, distributed and historical, what sort of gazetteer(s) shall we make? Gazetteers serve two purposes: resolving place names to geographic locations, and providing some description of the places themselves. They have been defined most simply as place name directories, having a minimal entry that includes "at least one name, one feature type, and one location" (Hill 2006). However this definition is more normative than extensional, because the range of information found in existing gazetteers is broad. Locations might be a set of point coordinates, a bounding box, or a polygon rendered at any resolution. Descriptions can range from a simple designation of type to a detailed account of the place and its history, including text, tabular data and images. Most gazetteers lie somewhere between. In many print atlases, a gazetteer at the back of the volume simply lists place names, geographic coordinates and a page location. However in China, as the second chapter in this volume notes, *local gazetteers* (also known as *local histories*) have since the Song Dynasty been the repository of a great many details about places, at scales from provinces to towns to temples and rivers (Mostern 2008; Wilkinson 2013; Hymes 1996).

Given their generalized definition as information resources locating and describing places, gazetteers are, at a theoretical limit, a means for integrating much of what is known about places. This range of potential breadth and depth means the task of designing any particular digital gazetteer must begin with questions: What sort of software system will it be part of? More particularly, who are its prospective users and what are their purposes? At this stage in the development of the genre we should also seek highly general design principles that are applicable for most or all digital gazetteers.

Humanities scholars and information scientists (including the present authors) have begun to imagine a general-purpose distributed digital historical gazetteer with extensive and extensible place descriptions, global both in scope and in editorial participation. Such a system could be queried not only from a web interface by its human users, but programmatically by software applications. The primary users of this system will be its primary contributors as well: historical scholars, archivists, and cultural heritage specialists in the humanities and social sciences.

A list of requirements for digital gazetteers from humanities scholars' perspective outlined recently by historian Peter Bol (2011) appears in Table 1. In this chapter we show how a system using a *Linked Data* approach would address those requirements, in a prospective system that could realistically be undertaken now by a consortium of interested organizations and individuals.

It must accommodate multiple indicators of temporality, including (i) “start and end dates during which a feature is known to be valid;” (ii) “date of a map on which a place name appears;” and (iii) “date of a text...in which it is mentioned.”

It must be multilingual, for both place names and feature types.

It should “incorporate the historical gazetteers being created as part of local and national historical GIS projects, and ... volunteered information.”

It must be “...open-ended and cumulative;” i.e. extensible; able to grow in a completely flexible manner.

It must be built with a viable strategy for long-term sustainability.

Table 0.1 Requirements for Digital Gazetteers (Bol 2011)

As reported elsewhere in this volume, the Pelagios 3 project has taken concrete steps toward that goal. In order to realize it fully, however, the underlying modeling patterns must be able to capture information relevant in the humanities. One aspect of such information is the strong interdependence of spatial and temporal reference. Existing reference models and ontologies typically either focus on the spatial or temporal aspects, failing to reflect this interdependence. This chapter therefore proposes a new ontology design pattern for such *settings*, which can only be characterized when taking both spatial *and* temporal characteristics into account.

In Section 2 of this chapter we address the items in the Bol list, mapping each to functionality afforded by Linked Data methodologies. Then in Section 3 we propose two additions to the list and describe work we are undertaking individually and collectively to address them. The first of these is better joining time and space in data models. We present an overview and some details of an ontology design pattern for *Setting*, intended as a pragmatic aid to merging representations of Period and Place. The second added requirement, touched on briefly due to space constraints, is the representation and computation of imprecise temporal and spatial extents.

Linked Data and Humanities Requirements

The vision of transforming the World Wide Web to a Semantic Web, by linking structured data within documents rather than simply linking documents, was motivational but initially lacking in implementation specifics (Berners-Lee, Hendler & Lassila 2001). In a 2006 technical paper Tim Berners-Lee introduced the term Linked Data, which he described as being the "substance" of the Semantic Web. In it, he listed four rules for publishing data that arguably have made the grand Semantic Web vision simpler, more concrete, and more pragmatic (Table 2). These were later described as "a set of best practices for publishing and connecting structured data on the Web" (Bizer, Heath, & Berners-Lee 2009). The rules revolve around technologies represented by several acronyms—RDF, URI, HTTP, and SPARQL. There are several excellent resources available for learning what these terms mean (Heath and Bizer, 2011; Allemang and Hendler 2011). We present a very cursory introduction below, then discuss how Linked Data suits the domain of historical gazetteer development.

