UNIVERSITY OF CALIFORNIA

Santa Barbara

Representing Historical Knowledge in Geographic Information Systems

A Dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Geography

by

Karl Eric Grossner

Committee in charge:

Professor Michael F. Goodchild, Chair

Professor Keith C. Clarke

Professor Helen Couclelis

Professor Ruth Mostern

Professor Martin Raubal

December 2010

The dissertation of Karl Eric Grossner is approved.

Keith C. Clarke

Helen Couclelis

Ruth Mostern

Martin Raubal

Michael F. Goodchild, Committee Chair

December 2010

Representing Historical Knowledge in Geographic Information Systems

ACKNOWLEDGEMENTS

I wish to thank the members of my dissertation committee, each of whom enabled and facilitated this somewhat unlikely undertaking by a "non-traditional" student. Michael Goodchild and Helen Couclelis heard out a grandiose idea, accepted me into one of the top geography departments in the world, then patiently guided me towards finding the right questions, with collegiality, warmth and a light hand. Keith Clarke's encouragement and always-open door during my first year at Santa Barbara were a helpful welcome to graduate studies. Ruth Mostern has inspired me with her enthusiasm for the project of joining history and geography with maps and computation, and her friendship and support has been a gift. Martin Raubal has also been a supportive mentor, always available and willing to talk through matters technical, cognitive and logical. I consider myself to be a student of all five; each has helped me in becoming a scholar. Collaborating with Don Janelle on the TeachSpatial project has also been a great experience and thanks to him, I have a first academic position. Thanks also to Professor Dan Montello, for teaching me something about experimental science and doing some with me.

When I visited Santa Barbara before applying to its Department of Geography, Helen Couclelis suggested I also contact Mike Goodchild, given his interest in something called Digital Earth. As luck would have it, on my way off campus I passed Professor Goodchild outside Ellison Hall and recognizing him from a photograph, asked if he had a minute. In fact we spent thirty minutes chatting about Digital Earth, historical atlases, the Minard map and GIS databases; he's been as generous ever since and his incisive comments and advice have been invaluable.

I am also grateful to a couple of folks who matter-of-factly assured me this would be doable and took the time to explain why. Professor Emeritus Lew Lancaster, a distinguished scholar of Buddhism and founder of the Electronic Cultural Atlas Initiative at UC Berkeley, suggested I "get a Ph.D. from the best geography department you can manage." Check. Professor Jacquelyn Chase, a human geographer and urban planner at California State University, Chico, introduced me to the field and collaborated with me on my first digital historical atlas, *Chico Landscapes*. Several fellow graduate students helped me get going and keep going: Jeff Howarth, Julie Dillemuth, Alan Glennon, Dayna Quick, Kirk Goldsberry, Josh Bader and Ben Adams.

I am grateful to all of my family for their support throughout this journey, especially cousin Jessica and her husband Jeff, who in a way made it possible.

DEDICATION

This dissertation is dedicated to my father and my daughter. Nathan Grossner was an electronics engineer par excellence who literally wrote the book on pulse transformers, a hi-fi and music buff, and a pretty fair carpenter who once turned a big chicken coop into a family vacation house. He never lost faith when he might have, as he knew a thing or two about tenacity, and he would have appreciated everything about how these graduate studies came to pass and turned out. My daughter Talia has always been one of my best qualities. By that I mean she is an extraordinary young woman—now a gifted school psychologist and a great mom—who has inspired me in the ways that she looks forward and helps others.

VITA OF KARL E. GROSSNER

December 2010

EDUCATION

Bachelor of Science in Instructional Technology, California State University, Chico, June 2005, *summa cum laude*

Doctor of Philosophy in Geography, University of California, Santa Barbara, December, 2010

PROFESSIONAL EMPLOYMENT

Associate Specialist II, Center for Spatial Studies, University of California, Santa Barbara, 2010-

Graduate Associate, Center for Spatial Studies, University of California, Santa Barbara, 2008-2009

Teaching Assistant (Human Geography), Department of Geography, University of California, Santa Barbara, Winter, 2006

Founder and President, World Heritage Web, Inc., Santa Barbara, CA, a 501(c) (3) California nonprofit corporation, 2002 -

Freelance web design and development, ClearEdge|Design, 1999-2005

Network systems consultant, Berkeley, CA & Minneapolis, MN, 1993-1999

Architectural Computing Systems Manager, Interactive Resources, Inc., Point Richmond, CA, 1983-1993

AWARDS

University of California Graduate Opportunity Fellowship, 2009-1010

Jack and Laura Dangermond Graduate Fellowship in GIScience, 2009

National Science Foundation IGERT Fellowship in Interactive Digital Multimedia, UC Santa Barbara, 2005-2008

SUNY-Buffalo IGERT scholarship for 6th Vespucci Summer Institute on Geographic Information Science in Fiesole, Italy, June 2008

AAG-NSF Student Travel Award, International Geograpical Union meeting in Tunis, Tunisia, March 2008

PUBLICATIONS

Grossner, K. (2010). Event Objects for Spatial History. In R. Purves, R. Weibel (Eds.) *Extended Abstracts Volume, GIScience 2010, Zurich*.

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

Grossner, K. and Clarke, K. (2007). Is Google Earth, 'Digital Earth?' − Defining a Vision. In *Proceedings of the 5th International Symposium on Digital Earth*, Berkeley, CA.

INVITED TALKS

Modeling Human Activity in Historical GIS for Education and Research, ESRI/Redlands University Colloquium, October 2009

IssueBrowser: The First Spatial History of the 2008 US Presidential Election Cycle, ThinkSpatial Brownbag Series, Center for Spatial Studies, UC Santa Barbara, November 2008

CONFERENCE PRESENTATIONS

Grossner, K., 2010, *Contested Knowledge in HGIS*, Social Science History Association (SSHA), Chicago, IL

Grossner, K., 2010, *Event Objects for Spatial History*, GIScience 2010, Zurich, Switzerland

Grossner, K., 2010, *Representing Historical Knowledge in GIS*, Interactive Demo, Space-Time Modeling and Analysis Workshop, Redlands GIS Week, Redlands, CA

Grossner, K., 2009, *Modeling Human Activity in Historical GIS*, Social Science History Association (SSHA), Long Beach, CA

Grossner, K., 2009, *Towards an Integrated Model of Human Activity* , Doctoral Colloquium , Conference on Spatial Information Theory (COSIT), Aber W'rach, France

Grossner, K., 2009, *An ontology-driven data model of the 2008 US Presidential Election Cycle*, Association of American Geographers (AAG), Las Vegas NV

Grossner, K., 2008, *The First History of the 2008 US Presidential Campaign: Modeling and Measuring Election Discourse*, Social Science History Association (SSHA), Miami FL

Grossner, K., 2008, *An Ontology-Driven Data Model for a Digital Historical Atlas*, Historical GIS 2008 (HGIS2008), University of Essex, Colchester, UK

Grossner, K., 2008, *Defining a Digital Earth System*, International Geographic Congress (IGC), Tunis, Tunisia

Grossner, K., 2008, *Spatiotemporal Primitives for Digital Historical Atlases*, Association of American Geographers (AAG), Boston MA

Villacorta, A., Grossner, K., Ventura, J., Hansen, A., Moxley, E., DeGuzman, J., & Peterson, M., 2007, *Spheres of Influence* (installation), ACM SIGGRAPH 2007 Art Gallery, San Diego, CA

Grossner, K., 2007, *Is Google Earth Digital Earth?—Defining a Vision*, International Symposium on Digital Earth (ISDE), Berkeley CA

Grossner, K., Goodchild, M., Clarke, K., 2007, *Defining a Digital Earth System*, Association of American Geographers (AAG), San Francisco CA

Grossner, K., 2006, *Is Google Earth Digital Earth?—Defining a Vision*, University Consortium for GIScience Summer Assembly (UCGIS), Vancouver WA

INVITED WORKSHOPS

LENS Summer Institute: Mapping People, University of Redlands, June, 2010

Visualizing the Past Roundtable & Workshop, University of California, Merced, March 2009

Spatial Thinking in Science and Design, University of California, Santa Barbara, December 2008

PROFESSIONAL AFFILITATIONS

Association of American Geographers, 2005-

Social Science History Association, 2005-
- Historical Geography Network co-chair, 2009-10

FIELDS OF STUDY

Major Fields: Geographical information science

Minor Fields: Spatial cognition; cognitive science

ABSTRACT

Representing Historical Knowledge in Geographic Information Systems

by

Karl Eric Grossner

A growing number of historical scholars in social science and humanities fields are using geographic information systems (GIS) to help investigate spatial questions and map their findings. The nature of historical data and historiographic practices present several challenges in using GIS that have been addressed only partially to date. For example, although events are inherently spatial and a fundamental construct in historical reasoning, there have been few attempts to create comprehensive data models for describing them. Likewise, computational representations of historical processes and narrative remain largely undeveloped. In this research, the emerging genre of digital historical atlas is presented as a broad use case and contextualized. Its representation requirements are detailed in novel conceptual and logical models of relevant geo-historical information constructs, presented as a generalized development framework. An event-centered and information-based spatial history ontology (SHO) was developed by adapting and extending an existing upper ontology (DOLCE). Its implementation in a spatial object-relational database populated with several historical datasets is described. Some important challenges remaining in this large, ultimately collaborative undertaking are discussed.

TABLE OF CONTENTS

TABLES

FIGURES

1 Introduction

> "Geography and history are different ways of looking at the world, but they are so closely related that neither one can afford to ignore or even neglect the other."
>
> (Baker, 2003:2)

> "Geography and history are not only analogous, but complementary and interdependent, bound together by the very nature of things."
>
> (Meinig, 1986:*xv*)

The best print atlases of world history are remarkable scholarly works. Typically authored by teams of historians, historical geographers and cartographers, their pages contain not only dozens of high-quality maps, but also narrative and analytical prose, timelines, statistical graphs, photographs and drawings (Figure 1-1). This dissertation research was initially motivated by the question of whether such atlases had a digital counterpart. If not, what technology would be required to enable their development? If so, how might they be enhanced and made easier to produce? The answer to the first question is that while some digital historical atlases do exist, the genre is still an emergent one; none approach the breadth and depth of the better print atlases. Concerning requisite technology, existing geographic information systems (GIS) are clearly a good starting point. They can provide the database-driven information storage and dynamic mapping to enable digital atlases to generate and display an essentially unlimited number of maps, as well as a wide range of spatial analytical operations. However, in order to achieve additional uniquely digital functionality, additional capabilities are called for. We can further escape the constraints of the static page to the extent our underlying data models permit simple formal descriptions of the things and occurrences we're concerned with.

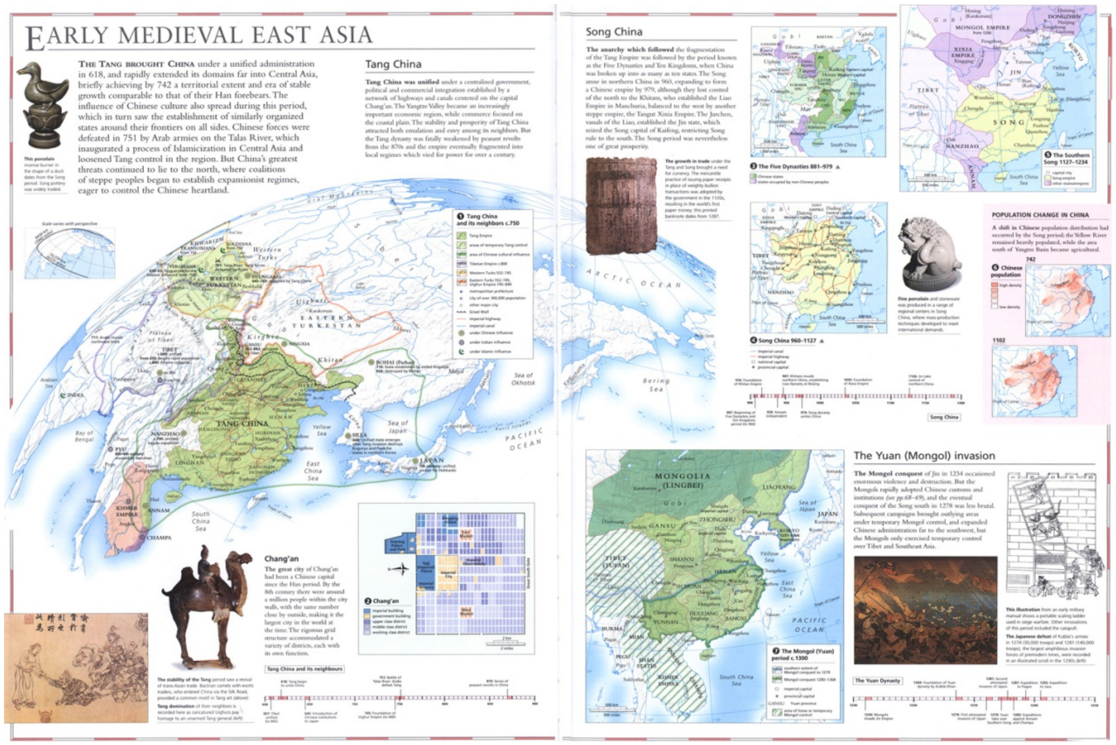


Figure 1-1. Plate from a world history atlas (O'Brien, 1999)

A growing number of historical scholars in social science and humanities fields are using GIS to investigate spatial questions (Gregory and Ell 2007; Knowles 2002, 2008). Quite apart from digital historical atlases, many of them have expressed requirements for systems having greater compatibility with historiographic practices, including by means of temporal and semantic extensions to typical GIS data models.

Modern GIS software is map-centric and, not surprisingly, has privileged the spatial attributes of geographic phenomena over the temporal and thematic. GIS was originally developed for the purpose of digitizing paper maps, and its earliest analytical functions stemmed from the concept of map overlays (Longley, et al., 2005). A distinctive perspective on GIS, and on geographic information science (GIScience) more generally, is that of automated, or computer-based analytical cartography (Clarke & Cloud, 2000). GIScience is said to be “the science behind the systems” (Goodchild, 1992; 2004), but that succinct formulation undersells another key point and its corollary: that it specializes the more general *information science* and that our expanding understanding of geographic information and its use should drive development of more advanced GIS software, made useful for more knowledge domains.

Although we can say that all geographic information has a spatial component, its temporal and thematic attributes are in many cases equally important; a more completion integration between all three planes is required in semantic data models for at least some GIS applications. Although many atlases are simply map collections, most modern print and digital historical atlases contain more than maps. Interactive timelines, directed graphs and animations will do a better job of depicting meaningful event sequences. Likewise, histograms, concept maps and other digital visualizations will be better at describing the dynamic, non-spatial attributes of many geographic phenomena. Spatiotemporally indexed multimedia objects can describe some essential qualities of Angkor Wat far better than its geometric footprint.

Galton (2005) has drawn a useful distinction between modeling systems (MS), as analytical software for simulation and prediction, and information systems (IS), as repositories of information with sufficient computational capability for some degree of inference, enabling sophisticated search and visualization. The requirements for these two classes of system have important differences, but as will be shown in this work, overlap to some degree as well.

This research develops novel conceptual and logical models of historical knowledge representation for GIS databases, intended to support both individually or collectively authored knowledge repositories such as the emerging genre of digital historical atlas, and standalone analytical applications in the domain of spatial history.[1]

1.1 Events and place

> "The conceptual vehicle by means of which historians construct or analyze the contingency and temporal fatefulness of social life is the **event**. Historians see the flow of social life as being punctuated by significant happenings, by complexes of social action that somehow change the course of history."
>
> (Sewell 2005:8)

Complex historical events are dynamic geographic phenomena: they comprise human activity associated with particular locations on the earth surface, and their participants' locations and attributes over time are integral to their analysis. Conceived as such, they represent a relatively unexamined domain for evaluating recent research on spatiotemporal data models in the field of geographic information science (GIScience).

Events are the central and most comprehensive container for information about

[1] Coined by Paul Carter (1987), spatial history "has become a broad umbrella term for scholarship that examines human experience of social and physical space" (Knowles 2008b).

dynamic geo-historical phenomena. To describe an event well is to account for its purpose and results, its participants' roles in component activities during some interval, its setting in terms of space-time locations and relevant condition states, and its relation to other events, including as elements of historical processes. The representation of large numbers of events along those dimensions will enable a powerful "faceted browsing" capability (*cf.* Gnoli, 2008; Yee, et al., 2003) and spatiotemporal analyses supporting the discovery of underlying processes. The power of events as information containers will stem in large part from typing them along their numerous dimensions. New methods in humanities scholarship have emerged from previous technological advancements (e.g. computational linguistics and visualization for the analysis of historical texts). I will argue that robust event representation can lead to novel, emergent historical analysis methods as well.

Geographer Doreen Massey describes place as "the meeting up of histories," (2005:4) a poetic view that might be readily explored computationally by modeling events. There is some interest within the GIScience community in developing computational models of place.[2] The description of a place in terms of what has happened there is a pivotal aspect of the envisioned digital historical atlases described in Chapter 3. Historian and political scientist William Sewell makes a similar observation regarding societal structure, which he calls "a product of the events

---

[2] An international workshop was held at the 2008 COSIT meeting in Aber Wrac'h, France, and followed by a 2009 special issue in *Spatial Cognition and Computation, (9)3*

through which it has passed," (2005:200) adding that "to understand and explain an event…is to specify what structural change it brings about, and to determine how [that change] was effectuated" (*Ibid.* p. 218). Computer scientists Westermann and Jain (2007) have described rich, navigable descriptions of social networks in terms of event participation.

The explicit representation of activity, events, and process—called "the spatiotemporal constructs of geographic dynamics" by Yuan (2009)—has been a significant thread in GIScience research for over two decades (*cf.* Hornsby and Yuan 2008; also, Chapter 2). If GIScience is to discover generic principles of representing geographic dynamics concerning potentially all disciplines, as proposed by Goodchild and Glennon (2008) and Yuan and Hornsby (2008), then intensive investigations into the dynamic aspects of natural and human phenomena for numerous and varied domains of interest are required. The particular requirements concerning historical event data have not been addressed comprehensively and are a core element of this research. Improved methods for the explicit representation of historical occurrences in geographic information systems will enable knowledge repositories capable of supporting the emerging genre of digital historical atlases, as well as development of novel spatiotemporal analytic methods by historical scholars.

So far, GIS representations of activity, event, and process have received surprisingly little attention in the historical social science community. One may ask whether this relative inattention to event modeling *per se* is due to methodological (historiographic) issues, or the technical challenges due to GIS deficiencies. A premise of this work assumes the latter: that the difficulty and expense in event coding is to blame, rather than lack of interest in events as objects of inquiry.

A significant proportion of the historical phenomena one might represent in a digital historical atlas can be abstracted to directional space-time paths, including journeys, migration, cultural diffusion and trade flows. Although flows for hydrological, transportation and utilities networks can be modeled with composite datatypes in commercial GIS, and extended to support other domains (Glennon 2010), temporal aspects are problematic. Some progress in modeling paths and flows for the historical domain has been made in this research (§6.3.7).

GIS data models have naturally privileged the spatial attributes of geographic data, and despite the attention given to temporality since the 1980s, we still hear that GIS doesn't "do time well." It is also the case that the *what* of data—the thematic attributes of objects being modeled—is underspecified, in part due to the inadequacy of data models. Research on geospatial ontologies addresses that set of issues. In this dissertation, event-centeredness means joining representations of the *when* and the *what* in semantic models for temporal constructs.

I present a simple conceptual model unifying space, time and theme for the historical knowledge domain, in which an event is composed of one or more activity-instances as well as sub-events, and may itself be asserted as belonging to one or more historic process or period. Events and historical processes can be complex, in the sense of being composed of many interrelated parts (sub-events or sub-processes), having significant and frequently non-contiguous spatial extension, and having multiple participants. Activity and state are normally modeled as being homogeneous, i.e. ongoing throughout an interval. All but state may have results or products, including other occurrences and artifacts, and all reflect or exert causal influences driven by human purpose. Finally, the observational data and propositional knowledge constructs concerning all of these aspects of occurrences are themselves the products of events, and documented or asserted in information objects of various kinds. Dynamics are captured in several key relations, notably participation and membership.

One can argue that virtually all phenomena could be referenced to discrete temporal entities, and I argue that by doing so, spatial, temporal and thematic perspectives may be integrated to a novel degree.

1.2 Problem statement

When the question was posed at a recent gathering of GIScience researchers, "what societal benefits has our research produced in the last twenty years," one response noted the ubiquity of interactive digital maps that "let people learn what is happening anywhere." This prompted an incisive comment, "no, we can learn what *is at* nearly any location, but not necessarily what *is happening.*"[3] This is not to say that researchers in some disciplines cannot model what is (or was, or may one day be)

[3] The occasion was the 20th anniversary reunion of the National Center for Geographic Information and Analysis (NCGIA) in Santa Barbara; the comment was by Andrew Frank.

happening with a growing array of spatiotemporal analytic tools (Weaver et al. 2006; Rey and Janikas 2006; Rey and Anselin 2007), but it does suggest that a) the robust treatment of geographical dynamics is fairly new, b) the means for representing results lags those for analysis and c) some knowledge domains are as yet underserved.

Conceptualizations of physical geographic processes and of human activity as rule-based agent behavior have understandably led to mathematical models and simulation algorithms as representation frameworks for geographic dynamics—*process objects* per Goodchild (2004). However, the domain of historical scholarship is largely unconcerned with deterministic models or the discovery of laws governing human activity.[4] Natural language and logic are far better suited to representing historical conceptions of dynamic phenomena in terms of Sewell's *event* and of *process*, as ". . . a series of actions or events; a proceeding; a succession of things in order" (Process, n.d.). That said, digital representations of such historical events and processes should afford many kinds of quantitative, qualitative and visual analyses along spatial, temporal and thematic dimensions. Analysis is necessarily preceded by description, so robust event description is a good place to start.

The suggested reinvention of the print historical atlas genre as geo-historical information systems and knowledge repositories requires both a conceptual framework and specific means for formally describing dynamic human activity that are compatible with historiographic practices. Towards that end, the following

[4] Methods used by some historical economists are a notable exception.

questions are posed and answered in this dissertation: (i) What may digital historical atlases be? (ii) What are their representational requirements? and (iii) What extensions to 'traditional' GIS data models will effectively meet those requirements?

1.2.1 Specific objectives

These research questions have led to the following specific objectives in this dissertation:

1. Enumeration and further definition of representation requirements for an emerging genre of digital historical atlas, including a set of competency questions with which resulting models may be evaluated;
2. Development of a general representational model—one that can be implemented in the object-relational spatial databases typically used in geographic information systems;
3. Development of a particular exemplar system demonstrating novel capabilities resulting from the representational model that meet most if not all of the discovered requirements.

1.3 Approach

The goal of this work is a general data model for representing and contextualizing knowledge about complex geopolitical events such as wars, and about human activity and processes as diverse as settlement, trade, cultural and technological diffusion, popular movements, and the rise and demise of states. Such representations of the dynamic structure of temporal entities and their participants will form a basis for embedded or attached reasoning systems. Assuming it will be useful to explicitly represent occurrences formally in object-relational databases such as underlie GIS, we can ask, how best to go beyond the normal case now, where an event is most often simply a geometry with a text string label and associated timestamp or interval? The simple answer: an extensible domain ontology for spatial history, expressing the meaning of those labels in an object-relational database schema.

The term ontology is used here in the computing sense, as "an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words" (Guarino, 1998:4). Technically, ontologies are formal specifications of such vocabularies in terms of entity classes (types), properties (also termed relations or roles), axioms (rules) and functions (operators)—i.e. "dynamic, object-oriented structures that can be navigated" (Fonseca, et al., 2002:232).

This structure permits creation of a *knowledge-base*: a collection of statements (reducible to <subject, *predicate*, object> tuples) and axioms, from which we may infer other statements using some subset of first-order logic such as OWL-DL, or the relational calculus of database queries. A knowledge-base is distinguished from a database in supporting the *open-world assumption*, essential to reasoning about the incomplete data and conflicting accounts of the historical domain. Databases are said to make the *closed-world assumption*, wherein what is not known is false (Sowa 2000). It is argued in this work that open-world knowledge-bases may indeed be implemented in object-relational database software, by (i) allowing null values to be interpreted as unknown, and (ii) permitting entries with duplicate values in field sets that in closed-world systems might comprise unique keys.

The *ontology-driven information system* described by Guarino (1998) has a four-part architecture comprising a *top-level* (often termed *upper*) ontology, generic *domain* and *task* ontologies and *application* ontologies dependent upon the three levels above. The development method for this dissertation research is *ontology-driven* in the sense of seeking to identify or create a suitable upper ontology, then experimentally extend it for a newly theorized knowledge domain. The result may ultimately contribute to a growing global library of "foundational ontologies" as envisioned for the *WonderWeb* infrastructure (Masolo et al. 2003).

Ontologies in this sense should be the basis for GIS databases underlying digital historical atlases because they help address issues of scope, interoperability, and the nature of historical data and methods. As will be shown in §4.2, print historical atlases come in a wide range of information types and perspectives. We should expect the range for digital historical atlases will be wider yet. Ideally, their underlying databases will have as many re-use and 're-purposing' characteristics as possible, so that development costs might be recovered to some degree, and for the collaborative potential provided.

The challenges presented by historical data also suggest the appropriateness of an *ontology-driven GIS* (ODGIS) as proposed by Fonseca, et al. (2006): "The requirements of a next-generation GIS that can be fulfilled by an ODGIS architecture [include] the ability to support representations of incomplete information, multiple representations of geographic space, and different levels of detail" (p. 236). The approach taken in this work presumes we should design a theoretical system at the limit first, and tackle implementation issues afterwards. Once the problems associated with formal-logical representation of multiple conflicting statements, vagueness, and uncertainty are addressed, the visualization and other challenges that inevitably result will become clearer.

Given an alternate definition of ontology as “a neutral and computationally tractable description or theory of a given domain which can be accepted and reused by all information gatherers in that domain,” (Smith & Mark, 2001, p. 594) we must grant that human history is far too large and complex a realm of inquiry to be described in a single theory; after all, one can say broadly that historical information describes everything having occurred before now. In such a vast information space, there will be “a multiplicity of complementary ontologies – distinct perspectives on reality, each one of which is veridical” (Smith and Grenon 2004). I take “perspectives” here to include variations in thematic, spatial and temporal extents, as well as in terminology, granularity, and completeness.

The notion of a single domain ontology for all historical research and education is extremely doubtful and undesirable in any case. We require an extensible ontological framework that can model numerous perspectives, making no claims of completeness for the domain. Models (including semantic) are useful constructs for discovery, analysis and discussion—nothing more.

If any number of ontologies may be correct, each expressing a particular viewpoint or perspective on reality, can the solutions proposed in this dissertation be evaluated? The answer is two-fold and partially deferred: the <u>approach</u> is successful if it ultimately adopted by developers of digital historical atlases that find extensive use in research and educational settings. The <u>execution</u> is successful in this case if a resulting exemplar system can (i) represent the most important knowledge constructs for the domain, and (ii) provide sufficient computational reasoning capability to infer answers to the questions system users can be expected to ask.

The choice of a top-level ontology entails commitments—to a logical temporal framework for example (§5.3)—but the relationship between top-level and domain ontologies is seen to be a two-way street by their authors (Masolo et al. 2003; Grenon and Smith 2004; Crofts et al. 2008; Niles and Pease 2001), in that the viability of a top-level ontology is assessed by testing its ability to support multiple disparate domains. This work constitutes such a test.

This ontological framework will not attempt to classify or describe all historical information. The intent is that it will:

1. Stem from an existing upper ontology, acting as a bridge to (and providing interoperability between) successively finer-grained domain ontologies.
2. Be extensible, permitting implementation of narrower controlled vocabularies for individual digital historical atlas and historical GIS projects having relatively narrower thematic, spatial and/or temporal extents.
3. Provide sufficient logical inference capability to support faceted browsing and to enable a class of queries and related application features I have termed *historiographic zoom*—the coordinated navigation through and between the three dimensions of geographical inquiry: space, time and theme.
4. Be informed by cognitive principles to some degree. After all, humans are the

users of GIS and of digital historical atlases.

Finally, however real the entities historians study are, the ontology for spatial history developed here is of one of information, and not “the real world” per Frank (2003) and Couclelis (2010).We wish to model historians’ conceptualizations, as we understand them via language.

1.3.1 Theoretical foundations

Given these stated objectives and approach, this research in a sense merges and expands upon three earlier theoretical frameworks by others: the notions of *semantic reference systems* (Kuhn 2003); *event objects* (Worboys 2005; Worboys and Hornsby 2004), and *information ontologies* (Frank 2003; 2007, Couclelis, 2010). These are discussed briefly here, and in corresponding sections of Chapter 2.

Kuhn has argued convincingly (2003; 2005) that *semantic reference systems*—multi-level frameworks supporting multiple ontologies—are needed in order that representations of thematic data approach equal computational footing with spatial and temporal data. In this work I presume that useful data models for the geo-historical information systems supporting digital historical atlases *must* follow from an underlying extensible ontological framework; that is, from a ‘spatial history-compatible’ upper ontology supporting unlimited theme- and project-specific extensions ‘beneath’ it.

Worboys has made a strong case for granting “first-class” object status to temporal entities, as one requisite for the “third stage in development of spatiotemporal information systems…a full-blooded treatment of change, in terms of events and actions" (Worboys 2005). Owing to the centrality of events and processes in historical studies, it is self-evident that conceptual and computational models corresponding to the historical senses of those terms are necessary.

Frank has drawn an important distinction between "philosophical ontologies" of reality and "information' ontologies" (2003). Both Couclelis (2010) and Frank (2003; 2007) have proposed models of the information- and knowledge-creation process, from the perception and measurement of 'raw' phenomena through simple object identification and classification, to theoretical knowledge constructs about complex processes. These conceptual frameworks support the requirements of systems addressed by this dissertation research—systems that explicitly represent assertions by multiple authors about historical reality in *possible worlds* rather than in a single, closed-world reality. In this domain, events and processes must be modeled as information constructs—assertions about event composition, theorized causal relations between events, and relevant settings—rather than 'real' temporal entities.

1.3.2 Use cases: analysis and repositories

The development trajectories for Galton's *modeling systems* and *information systems* mentioned at the outset have quite naturally differed. Theoretical investigations of their commonalities with respect to data models can contribute to realizing greater semantic interoperability amongst distributed systems (Kuhn 2003).

Although occurrences comprising historical human activity *are* geographic, the quest for law-like statements about them is controversial when compared to that for dynamic natural phenomena. Arguably, the latter are more reliably simulated in mathematical models and software programs. We can ask: do the requirements for systems addressing the 'natural' and 'human' cases differ, and if so, at what stage of scientific inquiry do they diverge? Both require representations of observational data, i.e. instances of a state, event or activity at a location and time with measured properties, whether it be ocean temperature and chemical composition, or the presence, identity and attributes of a particular historical individual, feature or artifact. Geo-statistical analyses can reveal patterns of structure, motion or change in both cases. A discovered spatiotemporal pattern or structure might be modeled, named and represented as a composite *event* (Worboys 2005), or a *process object* reified as a "formalize(d) knowledge of process, allow[ed] to become part of the digital environment" (Goodchild 2004). In the natural realm, such temporal objects are frequently a basis or outcome of mathematical, *predictive* models; the same is true for a significant segment of social science and urban planning research.

One divide then between requirements concerning natural and human phenomena is the relative strength of predictive claims due to the presence of human agency. Another concerns the scope of questions that get asked concerning human phenomena. Although "what's past is prologue," (Shakespeare, 1999/1604) the descriptive models and computational analyses of human activity by historical geographers and historians normally stop short of prediction, at understanding and explanation. Urban planners, another category of spatial information system users, create normative models for spatial design of systems and artifacts corresponding to function and goals. Notwithstanding some interesting exceptions (Heise 1991; Howarth 2007), purpose and cause are seldom explicitly modeled, although theoretical support exists in, for example, Allen's logic of actions and events (Allen & Ferguson 1994). It seems fair to say that by virtue of formal ontologies, Galton's IS class of information systems can increasingly include descriptive models directly supporting analytical if not mathematical methods.

To reiterate, a general model is the goal of this work, one capable of representing complex events and human activity in any context: geopolitical interactions, intellectual and technological diffusion, wars or the rise and demise of states. Such representations must begin from a high-level ontological framework, extensible and flexible enough to support domain ontologies of increasing detail beneath.

1.4 Methodology

The research problem and objectives in §1.2 describe a software design problem and a direction toward their solution, but in a very general sense. In fact it is a *software genre* design problem. The methodological approach taken is information-based, ontology-driven, and as detailed in §1.1, event-centered.

It is information-based in seeking to support particularly a broad class of spatiotemporal knowledge repositories and related software. To an extent it seeks to model stages of knowledge creation, from observation and measurement to analysis and synthesis at the service of human goals. The notion of information objects is integral to this—these can be both the subjects of analysis (speeches in one exemplar, §7.2.2), and those tables and documents storing cited measurements and assertions of process.

An ontology, as a core element of a *semantic reference system*, can provide a very useful level of reasoning even when using relatively few logical constructs: is-a and part-of relations, domain and range axioms (e.g. has-member(group, person)), and cardinality (at-least, exactly, at-most).This logic provides a way of describing and discovering spatiotemporal structure and pattern and is readily modeled in languages like RDF/S, OWL or SUO-KIF. It may be expressed in the algebra of relational database management systems (RDBMSs), and in procedures and scripts within an RDBMS (albeit more problematically). Guarino (1998) and Fonseca (2006), have discussed how merged top-level and domain ontologies can play a central role either at "development time," at "run time," or both. In the first case, they inform the conceptual and logical database design; in the second, they also serve as the data store for supported software. In this work a top-level ontology was a conceptual modeling tool, an extended version of which is implemented in a particular object-relational database schema.

1.4.1 Sufficient logic

Ontologies as formal descriptions of the entities and relations in a domain of discourse can be built with a range of expressiveness. That is, each will afford some degree of inferential reasoning capability. Historical knowledge representation is significantly constrained by data quality issues including sparseness, imprecision and uncertainty. This research therefore has limited goals with respect to enabling logical completeness system-wide. Rather, it is assumed that a great deal of benefit may obtain from relatively few logical formulations, including (i) the hierarchical and transitive identity (is-a) and composition (part-of) relations, and (ii) event participation and group membership relations. Further discussion of logical sufficiency appears in Chapters 6 and 7.

1.4.2 Engineering an information ontology

> "…indicate precisely what you mean to say…"
>
> (Lennon and McCartney 1967)

The process has followed these steps: (1) identify what is to be modeled; (2) build a conceptual model that accounts for all entity classes and relations of interest; (3) identify the top-level ontology best suited to the knowledge domain; (4) extend and modify that ontology as necessary by fitting the domain entities to it, ensuring a significant subset of the RDF/OWL logical expressiveness; (5) build a particular database on a schema expressing the new ontology, populate it with exemplar data and demonstrate novel utility with functions capable of answering a set of "competency questions."

The specific tasks within those steps are listed here, with references to relevant chapters:

1. High level description of the domains of interest and use scenarios (Chapters 1

and 3).

2. Identification of entities to be represented, as discovered in print historical atlases and existing historical GIS projects (Chapter 4).
3. Development of a high-level conceptual model of the domain being modeled, producing a set of *geo-historical information constructs* (Chapter 5).
4. Enumeration of 'informal competency questions' that users of digital historical atlases may wish to ask about those entities (Chapter 5).
5. From those questions, eliciting the core relevant relations between entities—both explicit and implicit (Chapter 6).
6. Evaluation of several upper ontologies as suitable candidates for supporting a domain ontology for spatial history (Chapter 6).
7. Fitting entities to be modeled to the chosen upper ontology, creating modifications and extensions as necessary (Chapter 6).
8. Expressing the resulting ontology (SHO) in an object-relational database, then loading and fitting varied exemplar datasets to the data schema to create a knowledge base capable of supporting possible atlas and analytical applications (Chapter 7).
9. Developing functions and materialized views capable of answering the competency questions from the ontology and instance data, providing a basic means of evaluating the ontology's effectiveness (Chapter 7).

1.4.3 Evaluation

Elements of an ontology development process described by Uschold and Gruninger (1996) are adopted in this work and reflected in the steps just enumerated. Steps 1, 4 and 9 refer to *scenarios* and *competency questions*. Scenarios are narratives of system functionality that can be thought of as broad use cases; several for this domain are outlined in Chapters 1 and 3. From Uschold and Gruninger (p. 113):

> "Given...motivating scenario[s], a set of queries will arise which place demands on an underlying ontology. We can consider these queries to be expressiveness requirements that are in the form of questions. An ontology must be able to represent these questions using its terminology, and be able to characterize the answers to these questions using the axioms and definitions. These are the informal competency questions..."

The questions developed for this work are listed in §5.1. The SQL functions that answer them appear in §7.1.3.

1.5 Contributions

This research contributes to theory and practice in two fields—geographic information science (GIScience) and historical GIS (HGIS), the development of which Anne Knowles has traced from "a compelling methodology" (2002:*ix*) to "a scholarly practice increasingly recognized as an interdisciplinary subfield within historical studies" (2008b). I synthesize and expand upon recent theoretical work in GIScience concerning representations of spatiotemporal phenomena, to meet the particular requirements of historical scholarship and education. Products of this research include an ontological framework for geo-historical computing and exemplar domain concept taxonomies of complementary granularity that together demonstrate a suitable foundation for a broad range of historical data modeling tasks.

Most HGIS projects to date have represented human activity in terms of *state*, i.e. conditions of some geo-referenced entities in one or several temporal snapshots. Events and historical processes are most often implicit: for example, population statistics imply birth and migration events and historical maps of transportation networks imply the various economic activities and processes related to their development. While the explicit representation of *activity*, *event*, and *process* has been a significant thread in GIScience research for two decades, it receives relatively less attention in the historical social sciences.

Digital historical atlases have the potential to advance geographic, historical and spatial literacy by providing compelling learning environments, and to help forge better connections between research and education by providing a venue for scholarly work presented in an accessible manner to students and the general public. Early results are promising, however new types of database representations will pose significant challenges to digital cartographers.

The approach is novel in these respects: (1) *historical-processes* can be modeled as theories of event relations, incorporating cause and purpose both explicitly and implicitly; and (2) *information objects* are prominent as both artifacts and in a sense as embedded metadata, documenting the observations, measurements and assertions of the knowledge construction process.

1.6 Outline of the dissertation

In Chapter 2, relevant research from several fields is reviewed. Chapter 3 further develops motivational application use cases, including the visionary Digital Earth and geolibraries, and particular digital historical atlas in development, Cultural Heritage Web. Chapter 4 begins a series of four sections of the dissertation that trace "the stuff of history;” sections §5.5, §6.3, and §7.2 follow a process of discovering, classifying and representing *geo-historical information constructs* (GHICs) in conceptual, logical and physical models. Chapter 4 also examines print historical atlases, existing historical GIS projects and historical scholars’ desiderata for future systems, producing a preliminary enumeration of entities to be represented.

Chapter 5 presents a conceptual model that consolidates many requirements for historical knowledge representation in GIS. In Chapter 6, a novel spatial history ontology (the SHO), which has been fit to the DOLCE upper ontology, is described. Chapter 7 shows how the SHO has been implemented in an actual spatial RDBMS, holding varied exemplar datasets. Chapter 8 concludes the dissertation with discussions of results and future work.

2 Background

2.1 Introduction

This dissertation helps to define an emerging genre of information system, the *digital historical atlas* and takes steps towards a particular realization. In Chapter 1, I rationalized the need for such systems, proposed research questions and objectives, and outlined a methodology for proceeding with the development of an event-centered *spatial history ontology*. In §2.2 through §2.6, I review the literature that bears upon those research questions and the chosen approach to answering them.

This work proceeds from several strands of research, including temporal GIS, geographic ontologies and event modeling, adding novel elements to a synthesis of ideas and methods from those domains. The broad goal is itself a synthesis—computational systems supporting Berry's methodological vision (1964) integrating *place*, *time* and *characteristic* as dimensions of geographical phenomena that may be viewed at various levels: the "*static structure* of frameworks in space and time," "c*onnectivity* of places, flows and interaction," and "*dynamic interrelated processes*" (p. 10). Variations on this goal have appeared as a steady stream in GIScience literature, as for example in conceptual frameworks from Peuquet (Triad; 1994), Yuan (Three-Domain Representation; 1999), and Mennis, Peuquet and Qian (Pyramid; 2000; also, Peuquet, 2002). Particular data models for physical phenomena followed from those, including for "applying knowledge acquisition techniques," (Yuan 1997), and for supporting "complex spatiotemporal queries" (Yuan 1999) and "knowledge discovery in databases" (Mennis and Peuquet 2003). The need for further developments has been expressed in, for example, Galton's desiderata for a "spatio-temporal geo-ontology" (2003) and in two successive UCGIS[5] research agendas (McMaster and Usery 2005; Hornsby and Yuan 2008).

2.2 Spatial history and computation

> "Geography and history are alike in that they are integrating sciences concerned with studying the world. There is, therefore, a universal and mutual relation between them, even though their bases of integration are in a sense opposite—geography in terms of earth spaces, history in terms of periods of time."
>
> (Hartshorne 1939/1996)

An informal survey of several recent issues of the *Annals of the Association of American Geographers* revealed the topics of more than one-third of the articles to be historical. That is, they involved temporal as well as spatial analysis, with such topics as surname migration in the UK, the several impacts of a single 19th century Midwest canal, and a history of corporal punishment in Cornwall. Even so, the number of self-described practitioners of Historical Geography is quite small, and Geographical History even smaller. Baker (2003) has noted geographic perspectives in historical practice, have been narrowed in many cases to concepts of spatial variation and cartographic methods.

The philosophical underpinnings for this work begin with Kant's characterization of geography and history as co-equal, holistic, integrative sciences concerned with

[5] University Consortium for Geographic Information Science

spatial and temporal explanations of natural phenomena, respectively (Livingstone, 1992; Barnard, 2001). Fernand Braudel, a prominent voice from the *Annales* school of French historians, repeatedly expressed affinity for geographic perspectives in historical scholarship, declaring even that "spatial models are the charts upon which social reality is projected, and through which it may become at least partially clear; they are truly models for all the different movements of time (and especially for the *longue durée*), and for all the categories of social life" (1969/1980:52). The view of history as a science—at least a scientific undertaking—might seem counter-intuitive in recent times, but Novick (1988) has traced a robust objectivist heritage in the field, noting that "orthodox historical method…decree [s] that any reputable generalization had to be consonant with all the discoverable evidence" (p. 583). I am also inspired by Sauer's brand of regionalism and cultural landscape studies, discussed in for example *The Morphology of Landscape* (1925/1996). Sauer describes how examination of the human material record reveals evolving, inherently historical social structures. One wonders how such investigations of dynamic social structure would have played out with the computational tools of this digital age.

Historian John Gaddis has argued for his discipline's intrinsic spatiality and scientific rigor, in describing a conceptual framework that will be familiar to geographers and lends itself well to computation: "If time and space provide the field in which history happens, then structure and process provide the mechanism" (2000, p. 35). He draws parallels with geologists and physicists who also study unseen phenomena and seek parsimonious explanations: "[…] historians too start with structures, whether they be archives, artifacts or even memories. Then they deduce the processes that produced them" (p. 41).

Regarding the undertaking of "representing historical knowledge," some statement about perspective and scope is called for. History as a discipline is defined in the Oxford English Dictionary (history, n.d).as:

> 3. That branch of knowledge which deals with past events, as recorded in writings or otherwise ascertained; the formal record of the past, esp. of human affairs or actions; the study of the formation and growth of communities and nations.

There are many perspectives on historical events—positivist, neo-Marxist, feminist, post-structuralist, post-modernist, and so forth—but historians by-and-large share that element of the professional practice found in a second definition, for (*small h*) history:

> 2. A written narrative constituting a continuous methodical record, in order of time, of important or public events, esp. those connected with a particular country, people, individual, etc.

That reference to seemingly simple chronicles omits important aspects of true historical works, which are ordinarily interpretive and go beyond delineating a factual record to making arguments for causation or purpose, for example. For Sewell (2005), "chronology is crucial because it tells us within what historical context we must place the actions, texts, or material artifacts we are attempting to interpret or explain" (p. 11).

The digital historical atlases and historical GIS projects enabled by this research will support interested historical geographers, as well as the growing number of historians and social scientists (and their pupils) who view their fields as inextricably intertwined with geography. Most of them would probably self-identify as positivists, but they could easily have other, overlapping philosophical or political perspectives, including any of those listed earlier. Narratives derived from a "formal record of the past," should benefit from the analyses and visualizations computational systems offer.

This research seeks to model historical narratives, interpretations or arguments in only a skeletal manner—to enable creation of a navigable *factual substrate* which I hypothesize will facilitate knowledge discovery and may be cited to support interpretive works by historical scholars of any philosophical bent. An excellent example of this is Monica Smith's exposition of the way that polygonal areas drawn to represent Incan regions on maps were in actuality defined by networks (2005). A historical argument's full exposition will come in prose, which can be more readily presented as spatially and temporally indexed documents. In this, I draw further inspiration from Braudel: "I believe in the reality of a particularly slow-paced history of civilizations, a history of their depths, of the characteristics of their structure and layout." This historical perspective, which he termed the *longue durée*, requires that "…a whole vast body of documentation…be brought to light so as to be able to answer the new questions" (1969/1980:12-13).

Some historians and critical theorists in the social sciences will be incredulous at the notion of a politically neutral "factual substrate." There are valid concerns along these lines, but as noted, a growing number of historians, social scientists and geographers clearly feel otherwise and this work is a contribution to that community. Gahegan and Pike (2006) have shown that in fact, data models accounting for "situated" and contested knowledge constructs are critical to approaching the sought-after neutrality. Historian David Staley (2003) and others have argued persuasively that the acts of organizing, encoding and mapping historical information in a GIS themselves constitute useful scholarly practice. I argue that data structures can be independent of content or political message. If perfect neutrality in the labeling and classification of concepts were required for scholarship, libraries would be empty.

2.2.1 Historical GIS

Historical GIS seems to be gradually gaining credence as a distinct sub-field. The 2010 meeting of the Association of American Geographers (AAG) featured a day-long HGIS session track, organized by Harvard University historian Peter Bol and AAG Executive Director, Douglas Richardson. A similar track is planned for the 2011 meeting, as is a Space-Time Symposium that lists among its focus topics "historical time and HGIS," and "ontological frameworks." The Historical Geography Network of the Social Science History Association (SSHA) organizes numerous sessions focusing on HGIS, and it was a central element in fifty percent of the network's papers for the 2010 meeting.

Two significant contributions to defining the field have been published recently. In *Historical GIS: Technologies, Methodology and Scholarship* (2007), geographers Ian Gregory and Paul Ell provided a comprehensive view of the state of the discipline, and importantly, how the three elements in the title fit together both theoretically and at various scales in particular recent research projects. In *Placing History* (Knowles, 2008b) several recent projects are profiled in some depth by their principal investigators. Other contributing authors, including both historians and geographers, outline its potential and major research challenges.

As befitting "integrating sciences," there is a growing community of social scientists and humanities scholars who view spatiotemporal computing systems like GIS as invaluable research tools that are shaping the practice of historical scholarship in positive ways (*cf.* McCarty 2004; Mostern 2008; Staley 2003). For McCarty, although "there is a radical difference between what we know and what we can specify computationally," the struggle to construct useful models is itself instructive, in revealing "tacit and inchoate knowledge" (p. 256). On this view, the inherent requirements of GIS for certainty and precision, so often lacking in historical source material, may be as much a positive as a constraint—a crucible for characterizing uncertainty. Staley (2003) notes the utility of maps as "visual abstractions of primary sources," (p. 75) requiring the same selectivity as historians' crafting of words and sees some of McCarty's fortune in adversity, in the GIS capability to represent and explore counterfactuals.

2.2.2 Event models for history

For a number of HGIS projects, events are central, but without an underlying ontology, remain abstract nodes, i.e. identified by terms that are undifferentiated with respect to type, and capture little meaning. Studies of disease and crime exemplify this.

Shaw, et al. (2009) have developed the LODE data model, a lightweight ontology for representing events in the Linked Open Data paradigm of the Semantic Web. They examined event models within several existing ontologies, including DOLCE[6] and CIDOC-CRM[7]. Their simple set of modeled relations include agentive participation, non-agentive involvement or presence, and temporal and spatial location. The LODE model has not yet been related to the spatial analytic and mapping capabilities of GIS, but it does exemplify how the inherent centrality of events as information 'containers' means that a relatively limited set of relations will permit powerful search, browsing and inference capabilities.

The FinnONTO[8] project has produced an event-centered semantic "metadata model" for harmonizing heterogeneous schemas in the domain of material cultural heritage (Ruotsalo & Hyvönen, 2007). The core relations of that model—*place*, *time*, *participant* (incl. *agent* and *patient*), *goal* and *instrument*—are similar enough to the models developed in this dissertation to suggest future compatibility. The authors make a strong case for events as comprehensive information containers for the

---

[6] Descriptive Ontology for Linguistic and Cognitive Engineering (Masolo, et al., 2003).

[7] The International Council on Museums' International Committee for Documentation-Conceptual Reference Model.

[8] A product of the Semantic Computing Research Group at Helsinki University of Technology (www.kulttuurisampo.fi; www.seco.tkk.fi/projects/finnonto)

historical domain, a point this research is in complete agreement with (§5.2). Their work on events is not yet ontological.

Mostern and Johnson (2008) have explored event models as a basis for digital gazetteers and timelines. They show that place names and place-naming events hold or reference a great deal of historical knowledge and introduce the idea of a *historical event gazetteer* that would index large stores of historical knowledge about places to place naming events, situated in larger event chronologies. It also described a pilot project and listed several outstanding challenges facing those who would “model history with events.” The modeling of causation is a stated goal, but awaits a unifying ontology of event relations. They reviewed several approaches to event modeling taken by historians and found that even the most promising relevant projects at that time did not incorporate spatial information. Heise’s Event Structure Analysis (ESA) (1991) offers a methodology and associated software, *ETHNO*, for representing logical relations between events within larger, complex event sequences. ESA “creates sequences of social events that are grounded in chronology and linked together to create historically and temporally based causal interpretations of history” (Mostern & Johnson, p. 5-6). It has the potential for discovering and demonstrating causality but omits spatial information entirely. This was the case for two other projects reviewed, the Temporal Modeling Project (Drucker and Nowviskie, n.d.) and SemTime (Jensen, 2006).

Several projects by social scientists have undertaken to represent and analyze large quantities of historical events. Historical sociologist Charles Tilly (1995) encoded extensive data concerning 8088 events of ‘popular contention’ in Great Britain between 1758 and 1834; that dataset is used as an exemplar in this dissertation, and discussed in §7.2.2. Roberto Franzosi has developed and analyzed a similar kind of event data, concerning the 20th century rise of Italian fascism, and has written extensively on his methodology (2004; 2010). The Correlates of War (COW)[9] project, was initiated in 1963 by J. David Singer and continues to develop and maintain data about international relations. As the title suggests, it particularly concerns armed conflict. The Penn State Event Data System[10] (formerly KEDS), began in 1991with similar purpose; it is a “program for the machine coding of international event data using pattern recognition and simple grammatical parsing.” Oddly enough, most of the research based upon these data does not investigate spatial questions particularly. That said, they are a valuable resource for expert classifications of several kinds of geo-political events.

Historian Bruce Robertson’s Historical Event Markup and Linking (Heml) project (2005) “explores the possibility of a web-wide indexing of historical data” with the RDF/OWL ontology language, but its goals are self-constrained to what is readily visualized by current means, rather than the “most expressive schema for events.”[11] The approach taken in this research aims for much greater expressivity, and assumes

---

[9] http://www.correlatesofwar.org/

[10] http://eventdata.psu.edu/

[11] Personal communication, July 13 2008

that the cartographic challenges it surely generates will be met in the future.

2.2.3 Desiderata for HGIS

> "Even the most advanced, sophisticated and visually multivalent digital systems still are not able to integrate either in their digital models or their visualization, the 'ticker tape of events' together with historiographic insights. They do not have general models of historical process and narrative. They put all of their sophistication onto the visualization front end […] and the models exist for the purpose of driving the visualizations. They're not bad […] but they are only a sub-domain of the potential models for describing history, and in particular they lack the interpretive character that is essential to the discipline."
>
> (Mostern, Grossner & Johnson, 2009)

Observers in the emerging HGIS field note that while standard GIS software such as the ESRI suite[12] have been useful in many projects, by addressing some conceptual, theoretical and technical challenges we can extend its usefulness.

A general approach to geo-historical knowledge representation must meet several requirements beyond event-centeredness. Gregory and Ell (2007) note a significant milestone in the application of areal interpolation to create consistent time series data across changing historical boundaries in national HGISs. They argue the next major goal should be historical place-name gazetteers, as "virtually all humanities and social science data resources can be referenced by geographical information." Projects like the Stanford Google Library Project[13] and the Early European Books Online™

---

[12] Environmental Systems Research Institute, commercial developers of ArcGIS software

[13] http://www-sul.stanford.edu/about_sulair/special_projects/google_sulair_project.html

project of ProQuest[14] indicate this is well under way. The 2011 meeting of the Association of American Geographers will feature a three day track of papers, panels and workshops devoted to the topic of historical gazetteer development. All of the HGIS commentators cited have expressed agreement on these general points:

- Geo-historical methods for visualizing and analyzing change in historical data can be improved considerably.
- Historical scholars should undertake "a clear look at how the discipline has traditionally handled uncertainty and error in data," (Gregory and Ell 2007) to better present requirements to the GIScience community.
- Given the great expense associated with creating spatial historical datasets, more collaborative efforts are required. This leads to an important requirement for interoperability between data repositories, facilitated by metadata standards and ontologies for negotiating meaning between systems.
- GIS is not a hammer that makes every spatial historical question a nail; the integration of GIS technology with other quantitative and qualitative methods is to be encouraged.

Historical scholarship entails the examination, analysis and interpretation of multiple accounts of the same phenomena. The systems motivating this work must represent conflicting sets of assertions, to enable computational and visual comparisons of their dynamic structure. Without venturing too deeply in philosophical waters (yet!), we can say it will represent multiple *possible worlds*, in the sense that any database and its underlying ontology represent some (normally singular) view of reality. Any given

[14] A plan to digitize roughly one million books printed between 1450 and 1700 (http://www.dc4.proquest.com/en-US/catalogs/databases/detail/eeb.shtml)

study will encode those event attributes that reflect its author's hypotheses.

Historical data models must accommodate multi-valued logics, including for probabilistic or fuzzy propositions and permit null values to mean "unknown," and not "false." Vague data is commonplace in historical studies, so representing estimations of data quality will be essential. Additional considerations concerning historiographic practice and methods are discussed in §4.1.

2.3 Knowledge Representation

Knowledge representation (KR) is a sub-field of artificial intelligence (AI), itself a sub-field of computer science, and is concerned broadly with methods for encoding human knowledge for computational systems in order to permit some degree of automated reasoning. Comprehensive surveys of the field are provided by Sowa (2000) and Brachman and Levesque (2004); the former also traces its philosophical underpinnings to Aristotle, Leibniz, Kant, Pierce and Whitehead. Sowa (2000) has described the field this way (2000, p. xi-xii):

> "Knowledge representation is a multidisciplinary subject that applies theories and techniques from three other fields: (1) Logic provides the formal structure and rules of inference; (2) Ontology defines the things that exist in the application domain; (3) Computation supports the applications that distinguish knowledge representation from pure philosophy […]
>
> Without logic, a knowledge representation is vague...without ontology, the terms and symbols are ill-defined...without computable models, the logic and ontology cannot be implemented in computer programs. Knowledge representation is the application of logic and ontology to the task of creating computable models for some domain."

In GIScience research on representation, sharp distinctions between geographic *information* and geographic *knowledge* have been drawn, particularly with respect to general models integrating space, time and theme (see §2.5) and ontologies for data integration, semantic interoperability and reasoning (§2.4). In both cases, GIScience research draws on the computer science literature concerning KR formalisms (§2.3.1) and cognitive science literature concerning mental representations (§2.3.2).

An important concept in KR is the *knowledge base* (KB)—a set of sentences or propositions representing explicit facts, from which things can be inferred. According to Sowa (2000, p. 495), a KB is "an informal term for a collection of information that includes an ontology as one component...[and] may contain information specified in a declarative language such as logic or expert system rules...[and possibly] unstructured or unformalized information expressed in natural language or procedural code."

KBs are commonly distinguished from databases principally by virtue of adhering to an *open-world* assumption that permits some or even all facts about the knowledge domain to be "unknown or unprovable" (*Ibid*, p. 378). Strictly speaking, databases are not permitted null values under the relational model, nor the three-valued logic of 'true,' 'false,' and 'unknown.' In practice, relational database management systems (RDBMSs) *are* able to store null values, so a dependent application's layers can be designed to interpret these as 'unknown.'

2.3.1 KR and formalisms

For Brachman and Levesque (2004), knowledge representation is more narrowly "the field of study concerned with using formal symbols to represent a collection of propositions believed by some putative agent" (p. 4). This definition highlights both the means for expressing Sowa's "computable models for some domain," and the core epistemological premise dividing information and knowledge, which must be made explicit in systems purporting to represent historical knowledge. A knowledge base for history must permit conflicting truth statements as assertions of knowledge ("justified true belief"[15]) and must not attempt to enforce logical consistency—a seeming contradiction. An individual project team or a community of historical scholars might agree upon the correct classification of an individual thing or event, and seek to prevent contradiction. However, the global knowledge repositories envisioned in Chapter 1 must allow for individually defined *possible worlds* having some variation in vocabulary and possibly contradictory facts.

The formal systems used to develop ontologies for computational applications stem from first-order logic (FOL), "an artificial language, totally under our control, with none of the maverick and unpredictable ambiguities that pervade ordinary language" (Galton 1997). According to Brachman and Levesque, "traditional first-order logic does not provide any tools for dealing with compound predicates […]" (2004, p. 157) and these are certainly required in this work. Numerous extensions to—and subsets of—FOL have been developed (along with notation systems), each having more or less expressivity. These include OWL, Conceptual Graphs (Sowa,

---

[15] The commonplace analysis of knowledge referred to as "JTB;" *cf.* Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/knowledge-analysis/#JTB

2000), and KIF[16]. Description Logics (DL) do allow the definition of complex concepts with unlimited first-order predicates and their use as "atomic" elements in truth sentences (Brachman & Levesque (2004).

Several versions of The Web Ontology Language (OWL)[17] have been developed by the W3C consortium around its Resource Description Framework (RDF),[18] a model for building endless graphs of simple <subject, predicate, object> triples. The ontology developed in this dissertation has been expressed in an OWL-DL version. In Chapter 6, proposed extensions to the DOLCE ontology are written in the standard FOL notation used in that project's documentation (Masolo, et al., 2003).

The implementation in a RDBMS of an ontology expressed in OWL-DL is a significant challenge only partially achieved in this work. SQL, the Structured Query Language used in such systems, is based on relational algebra, which is an FOL offshoot. In fact, SQL queries are themselves logical expressions of concepts, defined extensionally as sets of individuals. Relational model purists, including its originator, Codd (1970) and leading defender, Date (2003), have critiqued SQL for not strictly adhering to certain model precepts, but its practical utility and widespread implementation are undeniable.

---

[16] Knowledge Interchange Format; http://www.ksl.stanford.edu/knowledge-sharing/kif/

[17] http://www.w3.org/TR/2009/REC-owl2-overview-20091027/

[18] Resource Description Framework: http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/

2.3.2 Cognitive science

Cognitive science, as an interdisciplinary field that involves human knowledge representation and artificial intelligence in some degree, informs this project in a few ways. An important part of ontology design is defining taxonomies of entity categories—commonly referred to as classes. Arguably, the criteria for categorization in computing systems for human use should correspond with our best understanding of human cognitive practices, insofar as possible. Rosch (1978) introduced the widely cited *base-level categories* and *prototypes* as the fundamental bases for classifying concepts, entirely compatible with hierarchical structures of super- and sub-ordination. The process of enumerating the kinds of entities found in this domain of interest (§4.4) has entailed informally locating terms corresponding to base categories, per Rosch (e.g. wars, buildings, artists) and stepping both 'upward' to more general terms (conflict, artifacts, persons), then 'downward' to narrower ones (revolutions, churches, sculptors).

Peuquet has explored the cognitive processes involved in knowledge creation in considerable depth, at geographic scales and more generally (2002). She asserts that "it should certainly be expected that the *form* of how knowledge is stored reflect the *process* of knowledge acquisition" (p. 203), and furthermore that "how [the] data are represented is central to how a problem can be solved, and to the ease or difficulty of arriving at a solution. Indeed, in an automated context, the database representation drives the visual representation" (p. 211). This viewpoint is reflected in the Pyramid Model of Mennis, Peuquet and Qian (2000), which differentiates between a "data component" and a "knowledge component" in geographic information. The ontology and data model developed in this dissertation reflects a fundamental agreement with this view of a critical distinction for knowledge repository applications in the historical domain.

In Lakoff and Johnson (1980), Lakoff (1987) and Johnson (1987) the theory of image schemata is developed, describing mental constructs that originate from human sensorimotor or bodily experiences. These constructs have been called "recurrent patterns, shapes and regularities in our actions, perceptions and conceptions [. . .]" (Freundschuh & Egenhofer, 1997, p. 362). This almost primal grounding of human visuospatial reasoning mechanisms distinguishes it from other cognitive practices (the strictly verbal, for example). Lakoff (1987) illustrates by copious example the prevalence of embodied spatial metaphors used for understanding many kinds of phenomena. By all accounts, mapping—and geo-visualization generally—are powerful aids to spatial reasoning, and evidence suggests that the "processing of and memory for maps differs from that of (even) other visuospatial forms" (Taylor 2005, p.295).

By providing a robust framework for describing events and historical processes, this dissertation is seen as complementary to investigations of data structures I call *spatiotemporal complexes*, a notion to be explored in future work. These are similar in some ways to image schemata, and may possibly be linked directly to a visual vocabulary for the basic entities to be represented in maps, concept graphs and electronic sketches. This parallels and follows Kuhn (2003; 2007), who suggests ontologies could locate conceptualizations in semantic reference systems in similar fashion to how places are located in spatial reference systems. The potential connections between image schemata and the formalization of conceptual spaces for geospatial computing have been explored by several researchers (cf. Adams & Raubal, 2009; Kuhn, 2007a; Raubal, 2004; Frank & Raubal 1999).

2.4 Ontologies

The term ontology (small 'o'), is distinguished here from the branch of philosophy concerned with the nature of existence. An operative definition used in this work comes from Guarino (1998): "an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words" (p. 4).

Berry's Geographic Matrix (1964) is frequently cited for expressing geographers' concern with space, time and theme[19], or the *where*, *when* and *what* of geographic phenomena, but in subsequent attempts at integrating those perspectives in GIS data models, both time and theme have been under-specified. The need to better handle temporality for dynamic phenomena has received considerable attention, and considerable progress has been made (§2.5.2). Deficiencies in the handling of thematic data are unsurprising, because for the purposes of most mapping applications and spatial or spatiotemporal analyses to date, representing *what* an

[19] The matrix axes were in fact, space, time and characteristic

entity is in any detail—beyond that geometrical shape or mathematical symbol sufficing for measurement, a class designation, and a unique label—has been simply unnecessary. In an atomic field view, theme is reduced to the value of a single attribute. For this reason the highest level entities in the typical GIS data model are abstract: points, polylines, polygons, polyhedra, cells and surfaces.

There has been an increasing focus in the GIScience community on ontologies for several reasons. As the number of large data repositories has multiplied, the need for interoperability between them has been increasingly recognized due to the potential for both economic benefit and perhaps better science (Kuhn 2003). There is also a growing impetus for large-scale, distributed knowledge stores—Galton's IS category from Chapter 1. This is discussed further in Chapter 3.

A UCGIS-sponsored report on the "emerging theme" of Ontological Foundations for GIScience (Mark, et al. 2005) distinguishes three types of research in that subfield. These are: 1) formal descriptions of geographic reality at various scales, from the nature of objects, events and processes to the particular entities populating individual sub-domains; 2) methods and tools for using geo-ontologies in GIS; and 3) work in "eliciting geo-ontologies from human subjects."

The fact that *where* (places) and *when* (events, processes) are elements of *what* is fundamentally important in the historical domain. As noted earlier, gazetteers are essential components of HGISs although their development is at an early stage. Events are the natural currency of historiography and it seems evident we would want to classify them in any number of ways.

There has been relatively little research on ontologies for spatial history. Two exceptions are the work of Janowicz (2009; 2006; also, Janowicz & Keßler, 2008) and Shaw (2010; also, Shaw, et al., 2009), whose event model is discussed in §2.2.2. Janowicz relates gazetteers to ontologies generally and particularly addresses issues concerning historical places, but not events. He has shown how the critical goal of establishing "same-as" relationships between entries can be greatly aided by using ontology-driven similarity measures.

CIDOC-CRM, an ontology for the domain of museums, is used in this research and discussed extensively in §6.2.3. Like the FinnONTO project cited earlier (§2.2.2), its developers faced many of the same issues as those presented by digital historical atlases, including those of data sparseness and quality. At this writing CIDOC-CRM had not addressed spatiality in any depth; the FinnONTO group has recently recognized this requirement and begun implementing a mapping interface[20] (Kauppinen, et al., in press).

2.4.1 Ontology engineering

The concept of *ontology-driven information systems* (ODIS) for a broad range of application domains has been advanced by Guarino (1998), Fonseca (2006), Bittner (2007) and others. Its application to system design for GIS has been discussed by

[20] http://www.kulttuurisampo.fi/explore.shtml

Fonseca, et al. (2002) and Frank (2003).

ODISs are essential for the effective integration of the three perspectives on geographic phenomena. That is, data models attempting to capture meaning should emerge from an upper ontology that is extensible and flexible enough to support domain ontologies of increasing detail beneath.

There is considerable overlap in the methods used for ontology engineering and data modeling. In this dissertation research an amalgam or synthesis of them are used (§1.3.2). My effort to identify important *geo-historical information constructs* is similar to the *ontology design pattern* approach described by Gangemi (2005), in building from a set of use cases a set of relevant recurring questions the system will be expected to answer. Intuitive groupings of questions lead to "a minimal semantic characterization, and its formal encoding" (p. 267). Gruninger and Fox (1994; 1995) have presented a methodology for using what they term "competency questions" in both defining axioms for an ontology and assessing their adequacy. In their model, adopted in part in this research (§5.1), "motivating scenarios" are mined for informal competency questions which are then formalized (minimally) in FOL terms. These questions may be used in turn to evaluate completeness.

Enumeration of the domain entities and relations to be represented is a common early step in database and software design methodologies regardless of modeling paradigm—Extended Entity-Relationship (EER), Object-Oriented (OOP), etc. (Simsion & Witt, 2005; Arctur & Zeiler, 2004; Larman, 2002). The same is true for ontology design (Gangemi et al., 2002; Allemang & Hendler 2007). The generic description of the conceptual modeling phase for databases from Elmasri and Navath (2007, p. 424) is applicable across the board:

> "…we must identify the basic components of the schema: the entity types, relationship types, and attributes. We should also specify key attributes, cardinality and participation constraints on relationships, weak entity types, and specialization/generalization hierarchies/lattices."

2.4.1.1 Evaluation

Ontologies may be evaluated with respect to either (a) their fitness for a given application (Which shall we adopt?); (b) for the effectiveness of a new design (Does the ontology just developed suit its intended use?); or (c) formatively(Are these the right subsumption relationships?).

In the first case (see §6.2) criteria will usually focus on elements of (i) expressiveness: is the representation language sufficient; is modality (necessity/possibility) required; and (ii) ontological stances or commitments: is it *descriptive* (aiming to capture linguistic, commonsense notions) or *prescriptive* (more strictly realist); is it 3D (differentiating endurant/perdurant) or 4D (assuming a unifying space-time); is it *multiplicative* ("expressively profligate," permitting distinct co-located entities, e.g. a vase and the clay it's made of) or *reductionist* (the vase *is* the clay). Other factors impacting ontology choice can include its licensing arrangements, its maturity in the marketplace and its modularity, as discussed by Semy, et al. (2004). Brewster, et al. (2004) suggest statistical comparisons of the vocabularies of the domain to be modeled and the terms in a possible ontology, termed “data-driven ontology evaluation.”

For case (b), at the latter stages of an ontology-driven development process, data models can be tested for their ability to deliver expected results with an appropriate query language. This is undertaken in this work by means of “competency questions” mentioned earlier (see also, §1.3.6).

In case (c), the OntoClean methodology (Guarino & Welty, 2002) is a uniquely formal approach to validating hierarchical taxonomic relationships along dimensions of property “rigid-ness” and criteria for identity and unity.

2.4.2 Upper ontologies

Many researchers view an upper ontology (alternately, ‘top-level’) as a necessary starting point for data model design, particularly in wide-ranging domains and where data interoperability is a concern. The geo-historical systems envisioned here fit that description. The most frequently cited upper ontologies at present include DOLCE, CIDOC-CRM, BFO and SUMO. Each has distinct goals and stated philosophical foundations. DOLCE and CIDOC-CRM are discussed at length in Chapter 6. The others are listed and briefly described here.

BFO (Basic Formal Ontology) “…is a theory of the basic structures of reality…” (Grenon & Smith, 2004: 139). The authors acknowledge a “need to accept a multiplicity of perspectives upon reality which may be skew to each other” (*Ibid*.). BFO makes the same first-order division between things that are in time and things that happen in time (*endurants* and *perdurants*, respectively) that most if not all upper ontologies do. It takes the division a step further than others in defining distinct ontologies for each—*SNAP* and *SPAN*—and a set of 'trans-ontology' relations for reasoning about the phenomena that involve entities from both (e.g. participation, creation, destruction).

SUMO (Suggested Upper Merged Ontology), as its name suggests, has merged a number of ontologies in a single large ontology—over 20,000 terms and 70,000 axioms at this writing (Niles & Pease 2001). Its native first-order language is SUO-KIF, and an OWL translation is available. The SUMO "core" consists of 630 classes, 217 object properties and 28 data properties. SUMO is a descriptive linguistic ontology, in that it "provides definitions for general-purpose terms" (p. 2). It differentiates temporal from non-temporal entities (process and object respectively), with both subsumed by the class, Physical Entity. The large SUMO taxonomy is a useful resource for building domain-specific taxonomies, but it was ruled out as a possible upper ontology framework for this work due to what I view as a large number of confusing and counter-intuitive subsumption relations. For example, in SUMO, the concept "word" is appears in the following hierarchy: entity > physical object > self-connected object > content-bearing object > linguistic expression > word. It is instructive to consider that SUMO was constructed by significantly sized community of interest is is apparently of use to many. I make no claims about its worthiness or lack thereof, only report an insufficient comfort level on my part to proceed with it.

2.4.3 Information ontologies and the Matrix

Galton (2005; 2007) has put in clear relief the ontological issue of whether it is reality or human conceptions we are representing. The answer is often both, but in varying proportion depending on the purpose of the intended system. In (2005), he traced a path between Berry's Matrix (1964), Peuquet's Triad Framework (1994), Yuan's Three-Domain Representation (1997; 1999) and the Pyramid Framework of Mennis, Peuquet and Qian (2000). The Pyramid Framework posits a division between the observations and measurement of a "data component" and the derived constructs of a "knowledge component," linking them to what he argues are distinct requirements and development paths for modeling systems versus information systems.

In *Representations of Space and Time*, Peuquet (2002) elaborated on the Pyramid Framework as an approach to modeling geographic entities including locations (fields), objects and events in space-time. Referring to its fundamental epistemological stance, Galton (2005) noted that the line between low-level "raw" data and high-level interpretative concepts is not crisp or uniform, but that information objects of human creation, even in the case of storms or epidemics, are aggregations made from a particular perspective, with particular explanatory goals.

Computing systems for representing history must certainly differentiate authored, conceptual aggregations from their component empirical observations, as these types of information are "gross(ly) different in character" (Galton, 2005:301). This view corresponds closely with the notion of information ontologies (Couclelis 2010; Frank 2003). The historical domain is certainly real, but comprises information about multiple *possible worlds* of conflicting classifications and assertions about geo-historical phenomena. Frank (2003) draws this distinction sharply and describes a multi-tiered ontology that in a sense models the knowledge construction process. Frank differentiates between “philosophical ontologies” of reality and “information ontologies,” and proposes blending them (2003). Indeed, the line drawn between data and knowledge is not sharp; it is more akin to a multi-layered (i.e. tiered) membrane (Couclelis 2010). Frank's Tier 0 (physical reality) and 1 (human observation and measurement of it) are arguably below the data-knowledge line. The identification and classification of objects in Tiers 2 and 3 are human cognitive and social processes that take us, in stages, above it.

Couclelis (2010) has presented a system of "ontologies of geographic information," that formalizes to some degree what identify in Chapter 1 as an important theoretical element of this research. I view it as a progression from Frank’s ontology Tiers (2003), more finely grained in some respects. Seven generative Levels are described, tracing the cognitive processes of knowledge generation and representation from perception, to observation and measurement, to classification, and to the development of *geographic information constructs* of increasing complexity. I view that work as fundamentally compatible with the spatial history ontology (SHO) developed herein, as both are concerned with: (i) representations of information constructs versus reality; and (ii) accounting for human purpose at two levels—that related to the objects and processes being examined, and that of the information construct creators. As discussed in Chapter 8, the Couclelis framework will be helpful in further elaborating the SHO, which may in turn illuminate the particularly temporal *geographic information constructs*, which are so far undeveloped.

The distinction between reality and information *about* reality makes for some interesting modeling challenges, and leads to at least one ontological cul-de-sac (see §5.3). Many of the (possibly) shared conceptions in an information ontology, including complex events, historical processes, and named eras or periods, are on the philosophical view endurants that concern perdurants, and not perdurants themselves. An historical process as an asserted complex conceptual object may be represented by its author with prose and statistics or, following this research, in abstract computational terms, but in any case that process object exists through time and may undergo change. It can also be viewed as a dynamic temporal entity, having temporal parts (i.e. events and activity instances). So the answer to whether processes and periods are temporal things OR persisting things—SPAN or SNAP—seems to be "yes." Whether XOR is true is another matter; further discussion appears in §5.3. Galton has identified *dual-aspect phenomena* (2003; 2005), and in (2007) proposed to replace the endurant/perdurant distinction entirely with one a division between an Experiential world of objects and processes (EXP) and a Historical perspective of events (HIST).

In the logical model presented in Chapter 6, some decisions are taken. What is developed in this dissertation is an information ontology—specifications of conceptualizations, per Gruber (1993)—which are usefully organized as temporal or enduring, but may not correspond neatly to the most usual philosophical distinction.

2.5 Space, time and theme in GIS

Nearly a decade ago, Donna Peuquet observed that "…the GIS field (is) stuck in a primitive and artificial ontology of points, lines, polygons and pixels in computer representation," (2002:268) and this remains fundamentally unchanged. A large proportion of GIS applications principally support spatial reasoning, for example about proximity, distribution, density and topological relations. Entities are therefore usefully classified immediately by geometry: points, lines, polygons and calculated surfaces. Of course process is every bit as important as form (Goodchild, 2004) and scientific inquiry regarding geographic phenomena almost always has important temporal considerations, but by most accounts, GIS has not provided sufficient tools for many applications.

Efforts to extend geographic data models began with a focus on adding time, moving from *spatial* to intrinsically *spatiotemporal*. Characterized initially as achieving a "temporal GIS" (Langran 1992), that goal is now more frequently referred to as modeling the "dynamics in geographic domains" (Peuquet 1994; also Goodchild and Glennon 2008; Yuan and Hornsby 2008; Hornsby and Yuan 2008). The progression and breadth of approaches has been traced in (Frank 1998; Peuquet 2002; Worboys and Duckham 2004; Worboys, 2005; Yuan and Hornsby 2008).

Conceptual data models for geographic information systems (GIS) have been classified most broadly into two types: object and field (Couclelis 1992), representing discrete or continuous *spatial* data respectively—that is, the *where* of geographic phenomena. The object view is concerned with locations of geographic features normally regarded as discrete objects, such as rivers, mountain peaks, cities and roads. The field approach is appropriate for phenomena conceived as continuous surfaces, derived either from point samples (e.g. elevation or temperature) or from remotely-sensed imagery (e.g. land cover or atmospheric dust). This division necessarily plays out in all representation models.

All of the entities mentioned have both spatial and temporal attributes, and may be readily represented as spatial things with temporal attributes (objects, fields, field-objects, etc.), as necessarily both spatial *and* temporal (Grenon and Smith 2004), or as integrated space-time things (Yuan 1999). There have been several approaches suggested for unifying field and object views. Goodchild, Cova and Yuan (2007) have presented an "atomic" theory of geographic representation as an abstract foundation to support both relational and object-oriented database implementations. The information ontology frameworks proposed by Frank (2003) and Couclelis (2010), discussed in §2.4.3, both incorporate field and object perspectives on geographic data.

The question of how temporality is represented in GIS (i.e. added to essentially spatial data models) seems to be an ontological one, therefore closely linked with the third axis of Berry's Matrix, *characteristic*, or theme. However, research on geospatial ontologies and temporality in GIS has proceeded largely on distinct paths. There don't appear to be easy, general answers to questions of whether time is an attribute (an interval of existence or validity), or whether representing discrete temporal entities (actions, events, processes) is more generally useful. Indeed whether this is even an either/or question.

Some emerging requirements, including from digital historical atlases, have prompted Kuhn's (2007b) observation: "users of geographic information should be able to refer thematic data to semantic reference systems, just as they refer spatial (temporal) data to spatial (temporal) reference systems." It seems evident that ontologies will be instrumental in integrating reference systems across dimensions of space, time and theme.

2.5.1 Change objects and events

Worboys' view that "the next real breakthrough in computer modelling of geographic phenomena comes when we move from an object-oriented to an event-oriented view of the world" (2005: 2) is fully supported in this work. Worboys delineates four stages in the development of temporal GIS: Stage Zero (static GIS), Stage One (temporal snapshots), Stage Two (object change) and Stage Three (events and action). Stage Three systems entail "…a full-blooded treatment of change, in terms of events and actions," (2005: 7) and require at a minimum treating occurrents as "first-class entities" (*Ibid*. 24), corresponding to the object-oriented programming model and analogous to physical objects. This dissertation develops a theoretical and formal basis for a particular kind of 'Stage Three' spatiotemporal information system by treating complex occurrents (event and process) as 'first-class' computational entities, in a semantic model.

Work corresponding to Worboys' Stage Two systems has modeled events as instances of object change—change in identity, in position and attributes. Hornsby & Egenhofer (2000) introduce *change objects*, and a visual "Change Description Language" to model 81 conditions of identity-based change (creation, destruction, reincarnation, etc.) related to existence states. There is a natural correspondence between *change objects* and a more usual term, events, and the division between Worboys' Stage 2 and Stage 3 is not sharp. Hall & Hornsby (2005) investigated the automated ordering of events from the temporal logic of thirteen standard temporal interval relations. The methodology is interesting for its event-centeredness, but reasoning about event relations is limited to spatially plausible sequences—an approach that seems inapplicable to the non-linear spatially extended events in the historical record. Hornsby and Cole (2007) have identified and formalized three patterns of object movement event: *repeating*, *collocating* and *reiterating*.

Other recent formalizations of abstract spatial change types, include for motion (Klippel, et al. 2008), change of attributes like size and structural properties of distribution (Yuan, 2009), *generalized space-time paths* (Shaw, et al. 2008) and the membership relations and movement of *dynamic collectives* (Galton 2005) and *collectivities* (Wood & Galton 2008a; 2008b). The question of whether any of these change event formalisms might be applied to the typing and analysis of historical events is worthy of future investigation.

Worboys and Hornsby (2004) introduced the Geospatial Event Model (GEM) to explicitly model objects *and* events-as-objects in *geo-settings*, and in several important respects this dissertation research follows on from that work, addressing:

- The object-like nature of events with respect to classification, parthood, subsumption and attributes.
- The participation in events (perdurants) by objects (endurants), in roles.
- The notion of event-event relations beyond simple temporal sequencing and containment.
- The conceptual framework of settings contextualizing events. *Geo-settings* are a powerful construct, but their formalization is strictly spatial-temporal. A setting in historical terms must also include the thematic dimension.

The investigation of domain entities and relationships in the early stage of this research arrived at similar constructs, among others. This dissertation seeks to advance them particularly for the historical domain. The GEM paper called for further development and this research is a response.

Worboys has postulated a “pure event-oriented theory of space and time” (2005, p. 1), but I’ve stopped well short of the proposition that “everything is an event.” In simplest terms this work supports modeling the participation of persistent objects (endurants) in temporal entities (perdurants), and assertions of relevant states.

Several recent works suggest additional motivation for modeling geospatial information in terms of activity, events and process: Yuan (2009) lists them as “the spatiotemporal constructs of geographic dynamics,” and the principal elements of a framework for knowledge discovery in databases (KDD). Shaw and Yu (2008) have developed a GIS data model of activities, events and *projects* for hybrid physical-virtual spaces of individuals' activity in a time geography framework. Mostern and Johnson (2008) have described a prototype historical event gazetteer, navigable in coordinated map, timeline and relationship browser views. They note the need for a formal foundation, and this work responds in part to that call.

Life and Motion of Socio-Economic Units (2001) is an important collection of articles relevant to these topics. Spatial socio-economic units (SSEUs) are defined broadly as "spatial units in geographic space that are the result of some social, cultural, economic or behavioural process" (Frank et al., p. 9). As such, they comprise a large percentage of the non-event items of interest for historians, including polities like kingdoms, empires and countries, administrative districts and regions defined by shared attributes, such as demographic characteristics or land-use categories. The histories of SSEUs can be represented in terms of events. Also in that volume, Worboys (2001) defines events as instances of change (or composites thereof) and argues that both events and SSEUs can be modeled as objects in OOP terms, but allows that "events and objects (i.e. things, in the non-programming sense) belong to distinct categories" (p. 134). He also speculates that change might be usefully classified in terms of the force/compulsion image schemata of Johnson (1987). Yuan (2001) applies her "three-domain model" for representing geographic information objects to the problem of SSEUs, a conceptual precursor to my own notion of *historiographic zoom* between and across the spatial, temporal and thematic dimensions (Chapter 1).