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

Table 0.2 Four rules for publishing Linked Data (Berners-Lee 2006)

The Resource Description Framework (RDF) refers to a semantic data modeling language from the World Wide Web Consortium (W3C), which enables representing a very large range of information in sets of statements having the essential "triple" form of <subject> <predicate> <object>. In Semantic Web parlance, everything we might want to link to or refer to is a "resource" including abstract concepts like Feature, and instances of concrete things-in-the-world like people, geographic features (e.g. Athens, Greece), and artifacts of all kinds including web documents. RDF subjects and predicates are always resources having unique web locations expressed as Uniform Resource Identifiers (URIs) and retrievable via the HTTP protocol. RDF objects can be a resource with a URI, or a literal value, such as "Athens". SPARQL (recursively, SPARQL Protocol and Structured Query Language) is a standard means for querying repositories of RDF statements, which are often called "triple-stores."

The RDF model and Linked Data framework can play an important part in meeting humanists' desiderata for digital historical gazetteers as outlined by Bol (2011). To discuss how, we rearrange Bol's list to discuss its key points in terms of *extensibility*, *multivocality*, *integration*, and *sustainability*. An important overarching factor is that Linked Data technology inherently supports semantic, ontology-based data models. A semantic data model is one that explicitly represents intended meanings of the concepts and relations in the data. This means that the important attributes of places can emerge flexibly and “bottom-up” from place data as it is accumulated from many projects. These formal semantic representations can be more or less expressive, according to what the community of interest finds most useful over time. The Linked Data paradigm is then a formal means for harmonizing contributions—of data and models—from very many sources, and for those sources to negotiate over time the vocabularies and schemas that allow fruitful sharing of data and derived knowledge. This is not to say there will not be significant challenges, only that Linked Data technology is a good basis on which to proceed. Early results, for example from the Pleiades and Pelagios projects,[2] bear this out.

Extensibility

Some basic characteristics of the RDF model underlying the Linked Data approach accommodate several of the explicit and implicit requirements for historical gazetteers in the BoI list. Owing to its graph structure, the RDF model will permit any number of entity classes and relations to coexist, and allow contributors to harmonize similar ones in ontologies, with statements about logical relations between them (such as equivalentClass, equivalentProperty, subClass and subRelation). For example one collection of place names gathered from maps may have a relation *hasMapSource*; another with names gathered from texts, a similar *hasTextSource*. By creating a *hasSource* relation and stating the first two are sub-relations (see Figure 0.1), a query on sources would return records from both maps and texts. Many contributors to a global Linked Data Historical Gazetteer (LDHG hereafter) will have large existing data sets with potentially distinctive relations reflecting idiosyncratic interests or legacy data-gathering practices. In this way the structure of a LDHG will be extensible and able to “grow in a completely flexible manner.”

```
a:hasSource          rdfs:type             owl:ObjectProperty ;

a:hasMapSource       rdfs:type             owl:ObjectProperty ;
                     rdfs:subPropertyOf    a:hasSource .

a:hasTextSource      rdfs:type             owl:ObjectProperty ;
                     rdfs:subPropertyOf    a:hasSource .
```

Figure 0.1 Recording Sources

The same advantages hold in the case of feature types and entity classes generally. Figure 0.2 shows how a dynasty may be treated as a kind (sub-class) of Period in one system, a sub-class of Place in another, and both in a third. This example highlights the fact that in all three cases a Dynasty *could* have a spatial location, a temporal location, or both, depending on the author's perspective. A *Setting* pattern that addresses this specifically is discussed in Section 3.

```
a:Dynasty            rdfs:type           owl:Class ;

                     rdfs:subClassOf     a:Period ;
                     rdfs:label          'Dynasty'@en ;
                     rdfs:label          ‘王朝’@zh .

b:Dynasty            rdfs:type           owl:Class ;
                     rdfs:subClassOf     b:Place .

c:Dynasty            rdfs:type           owl:Class ;
                     rdfs:subClassOf     c:Period ;
                     rdfs:subClassOf     c:Place .
```

Figure 0.2 Identifying Dynasties

The authors of each dataset (namespaces a:, b:, and c: in the example) can choose to add an equivalentClass statement to join their Dynasty data with others' data in the larger system. All instances of Dynasty data will then be returned in the same logical set at query time. Because all classes, relations and instances retain their identity as members of a distinct namespace, the global graph can be flexibly partitioned at the application layer to include or exclude any contributing dataset.

Multivocality

We use the term *multivocality* here referring to another kind of extensibility made possible by the RDF graph model. Besides asserting differing definitions for classes of things like dynasties, places and periods, multiple contributors can also assert differing literal values for the same attribute of a particular entity. Take for example the start and end dates of a feature's existence or a place name's validity. We can expect that over time multiple conflicting temporal assertions for any given place name will enter a global distributed graph. In a historical knowledge-base like a LDHG, we want to allow all plausible statements. In the case of dates, application developers might want to average them as a temporal centroid, or to calculate a period based on their union or intersection.

One way to accomplish this is by modeling not Places (or spatial Periods) but Place *records*. A Place record is a set of authored assertions about a place. We need to record facts about the assertion as well as its subject: the document it appears in, the date it was made, perhaps a certainty level or probability. With reification in RDF this amounts to a collection of statements, which could take the form of Figure 0.3. The arguably more efficient and expressive *named graph* approach is in Figure 0.4. Both approaches are workarounds from a semantic point of view, though, since the semantics of the meta-information about the statement – such as the source – is not explicit (Trame, Keßler and Kuhn 2013).

```
z:pr_123  a                rdf:Statement ;
    rdf:subject            z:Western_Han ;
    rdf:predicate          rdfs:type ;
    rdf:object             z:Place ;
    z:citedIn              'ISBN_9780674056022' .

z:pr_124  a                rdf:Statement ;
    rdf:subject            z:Western_Han ;
    rdf:predicate          z:hasStart ;
    rdf:object             '206 BCE' ;
    z:citedIn              'ISBN_9780674056022' .

z:pr_125  a                rdf:Statement ;
    rdf:subject            z:Western_Han ;
    rdf:predicate          z:hasStart ;
    rdf:object             '208 BCE' ;
    z:citedIn              'ISBN_123456789' .
```

Figure 0.3 Documenting the Assertion via Reification

```
chinaGaz:place_123 {
    z:Western_Han_Dynasty ;
    z:hasStart                 '206 BCE' ;
    rdfs:type                  z:Place ;
    skos:preferredName         'Western Han Dynasty' . }

chinaGaz:place_123             z:citedIn 'ISBN_9780674056022' .

chinaGaz:place_124 {
    z:Western_Han ;
    z:hasStart                 '208 BCE' ;
    rdfs:type                  z:Period ;
    skos:preferredName         'Western Han' . }

chinaGaz:place_124             z:citedIn  'ISBN_123456789' .
```

Figure 0.4 Alternate names and classes (types) in named graphs

Name variants, including for language can also be handled in a couple of ways. A LDHG must allow us to state more about a given name variant than that another string of characters is equivalent to the preferred name. A place name will always be in a language, and may be associated with a time period.

Classes and relations might differ only in the language used for their names. This could be handled in a couple of ways: by stating equivalency (Figure 0.5) or assigning multiple labels (Figure 0.6)

```
abc:Country          rdfs:type                  owl:Class ;
                    owl:equivalentClass        def:πατρίδα .

abc:liegtBei         rdfs:type                  owl:ObjectProperty ;
                    owl:equivalentProperty     def:locatedAt .
```

Figure 0.5 Equivalent classes and properties

```
:PopulatedPlace      rdfs:type                  owl:Class ;
                    rdfs:label                 'populated place'@en ;
                    rdfs:label                 'χώρα'@el .
```

Figure 0.6 Alternate languages

Integration and sustainability

These two requirements are intertwined and supported by the Linked Data approach. Due to the high cost of historical gazetteer development and the necessity for focused and localized expertise in creating them, any attempt at global or regional coverage will require a federated architecture. The Pelagios 3 project has demonstrated that many gazetteers can co-exist in a single federated system, allowing for query results that are global in both space and time. Contributing project gazetteers might have any combination of spatial and temporal scope.

In a federated system, centralized functionality would be limited to directing queries to the appropriate store(s) and some means of coordinating mappings between contributors' distinctive ontologies. The Linked Data approach is by its nature distributed, so the match to this requirement could not be closer.