2.5.2 Temporal reasoning

Any formal model of temporal entities must include, or be adaptable to, one or more methods of temporal reasoning. In this research, the goal is a modest extension to Allen's interval relations (1983), to account for indeterminate start and end points. Greater expressiveness has been an ongoing focus of computer scientists for many years and range in attempted expressiveness from event calculus of Kowalski and Sergot (1986), to the Allen and Ferguson logical model of events and action (1994), to the TimeML specification language (Pustejovsky et al. 2005), which can capture natural language event references at a very fine resolution. Each of these approaches presumes actions specified as having determinate or determinable results.

On their own, Allen's seven interval relations (1983) (*before*, *during*, *meets*, *overlaps*, *starts*, *finishes*, and *equals*[21]) have ambiguous meaning. They could refer to incidental temporal relations along a single linear dimension (shared start and end points, durations and so forth) or have a deeper meaning, if for example *starts* is taken to be synonymous with *initiates*, which presupposes (or declares) a *part of* relation. At this early stage of the work presented here, Allen's intervals are purely temporal, and revealed with queries against start/end timestamps. Substantive relations between events, including *initiated*, *part of*, *ended* and *caused* are explicitly declared in a *historical-process* association class. At issue is how specific one must be about most event types for this class of systems. The Sisyphean task of integrating

---

[21] Counting inverses of the first six, there are actually thirteen.

a sophisticated expression of Allen's interval relations into a temporal SQL (TSQL2) for databases has been a major effort of Snodgrass (2000; 1992). At this writing, only the Oracle RDBMS has implemented this reasoning capability.

In this dissertation, a *period* datatype[22] for PostgreSQL systems that includes operators for Allen's relations has been adapted and extended (§7.2.4), by effectively bounding intervals with intervals in place of instants, and adding operators to calculate the semi-intervals of Freksa (1992). The exemplar database developed in this research is "historical," in that temporal attributes describe the period for which an assertion is true, or "valid" (Snodgrass, 1992) .Many of the functions described in §7.1.3 reify predicates as *fluents* in the sense of Kowalski & Sergot's Event Calculus (1986), returning conditions true at some time, *t*.

2.5.3 Time Geography

The introduction of time geography by Hägerstrand (1970) provided an important theoretical framework for reasoning about human movement in space. Doreen Massey's conception of place as "the meeting up of histories" (2005) seems entirely compatible with time-geographic principles and methods when the histories considered are of individuals. However, most applications utilizing time-geographic methods are not historical, probably due in large part to the relative scarcity of data at the level of detail normally studied. The historicity of "space-adjusting technologies"

[22] The pgChronos project of Scott Bailey (http://pgfoundry.org/projects/timespan/)

that can be studied with time-geography methods has been discussed by Janelle (1969; 2004) and others, as surveyed by Miller (2007).

The potential for applying time-geographic methods to historical inquiry have been discussed by Pred (1977), who suggested for example that "certain large-scale, historical-political developments can be reinterpreted in the context of knowledge about small-scale, time-geographic realities" (1977, p. 217). Several years later, Pred presented a theoretical discussion of place as "a process whereby the reproduction of social and cultural forms, the formation of biography and the transformation of nature ceaselessly become one another at the same time that time-space specific activities and power relations continuously become one another" (1984, p. 279). He saw time-geographic principles, methods and visualizations as fundamentally compatible to such investigations, and published several historical studies utilizing manual implementations of them (e.g. Pred, 1981).

Southall and White are historical geographers who characterized time geography as "something of a passing fad within human geography of the late 1970s and early 1980s," (n.d.) a fact which may be attributed to the lack of formalization and therefore computational systems. They suggest its potential value is principally as a visualization tool.

More recently, Miller has developed a "measurement theory for time geography," (2004) that does formalize most of its entities and relations, including space-time paths, prisms, stations, bundles and intersections. This corresponds to a considerable recent upsurge of interest and new computational approaches. Kwan and Lee (2003) have studied “gender/ethnic differences in space-time activity patterns” drawn from present-day individuals’ diaries. Shaw and Yu (2008; 2009) have introduced usable GIS tools in the context of their own merging of physical and virtual space within a time-geography framework.

The Spatial History Ontology introduced in this dissertation includes paths as kinds of places, further differentiated as trajectories and flows (§6.3.7). For the time being these rudimentary spatiotemporal structures can be queried in several useful ways, but a thorough investigation of their potential alignment with Miller’s formalization of Hägerstrand’s *space-time paths* constitutes future work.

3 The Digital Historical Atlas: Visions, metaphors and real systems

> "It would be utopic to think that even motivated knowledge engineers would be (in the near future) able/willing to represent their research ideas completely into a formal, shared, well-structured readable semantic network that can be explored like a decision tree: there are too many things to enter, too many ways to describe or represent a same thing, and too many ways to group and compare these things. On the other hand, representing the most important structures into such a semantic network and interconnecting them with informal representations seems achievable and extremely interesting for education and IR (information retrieval) purposes."

(Martin, et al., 2005)

> Start a huge, foolish project
> Like Noah

(Rumi, 13th century)

The work of this dissertation has been motivated by an envisioned web-based software application that is equal parts geographic and historical, a *digital atlas of world history*. There is a pressing need for an improved understanding of the historical context of today's geopolitical events among citizens and policy-makers. There are generally acknowledged deficiencies among students in geographic knowledge[23] and spatial thinking skills (National Research Council, 2006). The study of history is concerned with "why?"—interpreting and synthesizing empirical data and multiple accounts into plausible narratives to aid understanding for oneself and others. Both observational accounts and historians' narratives become part of the record. Geographic computing systems that enable the automated mapping of the

[23] The most recent National Assessment of Education Progress (NAEP) test results for geography (2001) showed that slightly more than one quarter of US K-12 students were rated "proficient" or better in geographic knowledge and reasoning skills.

factual substrate underlying assertions of historical processes would be an interactive and graphical venue for historical scholars, and for students and the general public, a window into academic research. To the extent we can argue about present circumstances or policy over a common set of facts, we are better off. Web-based search on historical topics can put a lot of material to hand quickly, but ‘Googling’ is a deeply flawed research method. It presents winners of a machine-run popularity contest, creating what I call perspective cul-de-sacs.

At this writing many of the technical requirements for digital atlas systems exist, in GIS software and spatial databases, while some important ones are missing. Although print historical atlases have been examined as a genre closely (Goffart, 2003; Black, 1997), there has been little investigation of theoretical underpinnings for their digital counterpart[24]. The goals of this dissertation as discussed in Chapter1 can be restated as follows: (1) to further define the digital historical atlas genre, (2) to identify the important theoretical and technical missing pieces, and (3) to contribute a few such pieces, including a method for representing knowledge of historical human activity in the sort of spatial databases used in commercial and open-source GIS. In sum, this work seeks to help reinvent the venerable print historical atlas as a digital geo-historical system.

There is a growing impetus towards large-scale globally collaborative knowledge-bases. The largest and most successful is Wikipedia, noteworthy for its value as

---

[24] One exception is Sieber and Huber (2007)

measured by use statistics,[25] and its remarkable level of participation in terms of "bottom-up" contributions from the general public. Wikipedia has grown increasingly reliable as more rigorous community editorial norms take hold, and it has become increasingly useful as its data becomes more structured. Examples of such structure include standardized fields for many types of articles (biographical, geographic, etc.), and the growing practice of georeferencing articles generally.

The ambitious Tim Berners-Lee vision of a Semantic Web would turn the entire Web into a structured knowledge repository. After halting progress a related, simpler concept, Linked Open Data (LOD)[26] has emerged with more realistic near term goals based on the simple <subject, predicate, object> of its RDF model. Initiatives such as DBpedia[27] and Freebase[28] are online databases that also aim to be comprehensive, using semantic web technologies such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL) to link Wikipedia content and other resources.

In the historical domain, large-scale systems have been limited to national historical GISs and data clearinghouses (see Appendix §9.4 for a listing). However, at this writing, proposals for a Global Historical GIS are in development by an

---

[25] Rated by Google as the 5th most visited internet site in August, 2010 with 310 million unique visits

[26] http://linkeddata.org/

[27] http://dbpedia.org/

[28] http://www.freebase.com/ (acquired by Google in July, 2010)

international research group. Individual historical analytical studies might one day be integrated with such large publicly accessible systems, linking research and education to a much greater degree than is currently possible.

Over the past five years, the awareness by academics and the general public of location as a powerful method for indexing information has increased dramatically. This is reflected in the expanding array of web mapping applications, location-based services (LBS) for vehicles and hand-held devices, and other GPS-driven technologies. As a result there is a growing interest in the geo-referencing of many kinds of documents, including historical texts. Still, location is under-utilized for indexing information systems generally, as compared to theme or concept. Google's plan to digitize many millions of books has encountered some legal snags, but is proceeding in a joint project with Stanford University.[29] Before long, a huge volume of textual information and material in other media will be effectively geo-referenced. The implications for information integration are enormous. More sophisticated systems to retrieve, navigate and explore such digital stores are required. The notion of historical event 'containers' that are inherently georeferenced seems promising.

Several metaphors have been used to conceptualize and distinguish systems for representing history in databases and in interactive software for research and education. There exist now a number of *digital atlases*, *digital libraries* and *digital*

---

[29] http://www-sul.stanford.edu/about_sulair/special_projects/google_sulair_project_faq.html

museums. Any or all of them could be enhanced significantly by the products of this research. A number of each have been surveyed (Appendix §9.4) and only one—CultureSampo[30]—addresses events formally (§2.2.1). Digital libraries are primarily concerned with enabling search and delivery of information objects (documents and imagery). Digital museums normally allow some virtual browsing of physical collections and in cases offer multimedia "exhibits," although oddly enough, maps are scarce in this genre. A few digital historical atlases are GIS-driven and interactive, but are typically limited to representing either point data for place names and artifacts and snapshots of the shifting boundaries of past states and empires. The most advanced digital geo-historical resources in several respects are national historical GIS projects such as Great Britain Historical GIS and the *China Historical GIS*[31] (CHGIS). *Vision of Britain through Time*[32] is noteworthy for being an innovative digital atlas built upon a national historical GIS. There are several impressive national HGIS projects, but the particular genre of digital historical atlas is still in a fledgling state.

3.1 The digital historical atlas as digital geolibrary

While there is no replacing the best print historical atlases—bound masterpieces of

---

[30] http://www.kulttuurisampo.fi/?lang=en

[31] The China Historical GIS project is centered at Harvard University; (www.fas.harvard.edu/~chgis/)

[32] Vision of Britain through Time is the web-based interface to the Great Britain Historical GIS Project based in the Department of Geography of the University of Portsmouth, UK. (www.visionofbritain.org.uk)

cartography and historical scholarship that can be leisurely browsed in an armchair—we can expect their digital offspring to surpass them in several ways. Instead of several dozen carefully crafted map plates, users of a comprehensive digital historical atlas could generate a very large number of maps for any combination of themes and spatio-temporal extents. Information and media objects of all kinds will be simultaneously indexed in spatial, temporal and thematic reference systems. Where appropriate, animations can enhance understanding of change over time. Depending on atlas authors' goals, many kinds of quantitative and qualitative analyses could be undertaken and visualized—the database supporting a digital historical atlas will support many kinds of visualization besides maps: timelines, charts, graphs, and schematic diagrams, such as for flow. This is not to say digital historical atlases will not lag in some respects. The cartography in high-quality print atlases is unlikely to be matched digitally in the near future, although recent progress in both desktop GIS and web-based mapping applications is encouraging. Also, print atlases frequently include essays of several hundred words, and reading text on screens is still more difficult than on the printed page (Garland & Noyes, 2004)

Many print historical atlases offer far greater depth and/or breadth of historical information than the digital efforts to date. The *Oxford Atlas of World History* for example, is a remarkable 360-page volume, produced by six editors with content authored by 45 scholars and several cartographers from a number of British universities (O'Brien, 2002). It contains 131 articles, over 300 maps, a 23-page timeline, a 31-page gazetteer of people, places and events, and many dozens of images and graphs. It is a graphical tour-de-force and an extraordinary reference resource for history students, though we can surmise less so for researchers. Simply replicating that breadth and depth digitally would be a major achievement, but initial goals should go much further.

Although remarkably dense, print atlases have practical limitations on the depth of information they may provide, so choices are made. Authors, cartographers and editors must decide which few maps, for example, best suit a two-page article on "The First Civilizations: Mesopotamia and the Indus Region 4000-1800 BC." One might show dynastic extents at a given point in time, another, the expanding commodity trade routes for a large portion of that era, and a third the plan of an ancient city as inferred from artifacts. A truly comprehensive digital historical atlas would obviate such editorial decisions and constraints. Driven by a GIS database, it could allow its users to generate unlimited thematic maps to answer particular questions, all extensively hyperlinked to a very large distributed store of articles, photos and other media. Visualizations produced by merged and/or overlaid thematic data layers are powerful knowledge discovery tools. They are being applied successfully in other fields, and hold great potential for historians.

3.1.1 Geolibraries

The term *geolibrary* was introduced by Goodchild (1998) to identify an emerging class of digital system, the "library filled with georeferenced information," as an important element of the growing global spatial data infrastructure. The geolibrary concept grew out of efforts at organizing and providing shared access to the digital holdings of map libraries and spatial data repositories.[33] These collections typically include scanned paper maps, satellite imagery, aerial photographs, and GIS data files. A National Research Council report (1999) solidified the concept considerably.

A geolibrary indexes its holdings by location, specified by one or more place names and a geographic "footprint" of any complexity, as well as by the traditional library catalog keys of title, author and topic. Many digital map libraries and spatial data clearinghouses are then by definition geolibraries, in that their holdings may be browsed and searched by location. It became increasingly apparent that methods developed for accessing those data could be applied to *any* digital information objects associated with locations specified by name or footprint. In principle any library would become a geolibrary if its holdings were comprehensively georeferenced. A key distinction was drawn between this "geographical information," as representations of the surface and near-surface of the Earth, and the vastly broader category of "georeferenced information," defined as any information referring to or about particular places.

In 2004, Goodchild suggested a next stage of progress in geolibrary development could be a transformation from simply delivering information objects (maps, imagery and other documents) to adding some ability to open and process those objects. Such

---

[33] The NSF-funded Alexandria Digital Library at the University of California, Santa Barbara is a prominent example, begun in 1994. http://www.alexandria.ucsb.edu

extensions are likely to blur distinctions between the geolibrary and other emerging classes of information systems, including digital atlases and georeferenced digital encyclopedias. According to this vision, the query answerable by existing digital geolibraries, "*what do you have* about that place?" would be extended to the vastly broader, "*what is so* about that place?" and even, "*what has been so* there?" and "what *might become so*?" More particularly:

- What is this place like, physically—its terrain, flora, fauna and climate—and how have they changed?
- What is it called now, and previously? What named regions does it belong to?
- Who lives here, when did they arrive and from where?
- What languages are spoken and what distinct cultural practices are and have been associated with this place?
- How have settlement patterns, food production and commercial activity in this region evolved over time?
- What places are like this place, in various ways, i.e. nearby in "attribute space" per Skupin and Hagelman (2003)?
- What important events have happened here and nearby, and why?

This range of questions has a spatial focus, but an advanced geolibrary or a digital historical atlas should be able to answer comparable questions from temporal and thematic perspectives as well. Two examples, drawn from Diamond's *Guns, Germs and Steel* (1999) might be: "during a given time period, what were the crops in cultivation, and where?" or, "trace the spatial and temporal history of the domestication of the horse."

Geolibrary development—particularly as distributed systems—has already raised numerous research challenges and is a valuable context for their solution: 1) interoperability for data sharing by means of metadata standards, authority lists and ontologies; 2) gazetteer development, i.e. the association of a particular spatial footprint with potentially numerous place names; 3) efficient spatial indexes for databases and related issues of scale, resolution and generalization; 4) georeferencing, including the parsing of textual material for named places and related temporal issues; 5) spatiotemporal query algorithms and 6) interface usability for spatial browsing and search informed by cognitive principles. Significant incremental progress in all of these areas has made the prospect of very large, distributed geolibraries increasingly realistic.

All are applicable to digital historical atlases as well, and we can add the following considerations.

- Databases are good at representing precise, uncontroversial data, and a digital historical atlas will store a great deal that falls in that category. However, it must also present multiple versions of the same facts, e.g. contested borders and conflicting observations of various kinds, and offer easy methods for maintaining and effectively presenting attribution to a scholarly standard.
- It must maintain a clear line between data and the conceptual knowledge derived from it, keeping the distinctions clear throughout.
- Place without time is only half the picture. Indexing place and time together—and representing process generally—is a focus of extensive research in GIScience, and will enable a new generation of systems for information retrieval and knowledge construction in a variety of disciplines.

At its furthest extension such functionality resembles the *Digital Earth* vision of systems providing access to all available digitized information related to any specified location.

3.2 Digital Earth

In a world of unknowable complexity, simple devices for grasping some sense of the whole are engaging, even poetic. The huge success of the Google Earth geo-browser following its release in 2005 has been due in large part to the ease with which an untrained user can virtually fly around the world, zooming in and out between rooftops and fields of interest. We can trace our own footsteps in a city, the paths of individual whales across the Pacific Ocean or the growing perimeter of a devastating flood. We can get the whole picture—albeit one hemisphere at a time. Or rather, the pictures we get are embedded in the "whole picture" of a contextualizing globe.

Scale and resolution are key considerations in any information system, just as they are in static maps. Goodchild, Yuan & Cova (2007) have said the plausible limits of spatial and temporal resolution for the domain of geographic phenomena on and near the Earth surface are one centimeter and one second, respectively. At that extreme of resolution, a truly comprehensive global GIS would approach the "mirror world" of Gelernter (1991) or the somewhat narrower Digital Earth vision (Gore, 1998)—a computing system providing, via exploratory tools and queries, access to what is known about the planet and its inhabitants' activities, now and at any time in history. Furthermore, it would accommodate modeling extensions for scientists to predict future conditions with what skill their algorithms might achieve. Although visionary systems like Digital Earth are likely to remain “a piece of technological fantasy” (Goodchild, 2000, p. 9) in the near term, research aimed at surmounting the numerous theoretical and technical problems they raise continues to facilitate piecemeal progress.

The Digital Earth concept has maintained significant currency as a framework for discussing spatial data infrastructure (SDI) research challenges (Craglia, et al., 2008). Digital Earth has become an umbrella term for the growing set of web-based geographic computing systems having a global or at least continental scope. The bi-annual International Digital Earth Symposia are well attended.[34] Novel and extraordinarily useful systems, albeit of significantly narrower scope than a complete Digital Earth, lie just around the corner (Grossner, Goodchild & Clarke, 2008).

3.3 Cultural Heritage Web

Digital historical atlases might be of any breadth or depth—from a single theme in one region and time period, to the global *longue durée*. Cultural Heritage Web (CHWeb) is a particular web-based software application for the exploration of human history and the world’s cultural treasures, under development by this author. The effort is ambitious in scope with respect to both breadth and depth, and is closely related to the largely unrealized conceptions of *digital geolibrary* and *Digital Earth*.

[34] Hosted by the International Society for Digital Earth (http://www.digitalearth-isde.org/)

Like them, it would index information by place, time and theme, and answer for its users not only "what print and digital resources are available about that place," but to a great extent, "what is known about that place?"

CHWeb could be simply characterized as "historical social science meeting 'citizen history,'" in time a large-scale, distributed and socially authored encyclopedic resource for historical researchers, K-16 education and the public at large. Its principal organizing framework will be designated cultural heritage sites of global, national or local significance, from the 731 cultural properties found on UNESCO's 2010 World Heritage List to the humblest historical highway markers found world-wide.

CHWeb is conceived as a distributed system with three 'tiers' of information, (Figure 3-1) along the lines of the digital earth system design proposed in (Grossner, Goodchild & Clarke, 2008). Level I will hold current base data having global coverage along with historical base layers as available. Level II will hold authoritative historical datasets and scholarly atlas projects for particular topics and spatio-temporal extents, curated by an editorial team. Level III is a public, Wiki-type layer intended to accept contributions from students and amateur historians, particularly about sites local to them.

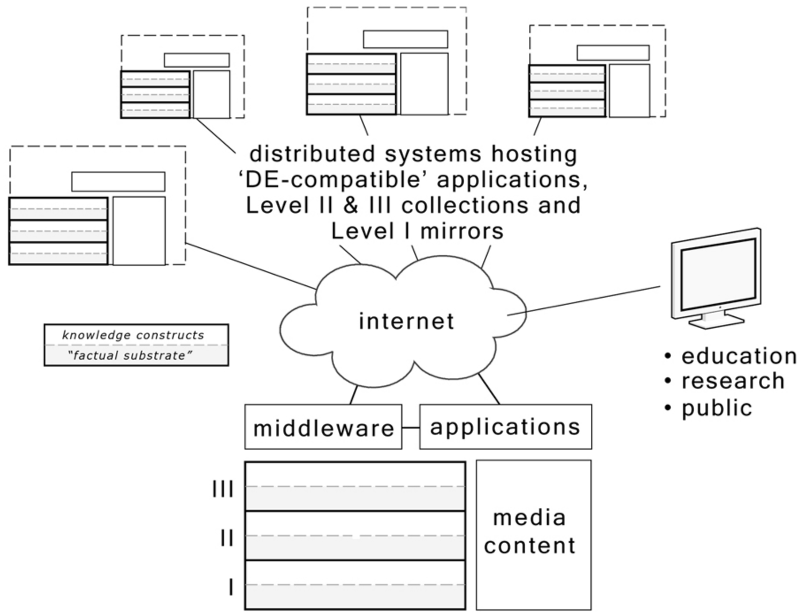


Figure 3-1. Data levels and distributed architecture for a digital historical atlas

Cultural Heritage Web is intended as:

- A large undertaking, integrating numerous distributed data and media sources and pushing the bounds of interoperability between disparate systems.
- Both a test-bed and exemplar for innovative approaches to representing historical and geographic knowledge in databases, in terms of events and processes. Historical places will be represented in terms of what has happened there—those relevant events and historical processes that led to the site's designation and/or protected status.
- A framework and venue for work on many research questions at the intersection of geographic information science (GIScience), geography, history and computer science.

More about motivation

Efforts to present scholarly work for general audiences can produce exciting and entertaining educational experiences, as demonstrated by the effectiveness and popularity of the best interpretive exhibits in major museums. As conceived, Cultural Heritage Web will borrow from that paradigm. History is effectively understood and conveyed with narrative, and heritage sites that have been designated as deserving conservation and greater public awareness represent excellent anchor points for compelling, complex stories involving events at and near those locations. They are in some sense curated objects already. Given widespread support and participation, it could grow towards becoming a digital museum of world history.

An atlas; a library; a museum—what's in a name? The envisioned system might be approached from any of those conceptual frameworks.

4 The Stuff of History I

> stuff, *n*. 3.a. The substance or 'material' (whether corporeal or incorporeal) of which a thing is formed or consists, or out of which a thing may be fashioned. (stuff, n.d.)

> "The art of ranking things in genera and species is of no small importance and very much assists our judgment as well as our memory. You know how much it matters in botany, not to mention animals and other substances, or again moral and notional entities as some call them. [. . .] This helps one not merely to retain things, but also to find them. And those who have laid out all sorts of notions under certain headings or categories have done something very useful."
>
> (Leibniz, 1996/1764:291-292)

The first objective listed in §1.2.1, "enumeration and further definition of representation requirements for an emerging genre of digital historical atlas," leads us to get more explicit about what constitutes historical knowledge for such systems. The entities ("stuff") we're concerned with fall into two broad categories: first, those things in the world that are the objects of historical inquiry (whether "corporeal or incorporeal"); secondly, those elements of historiographic practice that should dictate useful representation forms. These are, in cases, reified things in the world themselves—for example, the attribution of sources.

Digital historical atlases will obviously feature dynamic maps prominently, but this inquiry is not constrained to what is readily mappable or deemed "map-worthy." As noted earlier, existing print atlases tell history by numerous graphical and textual means; digital historical atlases are certain to expand that variety. Digital cartographers will be presented new challenges, including for example multiple conflicting geometries for the same entity, and probabilistic or fuzzy spatial-temporal boundaries. We should expect these challenges will be met and proceed to describe the entire domain of interest.

In my view, many elements of a possible "theory of geo-historical representation" are already in place. A concerted effort at classifying the spatiotemporal information in historical atlases and HGIS projects should make it possible to unify those theoretical elements within an ontological framework.

The development of conceptual, logical and physical data models for representing historical knowledge in digital atlases is the goal of this work, begging the questions: What is historical knowledge? And, knowledge of what? The "meta-level" answer to the first question appears in the theoretical framework proposed in Chapter 1, namely (i) "raw" data as observed and/or measured, (ii) asserted truth statements about things and happenings in the world derived from those data, and (iii) information about the human creators of those statements. Some further requirements of historians are presented in §4.1.

To begin answering the second question, I have analyzed a representative set of 17 print historical atlases (listed in §10.1), and I survey the genre in §4.2: their types; their content; their communication goals and the strategies employed to meet them. In §4.3 I review several historical GIS projects and self-described digital atlases. The objective in each case was to enumerate "the stuff of history"—those entities populating the maps, timelines, charts and diagrams found in those works. A preliminary classification of the entities found is presented in §4.4, informing both the choice of exemplar datasets (Chapter 7) and the conceptual model presented in Chapter 5.

4.1 Scholarly primitives

A number of historiographic requirements for a spatial history ontology are discussed in Chapter 1 and §2.2, in the context of envisioned digital historical atlases and desiderata for HGIS. Valuable additions to this come from China historian Ruth Mostern (2008), whose four "scholarly primitives for historical geography and spatial history" are a suggested framework for defining and specifying "design principles that are consistent with our (historians') disciplinary practices and the requirements of our research." These primitives are "(1) *attributing* sources extensively, (2) *historicizing* temporal change in spatial organization, (3) *contextualizing* the social, cultural and political basis for spatial organization, and (4) *modeling uncertainty and ambiguity* in time and space" (*Ibid*, p.45). Ontologies and data models for HGIS must account for requirements in all four categories. This research principally addresses issues in the first three, but I do necessarily provide some analysis of how all four aspects must fit together.

Attributing sources extensively

The study of history is concerned with 'why?'—interpreting and synthesizing empirical data and multiple accounts into plausible narratives to aid understanding for oneself and others. Both observational accounts and historians' narratives become part of the record. Primary historical sources often provide either subtly or starkly conflicting evidence that must be compared and otherwise evaluated, then selectively represented. Attribution is critical to differentiating disparate versions of conditions or events, because such provenance reveals to readers "the mentality that produced it, its inherent biases, and the points on which the information is most and least reliable" (Knowles, 2008b:13). In simplest terms, this means every "unit of historical information" (*Ibid.*) should somehow include an <*attested-by*> or <*asserted-by*> attribute.

Historicizing temporal change in spatial organization

Is a thing that changes location between t_1 and t_2 the same thing? What if several of its key attributes change as well? Eschenbach (2001) uses examples of Berlin and Germany to illustrate the complexity in cases and the multiple approaches that can be taken to represent such non-physical yet very real geo-historical entities: how many entities were involved in the post-WW2 division of the German Reich into East and West and back to modern-day Germany, and what has been the changing status of Berlin (and parts thereof) as the *almost* continual capital of multiple German states? The establishing, tracking and querying of unique object identity over time are key problems in modeling history. The case of Hagia Sophia in Istanbul is not unusual. It was the world's largest cathedral for almost 900 years, a mosque for the next 500, and declared a museum in 1935. Originally constructed in the year 360 CE, it was destroyed twice in fires before 562 CE when the current structure was completed. How many records might Hagia Sophia require in a table listing historic structures?

Contextualizing the social, cultural and political basis for spatial organization

Earlier, I suggested that given a universe of information with spatial and temporal attributes, one might browse or query a digital historical atlas at varying levels of resolution by specifying place, time or theme, individually or in any combination. What hasn't been mentioned is the ontological framework including individual domain ontologies that would be necessary to make the thematic dimension navigable. Subject headings in authority lists like the Library of Congress are useful for locating information objects that have been appropriately tagged, but lacking explicit information about relationships between named entities, they offer little in the way of meaning.

The contextualization referred to in this "scholarly primitive," and the browsing of digital atlases across space, time and theme (which I call *historiographic zooming*), will be enabled by query expansion algorithms that require an explicit, extensible ontological framework. If a query term is in a sense "understood" by a system by means of its known relationships with other concepts, the kind of contextualizing referred to here is made possible. Given extensibility and robust means for attributing sources, we can ensure no one particular theory of reality dominates if that is a goal. As with historical work in any medium, a computational model is whatever its authors or audience make of it—in any case: "not the past, it is a useful device for thinking about the past" (Staley, 2003:113).

Modeling uncertainty and ambiguity in time and space

This is one of the more vexing issues facing designers and users of HGIS software. Attribution of sources can address some problems of uneven data quality and multiple conflicting accounts, and historical gazetteers can at least give users all known facts regarding place names. The traditional practice in written history of exhaustive notes can be readily implemented in databases. Imprecision and vagueness are more problematic, and there is a significant interest and literature about the use of fuzzy sets, probability models and the cartographic considerations for visualizing them. I address this in limited fashion in this research, by providing in the exemplar database a stand-in approach for specifying indeterminate dates.

4.2 Print historical atlases

> "There is no guide book on how to produce an historical atlas: the field is too diverse and inchoate. Instead, each historical atlas is an individual work, reflecting the intellectual assessments and priorities of its creators […]"
>
> (Black 1997:133)

The print historical atlas is a venerable class of cartographic product, the definition for which I accept here as a collection of maps "applied solely to the study and teaching of history" (Goffart, 2003:8). The print genre has been examined well for complementary time frames by Goffart (2003), for the period from the 16th to 19th centuries, and Black (1997), for the 19th and 20th. Between them hundreds of atlases were analyzed for their changing content, as well as the identities and goals of their sponsors or authors. For Black, these have reflected the changing "nature of our understanding of space and of spatial relationships" (*ix*). Goffart characterizes atlases as visual evidence of the close association between geography and history with an intentionally European focus. Given this base of scholarly analysis, I sought simply to inventory the structure and content of a representative sample of modern atlases with sufficient detail to inform the data modeling tasks ahead (see References, §10.1 for a list).

Although the earliest historical atlases date to the 12th century in China and the 16th century in the West (Black, 1997), Goffart notes:

> "The enterprise as a whole (making historical atlases), however worthy its products, has steered clear of self-examination, either in the past or in the present. Solitary experimentation has been preferred to incremental, group-based improvement. There is no literature for potential atlas-makers. 'Historical geography,' a currently flourishing branch of geography with several learned journals and many interests, does not specialize in the study (or the making) of historical atlases and the potential amelioration of the genre" (p. 6).

According to Sieber & Huber, "powerful digital atlases should not only be able to analyse, process, and model multi-dimensional and spatio-temporal data, they should also focus on excellent graphics and high cartographic quality," (2007:165) if information is to be conveyed effectively. Two contrasting approaches to digital atlas authorship are described by Craglia and Raper (1995), "multimedia in GIS" and "GIS in multimedia." The first seeks to embed sophisticated cartographic functions within the GIS software environment; in the latter, atlases are developed within multimedia authoring environments such as Adobe's Director or Flash, or the open-source Lazlo, and necessary GIS functionality assimilated within them.

4.2.1 Atlas types

Print historical atlases are frequently, but not always explicitly named as such, as in "Historical Atlas of ___" or "Atlas of ___ History." For some themes or topics, atlases are necessarily historical but not named that way, as for example, *Atlas of the Languages of Suriname*, *Atlas of Shipwrecks and Treasure*, or *Atlas of Great Lakes Indians*. Many ordinary (i.e. simply geographic) regional or national atlases have significant historical sections, reflecting the view of regional geographic description and analysis as necessarily historical (*cf*. Sauer (1941) for example). The print and digital versions of the *Atlas of Oregon* (2001) exemplify this.

The highest level division of historical atlas types corresponds to the three commonly cited dimensions of geographic information, space, time and theme.[35] A coarse categorization of types (with some abbreviated titles) includes historical atlases of:

- *Geographic regions.* Far and away the largest category, these can be subdivided by scale, as for the world, global regions (e.g. Southeast Asia), countries, states/provinces, cities and towns
- *Time periods.* Most often these cover a particular spatial region over an interval (e.g. Scottish History to 1707) or for a named era (e.g. Classical Greece)
- *Themes*: As suggested by the titles mentioned earlier, the range of thematic subjects for historical atlases reflects the range of historians' inquiry.
  - *Cultural practices*. A broad spectrum, including religions and belief systems

[35] In Berry (1964) the term corresponding to *theme* is "*characteristic*."

(e.g. Islam, World Mythology); language; art movements; architecture; literature; artifacts (e.g. Shipwrecks and Treasure); political process (e.g. Rural Protest in Britain)

- *Event types*. Especially warfare (e.g. Napoleonic Wars, etc.); biblical events were the subject of a significant proportion of early historical atlases (Goffart, 2003).
- *Population groups*. Most often spanning considerable temporal and spatial extents, and include atlases of "peoples," nations and ethnic groups (e.g. The Jewish People, African-Americans) and individual lives (e.g. José Marti)
- *Sociographic topic.* Settlement, development and trade patterns and characteristics (e.g. World Population History); urbanism (e.g. Urbanism and Architecture) and globalization.
- *Ecological systems*. Distributions and dynamical history of flora and fauna distribution (e.g. U.S. and Canadian Environmental History)

The important spatial aspects of a large variety of historical topics have warranted the production of many hundreds of atlases, the economics of which are analyzed in some detail by Black (1997) . We can surmise their high production costs have constrained their number and scope. Black notes, "factors of cost and organization are such that it is usually the publisher, rather than the author, who takes the decision to begin a project, and who thus sets its parameters and provides a framework for its contents" (p. 133) . The fact that some atlases of world or national history have been profitable probably explains their numerical dominance. Narrower scales of theme or region translate to smaller audiences, and greater difficulty justifying expense. It is premature to speculate exactly how digital media will alter the financial equations involved, but the effect should be positive. Bodenhammer (2008) suggests that contingent on “continued progress in making the (GIS) technology more complete and easier to use,” we can expect a dramatic uptick in interest not only in the application of those tools for historical analysis, but as aids to narrative that may help in “integrating the multiple voices and views of our past” (p. 231). The latter purpose implies a growing role for the emergent genre of digital historical atlas, and an expansion of the above list of types going forward.

4.2.2 Content types and thematic structure

While there are clearly no templates for historical atlas design, there are trends, and structuring of their content tends to follow one of a handful of approaches. Financial constraints have seemingly had a large impact on what publishers, editors and authors even attempt; certainly more ambitious atlases would require larger budgets.

Whether an historical atlas is about a place (or all places—“the world”), a cultural theme, a particular time period or some combination, it is always more than “a collection of maps in a volume,” as the OED tells us. The seventeen atlases examined for this work include varying proportions of maps, text, photographs, illustrations, timelines, graphs, numerical charts, and schematic diagrams. Many have gazetteers, glossaries, and biographical directories. Explicit “references cited” sections are rare—more often there are bibliographies or suggested reading sections. A cursory look at major non-historical atlases unsurprisingly shows them to be far more map-centric, although there is a trend towards adding thematic material in illustrated essays separated from the traditional core map sets.

The text in historical atlases normally takes one of three forms: essays, short blurbs or extended captions embedded in or alongside maps and illustrations, and far less frequently, scholarly notes such as normally found in written histories. The prototypical structural design for a world history atlas centers on a series of illustrated essays (Figure 1-1, for example): two- or three-page spreads focusing on a particular blend of place, time and theme, including a text passage between of 300 and 2000 words, one or several maps, and one or two captioned photographs or illustrations of a person, site or artifact of the period and region. Other graphical elements, like timelines, charts and diagrams, are used to good effect in some atlases and not at all in others. The illustrated essays of most atlases of world history are organized temporally, into sections of roughly equal length, such as "The Ancient World, "The Medieval World," "The Early Modern World," and so forth. One interesting exception is the DK Atlas of World History (Black, 2000), which makes both temporal and geographic first-order divisions (e.g. "Eras of World History" and "Regional History," respectively). Essay topics can vary, even within the same volume, from the prosaic (e.g. "The Far East Since 1945") to narratives reflecting some complexity and literary aspirations, e.g. "The Spread of Islam," "Development of Complex Societies," and "Europe in Crisis."

Historical atlases of geographic regions smaller than "the world" and of cultural themes or eras show greater variety in design and structure, and normally have many fewer contributors.

4.2.3 Entities in atlases

The seventeen print historical atlases reviewed depict the following physical, non-physical and temporal things—on maps and in prose, timelines, charts, graphs, photos and illustrations:

Table 4-1. Entities in atlases

Peopled regions
Probably the largest category, constituting a basic organizing principle for much of what follows, i.e. the dynamic attributes (including spatiotemporal extents) and activities of human groups. These correspond to the spatial socio-economic units (SSEUs) discussed by (Frank, Raper & Cheylan 2001), a theoretical construct for "areal units that are used to describe social or economic phenomena" (Frank, Ibid:21) Polities: bounded zones of authority and/or administration; colonies Peoples; nations Alliances; zones of influence Religious administrative areas Regions of population characteristics: socio-economics, ethnicity, etc.
Extant settlements and settlement infrastructure.
Cities, towns, hunter-gatherer camps Waterworks; earthworks Transportation works: roads, railways, canals Built environment: monuments, significant buildings
Archaeological sites.
Settlement ruins; artifacts
Activities
Human practices, broadly speaking; also non-specific, ongoing actions by groups, e.g., insurgent activity in an area for an interval Technological: metalwork, ceramics, weapons, tool use Cultural: language, writing, religion, art Food production: agriculture, pastoralism, fishing Natural resource extraction and processing Commerce Manufacture Exploration Worship; proselytizing Armed conflict: skirmishing, rebellions Intellectual: science, engineering, philosophy Land use
Events
Particular occurrences, individual or composite, with time-spans from moments (e.g. an assassination or invention) to multiple years (e.g. a war or the post-impressionist art movement)

Armed conflict (wars, invasions)
Political (elections, laws, treaties)
Groups formed and dissolved (commercial, social)
Reigns
Inventions
Social movements and insurgencies
Migration
Natural catastrophes
Colonization/de-colonization; conquest
Individual/group journeys (e.g. expeditions, pilgrimages); life-paths

Time periods

These could be considered special cases of composite events (in cases vice-versa), differing only in scale.

Named eras or epochs, e.g. Renaissance, Hellenistic
Dynasties, e.g. Song, Middle Kingdom

Processes

Series of state changes, often with actions/events or artifacts as evidence and/or causal explanation

Growth; decline; "rise and demise"
Diffusion (of cultural practice, technology)
Transition (in cultural practice, technology)

Physical geography

Topography
Geographical features (hydrographic, peaks, volcanoes, caves)
Land cover
Climatic extents; sea level; glaciers

4.2.4 Authors and contributors

All five atlases of world history reviewed were edited by professors of history at British universities. While the numerous contributing authors represent several disciplines, a significant majority are academic historians (also from British universities). Geography is oddly under-represented considering atlases have map collections at their core. Other disciplinary backgrounds included archaeology, economics, international affairs, and area studies.

Representative historical atlases of regions (2) and countries (2) were likewise edited and/or authored by academic historians. At narrower geographic scales—state or county—and for atlases of periods and cultural themes, there was greater disciplinary variety in authorship, including architecture, urban planning, geography and American studies.

4.2.5 Motivation and Philosophy

Referring to his own *Historical Atlas of the United States*, historian Mark Carnes stated, "this historical atlas, like any historical atlas, offers a set of strong and clear interpretations" (2003:9). R. I. Moore, editor of the Rand McNally *Atlas of World History* (1992) states its single theme as “humanity's gradual progress from isolated societies to a world that is rapidly becoming a single global community,” and defends its admittedly western focus as owing to European and North American culture being the “foundations” of that community. An opposite tack was taken by Barraclough for the *Times Atlas of World History* (1993/1978:13), who describes a deliberate attempt to avoid Euro-centricity, focusing on "what is important rather than what seems important now;" with an "emphasis on the great world civilizations and their links and interplay […] we have not neglected the people outside the historic centres of civilization." Two ends of a spectrum, certainly. In these cases and more generally, editors of the major modern world history atlases seem to be fully aware that all maps have a viewpoint. To this point, Carnes observed “the power of maps is achieved by omitting extraneous detail” (p. 8). That is, by generalization—the compressing and condensing of information—maps impose order that is "suspect."

The primary audiences for these atlases are students, history buffs and the general population, so atlas editors also share a desire to tell stories in visualized historical *narratives* with the most advanced cartographic techniques available, and to go beyond mapping in integrating many other visual media.

4.2.6 Historical argument in atlases: historical process

> "Thomas Hobbes described the Roman Church as 'the ghost of the Roman Empire sitting crowned upon the grave thereof.' So it was elsewhere. The world's religions neither replaced nor transformed the ancient civilizations. They were the means by which the civilizations survived the ruin of their original political structures."
>
> *Rand McNally Atlas of World History* (Moore, 1992:44)

How are such sweeping, sometimes poetic expressions of historical analytic results supported in the visual material that accompanies them? What are the strategies and means for historical representation in this medium? The excerpt above appears in an illustrated seven-page, 2800-word essay titled "Religions and Civilizations." Its four large maps span five and one-half of the seven pages and are titled, "The Making of Byzantium," "The Conquests of Islam," "Religions of the Medieval World," and "The Byzantine Commonwealth." Two photographs of significant mosques are embedded as well. The legend for "The Making of Byzantium," includes: (i) territorial extents of empires, peoples and "Arab control;" (ii) dated battle events; (iii) undated military raids; and (iv) regions (termed "themes") and their capitals. The "Byzantine Commonwealth" map depicts the regions of Orthodox and Latin Christendom throughout Eastern Europe and Central Asia. The typical colored polygons are in this case supported by point locations of specific monasteries, sees and bishoprics with founding dates, as well as "centres of Byzantine artistic activity with dates of prominence." Routes of Viking trade and church province expansion are depicted with lines and directional arrows, respectively. Topographic relief is prominent for all maps.