Successful integration is what will make a global LDHG feasible. Contributing gazetteer developers will be responsible only for their own area(s) and period(s) of interest. Even so, the attention given to curating data sets fades over time, as their originators often move on to other work and priorities shift. This suggests there should be a strategy for long-term institutional maintenance of localized gazetteer data sets. Peter Bol and others have suggested this is a natural role for university libraries, and we concur. Each institution might have regions and historical periods of particular interest, owing to departmental strengths for example. We can expect universities to have a general interest in promoting and maintaining reputations for such strengths, and one way would be to have libraries maintain authoritative place records for their topic areas of particular interest (Stanford University Libraries 2011). This represents a distinct and significant expansion of traditional library services, so its uptake has been halting on the part of both library administrators and academics accustomed to such traditions. There is however a model for such large-scale institutional cooperation in the social sciences. The University of Michigan describes its Deep Blue digital archive as "a permanent, safe, and accessible service for representing our rich intellectual community." Not coincidentally, Michigan hosts the Inter-university Consortium for Political and Social Research (ICPSR), which has 776 member organizations worldwide (Sep 2012). Harvard's Institute for Quantitative Social Science Research (IQSSR) has developed the open-source Dataverse platform to manage its own collections and distribute that software framework for others to use and thereby join a Dataverse Network (there were 1,171 Dataverse instances housing data from over 58,000 studies in July, 2015). These examples suggest a workable model wherein a particular institution leads by undertaking a LDHG for some region of interest under the auspices of its library and the strong support of relevant departments. As noted earlier, at this writing there are several nascent initiatives involving major libraries.

Some further requirements

The list put forward by Bol provides an excellent start to framing requirements for LDHGs. However, we identified two additional elements which are perhaps more technical in nature, and for which there is considerable activity in several information science research domains. These are 1) improved means for formally representing the close association between places and temporal things like historical periods and events, and 2) improved means for representing and computing over indeterminate, vague, sparse, and otherwise uncertain temporal and spatial extents. In short, “joining place and period,” and “representing spatial-temporal uncertainty.”

Joining place and period in Settings

In this section we present an ontology design pattern (ODP) for *Setting*, with the aim of informing data models attempting to capture the way that places and periods are bound together. Gazetteers traditionally describe places. At minimum, a geographic gazetteer is a document or service which can resolve a place name to a geographic location, but gazetteers can also provide any sort of information about places their authors wish—from alternate names, to histories and rich descriptions of features and activities to be found there. In a sense, places have lifespans; they exist for some period of time, which may or may not be continuous. Their spatial extent can be time-varying during that lifespan. In fact, the temporal and spatial extents of places are bound together, although representing this circumstance in computer systems has been challenging.

In our view, places are described not only by their extents, the geographic features and artifacts found there, and their population characteristics, but by what has occurred there and what kinds of activity a place affords (or has afforded). Indeed, the artifacts and people present at a place are products of events, and the configuration of its geographic features is the product of natural events and processes. To the extent our records of events in globally shared information systems such as the Semantic Web include their *settings*, we can provide numerous dimensions for description and comparative analyses—not only of classes of events and of historical periods—but of places writ large in regions and landscapes and the *longue durée* of historian Fernand Braudel (1980).

A good example of the way place and time are bound together is the historical period. For historical scholars there is no single "Bronze Age." Rather, there is an "Early Bronze Age Southern Levant," a "Late Bronze Age Iran," and so forth. Historical periods are almost always geospatial as well as temporal. If the digital humanities community were to cultivate the habit of representing both the geography of events and periods, and the temporal attributes of places and other entities, geographic and historical information systems for humanities scholarship will significantly enhanced. This would be aided by having ontology design patterns (ODPs) that make this time-consuming modeling and encoding task more manageable and rewarding (Gangemi and Presutti, 2009). The intuition motivating our *Setting* design pattern is that both *Place* and *Period* have settings, an abstract element which has (at least) temporal and spatial defining components. The fact that places and periods are both spatial and temporal leads us to conclude that accounting for this is essential to historical gazetteer development.

Capturing the dynamics of geographic phenomena in computational models has been the subject of considerable research in the field of geographic information science for at least two decades (Yuan & Hornsby 2008; Hornsby & Yuan 2008). Frank's discussion of "types of times in GIS" (1998) exemplifies the recognition that more than a simple linear dimension must be considered. Many formal models for events, change, and process followed (including Hornsby & Egenhofer 2000; Galton 2005), but with negligible impact on GIS software, where time has been largely limited to filters and animating controls for a single timestamp field. Popular mapping platforms like Google Earth have provided similar functionality.

In their *Geospatial Event Model* (GEM), Worboys and Hornsby (2004) introduced the term *geosetting*, referring to a computational object describing the *situation* of an object or event—the where and the when of its existence, taken together. Perhaps owing to its relative complexity, GEM has not seen widespread implementation, but there remains a significant need for modeling the very commonplace concept of a *setting*, not least for historical gazetteers of places and of periods. Although the concept of setting can extend to any sort of contextual information about things and occurrences, we limit our view here to its spatial and temporal components, as these are the defining components that any other contextual information depends on.