The point taken from this case is that the broad concepts and causal arguments which are the typical fare of historians are, in historical atlases, demonstrated and supported by a select group of mapped entities. The digital representation of the "transformation of a civilization" entails representing select things; no two historians would choose the identical set and no two cartographers would choose identical means. Databases for large-scale digital atlases, potentially with multiple contributors and ongoing development, must accommodate a very broad spectrum of things indeed.

4.3 Historical GIS projects

Applications of GIS technology for historical scholarship to date fall in three broad categories: (a) large-scale national HGIS projects, such as the Great Britain Historical GIS (GBHGIS)[36] and the China Historical GIS (CHGIS)[37]; (b) reference GISs for narrower spatiotemporal extents, such as "Mapping Medieval Landscapes" (Lilley et al, 2005) or Siebert's "GIS Spatial History of Tokyo" (2000) and (c) projects aimed at answering particular research questions within even narrower extents—for example,

---

[36] See www.gbhgis.org; viewed 7 March 2008

[37] See www.fas.harvard.edu/~chgis; viewed 7 March 2008

“what are the likeliest locations to find Mayan settlements?” (Ford, Clarke & Raines, 2009); “what caused the 1930s U.S. ‘Dust Bowl’ event?” (Cunfer, 2008), or “What Could Lee See at Gettysburg?” (Knowles, 2008a) The first two categories are expected to support many and various projects of the third type. Bol has drawn the distinction between “research-driven GIS versus infrastructural GIS” (2007). Categories (a) and (b) are the latter. Large-scale historical GIS projects also can and do support a few early digital historical atlases. The Vision of Britain web site is one example—“a vast statistical atlas of Britain, organized by theme” that “provides a window into the Great Britain Historical GIS […]" (Vision of Britain, 2005).

The data models for HGIS projects in all three categories are relatively constrained in scope; each is concerned with a relatively small subset of the entity list for print atlases in Table 4-1. The GBHGIS database, for example, models administrative units, population statistics, settlements and cultural artifacts such as images, literary works (travelers’ journals) and historical maps (Southall 2007). Siebert’s relatively broader Tokyo Spatial History project (2000) models administrative units and named settlements as well, but also transportation networks and land cover/land use data.

To date virtually all historical GIS applications are created using a ‘snapshot’ temporal paradigm (*cf.* Worboys 2005; Yuan 1999) and differ from non-historical applications only in that date values (if any) are earlier. One exception is Frye’s

Boston 1775 HGIS,[38] which uses a novel citation data model (2008). The entities found in the HGIS projects reviewed (Table 4-2) are essentially a subset of those found in print historical atlases (Table 4-1). The most obvious difference between the two lists is the relatively fewer items within the Activities and Events sections for HGIS.

Table 4-2. Entities in selected Historical GIS projects

Peopled regions	
	Polities: bounded zones of authority and/or administration; colonies
	Population characteristics: socio-economics, ethnicity, etc.
Extant settlements and settlement infrastructure	
	Cities, towns, hunter-gatherer camps
	Waterworks; earthworks
	Transportation works: roads, railways, canals
	Built environment: monuments, buildings
Archaeological sites	
	Settlement ruins; artifacts
Activities	
	Inferred, through geo-referenced media depicting cultural artifacts
Events	
	Spatial footprint change
	Place name change
	Migration
Physical geography.	
	Features (hydrographic, peaks, volcanoes, caves)
	Land cover
	Climatic extents

4.4 Preliminary classification

Although activity and events are seldom explicit in HGIS to date, clearly human social activity forms the greater part of the historical domain of inquiry and a

[38] http://resources.esri.com/mapTemplates/index.cfm?fa=codeGalleryDetails &scriptID=16333

significant proportion of geographical investigations.

Table 4-3. Broad classes of historical 'stuff'

CLASS	PRINCIPAL REPRESENTATIONS
Areal regions	polygons
Artifacts	points; polygons; polyline networks; images
Human activity, aggregated	points; polygons; arrows; text
Events	points; arrows; timeline ticks; text
Eras	timelines; text
Processes	text; small multiple images; graphs; polygon sets
Earth features	points; polylines; shading
Persons and groups	implicit

In the summary list of mapped entities (Table 4-3), there is a natural first-order division between artifacts (including physical things like settlements and the non-physical regions), geographic features, and temporal things. A few things stood out following a preliminary mapping of atlas legend items and cartographic symbols to those broad classes. First, all artifacts are closely associated with events—at the least, their creation and/or destruction. Secondly, by relaxing any tie to specific objects or geometries, we see that much of what historical atlas makers are representing are conceptual entities in the temporal realm. A process of "growing hegemony" as the subject of an illustrated essay may comprise multiple cases of territorial expansion, with the temporal extension of their punctuated composite events, and evolving geometries and attributes of their participants. There are no neat seams between space, time and theme. It became evident that occurrences would be the most inclusive descriptive framework for the information to be modeled.

When viewed in terms of spatial and temporal structure, the temporal entities listed above fall into several broad classes to be modeled. Accommodating these types of data in conceptual and logical models will be an effective measure of generality: i) complex events having sub-events in common, ii) theoretical processes, iii) procedure-like processes, iv) individual space-time paths, v) collective complex paths and vi) flows. Print historical atlases depict all of these in analog fashion, and a successful logical model must support their digital representation. Exemplar datasets were chosen on this basis (§7.2).

A couple of important points raised by this enumeration of entities are tracked in the model development of Chapters 5 and 6:

- Although it is natural to view complex events and processes as temporal entities—after all they occur over time—in fact, most historical events and all historical processes appearing in these lists are human constructs. Ontologically speaking, they are endurants—non-physical information objects describing perdurants (§5.4; §5.5.4; §6.3.6).
- It is necessary to develop ways of classifying events and activity, perhaps formally. This is taken up in §5.6.1 and §6.3.8.

5 A Conceptual Model Clarifying Requirements

Chapter 4 described the first steps in the ontology engineering process set out in §1.4.2— enumerating entities to be represented and introducing some requirements of domain experts as to how. The requirements for a spatial history ontology (SHO) can also be expressed in terms of the questions a SHO-supported system might be expected to answer. Such questions reflect the necessary descriptive and query capabilities for an ontology. They also correspond to a set of "informal competency questions," discussed in §1.3.6, as constituting a baseline for subsequent evaluation. In §5.1 below I pose a set of generalized questions a digital historical atlas can be expected to potentially answer.

Taken together, the classified entities, historiographic requirements and competency questions have informed the design of a conceptual model of the high-level requirements for the SHO. This model has, in turn, driven both the choice of a supporting upper ontology and the design of a database schema for expressing it. In §5.2 the conceptual model and its schematic representation in Figure 5-1 are explained in some detail. Section §5.3 lists and describes a set of *geo-historical information constructs* which have been identified. These may be thought of as patterns, natural divisions or 'regions' of representation requirements. Section §5.4 consolidates some additional requirements concerning classification, space-time paths and indeterminacy that were introduced in Chapter 4.

5.1 Answering what questions?

There exists a large set of generalized questions one could ask of any *non*-historical GIS—including for example questions about attributes assigned to geometries representing administrative areas. These might concern demography, infrastructure, land use and land cover, soils, biota populations, etc. at a given time. Given instead a series of times, one can ask about changes to infrastructure reflecting economic activity and serving as markers for societal change. The following list of questions is a super-set of those—essentially, informal descriptions of queries and functions necessary to describe the dynamic processes that produced the spatial structures found in the GIS. They help to determine the required formal structures for the SHO, and ascertain the presence (or absence) of those in an upper ontology.

The following questions will be referenced in §7.1.3:

1) Where am I/where is this?
   a) Case 1: given an Earth location and a time value (instant or interval), what areas, functional regions, or administrative districts have contained this place?
   b) Case 2: given a place name/location pair, what other names does it have (or has it had)?
2) What has happened here? Given a place and optionally a time period,
   a) What specific events and what *kinds* of events (spheres of activity, e.g. economic, political, military, artistic) have occurred here?
   b) What kinds of activity, variously defined (e.g. particular classes of actions—creative, militaristic, etc.) have occurred here?
   c) What written or visual works have referred to this place; particularly, to

events occurring at or near this place?

3) Given an event (or set of events), who participated, performing what activity, in what role? That is,
    a) What individuals; what organizations or other groups of persons?
    b) What *kinds* of participants?
    c) Was their activity ongoing throughout, or for some period during the event?
4) Given a named person,
    a) What events (and kinds of events) constituted their life? What has been their geographic life-path?
    b) What organizations or other groups of persons did they belong to, when?
    c) What were the products of their life's activity?
5) Given a named group or organization, what has been its membership over time?
6) Given an event or time period, what was happening contemporaneously elsewhere (in one or more spheres—political, economic, artistic, etc.):
    a) at a particular other place
    b) within some distance
    c) at some functionally related place(s), e.g. other democratically governed countries; in places that were trading partners
7) What kinds of events are there, i.e. what classifications of events have scholars in this domain made?
8) Given some such event type, where and when has it occurred? Where most or least frequently?
9) How did the "internal structure" of a complex event change over time?
    a) In terms of its constituent activity types?
    b) In terms of the type and number of its participants?
    c) In terms of its participants' purposes?

5.2 (A case for) event centrality

In the introduction to this dissertation, I suggested that the *event* is the most comprehensive computational container for information about historical human activity. The following arguments are intended to support that assertion.

The inventory in Chapter 4 showed that most of the entities represented in print historical atlases are either events or theorized processes. Of the remainder, most were participants or somehow involved in events—building construction or human births for example. An examination of the prose in those atlases revealed the centrality of verbs and verb phrases, as *case grammar* theoreticians since the late 1960s would have predicted. According to Cook, “in theoretical linguistics, case theory is used as an approach to sentence semantics […] in a predicate calculus framework” (1989, p. *ix*). To the extent that language and the deep structure of grammar reflect how humans conceptualize and reason about phenomena (a great deal, many would argue), case theory models have demonstrated the fundamental centrality of action, or ‘happenings’ in our sentences about what goes on around us. If verbs are central to all sentences (they are) and the purpose of sentences is to make propositions about reality (it is), can there be any doubt of the value of event-centeredness in information systems? If we make a simple graph of the things in atlas maps—say events, people, material objects, and settings—we see the only object directly connected to all the others is the event. Virtually all relationships between persons, things and places are a function of, or mediated by, events (Figure 5-1).

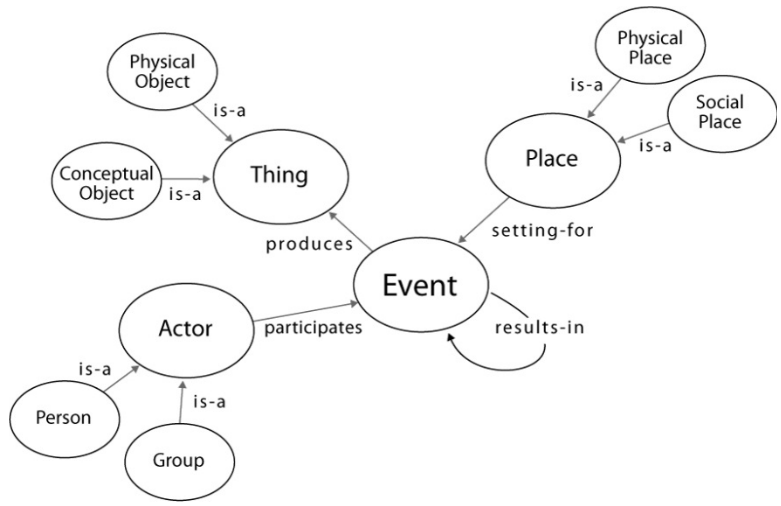


Figure 5-1. Event centrality

A more pragmatic argument has been made with respect to the CIDOC-CRM, an ontology developed for documenting material cultural heritage artifacts in the domain of museums which has been investigated and used in this research (§6.2.3). Doerr and Iorizzo (2008) describe how its event-centered approach permits "a picture of history as a network of lifelines of persistent items meeting in events in space/time […]," an "extraordinarily powerful" model supporting a "surprising wealth of inferences" (p. 5-8).

5.3 Temporal substance and temporal information constructs

I have been speaking quite naturally about events because we all know what they are; at least there are two common senses of *event*: instances of change, and a more complex set of happenings, like an election, a conference or a war. That said, Allen and Ferguson have stated, "…the world does not really contain events. Rather, events are the way by which agents classify certain useful and relevant patterns of change"

(1994). This points to a riddle faced by anyone who would use an ontological approach to data modeling: are we representing some imagined reality, human conceptions of it, or semantic structure in language? The answer for this work is, conceptualizations by means of language. While this might seem arcane, and I might have preferred to avoid it, I have found it necessary to make a few Ontological commitments—at least to a temporal reality.

The terms used for temporal entities—particularly *activity*, *event* and *process*—have been defined variously by researchers in many fields, including knowledge representation and linguistics. This circumstance is analyzed in some depth by Worboys (2005) and Galton (2006). The position taken in this dissertation is that there is no single correct set of definitions. Individual investigators can adopt one set, or as I have done, develop a formal synthesis that suits their representation goals. This section presents a conceptual model for temporality, which is then expressed formally in Chapter 6 and implemented in the exemplar database described in Chapter 7.

Worboys has proposed an effective approach to representing events as computational "objects," showing that this metaphorical association with physical objects holds true in important ways: both event objects and physical objects can be classified in subsumption hierarchies and have instances; both can have temporal parts; and both have relations with other objects at both the class and instance levels (2005).

I propose to extend this metaphor, in suggesting that just as physical objects are composed of one or more material substance, discrete temporal objects like events are composed of activity. At least it will be useful in the design of certain information systems to represent things as such, whether or not *in actuality* there is something in the temporal realm corresponding to matter in the physical realm. This basic formulation is not novel. Supporting arguments are presented in the immediately following §5.3.1. A somewhat novel formalization in terms of the participation relation is described in §6.3.2.

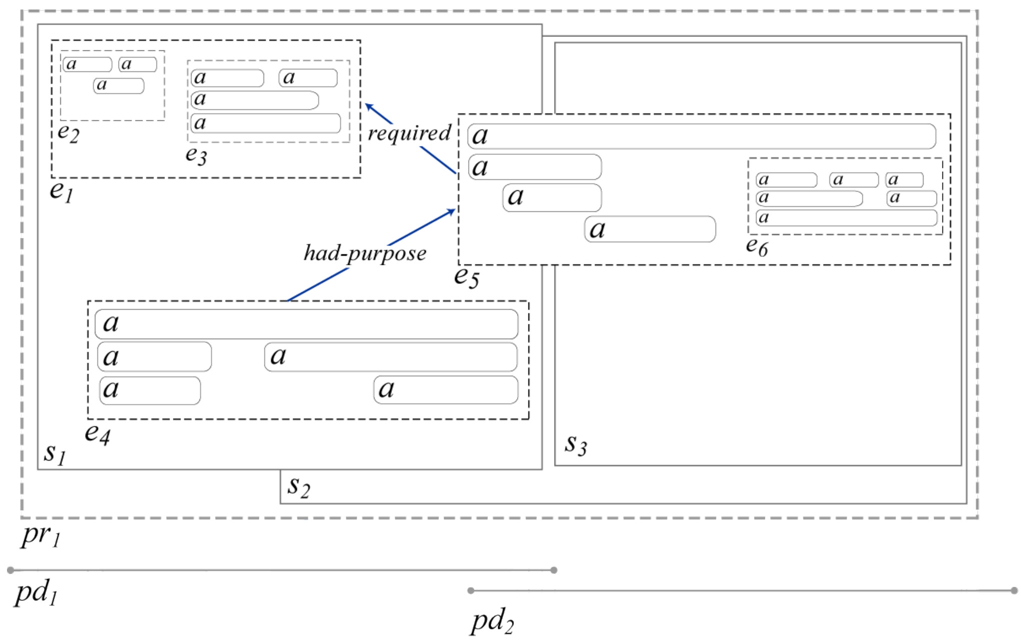


Figure 5-2. Asserted historical-process *pr*$_1$ spanning periods *pd*$_1$, *pd*$_2$, having component activity instances a, events e$_n$, states s$_n$ and the relations *required* (e$_5$, e$_1$); *had-purpose* (e$_4$, e$_5$).

Physical objects are composed of matter, which naturally occurs as kinds. There are solids, liquids and gases: wood, stone, water, hydrogen and so on. Material things-in-the-world—*physical objects* (man-made or not) and *earth features* (rivers, mountains, etc.)—are composed of kinds of matter, in varying proportions. We usually say amounts of matter are homogenous, but whether they are or not is really a function of the scale at which they are examined. New “atomic” particles are discovered regularly.

In the simple model proposed here (illustrated in Figure 5-1 and formalized in §6.3.2), events are composed of one or more instances of activity (human or natural); compound “macro-events”[39] are composed of sub-event parts and in cases, non-event activity instances. In this way, activity can be considered temporal substance. Sticks are made of wood; this walking stick, some wood and (its handle part) some brass. A political rally might consist of instances of gathering and chanting activity and a speech event—itself composed of speaking and applause activity. The parts to be represented will depend on data resolution and the desired granularity of analysis.

More complex aggregations of activity instances and events, including *historical-process* and *period* follow from this. Both are decidedly human constructs. This is the “troublesome seam” referred to in Chapter 1, between reality and human conceptions (which are themselves quite real). For this domain, I define historical-process as a non-temporal entity—a *theory of event relations* (§5.5.4) describing a meaningful sequence, not unlike the composite event in Worboys (2001:134): instances of change bound "in some sort of unity or unifying principle." As depicted in Figure 5-1, these relations could be topological (e_4 *overlaps-with* e_1), mereological (events e_2 and e_3 *part-of* e_1), or telic (e_4 *had-purpose* e_5) and (e_5 *required* e_1). The basic ontological

---

[39] A term borrowed from historical sociologist Roberto Franzosi (2010).

presumption is that such descriptions are of real phenomena, which are temporal entities. This formal dependence is explained in Chapter 6.

An historical period is roughly analogous to a spatial region. Boundaries in either case may depend upon empirical criteria and be uncontroversial (Reagan Era, Lakes District), or be subjective and contested (The Enlightenment, Palestine).

To round out the space, time, theme triad it would be interesting to try extending the object metaphor to concepts—to ask whether complex mental objects are composed of one or more bits of conceptual substance. Is a screenplay composed of conceptualizations (more or less shared) of physical and mental objects stitched together in asserted relations? That notion lies outside the present scope (the metaphor in use already!), although it is touched on again in §8.3.4 as future work.

It is necessary to define the key temporal terms appearing in Figure 5-1:

There is <u>***activity***</u> (*a*), those real world ‘goings-on' we may consider primitive temporal substance, analogous in many respects to material substance. Activity sub-classes distinguish non-specific ‘occurring’ as either <u>agentive-activity</u> or <u>non-agentive-activity</u>. Agentive activity is associated with sentient or robotic agents—that which humans, other animals or robots engage in. In English, most classes of agentive activity are referred to with the gerund or ongoing form of action verbs. They may correspond to physical motion (farming, speaking, moving, writing, fighting, etc.) or be non-physical (mental, legal, even state-like), e.g. inventing, owning, loving, etc.). It is important to note that terms for activity are routinely used at various granularities. 'Farming' can be seen as aggregating types of activity we don't happen to be modeling; speaking is more nearly atomic. Non-agentive activity includes such environmental phenomena as physical flows, erosion, seismicity, etc.

Any set of one or more activity instances given temporal bounds is an event. It may be desirable in the future to differentiate a kind of aggregated activity (aggActivity) object, to cover such cases as "maize farming in pre-Columbian Mesoamerica" or "21st century glacial retreat." These hardly seem like events in the usual sense.

Therefore, an ***event*** (*e*) is an aggregation of one or more *activity-instances* and zero or more *sub-events* standing in *part-of* relation to it. Events may be infinitely complex, dependent on the granularity of the data or scale of representation. The correspondence with material substance is maintained—that stick is composed either of wood, cellulose fibers or C, H and O atoms according to application requirements. A given battle in war may be usefully represented as an individual event or myriad sub-events—each with known duration and composed of activity occurring either "some time during" or "throughout" it. The spatial extent of an event is the union of its component sub-event locations. Similarly, the temporal extent of an event is the union of its components' intervals.

A ***state*** (*st*) is a set of observations and measurements: the values of one or more properties of one or more endurant, asserted to be valid at some moment(s) of observation and measurement and/or 'valid enough' during some interval to be useful. Decennial census figures are often asserted to represent the attributes of a population for a ten year period; far longer 'valid enough' periods are commonplace in historical research.

A ***historical-process*** (*pr*) is a collection of two or more *events*, zero or more *activity* instances and zero or more *states*, having asserted *influenced-by* relations (or sub-relations thereof) amongst each other. In other words, it is a theory of event relations. In several other formulations *process* is roughly synonymous with the sense of *activity* used in this research (Vendler, 1957; Galton, 2006; Galton, 2009; Grenon & Smith, 2004; Niles & Pease, 2001). For that reason, in this work the term *process* alone will be set aside to avoid confusion with the non-temporal *historical-process* we need to model.

A historical ***period*** (*pd*) is simply a named temporal interval. Its bounds might be asserted as calendar dates, event instances or some function (e.g. the bounds of the temporal sum of events having participants of some type).

Events, historical-processes and periods all have temporal bounds, locations, participants and results. All can be described in terms of *n*-ary predicate functions on those dimensions, at levels of granularity supported by the data. For example, $r\ (e_i)$ might return entities having a *result-of* relation with any components of e_i, and $l\ (e_j)$ the union of locations specified for all parts of e_j.

5.3.1 Discussion

Galton (2007) has suggested that “events are made of ‘process-stuff,’” just as objects are made of matter (p. 327). The same holds true for Bach's algebra of events (1986), in which events are formalized as "bounded bits of process" (p. 9). Atomic events are entirely characterized by their process component. Complex, or “plural” macro-events are composed of child sub-events and may be characterized (i.e. typed) by the various process components of those children. This intuitive arrangement is adopted for this work. Note that Galton’s and Bach’s ‘process’ corresponds closely to ‘activity’ in this research.

The argument that occurrences are ultimately composed of activity is further supported by the distinction from action theory, between *activities* and *performances*; As noted by Simons, “performances *take* (a certain) time [. . .] whereas activities *go on for* a time” and furthermore, that “performances, but not activities, may be brought to completion” (2000, p. 132). The key point here is that with respect to digital representations, a performance (clearly, an event) is one or more bits of activity we have chosen to discretize. Galton concurs on this again in (2009): “a durative event is made of processes […or] composed of a number of distinct process chunks representing different phases of the event as a whole” (p. 4).

There is a related mereological question as to whether activity (the elements of an event that are not themselves events) are parts or constituents. Assuming the following declaration of Simons is true, “a *temporal part* of an occurrent is a part containing all simultaneously occurring parts of it” (2000, p. 132), then instances of activity which have unspecified duration cannot be temporal parts of events. This DissertationWriting event has several temporal parts: Draft01Writing, Chapter4Writing, etc. There are a few kinds of activity (writing, reading, discussions) and numerous instances thereof which comprise each temporal part. Any such activity instances having specified temporal attributes (start, end) are themselves events—sub-events of the larger DissertationWriting event. Yet in many cases we wish to describe events with respect to the kinds of activity they were composed of. It seems very natural to describe this relationship as *composition* or *constitution*. Statements of event composition are naturally made in the context of a complex participation predicate: *participates*(actor, event, activity, role, duration). In this case, allowable values for duration are limited to "some time during" or "throughout." Again, if specific start and end times *within* the event duration are to be represented, the activity becomes a separate sub-event. This brings home the earlier point of Allen and Ferguson (1994)—that some activity is an event if and when we say it is.

According to most accounts of Mereology, events do indeed have temporal and spatial parts; a baseball game event is composed of potentially myriad sub-events: a home run, a strikeout, etc. We can talk about (and formalize) its spatial parts as well: e.g. that which happened on the playing field vs. in the stands.

I argue that in a commonsense way, any given event is composed of one or more instances of activity (i.e. "happenings"), which are not themselves events. A strikeout is composed of some throwing activity, and possibly bat-swinging. A class session is composed of some lecturing, lots of note-taking (one hopes) and discussion. When it comes to representing events in digital systems, the same issues of scale and granularity are encountered as we find with physical phenomena. At what point does “some activity” become an event? I'd say at the same point “some wood” becomes “a stick”—when we define its boundary. This can easily become a linguistic and philosophical thicket[40], because all of our terms for temporal things do double or triple duty, dependent upon scale. Activity might refer to the finger motion of a baseball catcher in giving signs before a single pitch, or all of the warm-up throwing and catching going on during a game (the “bullpen activity”). We can talk about “Central American economic activity” during a month, a year or a decade, and in each case be referring to myriad complex events and untold individual activity instances.

Theoretical speculations about temporal reality and the resulting linguistic permutations are interesting and possibly worthwhile, but in the interests of a simple pragmatic approach to modeling historic occurrences, this dissertation proposes the following modeling strategy; its relation to reality is unknown. It relies on the primitive relation of *participation*. We say that the participation of an endurant in an event is constituted by one or more types of activity, possibly in some role and either *throughout* or *some-time-during* that event. If we know (and choose to specify) the particular sub-interval of time for that activity, it becomes a sub-event. That is, we choose to discretize one or more particular instances of activity as events. Alternatively we can say that an actor’s participation in an event is characterized by some activity; that *activity*, like *role* is an attribute of participation, a predicate with

[40] see the *Free Process Theory* of (Seibt, 2004) and Galton (2007) for example.

four terms:

participates (endurant, event, role, activity, duration)

There needs to be flexibility in how events are represented and the participation relation is the formal means. We can say only that someone had the role of “speaker” at an event or that they had the role “honoree” and engaged in speaking activity. In fact, a participant can have multiple roles and engage in multiple kinds of activity. When we want to assign a temporal interval to some activity, and possibly associate it clearly with another bit of activity, it becomes an event with two participants, in roles performing some activity type.

Several taxonomies of activity type are possible, and are application-dependent. At an elemental level, activity might be classed according to whether it principally involves motion (traverse, gather, diffuse, disperse), growth and contraction (physical or metaphorical), interactions (exchange, transfer), creation (produce, birth, manufacture) or destruction. Human activity can be usefully classified by purpose or “sphere” at any level of granularity, for example: economic, political, artistic, military and so forth. These issues are taken up further in §5.6.1 and §6.3.8.

5.4 Facets of historical description

After examining print historical atlases and HGIS projects, I selected several exemplar datasets (§7.2) and using a process akin to that in case grammar analysis (Cook, 1989) and echoed in (Kuhn, 2001), I located verbs, verb phrases, events and linguistic structures tying events and other temporal entities to each other and to other kinds of entities. A preliminary analysis of relations proceeded from those steps, the representation requirements discussed in Chapter 4, and the questions in §5.1. The result is illustrated in Figure 5-1. Having made a case for event centrality, their position in one corner of the figure may appear counter-intuitive; this will be explained presently. The items appearing in ellipses represent facets of historical description; some correspond to computational objects that will be specified at a later stage, others do not.

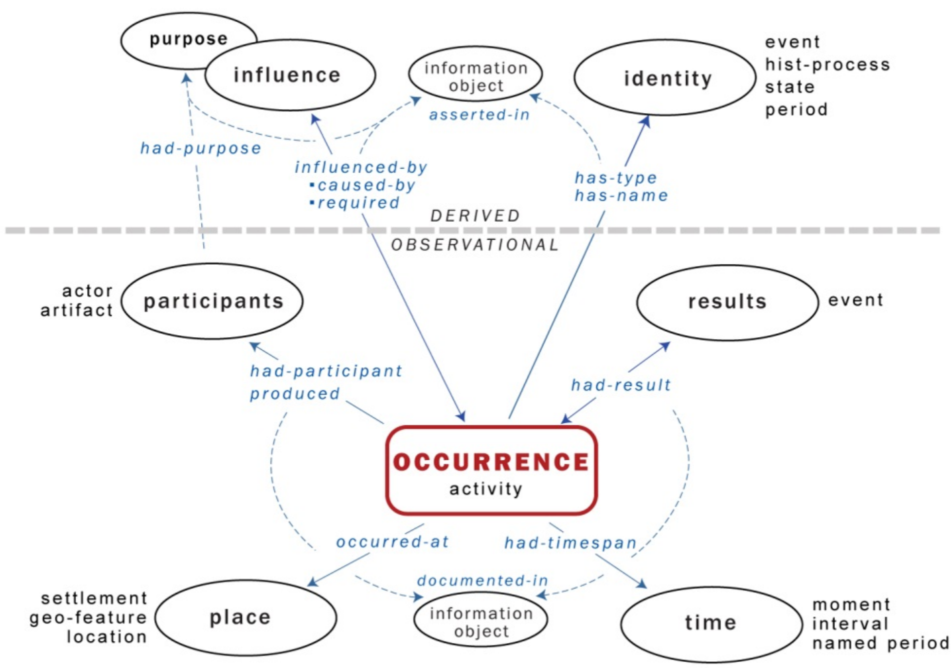


Figure 5-3. Conceptual model of historical knowledge representation requirements

The model serves both to inform the choice of a compatible upper ontology and to illustrate the requirements for domain ontologies and data models beneath it. *Occurrence* is the central organizing principle for representing the fully contextualized dynamics of human activity, which includes—in somewhat circular fashion—the production of information objects about human activity. Occurrence in this model stands for the fact of dynamic phenomena, of things happening. Particular discretized instances of activity, events, states, and historical processes are indicated near "identity."

To describe an event comprehensively is to represent a very broad spectrum of knowledge. Events are situated at one or more place and time; they have participants in roles ranging from simply *present-at* to any level of specificity, often performing some activity; they have results or products, such as state changes, other events, creative works, ideas, or human offspring. These facets of description—place, time, participant, result—are shown below a line drawn to differentiate observational data and derived knowledge. This follows the *Pyramid Framework* of Mennis, et al. (2000) and Peuquet (2002), and reflects an important epistemological consideration. In very general terms, information below the line can be considered a factual substrate, and assertions in this category are ordinarily documented in databases and authority documents of various kinds; as there are frequently multiple and differing accounts, representing their provenance is essential.

Above the observational/derived line are those knowledge objects generated by human analysis and interpretation—"higher-level semantic abstractions that can be derived from that (observational) data" (Mennis, et al 2000 p.502). Foremost of these are declarations of identity[41], including the class or type of occurrence, and its asserted names or labels. The most basic class of occurrence is activity, shown below the line as the only declaration about temporal reality made here; i.e. there *are* dynamic phenomena. As the extended discussion in §5.3 argued, events, historical processes and even states properly belong above the line, as they are human constructs—discretizations of activity. Moving down the hierarchy in classifying kinds of events, the subjectivity persists: one observer's "insurgency battle" is another's "terrorist act;" classification can have empirical criteria, but even then the choice of dimensions is always subjective. The Correlates of War project[42] has developed precise definitions for encoding the attributes 'militarized disputes' (i.e. intra-, extra- and inter-state wars). In the physical geographic domain, several authoritative storm scale systems use differing measures and thresholds to classify severe weather.

Influences upon phenomena—including purpose and causation—appear in this model because they are common products of historical and normative analyses; they are naturally placed above the line. The degree to which a database can *explicitly* model these assertions and discovered relationships is an open question. Certainly if they do not appear in conceptual models at this stage, they will never make it into subsequent data models. The fundamental requirement for attributing sources in both

[41] Identity for Golledge (1995) is one of four "primitives of spatial knowledge." Location, magnitude and time are the others.

[42] http://www.correlatesofwar.org/

observational and derived (data and knowledge) cases is indicated by dashed directional arcs in both regions of the figure.

Each element of Figure 5-3 is now discussed in greater detail; ellipse elements appear [bracketed]:

Occurrence—the fact of activity—appears in a central position below the line, indicating an ontological commitment to a physical reality that "exists in space and evolves in time" (Frank 2007: 408). Discretized and aggregated activity (*event*, *hist-process*, and *state*) appear above the line, alongside [Identity] as they are to be modeled as human knowledge constructs, as opposed to raw observational data. The distinction between events as information constructs versus unequivocal temporal facts is a central ontological issue; relevant discussion appears in §5.3. While a historical process such as “democratization in 19th century Great Britain” can be said to have factually occurred, which events constituted the Swing Rebellion of 1830 (§7.2.1) and which relations between events of that period were particularly consequential is a matter of scholarly interpretation and analysis. There are several plausible accounts, and it would be interesting to analyze their differences visually and structurally.

Place and time. Activity instances or simple events usually have contiguous space-time locations, although their representation will be dependent on either the granularity (resolution) of data or the scope of queries. Counter-examples include a telephone conversation (e.g. between two heads-of-state) or a webcast seminar joining distributed participants in a virtual meeting room. By contrast, complex events, historical processes and periods normally have discontinuous or punctuated spatial and temporal extents that are the union of their components' extents.

Participation. Occurrences have participants, performing activities in roles, for intervals. Participants may be sentient actors, or artifacts which are either products or simply present and somehow integral to the occurrence.[43] Roles might be specified at one or more level of abstraction, from subject/object (agent-patient), to activity types (Leader, Speaker, Observer, etc.), to job titles. Activity can be specified by any action verb.

Results. In this model, *activity* is the base temporal element with which more complex temporal information constructs are composed. Defined as, "exertion of energy or influence" (Activity, n.d.), activity implies change of some sort (identity, position, size, or other attributes). Note that change is not modeled explicitly in this work. It follows that all occurrences other than state have results. Occurrence results might be asserted explicitly (<e_1 *had-result* r_1>), inferred using reasoning engines or discovered by means of ad hoc queries or functions (*results* (e_1)).

Identity. There are two broad aspects of identity to be addressed: that conferred by membership in one or more class sets (animal123 *has-type* Dog), and declarations of uniqueness or individuality (animal123 *has-name* ‘Buster’). In the first case, we are

---

[43] E.g., the presence of the Lincoln Bible at the inauguration ceremony of US President Barack Obama

concerned with categories as shared conceptualizations, with the means for specifying their factual (i.e. observed and measured) basis in formal and semi-formal terms, and with their labeling. The extensible ontology sought here must provide a set of basic logical constructors, including *is-a* and *part-of* relations and axiomatic constraints such as *domain*, *range* and *cardinality*, in order that a) event and process types may be discovered and asserted and b) individuality may be tested.

In the second case, although database systems can easily assign unique ID numbers to any entity or reified tuple, in socially authored systems we will need to discover when two objects that have not been registered as identical are.

Doerr and Iorizzo (2008) have noted that co-reference—the assertion that two represented things are in fact the same thing—is an extremely valuable "element of knowledge." They propose a relatively straightforward approach to establishing identity for represented entities entailing the "semi-automatic maintenance of co-reference links." A similar method can be applied to aligning vocabulary terms within a domain. Such assertions of equivalence are readily maintained in a knowledge base, though possibly at high computational cost; a specification for such a methodology is outside the present scope of work.

Purpose and Influence. Human purpose is often implicit and multi-level. Houses of worship are built for the immediate purpose of affording prayer space, and for larger purposes of converting others to a religious faith. Demonstrating causality is notoriously difficult, even with tightly constrained experiments on solely physical phenomena. Nevertheless, human actions are always purposeful and we normally presume dependencies between occurrences. In the domain of historical research, assertions of purpose and influence are the product of systematic quantitative and qualitative analyses. An information system that models historical knowledge must permit representing such assertions, and the demonstration and visualization of their factual bases. A few fairly general sub-relations of *influenced* will go a long way, and might include: *required*, *initiated*, *furthered*, *hindered*, *terminated*, even *caused* (*cf.* Chapters 6 and 7; also, the GEM model of Worboys and Hornsby (2004)).

Information objects. The historical record is characterized by incomplete, conflicting and uncertain data. Interpretation at every level, from simple classification (was it a vase, a cup or a sacred vessel?) to causal and telic theories of event relations, is held in information objects. These must be attributed to the same effect as the *situated knowledge* formalisms proposed by Gahegan and Pike (2006) for geographical, non-historical knowledge. Provenance by means of embedded, "datum-level" metadata (Frye 2008; Plewe 2002, 2003) makes available information about how, and by whom, knowledge was created.

In Figure 5-3 a set of dashed arcs are drawn to indicate that properties of artifacts and occurrences, as well as relationships between occurrences, are documented or asserted in information objects. These range from records of meteorological observations, to tables of analytical results, to scholarly articles and monographs, to first person journals and cave art. Historical knowledge is often drawn from multiple conflicting accounts or datasets, and whether individually or socially authored, a geo-historical knowledge-base must permit the record-level (if not datum-level) attribution of sources. This reflects the crucial ontological premise intimated at earlier: a historical knowledge-base of this type represents *possible worlds*. If you ask ten historians which key events should comprise an account of The Cold War (not to mention the important relations between them), you will get ten non-identical sets. The implications of this for the formal logical model in this work are discussed Chapter 6.

5.5 The Stuff of History II: Geo-historical information constructs

Several sets of requirements have been outlined to this point. The task is now to identify computational objects that can meet them. The term *geo-historical information construct* (GHIC) borrows Couclelis' notion of *geographic information constructs* (2010). These are "information subsets" purposefully selected from a "comprehensive domain of possible data [. . .] semantically appropriate to each purpose" (p. 12). Purposes might be either those of the investigator in assembling the construct for analysis or argumentation, those of human agents in the context of activity being examined, or both.

A GHIC describes the arrangement or assembly of statements into complex objects that we routinely reason about within the historical domain. GHICs are a way of grouping related representational requirements in anticipation of formalizing them. The GHICs considered here include: groups and membership, historical event objects and participation, place, historical processes, attribution, and temporal location. These are somewhat like the *patterns* of Alexander et al. (1977), as interpreted by ontology developers (Gangemi 2005) and computer programmers: "important recurring designs in object-oriented systems" (Gamma et al., 1994:2). The GHICs described in the following pages are an interim summation of requirements for a spatial history ontology (SHO). In each case, a simplified UML class diagram illustrating key entity types and relations is followed by explanatory discussion. I begin introducing entity classes that will require a place in the SHO; these are indicated by abbreviations, e.g. "Functional Group (FGRP)."

In Chapter 6 each will be described formally, as elements of a proposed ontology. Their subsequent implementation in an exemplar object-relational database is described in Chapter 7.

5.5.1 Groups and membership (dynamic collectives)

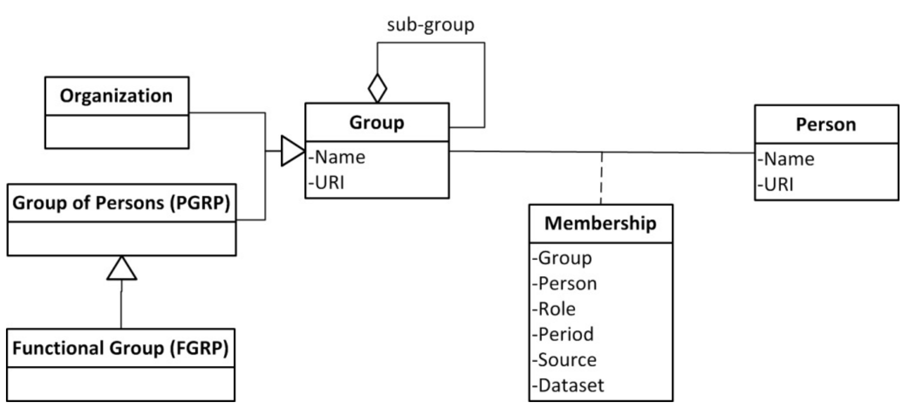


Figure 5-4. Groups and membership

Galton (2005; 2003) has described human groups as one example of "dual-aspect phenomena" having fluid composition and structure that can be alternately viewed as continuants (*dynamic collectives*) or occurrents (*collective dynamics*). In (2005:303) he notes, "observations of the behaviour of […] individuals constitute the 'raw data' from which knowledge of the collective as a whole emerges through conceptualization." In many cases, human groups are related to, or even a function of, non-material social entities like bounded administrative areas and service areas (e.g. a county's population); these have been called *spatial socio-economic units* or *SSEUs* (Frank et al 2001). Note that persons, groups, and membership are bound up with the notion of event participation, which is mentioned in this section briefly and discussed separately in §5.5.2.