A model for historical settings will ideally be composed of a few simple patterns, sufficiently general to encourage and facilitate interoperability, and to accommodate application-specific extensions as required. We have focused on Periods, but as Figure 7.7 shows, Places are their close cousins in many respects.

Setting

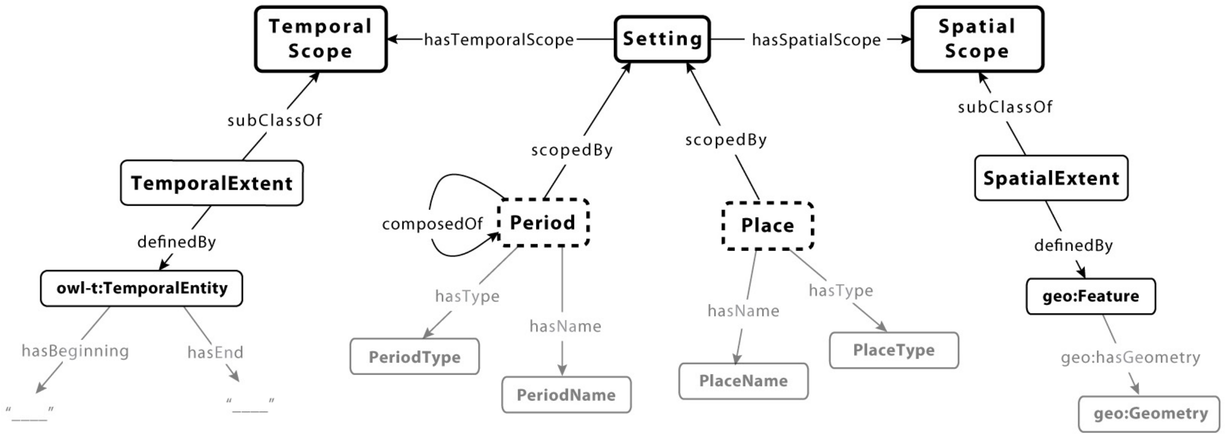

Figure 0.7 Conceptualizing Setting, Period and Place

The concept of a *setting* captures the spatial-temporal nature of answers to many "where" and "when" questions. A simple high-level pattern for *Setting* can be elaborated from spatial and temporal perspectives to effectively join place and time (Figure 1). Period and Place are commonplace information constructs, modeled variously to meet particular conceptualizations and computing application requirements. As their dashed-line representation suggests, our intention is not to specify their meaning here, but to indicate that regardless of what other definitional attributes they may have, their spatial and/or temporal extents can be scoped by a Setting. That is, any given Setting may apply to one or more Place or Period, and describe spatial scope, temporal scope, or both. Some definitional notes for the core elements of Setting follow:

A **Period** refers here to one class of non-eventive discretizations of time. Periods are most typically named timespans, conceived as containers for events or an interval of time (possibly discontinuous or multi-part) during which some condition state or states were true. Periods are frequently associated with or relevant to specific places. Examples include, ”Western Han Dynasty,” “Levant Bronze Age,” “Age of Enlightenment,” and “Christianization.”

Place refers here to a meaningful discretization of geographic space. An increasingly common modeling pattern for Place is as a context for attestations about the location of events, artifacts, and Earth features. Places ordinarily have names, with associated “valid times,” and may be actual or fictional. Examples include “the Levant,” “Addis Ababba,” “Middle Earth,” “The Confederate States of America,” and “the Amazon River basin.”

Setting refers here to the spatial-temporal context that scopes a Place or a Period. Settings are defined by their *SpatialScope* and *TemporalScope*, which can in turn be defined by spatial and temporal reference systems, respectively.

SpatialScope is a super-class of the metrical *SpatialExtent*, and potentially other scoping constructs such as natural language descriptions, as well as topological assertions such as for containment or adjacency, used to accommodate uncertainty (e.g. “in,” “near”).

TemporalScope is a super-class of the metrical *TemporalExtent*, and potentially other scoping constructs such as natural language descriptions, as well as qualitative descriptions of temporal intervals (Allan 1983).

SpatialExtent, a sub-class of *SpatialScope*, refers to a geometric description of a portion of geographical space (possibly non-contiguous), in terms of a spatial reference system, such as geographic coordinates. One means of defining SpatialExtent is as an OGC Feature.

TemporalExtent, a sub-class of *TemporalScope*, refers to a metrical description of an interval or set of intervals in time, in terms of a temporal reference system, such as a calendar.