The kinds of groups whose dynamic structure and behavior we are concerned with representing include physical groups of persons (PGRP), functional groups of persons (FGRP) and the non-physical social unit, or organization (ORG). We presume that all functional groups and organizations correspond to physical groups of persons; the identity of whose members may or may not be known or identified. An ad hoc PGRP might include: "some people," "a big crowd" or "20-30 sailors." Functional groups are necessary to specify and discover meaningful aggregations of persons without requiring declarations of identity. For example, set (group) membership could be a function of

- Legal or otherwise formalized membership in a role: *citizens*(country); *residents*(place); *employees*(organization); *members*(organization); *registrants*(conference); etc.
- Event participation: *participants* (event)
- Relationships: *friends-of* (person)

- Personal attributes: *has-profession* (x); *has-skill* (x); *has-age* $(20 > x > 12) \wedge$ *resident-of* (place, *t*)

Some examples from these categories include:

- Some participants in an event, as a collection of individual persons, who may or may not be individually identified, e.g. the mob that broke Mr. Smith's thresher at his farm in Kent on 20 Oct 1830 [PGRP];
- The human cargo of a particular slaving voyage; the expedition party of the monk Xuanzang as it was constituted during Fall, 630 CE. [FGRP]
- A spatially and/or temporally unified functional class, e.g. "central European farmers circa 1200 CE." [FGRP]
- A functional class having non-spatiotemporal unifying criteria, e.g. "French speakers." [FGRP]
- The historical holders of a formal office or role, e.g. all U.S. Presidents. [FGRP]
- A legal organization, e.g. the U.S. Senate; the Beatles; BP Corporation. [ORG]

The five exemplar datasets used in this dissertation (§7.1) present cases illustrating associated representation requirements (<u>underlined</u>):

1. <u>Ad hoc</u>: Participants in most contentious gatherings are not a single group, i.e. they are ad hoc, and could reasonably be labeled only "participants in Event423." Such a group might include "MR. SMYTHE, A FARMER," "A CONSTABLE," and "SOME (20-30) WORKERS."
2. <u>Functional</u>: Democratic candidates for president are a group only as a function of their common participation in the Democratic nomination race event. We can and do examine the motion of this group across the landscape.

3. <u>Complex sub-groups and transitivity</u>: The "Napoleon's Advance on Moscow, 1812" event had as participants some (group) parts of a larger group—the 10th Corps and Imperial Guard parts of the Grande Armée.
4. <u>Temporary membership</u>: Members of Xuanzang's expeditionary party came and went; most details of the changes in membership are not part of the historical record, however.
5. <u>Identity survives destruction of parts</u>: The 343 slaves aboard the Spanish ship *Intrepido* between 31 Dec 1827 and 14 Aug 1828, 208 of whom died in passage.

5.5.1.1 Discussion: mereological considerations

In his seminal work on Mereology, Simons (2000) demonstrates that the ontological nature of groups is by no means simple, even suggesting that although the notion of group is intuitively and linguistically unproblematic, there is an ontological question as to "whether there are groups." In English, terms like *class*, *group*, and *collection* are collective nouns that "...all serve the function of referring collectively to a number of objects while the noun phrase is singular" (p. 145). One of the issues Simons raises concerns group parts. The declaration is made,"...in the sense of '<' (part), only individuals, and not pluralities (classes with two or more members) can have parts, the role of plural reference in Mereology is a background one" (p. 150). However, Simons seems to describe Galton's *dynamic collectives* in this passage:

> (it is not) necessarily the case that an object composed of one or more others came into being by putting these together in some way. That may work well enough for some artefacts, but it does not work for organisms, as they are in natural flux, and what composed an object at its beginning may have little or nothing in common with what composes it later. (p. 238)

The task at hand is to identify a simple formal-logical way of expressing the following intuitive notions for the historical domain: (i) groups *have members*, in roles, during some period; (ii) groups also *participate in* events, and by extension historical processes; (iii) groups have sub-group parts which, like any endurant parts, can be temporary.

Domain ontologies for spatial history must certainly capture commonsense cognitive constructs such as found in the texts that are a principal data source for historical scholars. In the Text Encoding Initiative (TEI)[44] guidelines for encoding digital texts, which have widespread and increasing acceptance in the field of digital humanities, one core element tag is <personGrp>, which describes 'a group of individuals treated as a single person for analytic purposes.'

The literature on OntoClean, a methodology for validating the logical consistency of taxonomic relationships in ontologies, tackles some issues associated with groups (Guarino & Welty, 2009). The distinction is drawn between social entities (organizations such as companies, government agencies and clubs), which "*must* be somehow unified" and groups-of-people, which, it is claimed, "do not require a unifying relation, as we assume (they) can be . . . scattered in space, time or motivations" (p. 214).

---

[44]http://www.tei-c.org/

5.5.2 *Event objects and participation*

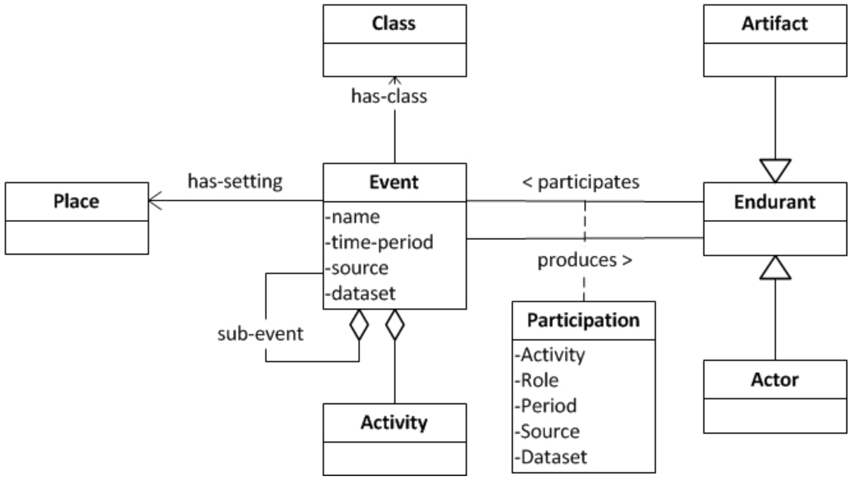


Figure 5-5. Event objects and participation requirements

In §5.3 I took the ontological position that there <u>are</u> events and that they are composed of activity and possibly sub-events. What kinds of events there are, or what a particular event instance is composed of is another matter entirely. It is necessary to model the way that human observers describe events; how we package bits of activity and sub-events variously, in Rashoman-like fashion[45]. The requirements for a simple, formal description of such historical event objects may be summarized as follows:

- Events are said to be *composed-of* activity and are most completely described in terms of participation, or involvement.
- Events have a specified duration; they occupy a (possibly punctuated) temporal 'region' typically represented as an interval or an array of sub-event intervals that could be generalized to a range, or period (see §5.4.2 for further discussion of requirements, and §7.2.4 for one implementation).

---

[45] In the 1950 Kurosawa film, three eyewitnesses to a crime give radically different accounts of what happened.

- Events may have sub-event parts, to any level of detail; these may often be considered as events themselves (e.g. Session 1a and Welcome Reception as proper parts of a conference). They might also be strictly temporal parts or phases (e.g. the first 2 days of the conference).
- Events have at least one participant, defined broadly; that is, any endurant involved in any way is a participant. This includes the endurant products of events (what came into being as a result) and things simply present.
- Events can result in other events, i.e. be asserted as temporal products in a sense. This and other event-event relations form historical-processes, which are discussed in §5.5.4.
- Event participants in agentive roles perform one or more activity, either *some-time-during* or *throughout* an event's duration.
- Event location may be asserted or calculated as the union of sub-event locations and generalized to a simple convex footprint or concave hull. Although in some formulations, event locations are a function of their participants' locations, that is judged here as an inefficient modeling approach because data is seldom presented that way. Rather, participant location at time *t* is a function of the location of events they may have participated in. Note that this permits someone being in two places at once; e.g. attending a county fair and talking on the phone to someone elsewhere, if the location of the phone conversation event is given as a path/flow having two point coordinates.
- Events having no agent participant (e.g. floods, earthquakes, hurricanes) are presumed to involve at least one human participant, a functional group of persons (FGRP) that were present at the location of the event, e.g. *residents*(Port-au-Prince, date).
- Events can be classified, specialized by sub-classes, and further typed in informal taxonomies.

5.5.3 Place

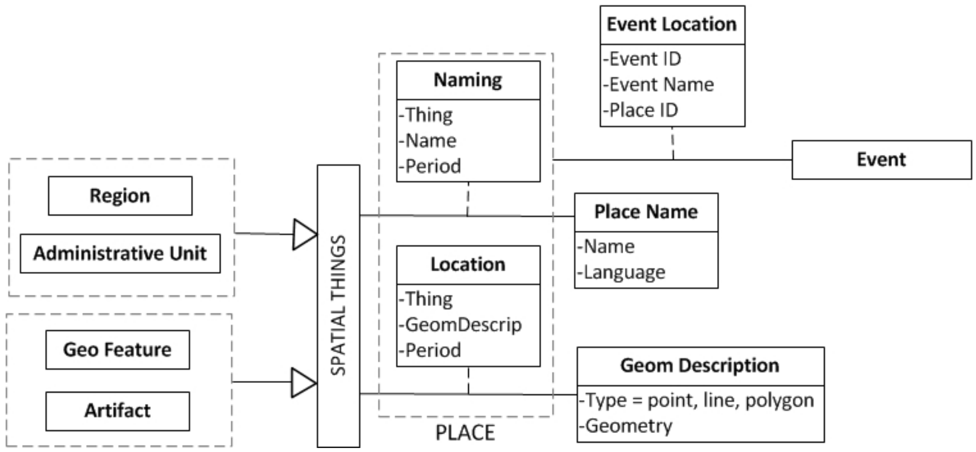


Figure 5-6. Representation requirements for historical places

In the conceptual model developed in §5.4, *place* refers simply to locations where events and historical processes (and their constituent activity) occurred. Places are not simply locations, however. Janowicz (2009:1) has called places "abstract entities used to structure knowledge and to ease communication," and outlined the many challenges in representing them for the historical domain. In historical and cultural geography scholarship, the term has been imbued with many deeper meanings[46]. While it is expected that a SHO-supported digital historical atlas will permit authors to convey varied interpretations of socio-historical places, the representational goals at this stage remain straightforward. This dissertation does not undertake a complete specification of a historical gazetteer (a large undertaking; *cf.* Janowicz (2009); Mostern & Johnson (2008); Janowicz & Keßler (2008); Janowicz (2006)). Rather,

[46] There is large literature on the topic of place. Although many theorists steer entirely clear of the notion of bounded places, the work of Massey (1997; 2005) and others seems hopeful in terms of assuaging their concerns.

places here are descriptions of the locations of things and happenings.

As Figure 5-6 indicates, the locations of endurants—physical and social objects—should be described by one or more tuple of <type, placename, geometry, valid period>. In practice, locations for many artifacts are given as that of another. For example, the Taj Mahal is an artifact of type mausoleum, located at "artifact of type settlement, having the current name Agra, India," or alternatively at [27.174799°N, 78.042111°E], which we might compute as being within an area geometry given as Agra's extent at some time.

Events and historical processes have locations as well, given as either named places or geometries. Named places include both physical and social objects: (i) human settlements of all kinds, and parts thereof (cities, hamlets, districts, neighborhoods, landmarks, etc.); (ii) administrative divisions of territory ranging from precisely described land parcels to nations, and collections thereof (e.g. Soviet Union); (iii) natural geographic features, including rivers, valleys and mountain ranges; (iv) man-made features such as dams, canals and significant buildings; and (v) functional regions and arbitrarily named areas (e.g. "the Industrial South," "Central Highlands," "North Waziristan").

Unnamed places appearing in maps of history are the locales of activity and events—areas or paths, typically—that are often imprecise and have not warranted a permanent appellation, e.g. "the path of Napoleon's 10^{th} Corps en route to Moscow in 1812," or "regions of bronze-working in the 2^{nd} millennium BCE." Data quality is a major consideration: in cases, place names will have to be system-generated, the spatial location given may be imprecise and the valid period of name use may be unknown. Furthermore, ascertaining that two such entries refer to the same place is a non-trivial problem, and important for socially authored works (as discussed briefly in §5.4).

The key attributes of places the SHO must attempt to account for are as follows:

- The locations of historical events are normally given as named places; names may also be system-generated, particularly for aggregated locations like trajectories.
- Locations of activity and historical processes can be calculated from event locations, as can the time-indexed locations of their participants. The locations of complex macro-events and historical processes are the union of their sub-event and collection-member locations, respectively.
- Places have unique identity insofar as possible and are typed, according to a feature-type thesaurus[47] as either named geographic features, geographic-scale human artifacts like settlements, social objects like named regions and administrative units, or as unnamed asserted location(s) of particular events.
- Where possible, a unique Uniform Resource Identifier (URI) pointing to a web-based authority record for the place should be supplied.
- Places (named or not) may correspond to any number of geometric descriptions, according to mapping requirements. These include points, polylines, and polygons (and collections thereof) at various resolutions, using various geographic coordinate systems.

---

[47] E.g. GeoNames (http://download.geonames.org/export/dump/featureCodes_en.txt); Alexandria Digital Library (ADL) Feature Type Thesaurus (http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/index.htm

- Named places should have a temporal attribute, i.e. names are valid for some period. Realistically, this data will be difficult to acquire.
- As place-names can have numerous variants in multiple languages, they should be attributed to sources where possible.
- The represented locations of things-in-the-world (physical endurants) may be either a function of their participation in some Creation or Change event, their spatial containment in some spatial region, or simply asserted, independent of events.
- *Paths*. There is a sub-class of unnamed place that is inherently temporal—locations of certain events constituted by motion or interaction activity. These are referred to here as *paths*, and have two distinct sub-classes: *trajectories* and *flows*. Further discussion of this appears in §5.6.2 below. Briefly:
    - A *trajectory-path* is a sequence of locations corresponding to movement across (or near) the Earth surface—the changing geospatial position of some entity over time. Examples include a military march, an expedition or a person's 'life-path.' Such events occur at a *trajectory-path* place/location.
    - A *flow-path* is a dyadic, directional, temporal geometry representing an aggregation of activity and/or events; its location is given as two points ("from/to," or "source/goal" coordinates pairs) and its connecting edge has no spatial reference. Examples include commodity trade, migration, and diffusion of cultural practices. Such events occur at a *flow-path* place/location.
- An enormous volume of data about place attributes (*states* in this model) is available and will be integral to certain analyses and provide valuable context for many event and historical-process representations.

5.5.3.1 Discussion

Digital historical atlases require robust historical gazetteers of named places, an important research topic in its own right (§2.2.3). Several significant gazetteer efforts are completed or underway, including those for national historical GISs (Great Britain, China, Germany, and United States) and projects such as Pleiades[48] and the Digital Gazetteer of Song Dynasty China (Mostern & Meeks, 2010).[49] Historical gazetteers are time-consuming to produce and therefore costly, however once completed they have potentially unlimited re-use when licensing terms permit. If the global community of historical scholars (granting there is such a thing) were to agree upon a standard formalization, a shared global historical gazetteer could be assembled piecemeal as individual projects are completed. The Place pattern developed here represents a 'placeholder;' a complete specification of a historical gazetteer standard is beyond the scope of this dissertation.

---

[48]http://pleiades.stoa.org/ - A portal to "historical geographic information about the Greek and Roman world."

[49] http://songgis.ucmercedlibrary.info/

5.5.4 *Historical processes (H-Process)*

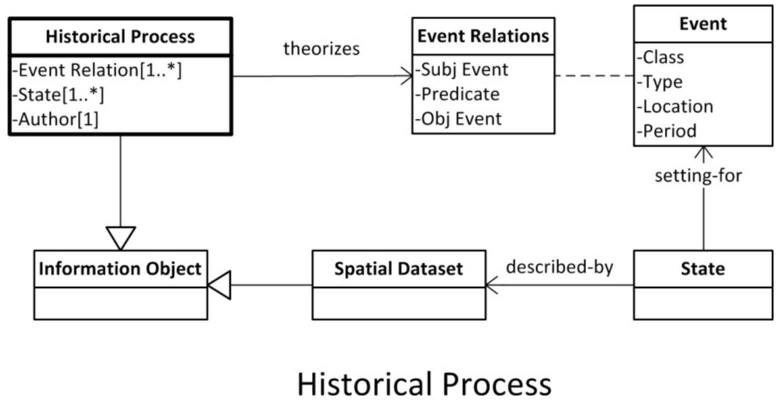


Figure 5-7. Historical process requirements

Although the term process has several meanings, as discussed in §5.3, a commonsense understanding used routinely in historical contexts is what I seek to capture here. Two OED definitions for process are relevant (process, n.d.):

> 4a. A narration, a narrative; an account; a story; a play; a discourse or treatise of any kind; an argument, a reasoned discussion, a disquisition.
>
> 8. A continuous and regular action or succession of actions occurring or performed in a definite manner, and having a particular result or outcome; a sustained operation or series of operations.

The use of the word "regular" in the second case is indicative of a procedure-like or strictly deterministic occurrence. These might be problematic for historians of course, but that said, the Election2008 exemplar in §7.2.2 is a fairly clear example of one. These definitions can be merged (with some license) to arrive at the following meaning for a *historical process* as used in this dissertation:

> A reasoned account or argument concerning a succession of events, in relevant settings, having a particular result or outcome; a theory of event relations.

In print historical atlases, maps and prose are used in tandem to convey knowledge about such historical processes (§4.2). Some very general terms for some include: formation, development, emergence, collapse, recovery, transition, expansion, dispersal, diffusion and ascendency. The degree to which such historical process types may be formally defined is an open question. In any case, digital historical atlases must permit the assertion of historical-process instances as particular sequences of events linked by sub-relations of *influenced* (e_2, e_1), perhaps to include *enabled*, *required*, *hindered*, *furthered* and so forth. More precisely, the SHO needs to model a historical-process as a collection of two or more *events* and zero or more *states*, having asserted *influenced-by* relations (or sub-relations thereof) amongst each other. Potential challenges in representing these in maps or timelines, although ultimately critically important, should not constrain definitions at this stage.

This model suggests the intriguing possibility that historical-process classes could be discovered in large historical knowledge bases by designing queries to search out event sequences having particular characteristics with respect to component event types, participant types, event results and so forth.

States

State is defined in §5.3 as a set of one or more properties of one or more endurants, assumed to be valid for some interval. In the SHO, states will appear only as elements of a historic-process. Examples (some are spatial datasets used in this research) include: census tables of U.S. demographic data, area class maps (e.g. land use, land cover), county election results, *1851 Ancient Counties of England* and the parish-level

Crop Returns for England, 1801. Each describes a set of arguably factual conditions that contextualize the events and processes in an associated knowledge-base (*possible world*) within the system; at least they are asserted to be relevant by the historical-process author.

5.5.5 Attribution

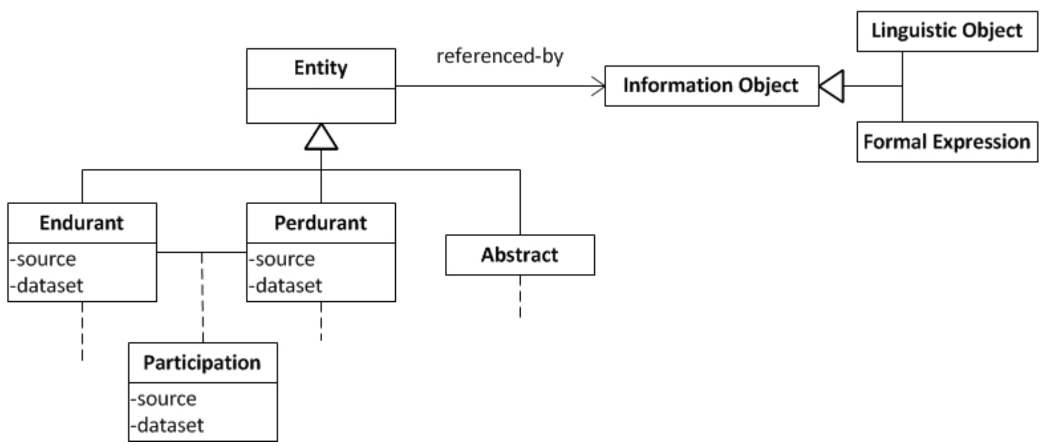


Figure 5-8. Attribution requirements

Digital historical atlases require robust means for attributing statements to sources and permitting system users to view such attribution as well as filter information by source. In this way authors can make explicit the provenance of a *factual substrate* underlying their arguments.

Two classes of atlas applications have been discussed in Chapter 3: individual atlases with one or several authors, and large-scale, possibly distributed systems that integrate multiple individual atlases. In both cases, a potentially unlimited number of references might be added. There are several ways an information object can reference an entity or an assertion about some entity. Possible specializations of the *references*(x, y) relation include *about*(), *lists*(), *classifies*(), and *represents*().

There are three levels of attribution requirements:

- The identification and virtual partitioning of component works within a larger atlas database; all statements belong to a dataset corresponding to a particular project or contributor—a *possible world* as discussed in Chapter 6.
- The attribution of sources; at the record level, statements are identified as coming from a source information object.
- Unlimited, flexible references; any statement in the knowledge-base may be referenced in any number of information objects, each of which may have authors and/or publishers.

These last are critical in socially authored or federated applications with multiple projects, but even in an individual atlas, an author may wish to represent multiple, possibly conflicting accounts of some phenomena. For example one of the exemplar datasets in this work contains multiple sets of 'march' trajectories of the various Corps within Napoleon's Grande Armée as they advanced on Moscow in 1812. One comes from a first-hand account by M. de Fezensac, an officer in the Imperial Guard (1970/1849); another is derived from multiple accounts by the authors of a meticulously researched military atlas and presented on a set of detailed maps (Esposito & Elting, 1999).

5.5.6 Historical periods

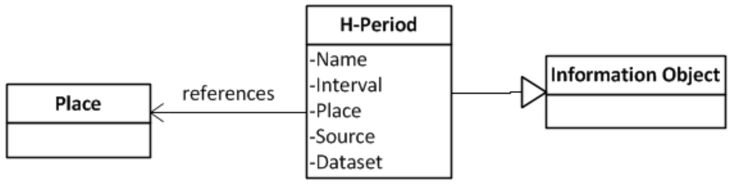


Figure 5-9. Historical period requirements

An *historical period* is, for the temporal domain, roughly analogous to *region* in the spatial domain; it is a named temporal interval. Galton (2005) and Frank (2003) have noted the fiat/bona fide distinction for spatial boundaries (following Smith, 2001) is graded rather than binary. That circumstance has a counterpart in boundaries for temporal regions. The Reagan Era and the Lakes District are relatively uncontroversial examples of period and region, respectively. There is likely to be greater variety in suggested boundaries for The Enlightenment or Palestine. The asserted boundaries for both spatial regions and historical periods may be empirically certain, generally agreed upon, or contested. For an historical period they might be (i) explicit calendar dates, (ii) calculated from the *start* and *end* attributes of certain event instances, or (iii) be relative times, like “before Event A,” “not after 1834,” or “some time between t_1 and t_2”. Only the last case is addressed in this research, by the introduction of a Period datatype that represents time intervals as bounded by intervals, as opposed to instants (§7.2.4).

We must also allow for historical-period parts, and for their having asserted spatial bounds—for example, the *Late Antique (AD 300-AD640)* period in Pleiades is presumably a part of an *Antique* period and meaningful for the Mediterranean world but not China.

5.6 *Further requirements*

Several important requirements fall outside of the GHIC rubric, but I track them here and through the logical and physical models of Chapters 6 and 7.

5.6.1 *Identity: Extensible classification*

Digital historical atlases will represent instances of any number of endurant and perdurant classes. A means for ordering them is necessary, in order to permit queries having a given class as a parameter. For example, the author of an art history atlas may want to compare aggregate footprints of artist birthplaces by style or school. Furthermore, entity class definitions and labeling are hopelessly subjective tasks; the notion of a single taxonomy for 'all the things in the world' is patently absurd.

In a multi-atlas system, each component project will be considered a logically distinct knowledge-base. Its author(s) will have distinctive taxonomies of entity types and predicates corresponding to their own research questions and encoding methods—essentially a distinctive project domain ontology. The exemplar database introduced in Chapter 7 includes the BRIT dataset, which has group types that include *aldermen*, *blacklegs* (strikebreakers), and *gentlemen*; also such broad classes of events as *delegation* and *gathering* as well as narrower ones like *strikes and turnouts* and *market conflicts*. To be incorporated within the larger system, each project's ontology will have to be integrated within the SHO.

We require a means for accommodating any number of semi-formal taxonomies, thesauri and controlled vocabularies developed by individual investigators or communities of domain experts, subsumed under the generalized high-level categories of the SHO. Some degree of inferential reasoning should be possible on them, though perhaps less than on the upper-level. In §6.3.8, an approach adapted from CIDOC-CRM for permitting infinitely extensible controlled vocabularies is described.

5.6.2 Primitives of temporal geometry: ST-Paths

The geometric primitives used to represent spatial properties include points, lines and polygons for object-based models, and grid cells or voxels—in the 2D and 3D cases respectively—for field-based models (*cf.* Couclelis 1992). Temporal primitives are normally considered to be instants and intervals on a single linear dimension, notwithstanding alternate "types of times," such as cyclical time or branching time (Frank 1997). Couclelis has asked: "are there comparable primitives for space-time, and if so, what are they?" (2009). The strongest candidate is the space-time path of Hägerstrand (1970), and that time-geographic construct has been finding its way into GIS data models recently (Shaw, et al., 2008; Shaw & Yu, 2009).

In the most abstract terms, the *geo-atoms*, and related *geo-fields*, *geo-objects* and *geo-dipoles* in a developing general theory of geographic representation (Goodchild, Yuan & Cova, 2007) provide possible answers from a fundamentally field-based perspective: "a geo-atom is defined as an association between a point location in space–time and a property. We write a geo-atom as a tuple <x,Z,z(x)> where x defines a point in space–time, Z identifies a property, and z(x) defines the particular value of the property at that point" (p. 243). On that view, objects are derived in several ways from aggregations of geo-atoms.

I have chosen to extend the Couclelis questions and approach them inductively, by examining a large number and variety of spatiotemporal entities in the domain of human history and seeking commonalities in their most basic properties. I adopt Kuhn's relativistic sense of the term *primitive*, as simply "building block(s) [which are] not necessarily atomic" (2007). The present goal is locating entities that allow efficient and highly usable methods of formalizing conceptual knowledge in databases for a particular if broad genre of computing system, as opposed to identifying smallest individual parts.

Arrows are omnipresent in print historical atlases. Arrows can denote:

- paths of specific individuals and groups, i.e. journey events such as explorations and particular military marches;
- diagrammatic trajectories of military activity as in non-specific inter- or intra-regional sorties, incursions and "pressure" over time;
- directional flows of anything over time, as specific events or non-specific trends, e.g. migration, trade activity, diffusion of technology, practices, ideas, ocean and atmospheric currents, etc.

Kurata and Egenhofer (2005) have presented a formal framework that includes an "Action" class of arrows relevant for atlases. In this dissertation, the immediate solution to modeling movement and exchange data that might generate directional arrows is found in the GHIC of ‘Place’ (§5.5.3), in which Path is a kind of Place, specialized by Trajectories and Flows.

5.6.3 Indeterminacy: probabilistic locations and periods

A significant proportion of historical data are estimates, often from contested textual accounts or maps. Place locations—the points, paths and area boundaries referred to in §5.5.3 above—can be indeterminate and potentially qualified as probablistic or fuzzy. This constitutes a significant challenge for the historical domain (see for example Plewe 2002, 2003). Similarly, temporal attributes may be vague (“latter part of the 13th century”) or otherwise uncertain (“probably in late 1830”). As discussed earlier, there are three aspects, or planes, of geographic information and analysis—spatial, temporal and thematic. Plewe has argued that “the nature and form of uncertainty is…approximately equivalent in the three aspects…” (2002: 431), which suggests similar or compatible measurement and representation strategies, making Kuhn’s vision of integrated reference systems for all three that much more plausible.

Most of these data quality issues, which constitute Mostern’s fourth “primitive of historical scholarship” (§4.1), are outside the scope of this dissertation. Indeterminate spatial locations are addressed only by permitting representations of multiple, possibly conflicting assertions, attributed to their sources. Of necessity, I undertake the implementation of a new Period datatype as an interim approach to representing indeterminate time-periods; it is discussed in §5.5.6 and §7.1.6.

Many of these representation challenges are being addressed in other geographic information domains. For example, see Zhang & Goodchild (2002) for a survey of uncertainty issues in geographic information; also, Liu, et al. (2009) on calculating probability distributions for locations of named places.

6 The Spatial History Ontology (SHO)

> "Perhaps the single most daunting methodological problem [. . .] is that of bringing formal considerations into some kind of meaningful contact with the plethora of phenomena found in the wild."
>
> (Simons, 2000:236)

An ontology development process was begun in Chapter 4 with the enumeration and characterization of domain entities, continued in Chapter 5 with a conceptual model consolidation of requirements in *geo-historical information constructs* (GHICs), and culminates here with a logical model for the domain—a Spatial History Ontology (the SHO). Its implementation in an object-relational database schema compatible with existing GIS software is the subject of Chapter 7.

In the interests of interoperability and the leveraging of good work by others, the SHO is subsumed under and extends an existing upper ontology, DOLCE[50] (Masolo, et al., 2003; Gangemi and Mika, 2003). Some elements of the CIDOC-CRM[51] ontology (Crofts, et al., 2008) are integrated as well; in fact, CIDOC-CRM was instrumental at an early stage of developing the SHO. Both DOLCE and CIDOC-CRM have qualities that are favorable for handling historical knowledge, as will be discussed in the following sections. The SHO is limited, or minimal, in the sense that like DOLCE and CIDOC-CRM, a high-level framework; it does not rigorously define taxonomies of events or artifacts for example.

[50] Descriptive Ontology for Linguistic and Cognitive Engineering

[51] International Committee for Documentation - Conceptual Reference Model

A modified "development time" approach to ontology design has been taken, wherein "semantic content expressed by the ontology (or ontologies) […] gets transformed and translated into an IS component" (Guarino, 1998*:*11) [52] The IS component in this case is a particular database schema. Some implications of that decision concerning logical expressiveness are discussed in §6.1.

In §6.2 the choices of DOLCE and CIDOC-CRM are explained. These are followed by the third 'Stuff of History' section (§6.3), which presents the formal expression of the GHICs described in Chapter 5, and a discussion of extensibility.

6.1 'Sufficient' logic

Most discussion of inference capability for knowledge-bases in the geo-ontology literature refers to systems implementing description logics (DL), such as developed for Semantic Web applications. The logical capabilities provided by the Semantic Web standards for schema definitions (RDF/S) and inference (OWL dialects), as listed in (Allemang and Hendler 2007), are shown in Table 6-1. In an object-relational database management system, class and relation definitions (the TBox of a knowledge-base) are achieved with schema design tools including tables, custom data types, check constraints, triggers and materialized views. Relational algebra, as a subset of first-order logic, can provide a useful subset of inference capabilities with structured query language (SQL). Others can be achieved using customized functions

[52] The approach is "modified" in the sense that an application-level task ontology has not been developed or integrated, as suggested in (Guarino 1998).

of any complexity, written in any of several languages. Note that the RDF/S term "Property" corresponds to "Relation" in the SHO.

Although the full expressiveness of OWL-DL in a native graph model is in theory desirable, at this stage it represents a bridge too far. In my view, the introduction of logical formalisms to historical knowledge representation in GIS should be attempted incrementally. Fortunately for those who want to make maps right away, a useful level of reasoning can be accomplished in a relational model, as Table 6-1 indicates.

Table 6-1. Logical capabilities compared

RDF/S or OWL capability	Achieved in RDBMS
Schema definitions, e.g. PoliticalEvent hasType rdfs:Class hasParticipant hasType rdf:Property	Tables for Class and Relation ('property'); all instance records have foreign key to Class and/or Relation
Class and property propagation (is-a; sub-relation) Property intersection (A *subproperty-of* B; A *subproperty-of* C → if x A y, then x B y, x C y)	Hierarchy established with parent_id value for each; recursive queries using SQL's `WITH`, `WITH RECURSIVE` and `UNION`
Class definitions	Table columns incl. custom data types (`ENUM`); check constraints incl. `NOT NULL; DEFAULT`
Property definitions - Domain and range - Cardinality	Check constraints
Transitivity of parthood, and is-a relations	Recursion, using SQL's `WITH`, `WITH RECURSIVE` and `UNION`
Complex classes, e.g. NewClass ≙ intersectionOf [ConditionA, ConditionB,…]	Materialized views
Differentiating individuals, e.g. owl:distinctMembers, owl:allDifferent	`UNIQUE` constraints, including primary keys
Instance checking	SQL: `SELECT…WHERE…`
Graph query patterns	SQL: `SELECT…WHERE…AND…AND…FROM…JOIN`
Reification; blank nodes	Association classes
Unions and intersections	SQL: `UNION` and `INTERSECT`

NOT READILY ACHIEVED

A number of OWL-DL inference capabilities rely on semantic reasoning software like Pellet with query languages like SPARQL, having functionality absent in SQL-based relation systems.

Domain and range inference, e.g. - maidenName rdfs:domain MarriedWoman - HillaryClinton maidenName 'Rodham' ⇒HillaryClinton is married	Constraining types in a relation, okay; not inference
Equivalence, e.g. sameAs; proper sub-parts of each other	An important requirement, this can be addressed with some difficulty: (i) an Equivalence table consulted in the course of many queries; (ii) making the URI field NOT NULL and providing built-in access to multiple authority sources.
owl:SymmetricalProperty owl:InverseProperty	Use cases for these are not obvious

6.1.1 RDF vs. relations

The RDF model—and description logics like OWL-DL—permit the ad hoc declaration of an unlimited number of predicates (‘attributes’ in the relational model and column headings in practice). The typical RDBMS schema assembles sets of predicates in tables corresponding to entity classes. For example, the unary predicates *firstName*, *lastName* and *birthDate* are ordinarily aggregated in a Person table (relation) under the assumption that all instances of Person have those attributes. For ‘closed-world’ application scenarios in business and scientific analytics, it is normal to make up-front decisions about which attributes of the entities being modeled matter; occasional adjustments and additions are unproblematic.

In a dynamic knowledge base the division between the ontology and instance data is not necessarily fixed. As suggested in the data mining and knowledge discovery in databases (KDD) literatures (*cf*. Yuan, 2009), new entity and relation classes may be discovered in the ABox and added (with attribution) to the TBox.

A large-scale digital historical atlas, particularly a dynamic, socially authored one, does require a relatively easy way to add the entity classes and relations pertinent for a given new project. It is hoped we can ‘grow’ the SHO as new exemplar projects are added and an expert community of atlas authors and historians develops. The requirement for compatibility with object-relational GIS data models is a significant constraint with respect to extensibility. It is fully acknowledged here that a graph model holds some advantages. Future research will aim at further integrating graph and relational models for this genre.

6.2 Choosing an upper ontology

There are several reasons why an ontology for spatial history should be implemented as an extension to one of the existing upper ontologies: (i) an upper ontology as a "development time" tool aids in a conceptual analysis of the knowledge domain; (ii) formal descriptions permit us to gauge the logical adequacy of any resulting database schemas (Guarino 1998); (iii) the viability of large, distributed geo-historical knowledge stores depends upon both the interoperability that shared vocabularies help provide, and the cost savings from developing universal base layers; and (iv) geo-historical information constructs in the emerging SHO requirements have led to perceived new requirements at the upper-ontology level.

DOLCE has been adopted as the upper ontology framework most suitable for supporting the proposed SHO. DOLCE is a dynamic project, in that numerous extensions have been written for it since its introduction and its development is seen by its authors as an ongoing process. Several of these are incorporated in the most recent version, 3.9 of DOLCE2.1-Lite-Plus (DLP397, 2006). These include the *Descriptions and Situations* (DnS) of Gangemi and Mika (2003), which has provided important elements for the SHO. A number of entity classes and relations from CIDOC-CRM are incorporated in the SHO as well.

6.2.1 Why DOLCE

DOLCE is self-described as descriptive and pragmatic—concerned with formalizing common-sense notions found in language. DOLCE assumes possibilist and eternalist views, “including in the domain of quantification all *possibilia*—all possible entities—independent of their actual existence…past, present and future.” (Masolo, et al., 2003:26). This particularly suits the requirements of SHO (§6.2.2).

The Basic Formal Ontology (BFO), often referred to as SNAP-SPAN for its spatiotemporal formulation (Grenon & Smith, 2004), was considered as a candidate framework for the SHO owing to its strong support for temporality. It was ruled out however owing to its central goal of providing a “formal ontology of reality” (Grenon, 2003). DOLCE’s realism is modal; its universe of *possibilia* makes it far more compatible with the conflicting evidence that historians wrestle with. The task at hand for the SHO is formally describing geo-historical information constructs: those scholarly assertions *about* reality that constitute history. With the addition of the DnS extension, DOLCE fits this requirement fairly well.

DOLCE has received attention for its potential applicability in GIS. Agarwal (2005) sees its cognitive and common-sense biases as compatible with many geospatial applications. Kuhn (2007b) suggested its Quality Regions category is a promising basis for metric conceptual spaces. However, data modelers committing to using the DOLCE framework will find a moving target. The most current documentation for various DOLCE versions (‘Core,’ ‘Lite Plus,’ ‘Ultra-Lite’) and extensions is typically found only within incomplete class and relation scope notes provided in OWL implementations. The subjective nature of ontology development is indicated by such statements as, "since the decision of designing an explicit description that unifies a perdurant depends on context, task, interest, application, etc.

[. . .] there can be indecision on where to align an event-oriented class;" also, "A different notion of event (dealing with change) is currently being investigated for further developments: being 'achievement', 'accomplishment', 'state', 'event', etc. can be also considered 'aspects' of processes or of parts of them" (DLP397, 2006).

There are almost always multiple possible logically valid approaches to model a given phenomenon in any ontology, a fact made clear at the OntologyDesignPatterns.org web site, where particularly challenging modeling issues are solicited and discussed by the community of practice. This ambiguity and flexibility has been borne out in the course of investigating DOLCE and CIDOC-CRM as a scaffold for the SHO. One simple example is that a birth date can be readily modeled as either an attribute of a Person, or an attribute of an Event of type Birth, having a person participant in a patient role.

All that said, the *attempt* to meet the requirements of this particular domain with upper ontologies has been useful and instructive. The goal has become to define a minimal logical structure for the historical knowledge domain that will *potentially* align with one or more upper ontologies, and DOLCE is most amenable. This conclusion is supported by the findings by Shaw, et al. (2009), which compared upper ontologies concerning event representation approaches particularly. It is important to note that alignment may ultimately amount to some degree of voluntary mapping of differing concept terms to identical or even similar referent concepts. This negotiation of meaning only becomes possible when both parties are sufficiently explicit about their own meaning.

The choice of DOLCE over BFO/SNAP-SPAN came down to these points noted by Grenon (2003:8), "DOLCE makes room for distinctions between abstract and concrete entities; it makes room for agents and intentionality. BFO is deliberately not committed to these distinctions."

6.2.2 Some DOLCE basics and notation

DOLCE's core "formal characterization" appears in the *Wonder Web D18 Deliverable* document (Masolo, et al., 2003:26-41). Its latest incarnation includes many additional categories and relations which are documented by Gangemi and Mika (2003) and in the OWL-format file for the latest version, (DLP397, 2006). The core documentation presents definitions and axioms for "Categories" and "Primitive Relations" of parthood, temporal parthood, constitution, participation, quality, quale, dependence and presence. Many of these are referred to in the following sections and in some cases repeated, but a complete listing of DOLCE is not provided for space considerations.

The FOL notation used in DOLCE is adopted in this work, as are the following conventions for denoting variables and constants:

- Constants denoting Universals: L, R, Q . . .
- Variables ranging on Universals: Φ, ψ . . .
- Constants denoting Particulars: a, b, c . . .
- Variables ranging on Particulars: x, y, z . . .

For every universal Φ, there is a possible world w, constituted by Φ: K (w, Φ)

Π_{DOL} is a set of “explicitly introduced universals” in DOLCE, {PT, ED, PD, AB…}.

Π_{XDOL} is a set of universals introduced here, extending DOLCE, {PER, A…}.

$\Pi_{SHO} = \Pi_{DOL} \cup \Pi_{XDOL}$

Relations within the text appear italicized: e.g. *references*(*x,y*); within logical expressions, relations appear in san-serif font: references(*x,y*). Core DOLCE relations are abbreviated e.g. Proper Part, PP(*x*,*y*).

Table 6-2. Logical expression reference prefixes

PREFIX	SOURCE
Dd, Ad	Definitions and axioms, respectively from core DOLCE (Masolo, et al., 2003)
dlp	Definitions and axioms from the Descriptions and Situations DOLCE extension, (Gangemi & Mika, 2003).
bfo	Derived from BFO literature, (Smith & Grenon, 2003).
sho	Newly introduced in this work

6.2.3 Possible worlds

The SHO is an *information ontology*. In the sense of Couclelis (2010) and Frank (2007), it is designed to model actors’ observations and measurements of a single, real world and their various assertions about it. For this reason, something like an <*according-to*> relation is required for all truth statements in a SHO knowledge base. Activity, events and processes do in fact occur but accounts of them vary. It is accounts of histories we wish to model.

The SHO will support applications where multiple, conflicting accounts of reality are represented side-by-side in order to be analyzed and compared. The SHO classes specified under DOLCE *categories* range over a domain D consisting of a finite set of possible worlds ($w \in \mathcal{W}$) representing the "maximal states of affairs of such domain" (Guarino, 1998:5). Such possible worlds can be understood as individual knowledge-bases (KBs) within a system, consisting of sets of statements contributed by a particular investigator, or specified as belonging to a particular project or source. Each author contributes a set of statements, each having some degree of certainty; in modal logic terms, either "*necessarily*, it was the case that […].", or "*possibly* it was the case that […]." and its variant, "due to SourceA, possibly […]." A comprehensive world historical atlas will for example have numerous authors, each citing various (and potentially conflicting) sources. The designation of statements (rows) as belonging to a particular set w_n tells us who compiled the project KB, or who acquired the statement and placed it in this compilation. All queries to this system can be restricted to return Boolean values or relations (in the relational model sense) ranging over one or more worlds, *w*. In the case where *w* is not specified, queries range over all known worlds in $\mathcal{W}$.