One means of defining TemporalExtent is as a TemporalEntity in the OWL Time ontology. Another is as a *timespan* in the Topotime model discussed in the next section.

The core elements of the Setting pattern, in turtle notation, follow in Figure 0.8:

```
set:Setting              set:hasTemporalScope  set:TemporalScope ;
                         set:hasSpatialScope   set:SpatialScope .

set:Period               set:scopedBy          set:Setting .

set:Place                set:scopedBy          set:Setting .

set:TemporalExtent       rdfs:subClassOf       set:SpatialScope ;
                         set:definedBy         owl-t:TemporalEntity .

set:SpatialExtent        rdfs:subClassOf       set:SpatialScope ;
                         set:definedBy         geo:Feature .

set:hasTemporalScope     a                     owl:FunctionalProperty .

set:hasSpatialScope      a                     owl:FunctionalProperty .
```

Figure 0.8 Core elements of a Setting pattern

In Figure 0.9, we model the historical period, “Middle Bronze Age Southern Levant,” an example from a Pleiades time period listing.

```
@prefix set: <http://somepatterns.org/setting/#>.
@prefix pl: <http://pleiades.stoa.org/vocabularies/time-periods/#>.
@prefix geo: <http://www.opengis.net/ont/geosparql/1.0#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix owl-t: <http://www.w3.org/2006/time#>.

pl:middle-bronze-age-southern-levant
    a                      set:Period;
    set:scopedBy           :setting_3210 ;
    set:name               'Middle Bronze Age Southern Levant (2000-
                           1400 BC)' .

:setting_3210
    a                      set:Setting ;
    :definedBy             :interval_0123a ;
    :definedBy             :feature_2345 .

:interval_0123a
    a                      owl-t:Interval;
    owl-t:hasBeginning     :interval_0123aBegin ;
    owl-t:hasEnd           :interval_0123aEnd .

:interval_0123aBegin
    a                      owl-t:Instant ;
    owl-t:inXSDDateTime    "-2000-01-01T00:00:00"^^xsd:dateTime .

:interval_0123aEnd
    a                      owl-t:Instant ;
    owl-t:inXSDDateTime    "-1400-12-31T00:00:00"^^xsd:dateTime .

:feature_2345
    a                      geo:Feature ;
    set:name               'Southern Levant' ;
    geo:hasGeometry        '{WKT geometry expression}' .
```

Figure 0.9 A Period instance with spatial and temporal scope

The example demonstrates how the setting is defined by instances of TemporalScope and SpatialScope sub-classes (owl-t:Interval and geo:Feature respectively). Intervals and features are in turn described by temporal or spatial expressions, in this case ISO-8601 dates, and a Well Known Text (WKT) string.

In Figure 0.10 we show how the Setting for a Period can have only a temporal scope. In the same way, the Setting for a Place can be defined by only a spatial scope.

```
@prefix set: <http://somepatterns.org/setting/#>.
@prefix cd: <http://chinesedynasties.org/periods/#>.
@prefix geo: <http://www.opengis.net/ont/geosparql/1.0#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix owl-t: <http://www.w3.org/2006/time#>.

cd:Han_Dynasty
    a                       set:Period;
    set:scopedBy            :setting_5432 ;
    set:name                'Han dynasty of China' ;

:setting_5432
    a                      set:Setting ;
    :definedBy             :interval_0246 .

:interval_0246
    a                      owl-t:Interval;
    owl-t:hasBeginning     :interval_0246Begin ;
    owl-t:hasEnd           :interval_0246End .

:interval_0246Begin
    a                      owl-t:Instant ;
    owl-t:inXSDDateTime    "-206-01-01T00:00:00"^^xsd:dateTime .

:interval_0246End
    a                      owl-t:Instant ;
    owl-t:inXSDDateTime    "220-12-31T00:00:00"^^xsd:dateTime .
```

Figure 0.10 A Period instance with temporal scope only

Representing spatial-temporal uncertainty

The Bol list of gazetteer requirements includes “multiple indicators of temporality,” and the Setting pattern above begins to account for that, in that places are closely (if indirectly) bound to the temporal extents of their setting. We have indicated the OWL Time ontology is one means for encoding temporal extents, but that schema has limitations that will constrain historical applications. For example, the range of *hasBeginning* and *hasEnd* are restricted to instants, but historical work may require beginnings or ends expressed as intervals, with an associated probability or indicator of certainty or confidence. A Linked Data historical gazetteer will require suitable ontology patterns and data standards to account for this and some other needs.