All universals enumerated in SHO are members of a single Domain noted earlier as the set Π_{SHO}. Each entity class and relation used in a SHO-based system is identified as belonging to its contributing domain. This formulation corresponds to the W3C notion of *namespace*. In this way, SHO is infinitely extensible. Its current constitution is listed in Tables 9-1, 9-2 and 9-3.

Given all this, must we explicitly introduce modal logic——possibility and necessity qualifiers? That is, are all statements explicitly either possibly or necessarily true? It quickly becomes tiresome to preface all statements with "In world *w*, [...]." By requiring a *w* attribute for all statements, application interfaces can be designed to make that qualification crystal clear and unobtrusive, letting the user choose "worlds" and understand that they are viewing or reading but one interpretation of possibly many. I do not take a position on how claims for authority or relative authority get made. At this stage I'm concerned with specifying encoding methods that permit those issues to be addressed at the application level. There is a discussion of this by Masolo, et al. (2003:8). Where the SHO will differ from DOLCE in this respect is that DOLCE "quantifies over a constant domain (of particulars) in every possible world" (p. 26) whereas SHO develops a TBox of merged namespaces (i.e. domains of universals) and multiple worlds of predicates; distinctions between worlds must be made clear at the application level.

6.2.4 CIDOC-CRM

The CIDOC Conceptual Reference Model (CIDOC-CRM)[53] combines an upper ontology similar in some ways to DOLCE with a fairly broad domain ontology originally developed for cultural heritage information in digital applications for the domains of museums, libraries and archives. CIDOC-CRM was developed over a 10-year period by a working group of the International Council of Museums to support applications for cataloging and contextualizing material cultural heritage and is now ISO standard 21127:2006 (Crofts, et al. 2008). A community of interest is working to

[53] See http://cidoc.ics.forth.gr; Viewed March 7, 2008

align the CRM with various emergent interoperability standards. As is the case with DOLCE and BFO, its authors encourage test cases such as these to assess its generality in the broader domain. Like DOLCE, CIDOC-CRM makes a high-level distinction between endurants and perdurants, which it terms *persistent items* and *temporal entities* respectively.

CIDOC-CRM was the initial 'scaffold' for the SHO because of its event-centeredness, its explicit modeling of information objects in documentation of artifacts' provenance, its strong support for the contextualization of material 'things' in terms of relevant temporal entities, and its rich set of relations (termed *properties* in CRM as well as RDF/S and OWL).

The CIDOC-CRM specifies 86 entity classes with multiple inheritance beneath the first-order divisions, *Temporal Entity* and *Persistent Item. Time-span*, *Place* and *Dimension* share that top level. It also specifies a set of 137 property declarations that define the inheritable relationships possible and/or necessary between classes. Together with properties like P11-participated-in, P107-has-current-or-former-member, P67-refers-to, and P15-influenced, these hierarchies enable a basic level of logical reasoning that can be expressed in a relational database system. Properties are further defined as holding for given *domains* and *ranges*. For example, *hadMember* holds for the domain of Group, with a range of Actor.

The SHO is not immediately "CRM-Compatible," achieved per Crofts, et al. (2008) by adoption of a core set of classes and properties, but an adequate mapping to

CIDOC-CRM is a worthwhile future goal.

6.3 The Stuff of History III: Extending DOLCE to SHO

In §6.2 the decision to adopt a DOLCE framework was rationalized due to its philosophical commitments. In this section I step through the *geo-historical information constructs* (GHICs) identified in §5.5 and show how they fit directly in DOLCE where that is the case; where they don't, I introduce and define the extensions required and Figure 6-1 indicates their relationship to the conceptual model in §5.4.

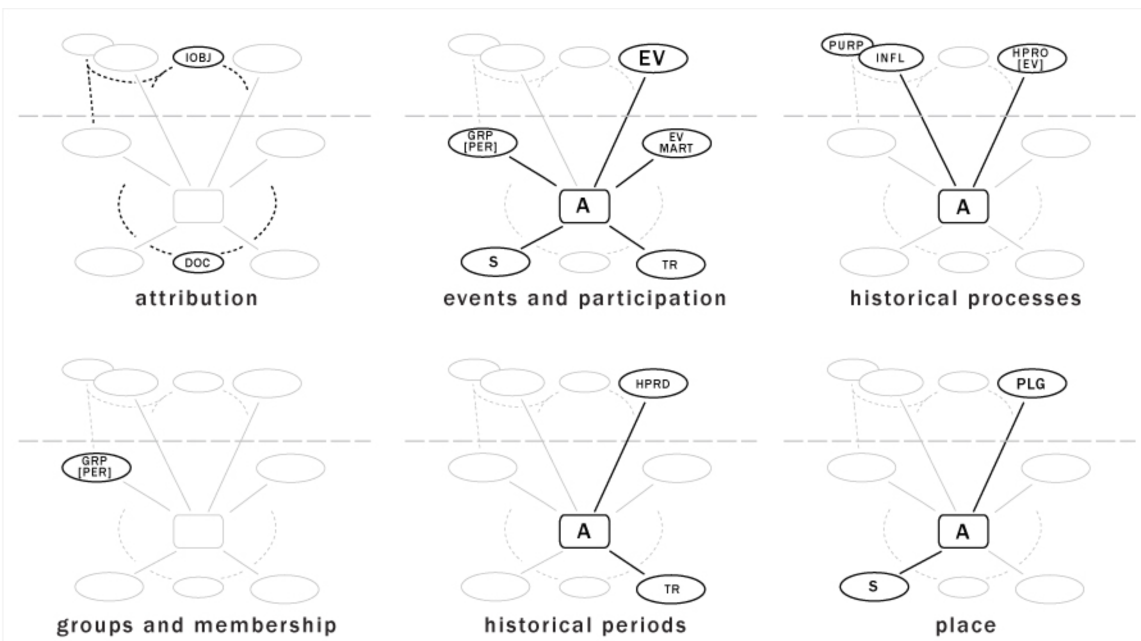


Figure 6-1. Geo-historical information constructs developed in the SHO

Note that the order of GHICs here has been altered from §5.5; I begin with Attribution because of its direct tie-in to the fundamentally important notion of possible worlds. Each section begins with a UML class diagram, followed by modeling decisions taken, including formal extensions to DOLCE for the SHO. These are followed in some cases by further discussion of the reasoning process involved.

The SHO builds upon and extends an amalgam of DOLCE versions. Of 93 core classes in the SHO, 70 stem from DOLCE Lite Plus (DLP397). Of the 23 new classes, 10 are derived from CIDOC-CRM and 13 are newly introduced here. The positions of new entity classes in the DOLCE taxonomy are shown in Figures 6-1 and 6-3, underlined. A complete listing appears in Appendix §9.1.

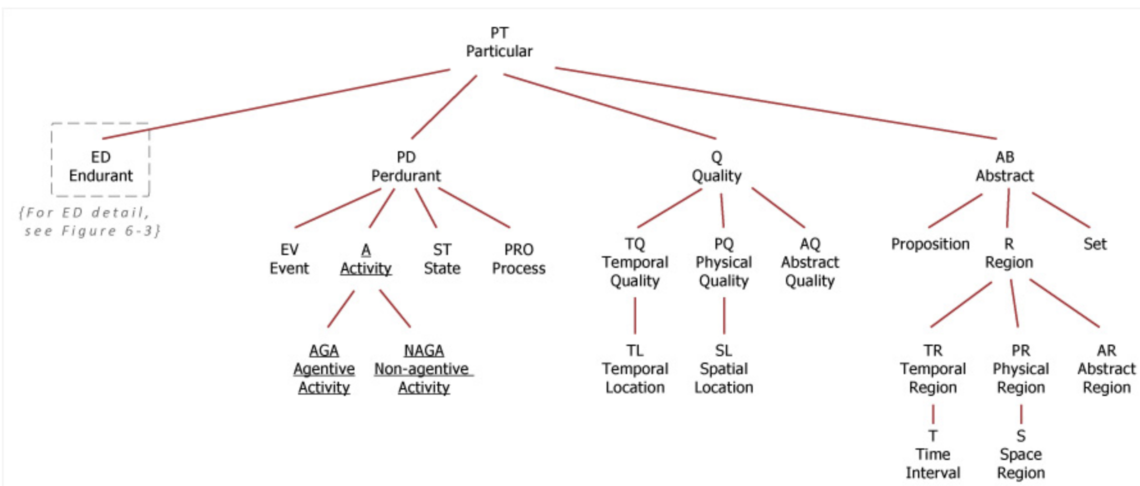


Figure 6-2. SHO under DOLCE: Perdurant, Quality, and Abstract categories

In the OWL-DL version of DOLCE, nearly all sub-classes of Perdurant, Abstract and Quality are defined only verbally, in scope notes. The same is true for many Endurant classes, including most core classes. The remainder are defined using owl:equivalentClass axioms that refer to and build upon the core classes and core relations. DOLCE's "primitive" relations (principally *subsumption*, *constitution*, *parthood*, *participation* and *spatial dependence*) have been thoroughly defined in first-order logic (FOL) (Masolo, et al., 2003). Note that a *relation* in DOLCE corresponds to *property* in both CIDOC-CRM and OWL (and *role* elsewhere). 'Relation' is the preferred term here, although in cases 'property' or 'relation/property' will used for clarity.

In CIDOC-CRM, classes and relations/properties are defined by fairly limited formalisms, and like DOLCE, intensionally in scope notes. Properties are defined with respect to domain, range and cardinality and in hierarchical and transitive sub-property relations with each other. Entity classes are defined only in terms of an *is-a* hierarchy with transitivity. The SHO follows this relatively minimalist paradigm shared by DOLCE and CIDOC-CRM (DOLCE'S primitive relations are the exception), concerned as it is with 'sufficient logic' and pragmatic results for a particular class of systems (§6.1). An immediate product is an OWL-DL representation of the SHO[54] having the $\mathcal{SHOIN(D)}$[55] expressivity of the DOLCE397 release.

The SHO is extensible, as described in §6.3.8: new sub-classes and sub-relations may be described in child XML *namespaces*[56], and selectively referenced by applications developed atop it.

---

[54] Available at, http://www.geog.ucsb.edu/~grossner/sho/sho-latest.owl

[55] A convention in Description Logics for denoting the expressivity of an ontology. SHOIN(D) = Property (Role) transitivity and hierarchy; inverse properties; "one-of" nominals; unqualified number restriction; concrete data types (well-defined concepts).

[56] A W3C recommendation standard: http://www.w3.org/TR/2009/REC-xml-names-20091208/

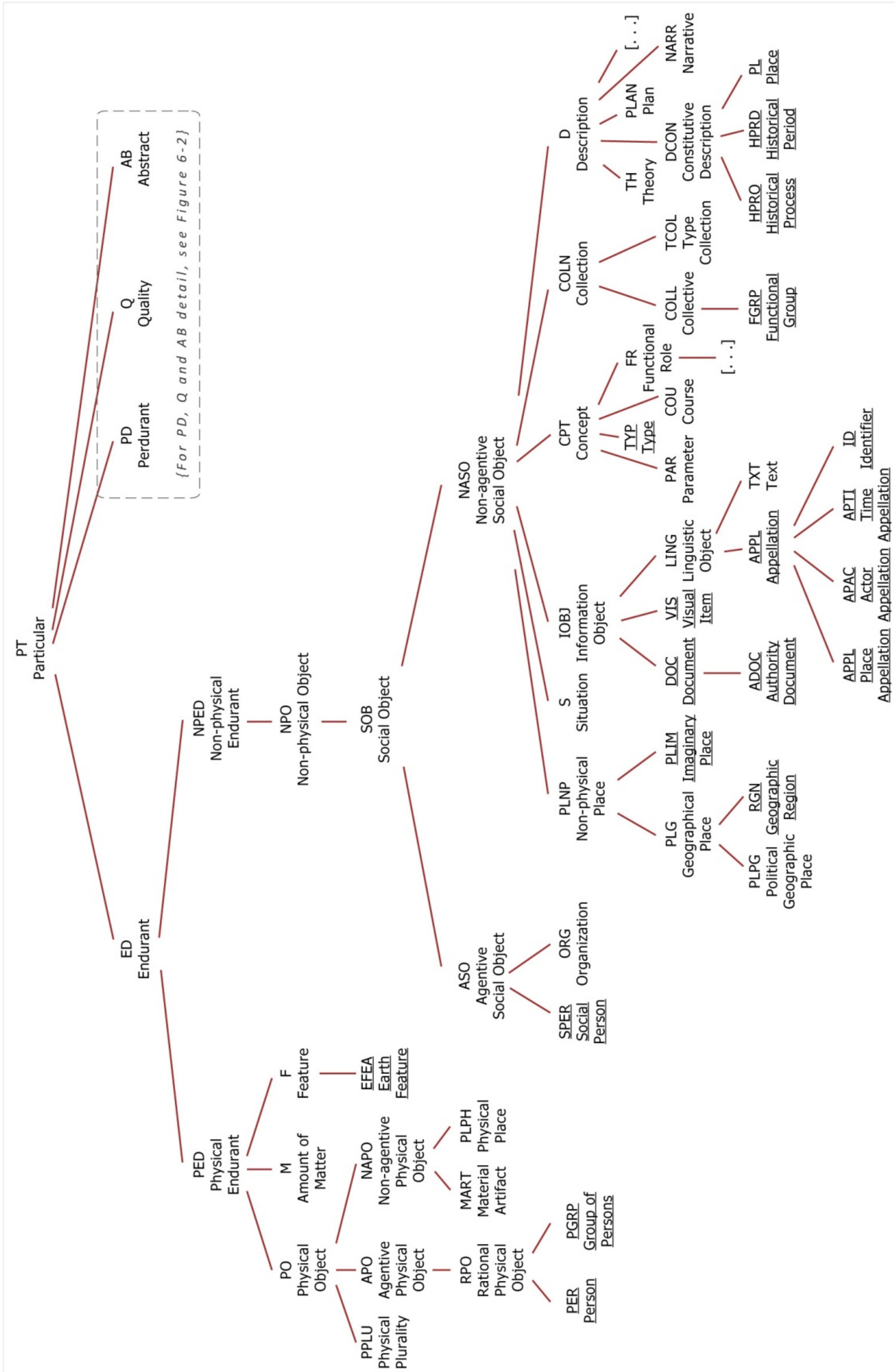


Figure 6-3. SHO under DOLCE: Principal endurant categories (new underlined)

6.3.1 Attribution

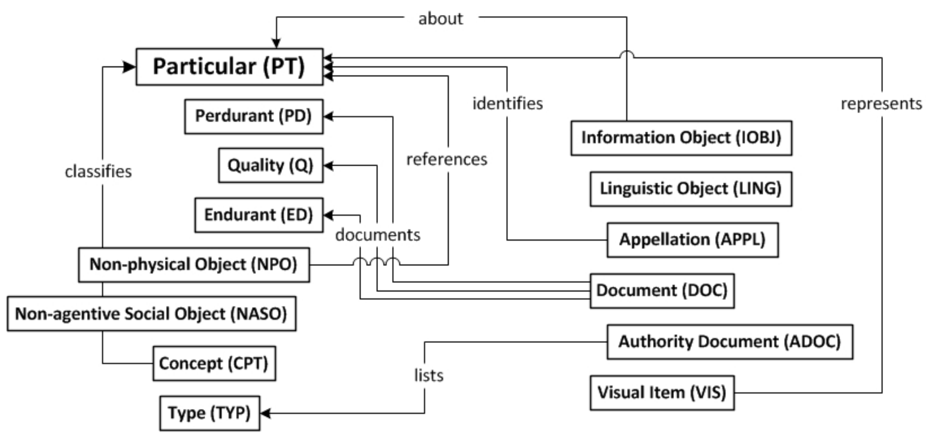


Figure 6-4. Attribution in the SHO

All knowledge-base statements must be qualified as sourced assertions about possible worlds, rather than as universal truths. The means by which a relevant 'world' is identified for each statement has been discussed in §6.2.2. The attribution of sources is readily modeled with DOLCE and CIDOC-CRM relations.

The DOLCE DnS extension has a *references*(x,y) relation with three child relations: *about*, *identifies* and *classifies*. In CIDOC-CRM, a comparable *refers-to*(x,y) relation has four children: *is-about*, *documents*, *lists* and *represents*. A merged hierarchy of these with source definitions is as follows:

Table 6-3. A merged *references* relation

RELATION	DESCRIPTION
references	"A relation holding between non-physical objects and all entities whatsoever (thus including non-physical objects themselves)" (DLP397);
about	"The relation between information objects and entities they are about." (DLP397)
identifies	"Being about an entity with the main purpose of conventionally naming that entity." (DLP397)
documents	"…intended for cases where a reference is …of a documentary character, in the scholarly or scientific sense." (CIDOC-CRM)
classifies	"…understood as a reification of a 'satisfiability' relation holding between elements of theories and elements of models." (DLP397)
lists	"…documents a source Authority Document (these 'define terminology or conceptual systems') for an instance of a Type." (CIDOC-CRM)
represents	"…establishes the relationship between a Visual Item and an Entity it visually represents." (CIDOC-CRM)

These are partially defined and incorporated in SHO, as follows (dlp1-5; sho1-5):

$$\text{references}(x, y) \rightarrow \text{NPO}(x) \wedge \text{PT}(y) \qquad \text{(dlp1)}$$

$$\text{about}(x, y) \vee \text{classifies}(x, y) \rightarrow \text{references}(x, y) \qquad \text{(dlp2)}$$

$$\text{identifies}(x, y) \rightarrow \text{about}(x, y) \qquad \text{(dlp3)}$$

$$\text{about}(x, y) \rightarrow \text{IOBJ}(x) \wedge \text{PT}(y) \qquad \text{(dlp4)}$$

$$\text{identifies}(x, y) \rightarrow \text{APPL}(x) \wedge \text{PT}(y) \qquad \text{(sho1)}$$

$$\text{classifies}(x, y) \rightarrow \text{CPT}(x) \wedge \text{PT}(y) \qquad \text{(dlp5)}$$

$$\text{documents}(x, y) \vee \text{lists}(x, y) \vee \text{represents}(x, y) \rightarrow \text{references}(x, y) \qquad \text{(sho2)}$$

$$\text{documents}(x, y) \rightarrow \text{DOC}(x) \wedge (\text{ED}(y) \vee \text{PD}(y) \vee \text{Q}(y)) \qquad \text{(sho3)}$$

$$\text{lists}(x, y) \rightarrow \text{ADOC}(x) \wedge \text{TYP}(y) \qquad \text{(sho4)}$$

$$\text{represents}(x, y) \rightarrow \text{VIS}(x) \wedge \text{PT}(y) \qquad \text{(sho5)}$$

In a SHO-compatible system, all tuples (rows)—including those of Parthood, Participation, and Membership—have at least one unique identifier.[57] Each tuple can thus be referenced to one or more source, down to a particular page in a book or

[57] SHO is intended for RDBMS-based applications; the relational model does require uniqueness, which can be implemented by either data-dependent or machine-generated keys.

article. Although DOLCE permits any non-physical object to reference any entity, most reference sources will be instances of information-objects or sub-classes thereof (texts, authority-documents, visual-items, etc.).

In the relational model, tuples (table rows) may be quite complex; in order to afford fine-grained attribution in implementations, efforts should be made to reduce that complexity via typical data modeling normalization methods (*cf.* Simsion, 2007). Systems fitting completely within the Semantic Web/Linked Open Data paradigm proposed by W3C do have as a goal reducing all statements to RDF triples (Berners-Lee, 2006). In that case, attribution could be made a datum-level. The SHO contemplates only row-level attribution at this stage. That said, the physical model developed for this dissertation (Chapter 7) is seen as ultimately compatible with RDF storage methods and inference engines, but implementing those components is outside the present scope.

6.3.1.1 Discussion

To develop more complete definitions for the relations in Table 6-2, we would have to define each class of Information Object formally, then make some formal ontological statement about their being conceptualizations of entities (cf. Guarino 1998)—in other words, tackling the notion of an information object as a complex conceptual object constituted by Concepts. This would ground all DOLCE's Social Objects (*SOB*) as perhaps non-physical discretizations of Mental Objects (*MOB*). This inquiry is beyond the scope of this dissertation.

The authors of DOLCE make clear their goal is not to describe reality, but on one view, Amounts of Matter (*M*) *does* ground the basic category of Physical Endurant (*PED*). In this dissertation I do introduce Activity (*A*) as a 'temporal substance' constituting Events (see §5.3; §6.3.2), but I have left the development of a similar construct for 'conceptual substance' to others. Therefore, I have introduced the relevant DOLCE and CIDOC predicates (i.e. relations/properties) and specify their domain and range, without venturing definitions. Their further meaning is bound in term definitions and scope notes in DOLCE and CIDOC documentation.

6.3.2 Historical-Event objects

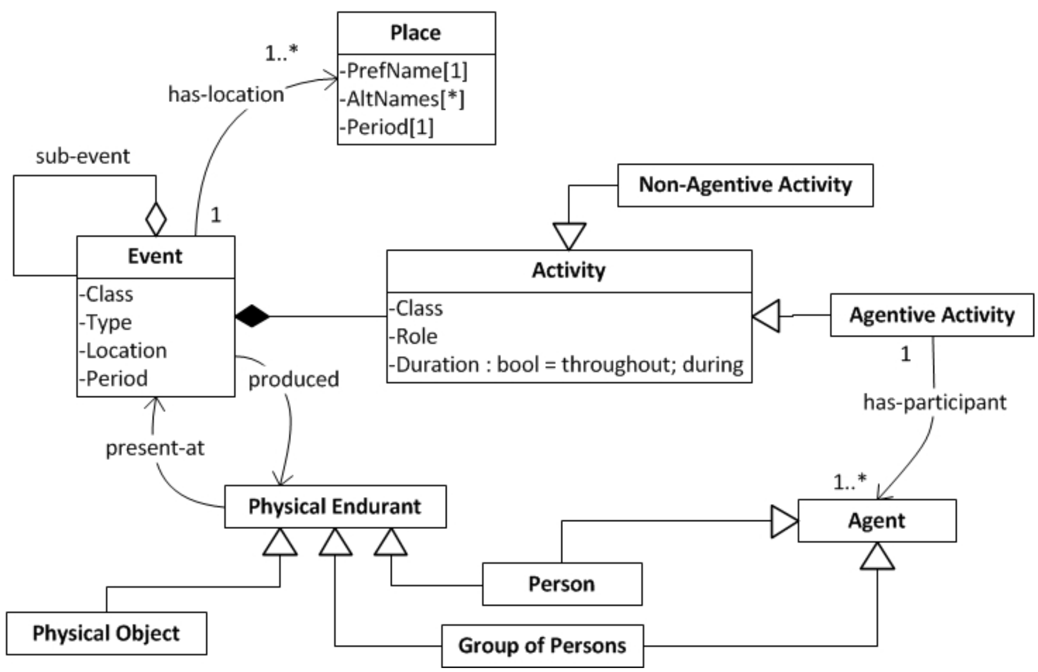


Figure 6-5. Event objects and participation

As discussed in §5.3, it will be advantageous to model events as being composed of the ‘temporal substance’ *activity*, and having sub-event parts—much the way physical things are constituted by material substances (Amount of Matter (*M*) in DOLCE). Simply put, an event is constituted by the activity performed by its participants, either throughout or for some interval within its duration.

An Activity class was introduced in DnS, as a subclass specializing Action (itself a sub-class of Accomplishment, making it a kind of Event), but no definition has been provided. It is the domain of one relation, *result*— i.e. an (some?) Activity can have a Perdurant (PD) result. In my view this concept is misplaced. As it appears in DnS, it’s an agentive event—an Action. I propose to replace it as follows. Activity (*A*) is introduced here as a Properly Subsumed Leaf (PSBL[58]) category of DOLCE’s Perdurant (PD), and itself subsuming Agentive Activity (AGA) and Non-agentive Activity (NAGA).

$$\mathrm{PSBL}(\mathrm{PD}, \mathrm{A}) \qquad \text{(sho6)}$$
$$\mathrm{PSBL}(\mathrm{A}, \mathrm{AGA}) \qquad \text{(sho7)}$$
$$\mathrm{PSBL}(\mathrm{A}, \mathrm{NAGA}) \qquad \text{(sho8)}$$
$$\mathrm{AGA}\ (x) \lor \mathrm{NAGA}\ (x) \rightarrow \mathrm{A}(x) \qquad \text{(sho9)}$$

Activity is defined formally in terms of Events and the relations Constitution (K) and Participation (PC). In DOLCE Events (EV) may be atomic (ACH, Achievement) or have sub-event parts (ACC, Accomplishment). Parthood is fully developed in DOLCE. The most important relevant definitions and axioms are listed here.

[58] cf. Dd9, Dd6, Dd7, and Dd4 in (Masolo, et al., 2003)

Atomicity is defined as the condition of having no parts, and Achievements are atomic:

$$\mathrm{At}(x) \triangleq \neg\exists y\ (\mathrm{PP}(y, x)) \qquad \text{(Dd16)}$$

$$\mathrm{ACH}(x) \rightarrow \mathrm{At}\ (x) \qquad \text{(sho10)}$$

Constitution and its derivatives, Direct Constitution (DK(x, y, t); x constituted by y during t) and Constant Specific Constitution (SK (x, y); x constituted by y for all t) are among the basic "primitive" relations holding between most DOLCE categories (others concern Generic and Spatial Dependence).

According to the characterization of activity as temporal substance presented in §5.3, we therefore say that all Achievements x are either *directly constituted* (DK) by some Activity y for some time t during the interval bounding the Achievement, or *specifically generically constituted* (SK) by some Activity throughout that interval (sho11).

$$\forall x(\mathrm{ACH}\ (x))\ \exists y\ ((\mathrm{DK}(x, y, t) \vee \mathrm{SK}(x, y)) \wedge \mathrm{A}(y)) \qquad \text{(sho11)}$$

In DOLCE Participation (PC) is a "primitive relation" holding between perdurants and endurants, and explicitly "not parthood" (Masolo, et al., 2003:22). Definitions are presented for time-indexed participation (Dd63, Dd64) but there is no formal ground definition; that is, the participation or 'involvement' of endurants in perdurants is a given, and formally refined according to whether it is constant ($\mathrm{PC_C}$) or temporary ($\mathrm{PC_T}$) with respect to "presence" (i.e. existence). Note, $\mathrm{PRE}(x, t)$ = Present (existing) at time-interval t.

$$PC_C(x, y) \triangleq \exists t\ (PRE(y, t)) \land \forall t\ (PRE(y, t) \rightarrow PC(x, y, t) \qquad \text{(Dd63)}$$

$$PC_T(x, y, t) \triangleq PD(y) \land \forall z\ ((P(z, y) \land PRE(z, t)) \rightarrow PC(x, y, t) \qquad \text{(Dd64)}$$

A new "temporary role" Participation relation is introduced to characterize the Activity constituting Events as a function of Participation (PC_{RT}):

$$PC_{RT}(x, y, r, t) \triangleq \forall y\ PD(y)\ \exists w(ACH(w)) \land \exists z\ [1..*](A(z) \land (GK(w, z) \lor SK(w, z)) \land \exists t\ (ql_{T,PD}(t, w) \land \exists t'(P(t', t)) \land \exists x(APO(x) \lor ASO(x)) \land \exists r(FR(r) \land playedBy(r, x) \qquad \text{(sho12)}$$

For clarity, (sho12) is described this way: all perdurants (PD) have at least one achievement (ACH, an atomic event) part, having a temporal quale ($t = ql_{T,PD}$) of some interval, and constituted by at least one Activity (A) instance; *and* some physical (APO) or non-physical (ASO) agent participates in that Activity, in some Role (FR), for some *t'* part of *t*. There are two cases to cover: (i) the activity is asserted as occurring *throughout* an event's duration, and (ii) the activity is considered to have occurred *some-time-during* an event's duration. In this way, the assertion of one or more activity instances as constituting an event is made in with the relation, Participates (PC_{RT}). We say in (sho12) that the time-period of participation is a part (i.e. is equal or a proper part) of the event time-period. That specification will be handled with a Boolean flag in the database.

Furthermore, according to DOLCE's transitivity axiom for Constitution (K), (Ad27), Accomplishments (non-atomic events) are constituted by the Activity of their Achievement parts:

$$K(x, y, t) \leftrightarrow \forall t'\ (P(t', t) \rightarrow K(x, y, t')) \qquad \text{(Ad27)}$$

In this way, events are classifiable by their dynamic structure, as described by the changing types and proportions of their constituent activity, participant types and roles. Roles have been defined in the DnS extension to DOLCE (Gangemi and Mika 2003) and are discussed further in the context of historical-process object, in §6.3.4.

<u>*Event products and results*</u>: In the conceptual model of §5.4, we say that events can produce artifacts and result in other events. Event-event relations are asserted in historical-processes (HPRO; §6.3.4). However, Event products are a function of participation: Material Artifacts (MART) participate as products (prod) or witnesses (witn; simply present):

$$\mathrm{PC}(x, y) \rightarrow \mathrm{ED}(x) \wedge \mathrm{PD}(y) \qquad \text{(dlp6)}$$

$$\mathrm{perf}(x, y) \vee \mathrm{prod}(x, y)\ \mathrm{witn}(x, y) \rightarrow \mathrm{PC}(x, y) \qquad \text{(dlp7)}$$

$$\mathrm{MART}(x) \wedge \mathrm{PC}_{\mathrm{RT}}(x, y, r, t) \rightarrow \mathrm{prod}(r) \vee \mathrm{witn}(r) \qquad \text{(sho13)}$$

If x is a material artifact and participates in some event, then its role is either product or witness (simply present).

<u>Spatial and temporal locations of Events</u>: In DOLCE, the location of an event is said to be a function of its participants' locations during its occurrence: "their (perdurants) spatial location *seems to* come from the spatial location of their participants" (Masolo et al., 2003:18; italics added); also definitions Dd85 and Dd86, p. 31). In practice, participant locations are more often a function of event locations.

We want to infer someone's location at time *t* from the location of events having intervals containing *t* in which they participated throughout. First, we say all events have a measured location, (partially) defining a Physical Place (PHPL):

$$\text{loc-sp}(x, y) \rightarrow PD(x) \land S(y) \quad \text{(dlp8)}$$

$$EV(x) \rightarrow \exists z(PHPL(z) \land (\exists s(S(s) \land ql_{S,\,PED}(s, z, t)) \land (\text{loc-sp}(x, s)) \quad \text{(sho14)}$$

Then, that if something is a constant participant in an event, its physical location at T(t) is that of the event:

$$\forall x(PC_C(x, y)) \rightarrow PD(y) \land \exists s(ql_{S,\,PD}(s, y, t) \land \text{loc-p}(x, s)) \quad \text{(sho15)}$$

There are spatial and temporal operations on events we will do in the RDBMS, where it is simpler: (i) we will compute the location of an Accomplishment (ACC) as the mereological sum of its sub-event locations; (ii) we will compute the (possibly punctuated) duration of an Accomplishment as the union of its sub-event time-periods, and the *range* of an Accomplishment as the period *spanning* all sub-event periods.

6.3.2.1 *Discussion: Activity vs. Process*

The sense of activity presented in this research seems at first glance to correspond to DOLCE's Process (PRO). Processes, as stative occurrences in DOLCE, are distinguished from events as being cumulative and non-homeomeric, i.e. a process "holds of the mereological sum of two of its instances," and has temporal parts that are not themselves the process. Running is given as an example of a process because *"there are (very short) temporal parts of a running that are not themselves running."* There is in my view unresolvable ambiguity here. If we call running a process, it is because it is being *modeled as* homeomeric. As the DOLCE authors and others have noted (see also §5.3), whether an occurrence is considered a homogenous process, an event or a state is a function of the granularity or scale of the data or desired type of analysis.

This intuition corresponds to the following statement in DOLCE documentation: "Further developments: being 'achievement', 'accomplishment', 'state', 'event', etc. can also be considered 'aspects' of processes, or of parts of them. […] the same process […] can be seen as an accomplishment (what has brought the current state that e.g. we are trying to explain), as an achievement…, as a state. . ., as an event" (DLP397, 2006). In the SHO, the notion being captured is that events are not so much 'aspects' of processes (which I call activity), as particular *discretizations* of them. This common-sense formulation is readily captured by describing the participation of endurants in events, as I have done. As argued earlier, events are human constructions, and in a SHO-supported system, all assertions of participation can be attributed to sources. Historical processes are human constructions as well, but far more complex, and the DLP extension for Descriptions and Situations is required to model them (§6.3.4).

6.3.3 States

State is minimally defined in core DOLCE (axioms Ad81, Ad77) distinguished from its fellow 'stative perdurant' Process as being homeomeric. That is, "all its temporal parts are described *by the very expression* used for the whole occurrence" (Masolo et al., 2003:24). All property characteristics of a state hold for all of its temporal parts. States get fuller expression in the *Descriptions and Situations* DOLCE extension (DnS), as elements of States of Affairs (Gangemi & Mika, 2003). Their role as elements of historical process objects in SHO is discussed in the following section.

6.3.4 Historical-Process objects

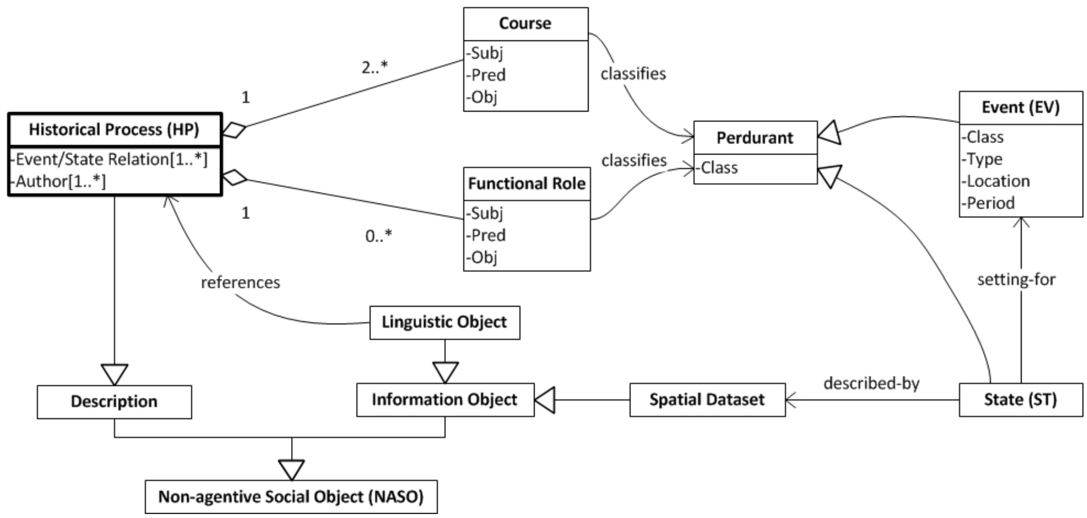


Figure 6-6. Historical Process (HPRO) in SHO

In §5.5.4, a historical process is conceptualized as a theory of event relations that can include assertions of relevant conditions as States (ST). They may be somewhat procedure-like (US Presidential Election cycle of 2008) or a sequence of contingent events (the democratization of 18th-19th century Britain), to cite examples from Chapter 7. The principal logical requirements are illustrated in Figure 6-6. States—most frequently base maps—can be any relevant set of geo-referenced attributes (in DOLCE terms, sets of valued Qualities) for the locales of a specified collection of events. In analytical applications, such attribute values might readily become terms in a spatial statistical model. This essential construct differs fundamentally from the existing Process category in DOLCE, which is being set aside for now. A historical-process can be modeled using elements of the DOLCE extension *Descriptions and*

Situations (DnS) (Gangemi and Mika 2003), with some adjustments and additions introduced here.

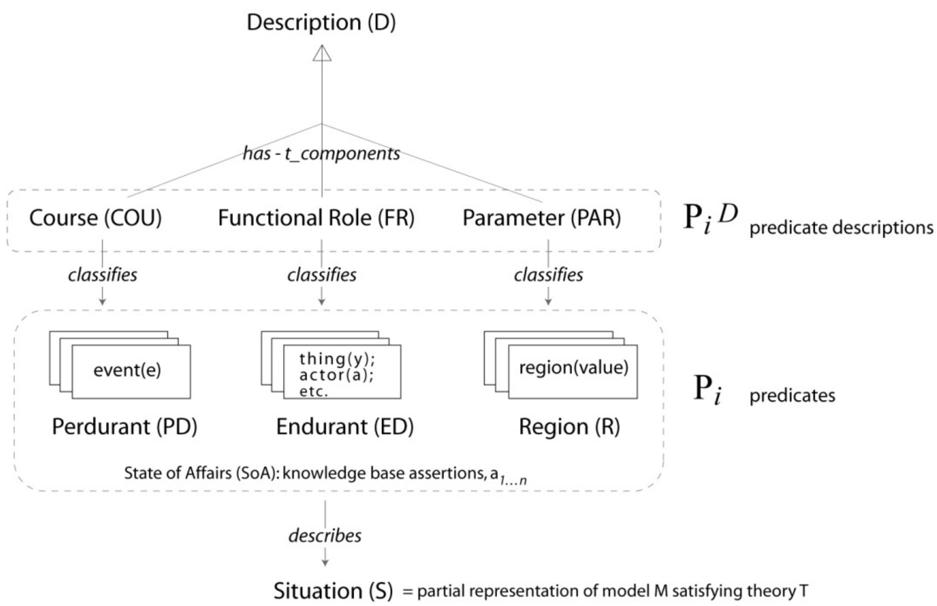


Figure 6-7. 'Descriptions and Situations' (DnS) as expressed in (DLP397)

In DnS a Description is "a social object which represents a conceptualization […] generically dependent on some agent and communicable. Descriptions define or use concepts or figures, are expressed by an information object and can be satisfied by situations" (DLP397, 2006). This achieves what Gangemi and Mika term "epistemological layering" (2003:694) that is neutral with respect to reality. In this way the underlying premise of the DnS framework supports the requirements of SHO. DnS includes a 'Theory' leaf category of Description, defined circularly as: "in a wide cultural sense: a theory about something, expressed in a rather systematic way…communicable in principle" (DLP397). On its face this seems an appropriate parent category for historical-process-as-theory, but since no formal expression for how Theory specializes Description is offered, it is set aside for now.

Therefore, the class Historical Process (HPRO; also Hist-Process) is introduced as a Properly Subsumed Leaf (PSBL) of Description (D).

$$\mathrm{PSBL}(\mathrm{HPRO}, \mathrm{D}) \tag{sho16}$$

A Hist-Process is an authored collection of knowledge-base statements concerning (i) relations between events, and (ii) state descriptions (normally, spatial data sets) that are the setting for, or somehow instrumental in, the process in question. The Hist-Process in SHO uses other aspects of DnS Descriptions in the following manner.

In DnS, a Course (COU) is a t_component element of a Situation Description (SD) constituted by a collection of selected Perdurant (PD) assertions (Figure 6-6). A Course *sequences* (i.e. classifies) perdurants. Assertions concerning Endurant roles and Parameter values are selected in similar fashion by Functional Roles (FR) and Parameters (PAR) respectively.

In DLP397, the relevant relations are children of *classifies*:

$$\text{sequences}\ (x, y) \rightarrow \text{classifies}\ (x, y) \wedge \mathrm{COU}(x) \wedge \mathrm{PD}(y) \tag{dns18}$$
$$\text{played-by}\ (x, y) \rightarrow \text{classifies}\ (x, y) \wedge \mathrm{FR}(x) \wedge \mathrm{ED}(y) \tag{dns19}$$
$$\text{valued-by}\ (x, y) \rightarrow \text{classifies}\ (x, y) \wedge \mathrm{PAR}(x) \wedge \mathrm{R}(y) \tag{dns20}$$

Concerning Courses, the authors' intent appears limited to enabling a temporally ordered listing of particular events without the means for explicitly describing the reason for their selection. This is inadequate for our purposes, so the notion of simply sequencing perdurants is expanded: a Course is a set of asserted relations between events; a Functional Role selects states (spatial datasets or individual assertions) as having a *setting-for* relation with a Hist-Process. Individual Parameters and valued Regions could conceivably be used in a Hist-Process description but there are no obvious cases in the exemplar datasets used, so they are not addressed yet.

The manner in which requirements for representing historical processes (§5.3.5) have been met in the SHO is described formally here, followed by an extended discussion of modeling considerations:

A Course (COU), as a component of a Hist-Process (HP), formally describes meaningful sequences of two or more events (EV) in one of several possible relations, and a Functional Role (FR) selects zero or more States (ST) in a *setting-for* role.

$$\mathrm{COU}(x) \rightarrow \exists y\,(D(y) \land \mathrm{t\_component}(y, x) \qquad \text{(dns15)}$$

$$\mathrm{FR}(x) \rightarrow \exists y\,(D(y) \land \mathrm{t\_component}(y, x)) \qquad \text{(dns16)}$$

$$\mathrm{HPRO}(x) \rightarrow \exists y(\mathrm{COU}\,(y) \land \mathrm{t\_component}(x, y)) \land \exists z[0,,*](\mathrm{FR}(z) \land \mathrm{t\_component}(z, y) \qquad \text{(sho17)}$$

We must account for three categories of event relations: (i) simple ‘temporal topology’ corresponding to Allen’s intervals (1983), (ii) parthood for complex events having ‘sub-events’ and (iii) influence.

Allen’s 13 interval relations are found in CIDOC-CRM and included in the SHO: starts (started by); finishes (finished by); equal in time to; occurs before (occurs after); occurs during; overlaps (is overlapped) in time with (by); meets (is met) in time with. Because all events have a time-period attribute, these relations can be calculated or discovered in temporal queries in the database, but there’s no harm in permitting their explicit assertion.

DOLCE models parthood well, following general extensional mereology. Non-atomic events are called Accomplishments (ACC), which may have Perdurants (PD) as parts (P) or as Constituents (this is the case for the new Activity (A) class (§6.3.2). Some relevant ground definitions and axioms include:

$\forall x\, P(x, x)$
Everything is a part of itself

$P(x, y) \rightarrow (PD(x) \leftrightarrow PD(y))$ (Ad2)
Perdurants have perdurant parts

$PP(x, y) \triangleq P(x, y) \wedge \neg P(y, x)$ (Dd14)
An entity is not a Proper Part of itself

$(P(x, y) \wedge P(y, x)) \rightarrow x = y$ (Ad6)
Entities are equal if they are parts of each other

$(P(x, y) \wedge P(y, z)) \rightarrow P(x, z)$ (Ad7)
Parthood is transitive

A complex event is an Accomplishment in DOLCE. We say that all Accomplishments have two or more Event proper parts (PP):

$$ACC(x) \rightarrow \exists yz(EV(y) \wedge EV(z)) \wedge ((PP(x, y) \wedge PP(x, z)) \quad \text{(sho18)}$$

Transitivity in generic parthood can raise difficult issues, i.e. in cases, transitivity should stop. A usual example given is that while my hand is part of me and I am part of UCSB Geography, it makes no sense to say my hand is part of UCSB Geography. These issues are avoided in DOLCE by means of Collections having Members (§6.3.5). In the case of events, there is no philosophical problem stemming from complete transitivity: each single Shooting sub-event in World War II is indeed a part of the macro-event. That said, implementation issues will obviously arise when the query is made, 'what events comprised WW II?'

Finally, we need to permit assertions of a variety of relations that may obtain between events or between events and various endurants. The SHO merges two sets of event relations, from CIDOC-CRM and BFO, under DOLCE's mediated-relation(PD, PD).

Table 6-4. Event-event; event-endurant relations

RELATION	DESCRIPTION
influenced	"…captures the relationship between an Event[59] and anything that may have had some bearing on it" (CIDOC-CRM); "a substance has an effect on a process" (BFO)
motivated	"…describes an item or items that are regarded as the reason for carrying out the Event" (CIDOC-CRM)
specific-purpose-of	"…relationship between a preparatory activity(/event) and the event it is intended to be in preparation for" (CIDOC-CRM)
general-purpose-of	"…an intentional relationship between an Event and some general goal or purpose" (CIDOC-CRM)
caused	an asserted causal relation between two Events or between an agent and an Event (SHO)
facilitated	"A substance plays a secondary role in a process." (BFO); similarly, an Event may facilitate the occurrence of another Event in a hist-process
initiated	"…a substance (e.g. agent) initiates a process…" (BFO); intuitively, an Event can initiate a macro-event or hist-process (SHO), e.g. a blockade initiates a war
perpetuated	"…a substance sustains a process" (DLP397); intuitively, an Event can sustain an ongoing macro-event or hist-process (SHO)
hindered	"…a substance has a negative effect on the unfolding of a process…" (BFO); intuitively, an Event can likewise hinder the unfolding of a macro-event or hist-process (SHO)
terminated	"…a substance (e.g. actor) terminated a process." (BFO); intuitively, an Event can terminate an ongoing macro-event or hist-process (SHO); e.g. an armistice terminates a war
setting-for	A relation of asserted relevance between a defined state such as a spatial table of attributes for the locale and a historical-process
result	In DLP397, the *result* relation holds between an action and any perdurant, specializing the *precedes* relation. This is very similar to *caused* above, perhaps redundant.

[59] The CIDOC-CRM term 'Activity' has been replaced throughout by the corresponding SHO category, 'Event.'

All have inverse properties (not spelled out here); domain and range axioms as follows:

$$\text{influenced } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(sho19)}$$

$$\text{motivated } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(sho20)}$$

$$\text{specific-purpose-of } (x, y) \rightarrow EV(y) \wedge (PD\ (x) \vee HPRO(x)) \quad \text{(sho21)}$$

$$\text{general-purpose-of } (x, y) \rightarrow PD\ (x) \vee HPRO(x)) \wedge TYP(y) \quad \text{(sho22)}$$

$$\text{caused } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(sho23)}$$

$$\text{facilitated } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(bfo1)}$$

$$\text{initiated } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(bfo2)}$$

$$\text{perpetuated } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(bfo3)}$$

$$\text{hindered } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(bfo4)}$$

$$\text{terminated } (x, y) \rightarrow (PD(x) \vee ED(x)) \wedge (PD\ (y) \vee HPRO(y)) \quad \text{(bfo5)}$$

$$\text{setting-for } (x, y) \rightarrow ST(x) \wedge HPRO(y) \quad \text{(sho24)}$$

$$\text{result } (x, y) \rightarrow EV(x) \wedge PD(y) \quad \text{(sho25)}$$

6.3.4.1 *Discussion*

In BFO, the only relations discussed having <SPAN-SPAN> signatures relate to causality and parthood (Smith & Grenon, 2004:288-289). Many of the relations designated as having <*SNAP Independent, SPAN*> signatures, for example regarding participation, seem intuitively to be assertions one might make in a historical context as also holding between events, or between events and H-Processes as conceived here (in BFO it is participation of substances in processes). In point of fact, Worboys (2005) has identified several participation roles appearing in SNAP-SPAN literature as potentially useful 'event-event' relations. They include *Initiation*, *Perpetuation/facilitation*, *Hindrance/blocking* and *Termination*. As both DOLCE and the SHO seek rudimentary formal representations of commonly used information constructs, we can skirt some ontological issues. For example, the *Initiation* of EventB by EventA may not be strictly possible. One could argue (as BFO does) it is a participant in EventA in some agentive role that does the initiating. For the purposes of SHO this is beside the point; in my view the level of explicitness the BFO approach implies is not a realistic possibility in this domain. Data is too sparse and budgets for historical projects too small.

All of that said, several relations listed by Smith & Grenon (2004) seem entirely appropriate additions to SHO, and they appear within the SHO as 'bfo1-bfo5' above. While SHO aims to be logically valid throughout and consistent with selected portions of DLP397, it must be stressed that by including a class or relation *having the same term label* used by some other ontology makes no claims about the SHO's logical consistency *with that ontology*. Rather, the use of like terms in many cases only points the way to possible future alignment.

There is a question of whether to permit event-event relations to be asserted (with source attribution) outside of an H-Process description, and then simply enumerated by the H-Process. I see no definitive answer for this. Either:

a) Assertions about event relations can be made in the context of an H-Process *or* independently of such a 'description of a model satisfying a theory,' or...
b) Assertions about event relations are *necessarily* theoretical and must be constituents of an H-Process.

As mentioned earlier, DOLCE is a moving target. In a DOLCE 'Ultra-Lite' version, Course was replaced by 'EventType' and Scherp, et al. (2009), have introduced EventType sub-classes, 'Cause' and 'Effect' to allow some expression of event relations. That approach is not compatible with the notion of a historical process as conceived here. Events are not themselves 'CauseEvents' or 'EffectEvents,' rather those are two of a large number of possible event relations that may be asserted.

6.3.5 Groups and membership: dynamic collectives

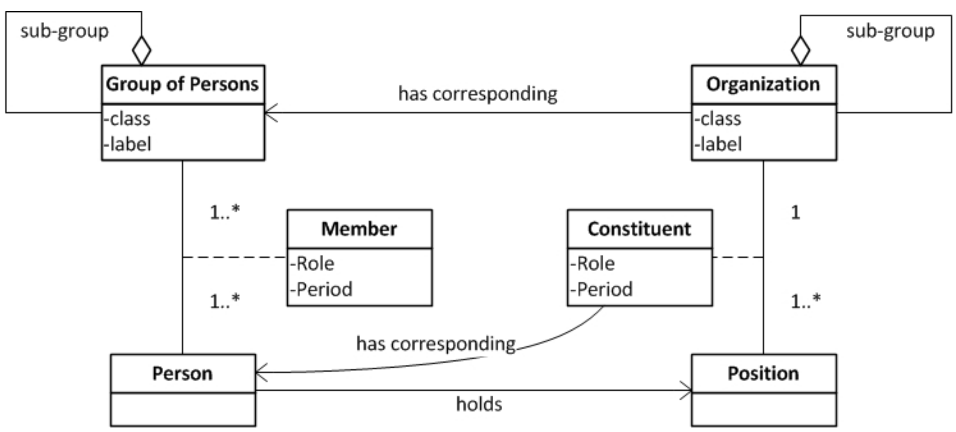


Figure 6-8. Groups and membership

The kinds of groups we are concerned with (from §5.5.1) include:

a) [PGRP]: a collection of event participants—individual persons and groups—who may or may not be individually identified, e.g. a mob, a ship's passengers, or an expedition party
b) [FGRP]: a functional class, e.g. 'central European farmers circa 1200 CE,' Bauhaus artists, U.S. Presidents
c) [ORG]: a legal or otherwise formal organization, e.g. the U.S. Senate, The Beatles, or BP Corporation

Category (c) is readily accounted for in DOLCE DnS, with Organization (ORG), a subclass of Agentive Social Object (ASO). Another DOLCE extension included in DLP397, 'SocialUnits,' proposes an elaborate taxonomy judged too complex for the SHO at this stage[60]. In any case, categories (a) and (b) are not addressed well in either DnS or the SocialUnits extension. The groups in (a) and (b) are collections of actual, certainly physical, humans. Core DOLCE and the DnS extension have neither physical persons nor groups of persons, committing only to rational-physical-objects (RPO), with rationality defined as "the ability to internally represent meta-descriptions (descriptions that have other descriptions playing roles used by them)." This could include programmed robots in some interpretations. DLP397 states explicitly, "a person in general is not constructed by this ontology" (2006). According to the requirements developed in Chapter 5, we must account for a class of Group that is collections of humans. Like the non-physical Organization, it would have time-dependent composition at time *t* and the spatial extension of its members at time *t*. Because the notions of a physical human being—and groups of humans—are so commonplace, they belong in the SHO.

The new classes Person (PER) and Group of Persons (PGRP) are introduced as Disjoint (DJ) Properly Subsumed Leaves (PSBL) of Rational Physical Object (RPO):

$$\mathrm{PSBL}(\mathrm{RPO}, \mathrm{PER}) \land \mathrm{PSBL}(\mathrm{RPO}, \mathrm{PGRP}) \land \mathrm{DJ}(\mathrm{PER}, \mathrm{PGRP}) \qquad \text{(sho26)}$$

In the SHO, Agentive Social Objects (ASO) include Organizations (ORG) and

[60] A rationale for that judgment appears in the Discussion section below

Socially Constructed Persons (SPER; Social Persons), which generically constitute (GK) them, and are in turn One-sided Constant Generic Dependent (OGD) on Agentive Physical Objects (APO).

The formal dependence of the social (non-physical) construct SPER on the physical endurant, PER (and indirectly by constitution, ORG on PGRP) is maintained:

$$\forall x(\mathrm{SPER}(x))\ \exists y(\mathrm{PER}(y) \land \mathrm{OGD}(x, y)) \quad \text{(sho27)}$$

For all social persons there is a corresponding physical person.

An Organization's (ORG) constitution by Social Persons (SPER) is made time-dependent, allowing for their dynamic structure:

$$\mathrm{ORG}\ (x) \land \mathrm{SPER}\ (y) \land \mathrm{K}(x, y) \rightarrow \exists t(\mathrm{T}(t)) \land \mathrm{DK}(x, y, t) \quad \text{(sho28)}$$

If an organization is constituted by a social agent, it is for some time-interval *t*.

$$\forall x(\mathrm{ORG}(x))\ \exists y(\mathrm{PGRP}(y) \land \mathrm{OGD}(x, y)) \quad \text{(sho29)}$$

For all organizations there is a corresponding group of persons.

The membership relations to be modeled include (i) a Person (PER) in a Group of Persons (PGRP) in some Role (FR), and (ii) a Social Person (SPER) in an Organization (ORG). The primitive relation appearing in DnS, member(x, y, t), covers the non-physical cases. Membership is defined as "being a (generic, temporary) constituent in a countable collection, for example: member of a society, bacterium in a colony, etc."

$$\mathrm{member}(x, y, t) \rightarrow \mathrm{ED}(x) \land \mathrm{COLL}\ (y) \land \mathrm{T}(t) \quad \text{(dlp8)}$$

To accommodate roles in membership, the relation r-member is introduced:

$$\text{r-member}(x, y, r, t) \rightarrow \exists xyrt(\mathrm{PER}(x) \land ((\mathrm{ORG}(y) \lor \mathrm{PGRP}(y)) \land \mathrm{FR}(r) \land \mathrm{T}(t)) \land$$

DK(x, y, t) and playedBy (r, x, t) (sho30)

This reads, if x is an r-member of y, in role r, at time t, then there is some person x, physical or social group y, role r and time-interval t such that y is constituted by x at time t, in role r

The dynamic constitution of ORGs and PGRPs (our *dynamic collectives*) is represented in time-indexed constitution and membership-in-role relations.

ORG(x) ∧ SPER(y) ∧ T(t) ∧ DK(x, y, t) → ∃z((PER(z) ∧ OGD(y, z)) ∧ ∃w((PGRP(w) ∧ OGD(x, w)) ∧ rMember(w, x, r, t) ∧ (∃r(FR(r) ∧ playedBy (r, w, t)) (sho31)

Functional groups will be discovered or inferred with functional queries leveraging the above formal relations, and reified or not, as an application-specific matter. For example, since for every government *g*, there is a corresponding PGRP whose membership changes over time, its dynamics could be investigated with a function of the form, *memberRoster* (*g*, *t*).

6.3.5.1 Discussion

The following scenario exemplifies a problem with the preceding strategy: although the London Symphony always has a conductor, first violin, etc., the membership of the Symphony at any given time—which individuals hold those positions—is in constant flux. The logical structure works only so long as those positions are always filled. That is, if the class Social Person is considered as a position in an Organization, which seems to intuitively fit, then when a position is unfilled (say for *t'* immediately after a death) then there is no physical person (PER) corresponding to the Social Agent 'London Symphony Conductor' for some time-interval, *t'*. The approach taken can be rationalized by saying that the PER corresponding to London

Symphony Conductor is an empty set (∅) at time *t'*. Whether this is adequate remains an open question.

Alternative approaches to defining groups have been considered. For example, in the OntoClean exemplar (Guarino & Welty, 2002), a Group is "an unstructured finite collection of wholes," whose instances are "mereologically extensional as they are defined by their members" (p. 209). These might be any sort of wholes, so a Group of Persons (PGRP) could be a sub-class of an abstract Group, as suggested in OntoClean, and subsumed by Abstract (AB) in DOLCE, alongside Set and Fact, which are undeveloped in that formalism to date.

Another possibility, which appears in SHO but will remain undocumented here, is to make Group of Persons (PGRP) a leaf category of Physical Plurality (Table 9-1) as well as of Rational Physical Object (RPO). That is, a Group of Persons (PGRP) *is-a* (is subsumed by) Physical Plurality (PPLU). Other instances of PPLU that come to mind are building ensembles, bookshelves, the trees on my property, wolf packs, a museum collection, or the artifacts discovered so far at an archaeological site.

The PGRP developed above corresponds to the Text Encoding Initiative (TEI)[61] entity *personGrp*, mentioned in §5.5.1 and defined as "(multiple) individuals treated as a single person for analytical purposes," is an important construct missing in DOLCE. In the SHO, we wish to be able to say a PGRP participated in an event. If

[61] http://www.tei-c.org/Guidelines/P5/index.xml

the particular membership is known for a PGRP, we can infer individuals' participation and analyses can extend to their attributes and other details of their life-paths (other event participation), etc. If individual members are not specified, analysis is still possible, given taxonomies of PGRP types, or at least along dimensions of *role* and *activity* attributes of a group's participation.

Like all endurants, a PGRP can have temporally indexed constituents. In fact it *must*. For that reason, it will be helpful to develop in the future a new Membership relation that narrows the Constitution (K) of DOLCE, and to test its efficacy in the context of the exemplar data used in this dissertation.

Whether groups (and organizations particularly) have locations is a modeling problem that has not been addressed in this work. For example, what is the location of the United Nations? Something simpler than ownership or occupation relations with buildings is desirable.

6.3.6 Historic periods

An H-Period (HPRD) is introduced as a Constitutive Description (DCON) that references a Temporal Region (TR) and a possibly a Place (PL).

$$\mathrm{HPRD}(x) \rightarrow \exists y(\mathrm{TR}(y) \wedge \mathrm{ref}(x, y) \wedge \diamond(\exists z(\mathrm{PL}(x) \wedge \mathrm{ref}(x, z)) \qquad \text{(sho32)}$$

6.3.7 Place: location, paths, flows and regions

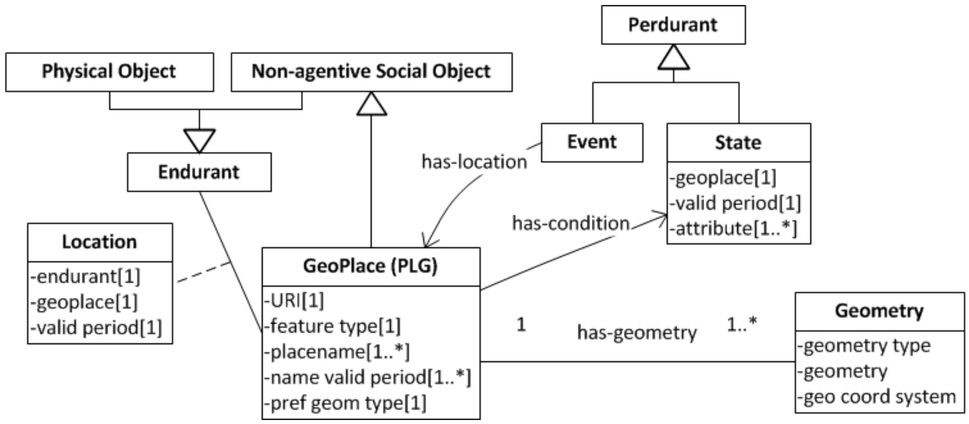


Figure 6-9. Place in the SHO

In 'core' DOLCE terms, Spatial Location (SL) is a Physical Quality (PQ) inhering in all Physical Endurants (PED). Its value ('quale') for any particular endurant is given as a Space Region (S). Space Regions may be described by geographic coordinates or geometric figures (either 2- or 3-D). Other kinds of regions include temporal (TR; §6.3.6) and abstract (AR).

The requirements for representing places discussed in §5.5.3 are considerably more involved, in that spatial location attributes are possible for social objects and events as well as physical things. The most recent DOLCE release (DLP397) includes two extensions, DOLCE-Lite and SpatialRelations, that elaborate place as Constitutive Descriptions of locations. Their use to account for SHO requirements is judged (at this time) as being far too complex. As noted in §5.5.3, this dissertation does not undertake a complete historical gazetteer, so an interim, simplified approach is presented here.

To begin with, we say a place is a *descriptive* construct "used to structure knowledge and ease communication" (Janowicz, 2009). Place (PL) is declared a Properly Subsumed Leaf (PSBL) of Constitutive Description (DCON), itself a Non-agentive Social Object (NASO).

PSBL (D, DCON) (dlp9)

PSBL (DCON, PL) (sho33)

Places define and describe both non-physical Geographical Places (PLG) and physical Geographical Objects (GOBJ), child categories for which include Earth Features (EFEA) and Material Artifacts (MART). Place descriptions include time-indexed locations and names, if they exist:

$PL(x) \rightarrow \exists y(PLG(y) \lor GOBJ(y) \land describes(x, y)$ (sho34)
Places describe either PLGs or GOBJs

$\forall x(PLG(x))\ \exists y(PHPL(y) \land DGDt_S(x, y, t)$ (sho35)
All PLGs are 'Temporary Direct Spatial Dependent' on a physical place at some time

$DGDt_S(x, y, t) \triangleq \exists \Phi, \psi(\Phi(x) \land \psi(y) \land DGDS(\Phi, \psi) \land x \approx_S <y, t>$ (Dd90)

$PHPL(x) \rightarrow \exists s(S(s) \land ql_{S, PED}(s, x, t)$ (sho36)
Physical places have time-indexed spatial location quales (river changes course)

$\forall x(PLG(x) \lor EFEA(x) \lor MART(x))\ \exists y(TYP(y) \land SB(x, y))$ (sho37)
All Place referents of a category are further Typed

$(PLG(x) \lor GOBJ(x)) \land \exists y(APPL(y) \land identifies(y, x) \rightarrow \exists t(T(t) \land identifies(y, x, t)$ (sho38)
If a Place referent has a name, that name is valid at some time

Places geometries are an application-dependent choice—one or more of the standard

OGC[62] types (POINT, MULTIPOINT, POLYGON, MULTIPOLYGON, LINESTRING, MULTILINESTRING) depending upon mapping requirements. A city is represented by a point in some cases, a polygon in others.

GIS geometries are Spatial Regions (SR) in the SHO. This can be handled straightforwardly in the RDBMS, as can multiple names for a single Place; tentative formal expressions for these follow:

$$\forall xyz(\mathrm{PL}(x) \land (\mathrm{ED}(y) \lor \mathrm{PD}(z))\ \exists s[1..*](\mathrm{S}(s) \land (\text{loc-sd}(y, s) \lor \text{loc-sp}(z, s) \quad \text{(sho39)}$$

$$\forall x(\mathrm{PL}(x)) \rightarrow \exists y[0..*](\mathrm{APPL}(y)) \land \exists t(\mathrm{HPRD}(t) \land \text{identifies}(y, x, t) \quad \text{(sho40)}$$

6.3.7.1 Paths and regions

The following constructs for Paths and Regions have been developed, however their formal definition for the SHO is incomplete.

Paths are spatial-temporal locales/settings for one or more Event. At this time, Paths are temporal geometry constructs, implemented in the exemplar database (Chapter 7) according to the descriptions below and in §7.1.2.5. Path is further sub-classed with Flow-Path and Trajectory-Path. An event location may be Place, a Trajectory-Path, a Flow-Path, or a Region.

Flow-Path: a Path described by an *ordered* pair of Places (from, to) and having its own temporal location (Q.TL). A Flow-Path spatio-temporally locates certain kinds of aggregated events. In the exemplar database, statistics regarding slave trade

[62] Open Geospatial Consortium

voyages for flag/route/period are aggregated into an event of type Commercial Exchange. Their “temporal geometry” in the database, necessary for generating maps or diagrams, are Flow-Paths, i.e. spatial point pairs with temporal attributes and magnitude.

Trajectory-Path: a calculated *sequenced bag* of Places associated with a complex Event, such as a journey (expedition, pilgrimage, etc.). A trajectory can be derived by selecting for the geometries and temporal attributes of all sub-event parts of a journey. In the case where the geometry of routes between ‘stops’ on a journey are not specified—either unknown or unimportant—a trajectory(e_1) operation can generate LINESTRING geometries for cartographic purposes. A Trajectory-Path may model only the from-to locations of a journey, or trace the specific route taken, if those routes constitute Events with associated Place locations. In the exemplar KB, the trajectory of Napoleon’s 7th Corps is represented by five ‘march’ events and two ‘encampment’ events—incompletely, as is common for historical events.

Regions: As conceived so far, regions are areas having either an asserted boundary or that of a collection of places. In the first case, a region would be a Type of Place, mapped to one or more geometries and one or more place names: a POLYGON or MULTIPOLYGON geometry bounding an area on the Earth surface having one or more place name appellations, valid for a period or period array. One or more of the place names may be the asserted locale for some simple or complex perdurant associated with any number of other attributes

In the second case, a region is a collection of places. For example, Appalachia is composed of 436 counties across 13 American states, according to the Appalachian Regional Commission. The composition might be achieved in the system variously, e.g. as a simple union of properties, including geometry, or as a hull or other spatial footprint.

6.3.8 Extensible classification

There are three ways the SHO can be extended—by developing new sub-classes in the sho: namespace or in a new, project-designated namespace, or by adding semi-formal project- or domain-specific concept taxonomies under the Type hierarchy. The latter approach follows a CIDOC-CRM convention. It is hoped that expert or community-derived controlled vocabularies will be developed to extend the Type category significantly.

Taxonomies of perdurant, endurant and relation types can be added as subclasses of Type (TYP), as semi-formal controlled vocabularies for representing application-specific detail. Type is itself a new sub-class of Concept (CPT), which is defined in DOLCE as "a non-physical object […] whose function is classifying entities…" (DLP397). Three categories of Concept are defined in DOLCE, Course (COU), Role (FR) and Parameter (PAR), and they are used in Descriptions (D) such as a Historic Process (HPRO) (*cf.* §6.3.4). Type is introduced in SHO here as a Concept (CPT), simply:

$$\mathrm{TYP}(x) \rightarrow \mathrm{CPT}(x) \qquad \text{(sho41)}$$

For the DRUMLIN database, I have developed a set of concept taxonomies that support the five exemplar datasets by expanding the categories, Event (EV), Organization (ORG), Role (FR), Agentive Activity (AGA) and Information Object (IOBJ). A listing appears in Appendix §9.2.

6.3.8.1 <u>Discussion: Possible event classes in SHO</u>

The Type taxonomies referred to above are an interim measure. It should be possible to extend the SHO classes Event (EV) and Activity (A) more formally, and I have begun some work along those lines.

GIScience researchers investigating 'change events' have identified the most elemental kinds of spatial change and developed formalisms to analyze and otherwise reason about them computationally: change of position (motion), change of identity (coming in and out of existence), change of shape (including growth and diminishment) and change of attributes (§2.5.1) .

We can ask whether activity (and events by extension) can be typed in those terms. The development of the SHO event taxonomy suggested two possible ways to approach event classification: by activity type and by "sphere of purpose." Many activity types correspond to two kinds of elemental change referred to earlier: identity and motion. A third, less elemental class emerges from the listings in Chapter 4 that we could call *exchange*, or perhaps *interaction* (Table 6-5). "Spheres of purpose," as indicated in Table 6-6, include Political, Cultural, Military, etc. and are a familiar way of organizing event in timelines, for example. It may be that a joining of activity type and purpose sphere will be most useful. A first take at that appears in Table 6-7. Gatherings, for example can occur in the context of any sphere of purpose.

This work is incomplete but seen as an important next phase of this research, as discussed in §8.31.

Table 6-5. Core activity types

creative	**destructive**	**motion**	**interaction**	**transformation**
create	destroy	journey	exchange	grow
form	dissolve	relocation	communication	merge
build	damage	dispersion	influence	add part
invent		diffusion		diminish
				remove part
				split

Table 6-6. Spheres of purpose

POLITICAL

- **electoral**
  - campaigning
  - voting
- **governing**
  - legislation
    - commerce (sic)
  - judicial
    - law enforcement
  - administrative
    - commerce
    - infrastructural
  - executive
- **contention**
- **ceremonial**
- **international relations**

MILITARY

- **war-making**
- **administrative**
  - infrastructural
  - commerce

ECONOMIC

- **commerce**
- **manufacturing**
- **extraction**
  - oil extraction
  - gas extraction
  - mining
- **agricultural**
- **economic crime**
- **infrastructural**

CULTURAL

- **artistic**
  - commerce
- **intellectual**
  - scientific
    - commerce
  - educational
  - contention
- **religious**
  - ceremony
  - administrative

BIOLOGICAL

- **fauna**
- **flora**

GEOPHYSICAL

- **geological**
- **hydrological**
- **atmospheric**
- **astronomical**

ceremonial	**sport**
criminal	commerce

Table 6-7. Purpose-independent event types

event types cross-cutting all spheres	
gathering	meeting, performance, convention, marshaling
creation	birth, construct, rebuild, manufacture, produce, split, merge
destruction	kill, destroy, transform, merge, split
motion	move, traverse, flow, disperse
interaction	exchange, communicate, play, battle, sex, allegiance
plan	invent, design
be somewhere	reside, stay, visit
growth	diffusion (spread), enlarge
diminishment	shrink, reduction

7 Prototype Database and Applications

A functioning prototype database (called DRUMLIN hereafter) was built using the open-source PostgreSQL/PostGIS[63] software, in order to investigate the extent to which the spatial history ontology (SHO) can be effectively implemented in a system compatible with a typical GIS architecture. After examining print historical atlases and historical GIS projects, several broad categories of geo-historical phenomena emerged: i) complex events having sub-events in common, ii) theoretical processes, iii) procedure-like processes, iv) individual space-time paths, v) collective complex paths and vi) flows. Print historical atlases depict all of these classes in analog fashion, and successful digital counterparts must as well. Exemplar datasets presenting a corresponding set of modeling challenges were chosen to provide an effective test of generality for the data model:

Table 7-1. Representation challenges and exemplar datasets

CHALLENGES	EXEMPLAR DATASET
(i); (ii)	Contentious Gatherings in Great Britain, 1758-1834 (BRIT)
(i); (iii)	US Presidential Election Cycle of 2008 (ELEC08)
(i); (v)	Napoleon’s advance on Moscow, 1812 (NAP1812)
(vi)	Transatlantic Slave Trade Voyages, 1514-1866 (VOYAGES)
(i); (iv)	The Pilgrimage of Xuanzang, 629-645CE (XUANZANG)

DRUMLIN is a historical GIS database that can support mapping, timeline, and graph visualizations for digital historical atlases, geospatial and textual analyses performed

[63] http://www.postgresql.org; http://postgis.refractions.net/

in desktop GIS clients and other software, and faceted browsing. The logical structure of the SHO has been partially expressed, sufficient to enable both explicit and inferential knowledge to be extracted with the relational algebra of SQL. It should be possible to integrate DRUMLIN with OWL-DL reasoning engines and SPARQL query interfaces in the future[64] but this is outside the present project scope.

The physical implementation of DRUMLIN is described in §7.1. In §7.2, the exemplar datasets are introduced. Their manner of fit to the data model, associated challenges they presented, and some analytical questions one might investigate with them are discussed. Figure 7-1 illustrates the key representational requirements carried forward from the logical model for six *geo-historical information constructs* (GHICs) developed in Chapter 6.

[64] For example, using APIs from Sesame (http://www.openrdf.org/) or Jena (http://jena.sourceforge.net/)

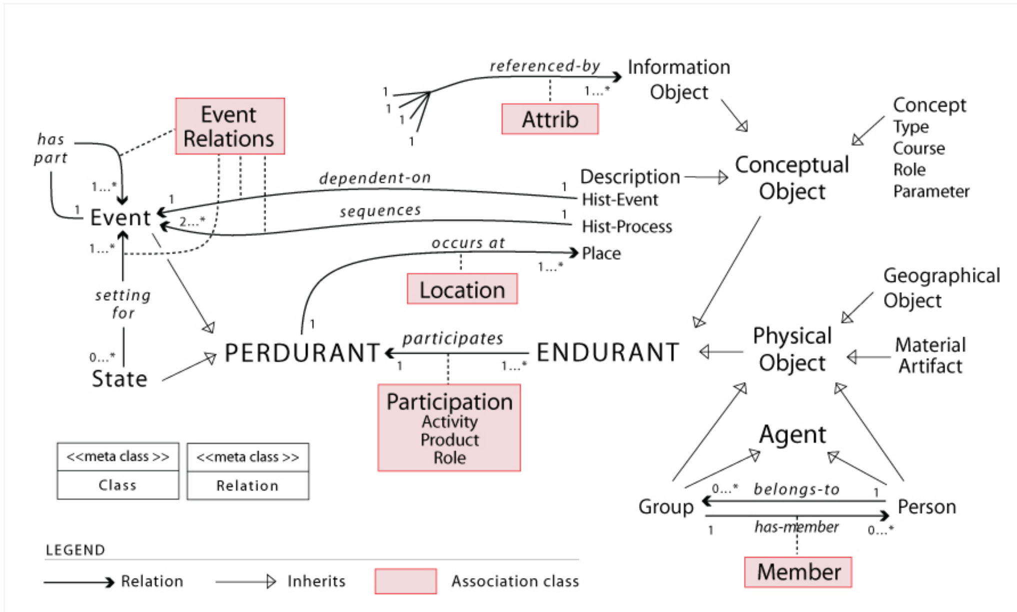


Figure 7-1. Schema requirements brought forward from SHO

7.1 The Stuff of History IV: Physical implementation

Logical statements from the SHO are represented within the DRUMLIN schema (Figure 7-2) by various means: table structure, datatypes, materialized views, functions and constraints. Borrowing terminology from description logics, in this model historical knowledge is held in:

- the 'TBox' of classes and relations defined by hierarchies (*is-a* and *sub-relation*), class attributes (column headings) and check constraints; and
- the 'ABox' of entity instances having asserted attribute values and relations between instances.

In DRUMLIN, the SHO taxonomies of entity classes and relations (a.k.a. properties) are stored in a [Class] table[65] and a [Relation] table with constraining *has-domain* and *has-range* properties. A significant proportion of instance data is stored in association tables corresponding to core relations concerning parthood, participation, membership, influence, location and attribution of source (Table 7-2).

Table 7-2. Association class tables in DRUMLIN

TABLE	RELATIONS
[particip]	participant(event, endurant, activity, role, duration) product(event, endurant)
[member]	member((organization ⋁ group-of-persons), person, role, time-period)
[relev]	influenced((event ⋁ endurant), event) part(event, event)
[evloc]	location(event, place)
[refer]	references(information object, particular)
[setting]	setting-for(state, historical-process)

7.1.1 Physical data model

The E-R diagram in Figure 7-3 represents the majority of the DRUMLIN table schema[66]. Database functions and views are described in §7.1.3 and §7.1.4 respectively. Some general points to note about the ontology's expression in the data model include:

1. Classes are defined:
    a. in parent-child hierarchies in the [class] and [type] tables; each class or

[65] Database tables are hereafter indicated in square brackets: [table_name].

[66] A few tables considered to be incidentally supporting do not appear.

type has a parent_id attribute. For example, Activity (A) has-parent Perdurant (PD), and Agentive Activity (AGA) has-parent Activity (A). These properties express the properly-subsumed-leaf (PSBL) property in DOLCE. The transitivity of this hierarchy is calculated in queries and query functions by means of a recursive UNION over "all children of class c." See for example the utility function f_subtypes (§9.3.17).

b. in cases, as tables (i.e. relations in the relational model). Table definitions can inherit properties/columns and express a modal logic of necessity and possibility in field definitions. Where fields have NOT NULL constraints, a property is necessary; others are possible. For example, the table for Events [e], inherits all properties of Perdurant [o] and adds its own type_id property (logically

$$\mathrm{Event}(x) \rightarrow \exists t(\mathrm{Type}(t))$$

Note that many definitional constraints will be enforced at the application level, e.g.

$$\mathrm{Event}(x) \rightarrow \exists y([1..*](\mathrm{Place}(y) \wedge \text{located-at}(x, y))$$

c. by extension; we can say that in World w the class of slaving captains, or "slavers" is all Persons (PER) who Participated (PC) in Events (EV) of type "Slaving Voyage" in the role "captain." In DRUMLIN this appears as the materialized view, v_slavers.

2. The taxonomy of Class categories can be extended (i.e. specialized) by adding new Type hierarchies, as described in §6.3.8. All entities are of a Class, and

can be specialized given a relevant taxonomy of Type.

3. Properties (relations) are defined in DRUMLIN

    a. in terms of an *is-a* hierarchy in [relation] table, functioning as with [class] above.

    b. in terms of domain and range; properties are constrained to holding for the class named in domain (and its sub-classes), with valid entries constrained to classes named in range.

4. Table inheritance functionality is used to group entities beneath two top-level divisions, [occur] and [contin], corresponding to the perdurants and endurants of DOLCE and SHO67. All instances of temporal things inherit a unique occur_id; instances of non-temporal entities inherit contin_id identifiers. All [contin] and [occur] 'child' tables inherit class_id fields, which are foreign keys to the [class] table. Most have a type_id which maps to [type].type_id for finer categorization in SKOS-like controlled vocabularies as discussed in §7.2.3.

5. As noted earlier, many ABox knowledge statements are stored in the association class tables (Table 7-2). Note that [relev] does double duty, storing (i) event/sub-event composition; and (ii) the influence relations between

[67] Continuant and occurrent are a personal preference, principally because of the obscurity of the term *perdurant*.

events asserted in historical-processes.

Figure 7-2. DRUMLIN schema, Entity-Relationship diagram

In order to implement the *possible world* logic discussed earlier (§6.2.2 particularly), all endurant ([contin]), perdurant ([occur]) and association class records have *source_id* and *dataset_id* attributes. The *source_id* field is a foreign key pointing to an information object in [inf] from which the data is drawn; the *dataset_id* differentiates between discrete projects within the larger database. DRUMLIN contains five distinct datasets (§7.2).

7.1.2 GHICs in DRUMLIN

The geo-historical information constructs traced through Chapters 5 and 6 are modeled in DRUMLIN in the following manner

7.1.2.1 <u>Attribution</u>

- All propositions (i.e. table rows) must be attributed to a project (dataset_id) and a source (source_id). Project/dataset corresponds to the notion of *possible worlds* developed in Chapters 5 and 6. The value for source_id is a foreign key to the infobj_id field the Information Object table [inf]. Rows have varying granularity: in DRUMLIN, some tables have many attributes/columns, some only a few. In an RDF model it would be possible to achieve datum-level attribution, but that is not the case here.

- Additionally, the [refer] table permits the assertion of any number of *references* relations between an Information Object (IOBJ) in [inf] and any tuple/row in the databse. Sub-relations of references include *about*, *lists*, *represents*, *documents*, and so on. Trigger functions can be applied by application level logic to enforce the various domain and rnage constraints for each. For example, only Visual Items (VIS) can *represent* something.

7.1.2.2 Historical Events and Participation

- Events are a sub-class of Perdurants and have their own table [e], inheriting several columns from [o] (named for Occurrence, a preferred term for perdurant). Events have preferred names (prefname) and a temporal attribute (period), which is an array of intervals, as described in §7.1.6.
- Events have one or more locations, represented in the [evloc] table. In cases, a location will correspond to some activity described in a participation record in [particip], but that connection is not modeled explicitly in DRUMLIN. The presumption is that if we know the spatial-temporal location of some activity, it becomes an event.
- Activity performed by Event participants (“speak, “march,” “fight,” etc.) is described in the [particip] table, time-indexed only as “some-time-during” or “throughout” an event. A [particip] instance may describe only a (subject, activity, role) or include an optional (predicate, object); *contin_obj* is any endurant and *role_id* points to a role from a vocabulary in the [type] table that can be extended by any given atlas application. For example, in BRIT08, a (PGRP) of 30 farmers might Petition a (PER) or (SPER) judge.
- Participation records also describe material results of events, with a *produced-by* relation.

7.1.2.3 *Historical Process*

- Historical processes are registered in the [hpro] table, inheriting [contin] columns such as class_id, dataset_id, source_id, wiki (a URI) and notes.
- The elements of a Historical Process are asserted in the [relev] table (related events). The relations possible between events include the mereological (*part-of*) and the telic (*influenced* and its children, including *motivated*, *had-purpose*, *caused*, *initiated*, *facilitated*, etc.). A query for influences could produce data to support a 'smart' timeline, as suggested in Figure 5-1.
- States (ST) can be asserted as relevant in Historical Processes; these are "measurement tables," typically having one or more geometry column and therefore mappable. State tables are registered in [state], where they inherit an occur_id key value. Such occur_id values can then be referenced in the [relev] records in the following manner: occur_id <*setting-for*> hpro_id. Records for Historical Processes therefore take the form, e.g.:

In [hpro]:

hpro_id	class_id	type_id	dataset_id	source_id	wiki	period[ ]	prefname
hpro1	HPRO	<null>	99	345	http:...	[[t][t2]]	Obama Rise

In [relev]:

relev_id	hpro_id	subj_label	subj	pred	obj	dataset_id	source_id
1	hpro1	speech23	ev01	initiated	event02	99	345
2	hpro1	appear234	ev03	facilitated	event04	99	345
3	hpro1	census2000	st03	setting-for	hpro1	99	345

7.1.2.4 *Groups and membership*

- In the SHO, three types of groups are defined: physical Groups of Persons

(PGRP), Functional Groups (FGRP), and Organizations (ORG)—a sub-class of Agentive Social Object (ASO). In DRUMLIN, all groups appear in a [grp] table and each record has a class_id of value of either PGRP, FGRP or ORG. The immediate goals are to be able to assert a group's (i) existence and class and/or type, (ii) participation in events, in roles, (iii) time-indexed membership, and (iv) sub-group parts, when they exist.

- The [grp] table inherits all columns from parent and grandparent tables ([actor], [contin]), including prefname, dataset_id, source_id, and wiki (a URI value).
- Individuals are represented in DRUMLIN in a [pers] table, also a child and grandchild of [actor] and [contin].
- Groups can have a parent_id value, indicating a *part-of* relation to another group. For example, all the infantry and cavalry corps groups in Napoleon's army have a parent_id value corresponding to the actor_id of the Grande Armée itself.
- A [member] table records all instances of membership by Persons (in [pers]), with a temporal attribute (period) and an optional role.

7.1.2.5 Place, including paths and regions

The manner that place and location are treated in DRUMLIN is considered a place-holder for a robust historical gazetteer, a general model for which has yet to be defined anywhere.