Spatial and temporal information found in historical sources is frequently uncertain in one or more ways, and therefore difficult to map or compute over. Temporal statements can be vague (“around 1745”), ambiguous (“1745 or 1746”), conflicting (“1745” per source *x*; “1746” per source *y*), or referential (“after the war”). The timespans of periods and events can be cyclical (“the summers of her youth”), intermittent (“the lifespan of Poland”), or have either a distinct or fuzzy duration within another timespan (“for about 3 months in 1745”).

In work that is ongoing at this writing, Grossner and Meeks (2014) have introduced an open-source temporal data model called *Topotime*,[3] which addresses particularly those temporal requirements. Topotime extends the GeoJSON[4] data format used extensively for representing geographic features in web mapping applications, along the lines suggested by the *Setting* pattern described above. A Feature in a *Topotime* FeatureCollection can be a place, period, or event of any complexity, i.e. a simple singleton or a composite with multiple time-indexed parts. The GeoJSON "geometry" element that describes spatial extent is joined by a sibling "when" element describing temporal extents. The "when" object comprises one or more timespan that can be specified with quads for *start*, *latestStart*, *earliestEnd*, and *end*. Each of those can be assigned the qualifying operator *about* (~) and can be either an ISO-8601 date expression (e.g. YYYY-MM-DD) or a pointer (before, after, or equals) to another feature in the FeatureCollection. Topotime will support Linked Data representation, using the emerging standard JSON-LD[5].

Topotime is work in progress, but the qualities described above are essential to modeling "multiple indicators of temporality" for digital historical gazetteers. Others agree, as evidenced by other temporal models in development. The CIDOC-CRM[6] ontology for cultural heritage applications has recently incorporated a construct not unlike our own *Setting*, a "Space-Time Volume."

Challenges and prospects

Some of the ideas we have presented in this chapter are not new. Solutions for many important gazetteer elements are either in development or have been proposed elsewhere. Perhaps the biggest challenge lies in operationalizing those elements *together* in a system of systems that is general enough to see widespread adoption. In proposing a general modeling pattern for *Setting*, we provide a conceptual framing for how place and period—so closely bound in reality—can be joined effectively in various and otherwise distinctive data models. Our work demonstrates that place and period can be understood as perspectives instead of distinct entities.

Consequently, it is not surprising that researchers have begun developing a gazetteer of historical periods[7].

Gazetteers describe and locate places. We believe many places can be well described—perhaps even best described—in terms of what has happened at them. We can even say that in an important sense places are a function of what has happened there. Although we have only included periods in our *Setting* pattern, it seems evident that other temporal entities, such as events and activity, occur at and during spatial-temporal settings as well. A semantic modeling pattern particularly inclusive of events will also be important for at least some gazetteers.

When we ask “what is that place like?” from a historical perspective, the answer can take many forms, depending upon what kind of place it is—whether it is for example a city, a neighborhood, a mountain, or a geographic region. If it is a populated place, we want to know what people have done there and what the products of those activities and events have been. This can include events leading to its current form, events resulting in artifacts that are present there (from buildings to art works to engineered infrastructure), events leading to the place’s naming and re-naming, the economic and cultural activities that typify the place and how they have changed, and so on. The prospect of digital historical gazetteers that represent well the way in which places and occurrences are wholly bound together is an exciting one. Linked Data will almost certainly be a key element to achieving them.

References

Allemang, D., & Hendler, J. (2011). Semantic web for the working ontologist: effective modeling in RDFS and OWL. Elsevier.

Bol, P. K. (2011). What do humanists want? What do humanists need? What might humanists get? In M. Dear, J. Ketchum, S. Luria and D. Richardson (Eds.). (2011). GeoHumanities. Oxford, UK: Routledge. pp. 296-308

Braudel, F. (1980). On history (S. Matthews, Trans.). Chicago: University of Chicago Press. (Original work published 1969).

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., Annoni, A., Bian, L., Campbell, F., Ehlers, F., van Genderen, J., Goodchild, M., Guo, H., Lewis, A., Simpson, R., Skidmore, A. and Woodgate, P. (2012) Digital Earth 2020: towards the vision for the next decade. International Journal of Digital Earth 5(1): 4–21.

Frank, A. (1998). Different types of "times" in GIS. In M. Egenhofer and R. Golledge (Eds.), Spatial and Temporal Reasoning in Geographic Information Systems. pp. 40-62, New York: Oxford University Press.

Galton, A. (2005). Dynamic collectives and their collective dynamics. In A.G. Cohn and D.M. Mark (Eds.): COSIT 2005, LNCS 3693, pp.300-315.

Gangemi, A., & Presutti, V. (2009). Ontology design patterns. In Handbook on Ontologies (pp. 221-243). Springer Berlin Heidelberg.

Goodchild, M. F. (1998). The geolibrary. Innovations in GIS, 5, 59-68.

Grossner, K. and Meeks, E. (2014). Topotime: Representing historical uncertainty. In DH2014, Proceedings from Digital Humanities 2014, Lausanne, Switzerland.

Grossner, K., Goodchild, M. & Clarke, K. (2008) Defining a Digital Earth system. Transactions in GIS, 12(1):145-160.

Heath, T. & Bizer, C. (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Hill, L. (2006). Georeferencing: The Geographic Associations of Information. Cambridge, MA:MIT Press

Hornsby, K. S. and Yuan, M. (Eds.). (2008). Understanding dynamics of geographic domains. Boca Raton, FL: CRC Press

Hornsby, K. and Egenhofer, M. (2000). Identity-based change: a foundation for spatio-temporal knowledge representation. International Journal of Geographical Information Science, 14(3), 207-224

Hymes, R. (1969). Writing places: Chinese local histories. Columbia Library Columns, vol. XLV.

Janowicz, K. (2009). The role of place for the spatial referencing of heritage data. Paper presented at the Cultural Heritage of Historic European Cities and Public Participatory GIS Workshop, 17-18 Sep 2009, The University of York, UK. Retrieved 14 Jan 2010

Kauppinen, T., Mantegari, G., Paakkarinen, P., Kuittinen, H., Hyvönen, E., & Bandini, S. (2010). Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. International journal of human-computer studies, 68(9), 549-560.

Keßler, C., Janowicz, K., and Bishr, M. (2009): An Agenda for the Next Generation Gazetteer: Geographic Information Contribution and Retrieval. Proceedings of 17th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, November 4-6, 2009, Seattle, Washington, USA, pp. 91-100.

Mostern, R. (2008). Historical Gazetteers: An Experiential Perspective, with Examples from Chinese History. Historical Methods: A Journal of Quantitative and Interdisciplinary History, (41)1: 39-46.

Mostern, R. & Johnson , I. (2008). From named place to naming event: Creating gazetteers for history. International Journal of Geographical Information Science, 22(10):1091-1108

National Research Council, (Ed.) (1999). Distributed geolibraries: Spatial information resources. Washington, DC: National Academies Press.

Southall, H., Mostern, R., and Berman, M.L. (2011). On historical gazetteers. International Journal of Humanities and Arts Computing. Volume 5, Page 127-145

Stanford University Libraries (2011). Stanford Linked Data Workshop Technology Plan. Retrieved 29 Dec 2012 from http://www.clir.org/pubs/reports/pub152/ LDWTechDraft_ver1.0final_111230.pdf

Trame, J, Keßler, C. & Kuhn, W. (2013) Linked Data and Time – Modeling Researcher Life Lines by Events. In T. Tenbrink, J. Stell, A. Galton, Z. Wood (Eds.) Spatial Information Theory. 11th International Conference, COSIT 2013 Scarborough, UK, September 2013 Proceedings. Springer Lecture Notes in Computer Science Volume 8116: 205–223.

Wilkinson, E. P. (2000). Chinese history: a manual. Cambridge: Harvard University Asia Center.

Worboys, M. & Hornsby, K. (2004). From objects to events: GEM, the geospatial event model. In M. J. Egenhofer, C. Freksa, H. J. Miller (Eds.) Geographic Information Science: Third International Conference, LNCS3234: 327–344.

Yuan M. & Hornsby, K. S. (2008). Computation and visualization for understanding dynamics in geographic domains: A research agenda. Boca Raton FL: CRC Press.

---

[1] PLEIADES, a community-built gazetteer and graph of ancient places: http://pleiades.stoa.org/; PELAGIOS, linking together the places of our past through the documents that refer to them: http://pelagios-project.blogspot.com/ See also Chapter 8 in this volume.

[2] ibid.

[3] TOPOTIME, for representing complex and/or uncertain periods and events: http://dh.stanford.edu/topotime/

[4] GeoJSON, a format for encoding a variety of geographic data structures; http://geojson.org

[5] JSON is JavaScript Object Notation; a Linked Data compatible form for JSON objects is in development; see http://json-ld.org

[6] CIDOC-CRM, a conceptual reference model (CRM) developed by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM); see http://cidoc-crm.org

[7] PeriodO (Periods, Organized) has completed a first stage of an ambitious period gazetteer; see http://perio.do/