- Events can have multiple locations, represented in the one-to-many association class table, [evloc]. Since events can have unlimited sub-event parts, each with its own location, the location for a complex "macro-event" is the mereological sum of its parts.
- The place_id attribute of an event corresponds to a single record in [place]—a table populated by ~350,000 records from the open source GeoNames database for the countries of interest in the exemplar data so far entered DRUMLIN. Although a logical requirement for permitting multiple time-indexed names for places is described in §5.5.3, that functionality is not described formally in Chapter 6, nor implemented in DRUMLIN. A GeoNames convention permits a comma-delimited list of alternate spellings and language variants to be entered in an altnames field.
- The [place] table has multiple geometry columns, as well as a prefgeom field that can be used by application logic to build spatial views for mapping. For example, cities in [place] may have both point and multipolygon data, with POINT as their prefgeom.
- A new class of place, Path—with sub-classes of Trajectory-Path and Flow-Path—can be specified in DRUMLIN, although formal definitions in SHO are incomplete. Paths are kinds of locations:
    - Trajectory-Paths are ordered sequences of point locations with connecting edges that generalize an actual route taken by some entity

across the Earth surface.

- A Flow-Path consists of two points and a time period array (note period arrays may have one or more elements). A Commercial Exchange aggregated event such as "British slave trade, 1754, Sierra Leone-Cuba" has a Flow-Path location of (8° 30', 11°30', [[1754-01-01, 1754-12-31]]). Measurement tables would list each event, a path_id, a commodity, and a magnitude. One can regard Flow-Paths as spatiotemporal channels (potential or actual) between places.

▪ Regions are handled in one of two ways in DRUMLIN:

- A polygon or multipolygon can be named in the [place] table, with a featcode value of "RGN."
- In the [region] table, a set of places can be asserted as comprising a region, e.g. according to one authority source, Appalachia is composed of a certain set uf US counties across several states. The North American free trade zone is composed of the territories of the US, Canada and Mexico.

7.1.2.6 Historical Periods

Historical Periods are simply named intervals, asserted in the [hper] table, as (prefname, period[ ], source_id, dataset_id). Queries on events, participants, activity, locations, etc. can be routinely constrained by temporal operations on Historical Period period[ ] values, such as like *contains*, *overlaps*, etc.

7.1.3 Functions

A great many operations are made possible by holding combinations of occurrent facets constant, including spatial and temporal regions/intervals, and all attributes of occurrent participants or results, including class and type.

Sixteen parameterized query functions were created for the DRUMLIN database to answer the "competency questions" listed in §5.1 (most appear in Table 7-3). Complete code for these appears in Appendix §9.2.

Table 7-3. Query functions in DRUMLIN (* = function 'supports' another)

QUESTION	QUERY	FUNCTION
2a; 6b; 6c*	what has happened at a place	*f_occurnear*(place_id, buffer) Everything that's happened within x distance of place in dec.deg.
6a; 6c	what has happened during a time	*f_concurrent*(period) Everything that's happened during a period overlapping the given period
3a; 3b; 3c	who participated, doing what	*f_particip*(occur_id) Simple list of activity performed by all participants in an event and sub-events
4c (with 4a)	what occurrences, artifacts, information objects resulted	*f_results*(occurrence [(place, period)])
4a	what's happened in a life	*f_activity*(actor_id) returns all events + activity performed by one actor; calls f_membership() as subroutine
4a	a person's life-path	*f_actorpath*(actor_id) Note: this relies on participates(actor,event) being constantly throughout.
9*	event location (union of sub-event locations)	*f_subevents*(integer) Event-subevents with locations
2a; 2b	nature of the 'goings on'	*f_particip*(occur_id) The structure of an event according to a dataset/source combo; subj/obj participants, roles, activity. E.g. 'select * from f_particip(20441,3,26009)' returns the assertions of (Esposito, et al., 1999) about Napoleon's advance

QUESTION	QUERY	FUNCTION
8	all instances of a particular type of event	*f_everhappened(*type_id*)*
9	measure of event products	*f_evmeasure*(event, measure, type) Returns results of any analysis of macro-event and sub-event products -- e.g. semantic analysis for speech text: 'select * from f_evmeasure(4000, 'semsig34', 'E55-3-6')' returns an issue 'profile' for all Dem speeches during 2008 election cycle.
5	members of a group, at some date	*f_members_date*(grp_id [, valid_date]) Returns all members of a group/org, on a date if given; E.g. 'select * from f_members_date(1019, '1862-01-01')' returns the US President on that date, Abe Lincoln
4b	group membership of a person	*f_membership*(actor_id) Returns groups a person has belonged to, when and in what role
2c	information objects referencing events at a place	*f_referenced(*place_id*)*
1a	the geometry of a region specified as a collection of places	*f_region*(integer) returns geometry of regions described as collections of places, e.g. census regions, Appalachia
		*f_subtypes(*integer*)* support function recurses Type taxonomy down from a given type_id

Several of these are called as sub-routines by others, and all could be called in possible functions not yet implemented. All can help to characterize a place in terms of what has happened there. For example, given *f_subevents*(*f_occurnear*(place_id, buffer)), a new function could ask about a place, 'who has been here, doing what?' Note that Question 1b, concerning multiple place names, is addressed in DRUMLIN only by direct queries to an AltNames field in the Place table; a better solution awaits a complete historical gazetteer.

Historical social networks could be developed with a *fellow_travelers*(actor_id) function. At one level, fellow travelers are those who have participated together in one or more events. For example, I have *some* connection to everyone who attended a given lecture with me; more tenuously to everyone at the first Paul Simon *Graceland* tour concert in Berkeley, CA one summer's eve in 1986. Obviously some metrics for strength of relation would be required. This benefit of an event-centered approach to formal-logical representations of human activity, historical or otherwise has been discussed by computer scientists Westermann and Jain (2007).

Having qualified participation by activity and role, we can imagine characterizing lives by locating them in activity *conceptual spaces* (see also §8.3.1): perhaps visualizing them along dimensions of creativity, destructiveness, movement, involvement in spheres (economic, artistic, etc.), even passivity or activity.

7.1.4 Materialized views

Each of these example views (Table 7-4) can be parameterized by functions to filter by dataset (world), source and in cases, class or type. So far, views are used primarily to create semi-permanent spatial layers for visualization within GIS client software and web mapping applications. Since queries are frequently hypotheses, views can be also used to assert or identify new complex classes by extension. For example, given a very large dataset, we might design a view to create a 'ViolentCreativePerson' sub-class based on activity, roles and event types.

Table 7-4. Materialized views in DRUMLIN

VIEW NAME	PURPOSE
`v_actors`	Returns preferred name of all persons and groups in a dataset (i.e. *possible world*) and count of events they are participants in

VIEW NAME	PURPOSE
v_event_typecount	Events in DRUMLIN by dataset (world), count, class and type
v_inf	Count of Information Objects (IOBJ) by class and type
v_issuesbyregion	Example of summarizing an 'issue-aboutness' measure for all speeches given in each of nine U.S. census regions; i.e. the aggregated results of speech-giving events
v_slavers	Creates a new complex class as slaving voyage event participants in the role of "Captain"
v_paths_napoleon	Generates a spatial table of all trajectories in the 'Napoleon advance on Moscow' event
v_paths_voyages	Generates a spatial table of all flow paths in the Voyages dataset
v_paths_xuanzang	Generates a spatial table of all journey segments
v_places_napoleon	Generates a spatial table with both point and polygon geometries in the Napoleon dataset (typ.)
v_relev	Returns all asserted event relations, including parthood and influence
v_sources_loc	Returns a count of Places in DRUMLIN by georeferencing source

7.1.5 SKOS-compatible taxonomies

The Simple Knowledge Organization System (SKOS) model is a W3C specification for controlled vocabularies with thesaurus-level logical structure. An SKOS implementation for enabling extensible concept hierarchies in DRUMLIN is partially completed. The [type] table can accommodate any number of domain- or author-specific vocabularies, each tagged as belonging to that domain (or author) namespace, e.g. 'brit:' or 'elec08:' So far, simple hierarchies corresponding to the Broader-Term (BT) and Narrower-Term (NT) thesaurus standard are implemented using a *parent-type* attribute. Recursive queries can navigate such hierarchies. Appendix §9.2 lists the current Type taxonomies. Several conceptual issues regarding classification requirements, and how this might evolve, are discussed in §6.3.8.

7.1.6 A 'Period' datatype

In §5.6.7 the problem of representing and reasoning about indeterminate spatial and temporal boundaries was introduced as a requirement for digital historical atlases and HGIS more generally. Although the development of a complete general solution is beyond the scope of this research, a means for handling indeterminate time periods is introduced here. The PostgreSQL database software permits the definition of new datatypes, and open-source developer Scott Bailey has developed and published a Period datatype[68], specified as a [start, end] pair of timestamps that can be used to describe a single interval or aggregated into an array (Period[ ])of any length. In other words, a Period [ ] may consist of many discontinuous Periods. An extensive set of operators and functions are defined that permit constructing Periods, generating Period[ ] arrays, and calculating intersections, unions, spans (termed 'ranges'), midpoints, containment, adjacency, and so forth.

In the DRUMLIN database, all temporal attributes for events have been converted to periods: (i) simple dates are converted to *periods* spanning 24 hours; (ii) intervals specified in the data by known start and end dates are converted to simple *periods*; (iii) temporal attributes given as month/year, single years or multiple years are likewise converted to simple *periods*, e.g. December, 1830 is effectively

[1830-12-01 00:00:00:00, 1830-12-31 23:59:59]

---

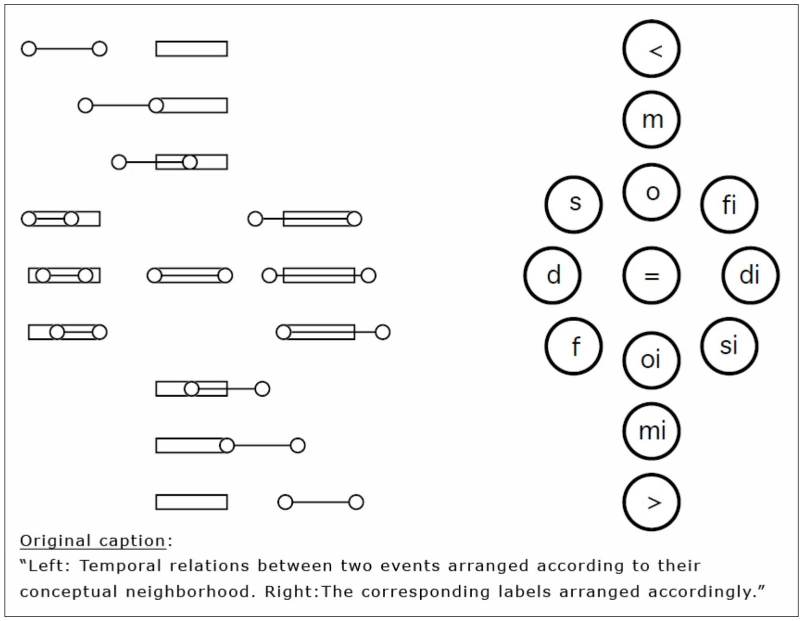
[68] The pgChronos project (http://pgfoundry.org/projects/timespan/)

Where start or end dates are vague or speculative, they can be specified as Periods themselves. For example, the interval "from the spring of 947 to mid-summer the next year" is represented as

[[0947-03-21, 0947-06-20], [0948-07-21, 0948-08-20]]

In a subsequent stage of development, probabilistic weights and confidence values will be added, so that queries about containment or overlap can return either multi-valued qualitative answers (like "maybe") or weighted indices ("with 0.7 probability;" or "with 0.9 confidence"). These values could then be translated to appropriate cartographic symbols.

Separately, some qualitative reasoning capability has been added for the Period and Period[ ] datatypes by creating operators corresponding to Freksa's semi-intervals (1992) (Table 7-3). These effectively summarize certain operations on intervals. Because all temporal attributes are given as periods, we can ask for example for "older contemporaries of *x*" (true if p1 starts before p2 starts and ends after p2 starts). A listing of these operators is given in §9.5.



Original caption:
"Left: Temporal relations between two events arranged according to their conceptual neighborhood. Right:The corresponding labels arranged accordingly."

Figure 7-3 - Semi-intervals for temporal reasoning, from Freksa (1992:14)

The Period and Period[ ] datatypes are not to be confused with the Historical Period (HPRD) class developed for the SHO in §6.3.6.

7.2 Exemplar data

In the following sections, for each of five exemplar datasets (Table 7-5) a simple map generated from DRUMLIN spatial views is followed by a brief description of the data, a discussion of modeling challenges, and some associated research questions.

Table 7-5. Exemplar data with supporting tables

SECTION	EXEMPLAR DATASET W/SUPPORTING TABLES
7.2.1	Contentious Gatherings in Great Britain, 1758-1834 (BRIT) -- Crop returns for England, 1801 -- Ancient county boundaries, 1851
7.2.2	US Presidential Election Cycle of 2008 (ELEC08) -- US 2000 Census w/2004 election results
7.2.3	Napoleon's advance on Moscow, 1812 (NAP1812) -- World Heritage sites, 2009
7.2.4	Transatlantic Slave Trade Voyages, 1514-1866 (VOYAGES)
7.2.5	The Pilgrimage of Xuanzang, 629-645CE (XUANZANG)

7.2.1 Contentious gatherings in Great Britain, 1758-1834 (BRIT)

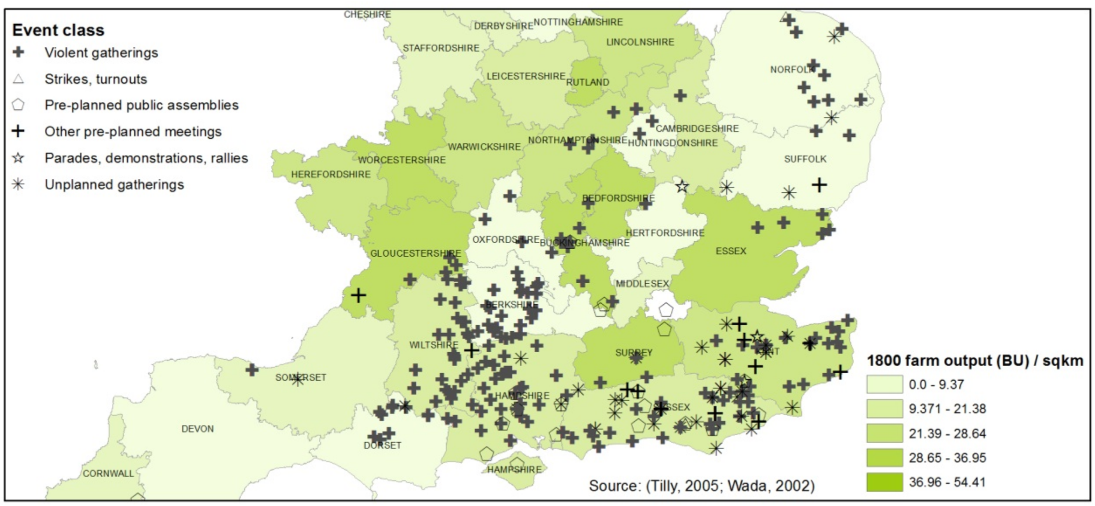


Figure 7-4. The Swing Rebellion of 1830; events by type, Aug-Dec and relative 1801 farm output (darker shaded counties were higher).

The *Contentious Gatherings in Great Britain, 1758-1834* dataset (BRIT) was compiled in digital format from contemporaneous written accounts in several British periodicals by the historical sociologist, Charles Tilly (1995) beginning in 1979. In Tilly's encoding scheme (Tilly & Schweitzer, 1980)—effectively, his ontology of contentious gatherings—*events* (8,088) are constituted by individual *actions* (50,875) undertaken by *formations* (27,184 instances of either individual persons or groups), at *locations* (2,377). The data was converted from a generally flat model to a relational database by Wada (2002), who has graciously shared it. Its structure presented several challenges:

- While many events have multiple locations, they are not matched with the event's individual component actions.
- Each event belongs to one of 15 types ("poachers vs. smugglers," "delegations" etc.) and principally concerns one of 43 issue categories ("corn laws," "forgery," "slavery-anti," etc.). Component actions, participants and locations were encoded according to hypotheses and methods particular to the investigator (Tilly) and the data. Such project-specific entities as 'machine breaking' and 'brawl' Events, "blacklegs" Groups and a "mobilization" Process have been assigned a *type_id* in DRUMLIN from a project-specific taxonomy added to the ontology as subclasses of Type;[69] For example, the narrow types Machine Breaking and Mob specialize the more general Violence and Gathering. Another researcher describing the presumably identical events of this period, Charlesworth (1983), used a different vocabulary—more commonly merging issues and activity type with for example "tithe reductions," "poor-law protest," "machine destruction" events. All such categorizations have nuanced meanings for their authors, which are normally spelled out in accompanying texts and should be maintained. Any harmonizing or merging of categories would preferably be undertaken by one or all of the authors themselves.

[69] A technique borrowed from the CIDOC-CRM (Crofts, et al 2008) and discussed in §6.3.8 and §7.1.5

- Events are composed of one or more action phases described by one of a large set of verbs, grouped in "verb categories." Only the 47 category verbs (e.g. *petition*, *proceed*, *request*) have been added to the [type] table; the larger list is omitted for now.
- In the original data, periodical sources cited are mapped to entire events, and not to individual actions.
- Point locations given for events are in six-figure Ordnance Survey grid coordinates. This intra-Parish level resolution (100m) provides useful detail within large cities, but is incompatible with more usual latitude-longitude point locations and as the digitization methods are unknown, its precision is suspect.

The principal questions Tilly asked of his data (1995) can be summarized as "how did the public's 'repertoires' of activity in British contentious gatherings during this period change, expand and disperse geographically—and to what end?" The DRUMLIN database can support animations illustrating that dynamic structure of the Swing Rebellion event for example—which has been modeled as a subset of gatherings during a 4-month period in late 1830.

Other research topics that could be addressed with this data include:

- Did particular kinds and levels of agricultural production correlate with Swing disturbances? To at least contextualize that, I have added a spatial dataset of parish- and county-level agricultural output for 1801 (Figure 7-4), which can be asserted as a relevant State in a Historical-Process.
- Was gathering frequency sustained throughout the period in some locations and not others?
- Other, non-Swing protests occurred during that period. Were they spatially correlated with the Swing event, and were they composed of similar activity "repertoires?"

- 206 gatherings in the period 1789-1834 were coded as “Anti-slavery.” We can ask what the characteristics of agricultural activity were in the locations where anti-slavery protests occurred. Another puzzling question arises: since the British House of Commons abolished the slave trade in 1807, what explains the continued protests? To begin answering that question, I queried the DRUMLIN dataset of trans-Atlantic slave trade (Voyages, 2008). British slave trade was extensive overall (Figure 7-9), but had virtually stopped by 1808 (§7.2.4).

7.2.2 Election 2008 (ELEC08)

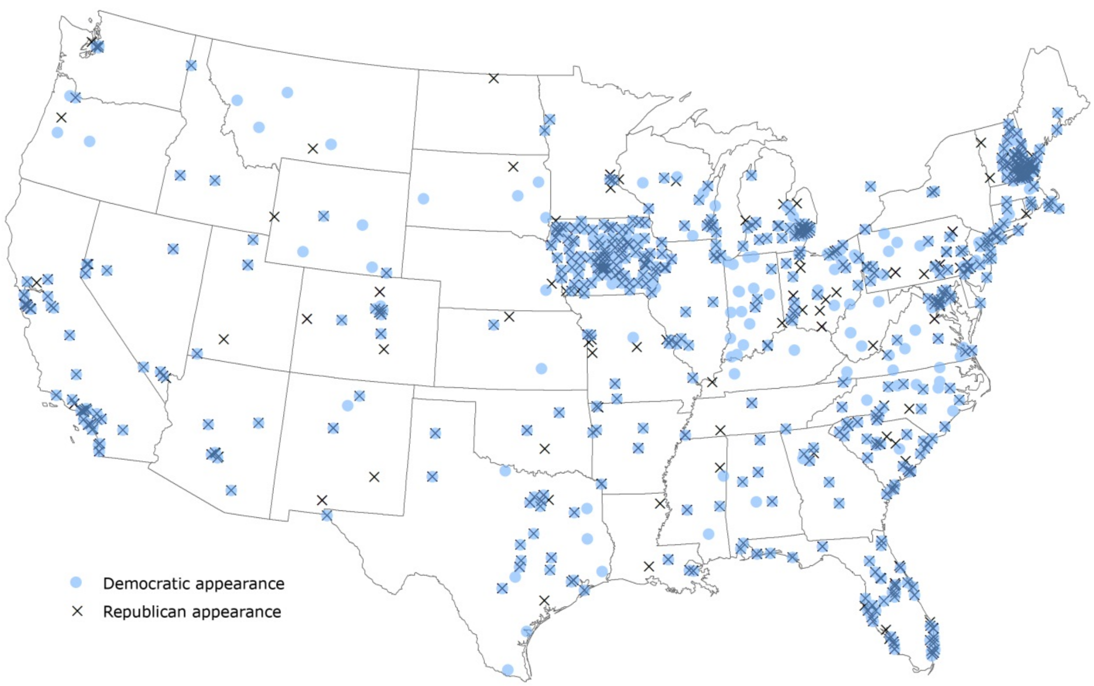


Figure 7-5. 3836 candidate appearances during 'Election '08,' 2007-2009

During the two-year period of 2007-2009, I harvested a large volume of data about 4008 events that were part of the 2008 U.S. Presidential election cycle. Although not historical (yet) this complex procedure-like process had a dynamic structure similar in many respects to election cycles in earlier times and elsewhere in the world. It also ‘exercised’ the model in several ways:

- It was constituted by complex macro-events (e.g. 2 nomination races, a general race, 14 individual campaigns) each having many sub-event parts (e.g. 3836 candidate appearances, 44 debates, 2 conventions)
- It had numerous participants, whose purposeful trajectories around the country we might wish to analyze.
- Many events (e.g. 427 speeches, 44 debates, 108 primaries and caucuses) had products (text) and results (statistical outcomes) that are also interesting objects of analysis.

Table 7-6 lists the entities of interest and notes those currently stored in DRUMLIN. A second pass was made later in the process, to contextualize the entities considered beyond the ‘Election 2008’ case by adding selected component sub-events of a related complex occurrence, the US Civil Rights Movement.

Table 7-6. Entities of interest for Election 2008 (* = in DRUMLIN)

EVENTS	ACTORS	INFORMATION OBJECTS	CONCEPTS
election cycle *	person *	written statement *	issue (declared) *
- public appearance *	- candidate *	spoken statement	issue (emergent) *
- issue a statement	- blogger *	news report *	
- debate *	- person of note *	commentary *	**PLACES**
- interview	- analyst	expert analysis *	location (city, county, state) *
- opinion poll	group	speech transcript *	region *
nomination race *	- the electorate	video metadata	
- enter; withdraw *	- campaign org. *	photograph	**STATE/CONDITION**
- primary; caucus *	- political party *	cartoon	demographics statistics *
- convention *	- government	opinion poll dataset	economic indicators
general race *	- news organization *		
- general election *	- think tank*		**MEASURES**
- electoral vote			issue ‘aboutness’ *
inauguration *			speech similarity *

The entities from Table 7-3 have been represented in DRUMLIN in the following general manner; classes (with SHO abbreviations) and types added to the Type taxonomy are capitalized; relations *italicized*):

Events

Event (EV) is one of four sub-classes of Perdurant (PD). An Election Cycle is a composite event type comprising party Nomination Races, a General Race, a General Election, an Electoral College Vote and an Inauguration. Most had numerous child events of shorter duration, modeled at various levels of granularity. Nomination Race and General Race events include myriad Political Gathering events, e.g. Speech, Fundraiser, etc. A Nomination Race includes multiple Primary and Caucus events, began with the first Enter Race event for a candidate of that party and ended with a Nominating Convention. Other kinds of events include the publication of Information Objects (IOBJ) and Measurements like polls, censuses or text analyses, the results of which can appear as instance attributes in any number of State (ST) or ‘measure’ tables.

Actors

Actors may be Persons (PER) or Groups (GRP; PGRP; ORG), which are sub-typed as Commercial Businesses, Advocacy Organizations, and so forth. Many actors have *member-of* relations in Groups, for intervals, further specified by Roles (FR) in a Type (TYP) taxonomy (leader, employee, etc.). Actors participate in Events during intervals by performing some activity, possibly in a role (debater, perpetrator, etc.).

Information Objects

An Information Object is—by inference only in DRUMLIN—the non-material creative product of an Event, with some Persons or Group in the Author role. Its sub-classes include Linguistic Objects (LING) further typed as Speech Transcripts, News Articles, Blog Posts, etc. and Visual Items (VIS) such as the content of a photograph or video. These are distinguished from the artifacts that carry them: the King James Bible is a Linguistic Object; a particular copy of it—the Lincoln Bible used in the inauguration—is Material Artifact (MART). Taxonomies of artifact types have not yet been developed in DRUMLIN.

Issues

Election issue topics, whether declared by candidates or discovered by computational linguistics methods, are descriptive entities that are not represented explicitly in DRUMLIN. They are however dimensions of analysis for Information Objects. Their proper place in the SHO is unknown at this time; one possibility is a 'Subject' sub-class of Description.

Places

The events of the 'Election 2008' Historical Process (HPRO) took place at locations modeled as Geographic Places (PLG) at the scale of cities, states and regions. Place Appellations (APPL) can correspond to both point and area geometries and are dynamic, in that multiple names can refer to the same place and the preferred name may change over time. Places like 'Appalachia' or 'The South' are the region subtype of Place, and not formalized in SHO (§6.3.7).

States/Conditions

A State (ST)—the particular condition of some entity we assert as being valid throughout some interval—is related to Activities and Events in the context of Historical Processes (HPRO). Examples of relevant states in the ELEC08 process include the demographics and political leanings of state and county populations, as measured by censuses and prior election results. The topography of a region could be a State: is political persuasion correlated to elevation?

Measures

Some measurements and analysis results are non-temporal, i.e. they are not States valid only for some interval. In the ELEC08 case, the results of semantic similarity measures like "issue aboutness" are asserted properties of various Information Objects. This database schema is extensible in allowing for any number of Measure tables, each with possibly unique dimensions, referring to one or more subjects.

7.2.2.1 Research questions

Two studies were undertaken with the ELEC08 data. In the first, the text of speech transcripts was compared with candidates' web statements and think tank analysts' reports about 34 campaign issues, using a cosine similarity measure. The specifics of the measure are not described here, but the product is relevant: for each document, there is a 34-value array, or *issue signature*, representing a document's position in a 34-dimension conceptual issue space. This enabled an analysis of geographic variation in issue 'aboutness' for the *IssueBrowser* project (Grossner, 2009). In the second analysis, the 34 dimensions were reduced to two with multi-dimensional scaling (MDS) in order to visualize the shifting of candidates' positions within the issue space over the last two weeks of the campaign. Striking patterns were discovered (Figure 7-6).[70]

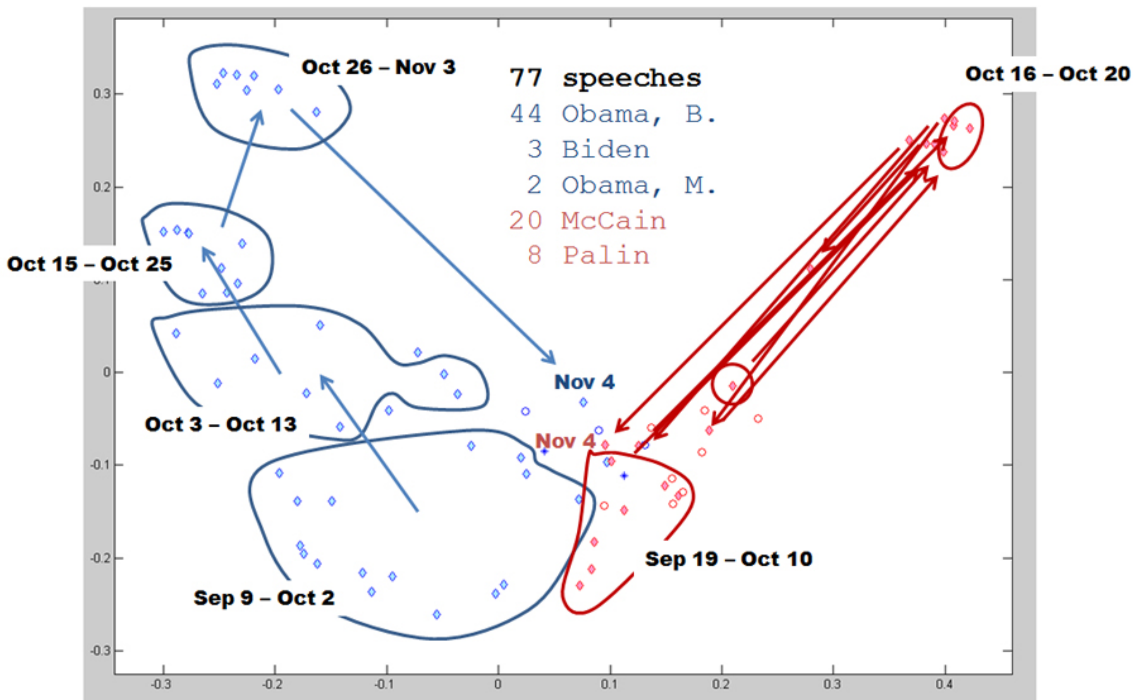


Figure 7-6. Movement through "issue space" at the end of Election Cycle 2008; the left side are the Democrats, the right side are the Republicans

Some additional questions one could address given the data and the ontology-driven data model include:

---

[70] Both studies are illustrated in http://spatial.ucsb.edu/events/brownbags/docs/2008-2009/Grossner-brownbag-presentation.pdf

- Whether and how results from the previous election in 2004 impacted the candidates' campaign stops in 2008. A dataset of 2004 results and demographics from the 2000 US census were added to DRUMLIN as relevant states in the Election 2008 historical-process.
- Which candidates were "average," and were there any outliers—along any number of dimensions: age, birthplace, life-path, language used, etc.
- How did the issue space change over the two-year period?

7.2.3 Napoleon's advance on Moscow, 1812 (NAP1812)

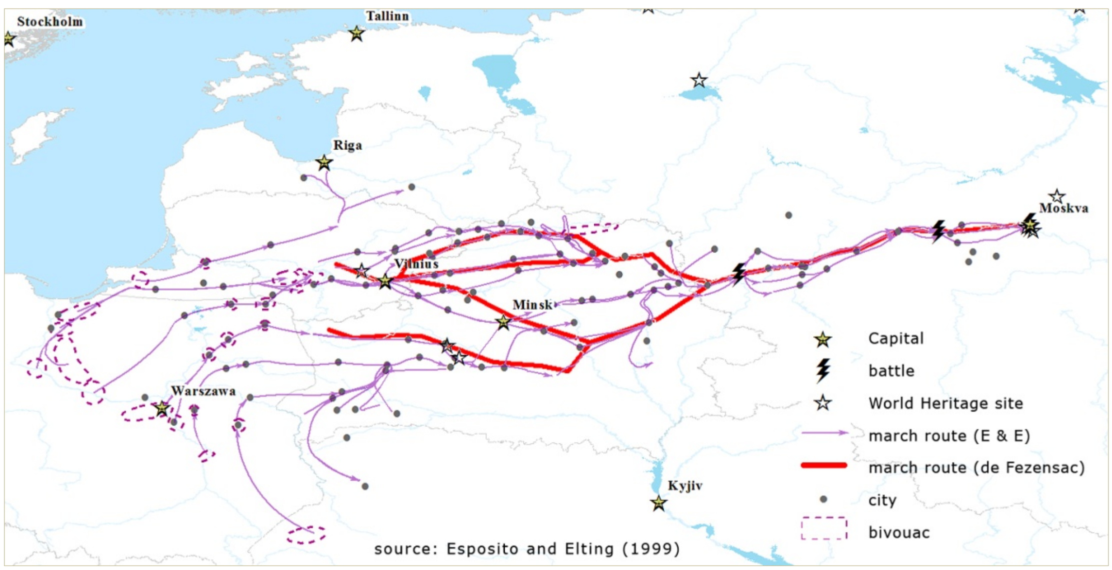


Figure 7-7. Napoleon's advance on Moscow, 1812, digitized from Esposito and Elting (1999)

One motivating question at the earliest stage of this dissertation research was, "what data model and specific GIS schema could reproduce the Minard map of the fate of Napoleon's Grande Armée during the 1812 Moscow campaign" (Figure 7-8). That graphic is justifiably celebrated for its concise and elegant display of several variables. As it turned out, several scholars have modeled what was *literally* represented in the map.[71]

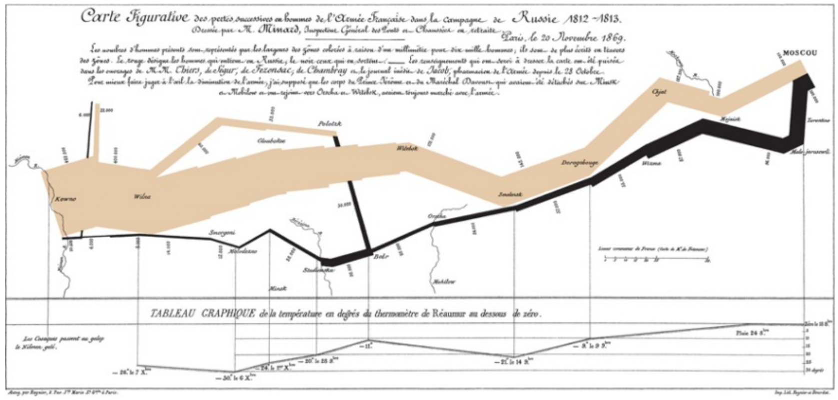


Figure 7-8. Carte Figurative (M. Minard, 1869) – "the successive losses in men of the French Army in the Russian campaign 1812-1813"

Upon close examination, the map data is interesting for what is missing. All of the activity of the Grande Armée was aggregated to a few flow lines. Events are present only insofar as they may be inferred: that the army was *here* on this date is obviously a very coarse generalization. So, the original challenge morphed and became more difficult: to model far more data about the campaign from at least two plausible accounts, such that they may be compared and explored further in a linked data repository. The inherent difficulties of this—sparse climate data, and conflicting accounts of events, actual paths marched and casualties—are the usual case with historical investigations.

The mapped results of what has been undertaken so far appear in Figure 7-7. The trajectories of fifteen army corps and Napoleon's Imperial Guard, along with their

---

[71] A survey of these has been compiled by M. Kraak, http://www.itc.nl/personal/kraak/1812/index.htm

approximate encampment locations have been digitized from detailed maps in a military atlas (Esposito & Elting, 1999). Several battles and the December, 1812 Fire of Moscow were added to this set of Military March and Bivouac events. Their participants are groups of the types Infantry Unit, and Cavalry Unit, sub-types of Military Unit. Group members include their commanders, about whom some biographical data is entered. The trajectories of and the Imperial Guard and the Armée as a whole according to a Napoleon aide-de-camp, M. de Fezensac, are displayed as well.

Not much in the way of analysis is possible with this data yet. However it does begin to answer the question, “what did Napoleon's trajectory across Europe and Russia *actually* look like, as opposed to what Minard has drawn?” Also, by adding the spatial dataset of UNESCO World Heritage sites as an overlay (§7.2.6) in a digital atlas application, one could ask what protected heritage sites lie on the now well-articulated, more nearly *actual* paths of Napoleon's Grande Armée? In such an atlas, one might browse the evidence of that military campaign existing presently on the landscape, at battle sites like Borodino, the Beresina River, Smolensk, and Moscow, or at heritage sites like the Mir Castle in Belarus, which was occupied and severely damaged.

7.2.4 Transatlantic slave trade voyages, 1514-1866 (VOYAGES)

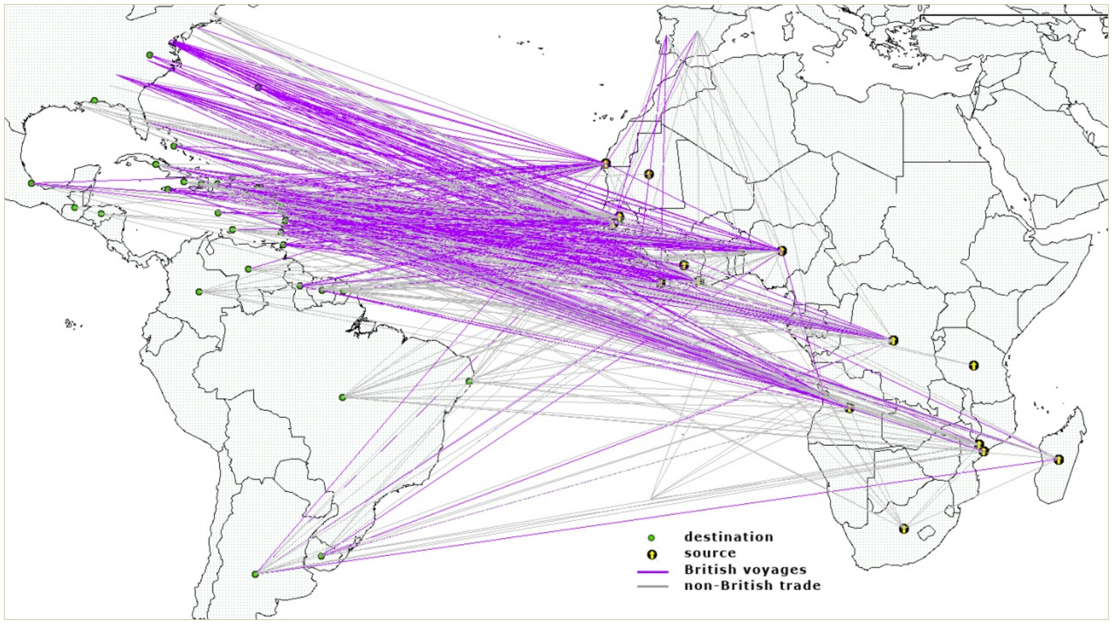


Figure 7-9. Transatlantic slave voyages: British-flagged ships 1758-1808

Flows are commonplace phenomena in print historical atlases. They are frequently mapped using arrow symbols with varying line width indicating magnitude. Typical flows include migrations and economic trade. Less obvious examples that might not correspond to quantitative data include diffusion of cultural practices, and communication or trade interaction generally. All of these are aggregations of multiple events, but commonly modeled as a single interaction.

The *Voyages: Transatlantic Slave Trade Database*[72] (Voyages, 2008), containing detailed records of 34,940 voyages between 1514 and 1866 has been imported into

[72] A project begun in 1993, centered at Emory University. The most recent website providing complete access to its data has been in operation since 2008: http://www.slavevoyages.org/

DRUMLIN. It allows us to model that trade activity in three ways: (i) as individual voyage events with named participants, including ship captains and (in cases) the captives;[73] (ii) as individual commercial exchange events—usually several per voyage; and (iii) as commercial trade flows, aggregated by time period, by ship flag, or by source and destination locations in any combination.

Each of these types requires a different geometry for digital mapping. Voyage events are trajectories comprising points and edges corresponding at least generally to a path on the Earth surface. Individual commercial exchanges would be point data in a GIS. Trade flows are dyads—point pairs that *are* geographically meaningful (although possibly centroids of large areas) with a magnitude value in relevant units. The Voyages dataset has provided a good test of the modeling approach taken for paths (§6.3.7.1): a path as a kind of place, having two subclasses, trajectories and flows.

In DRUMLIN, the Voyages data also serve to contextualize the BRIT dataset, in that a fair number of the protest events encoded were expressing the anti-slavery sentiment of the period. The data allowed me to map a simple answer to the question, "during the entire period of contention studied by Tilly, how extensive was English participation in the slave trade, relative to others?" (Figure 7-9).

---

[73] At the tail end of this period, after slave trade was made illegal in many places, a number of ships were intercepted in ports and on the high seas; there were trials held in Sierra Leone and Cuba, and freed slaves' names and originating locations (i.e. homes) were recorded.

The question can be answered several ways in DRUMLIN:

- A simple count of events of type Trade Voyage (E7-11-2-1) during the period 1758-1834, grouped by country flag (in this case embedded as a 2-letter code in an auto-generated event name). The top four countries were ["GB", 6338; "PT", 6149; "FR", 2565; "US", 1375]

```
SELECT substring(e.occur_name from 17 for 2) AS country,
count(*) FROM e WHERE e.type_id = 'E7-11-2-1' AND (e.yr > 1757
AND e.yr < 1835) GROUP BY country order by count desc
```

- In this case, 7479 Commercial Exchange events were derived from the ~35,000 total voyages, as point-to-point aggregated Flows for year/country. In DRUMLIN, the spatial data for all Events occurring in Path (Flow or Trajectory) locations are stored in a [path] table. A query such as the following joining paths [path], events [e], event locations [evloc] and a measurement table [m_slaveflow] will generate a spatial table that can be rendered in a GIS and symbolized variously, as seen in Figure 7-9.

```
SELECT p.contin_id, p.geom_line, el.path_id, e.occur_id,
p.startplace, p.endplace, e.occur_name, f.yearam, flag_ccode
FROM path p
JOIN evloc el ON p.path_id=el.path_id
JOIN e ON el.occur_id=e.occur_id
JOIN m_slaveflow f ON e.occur_id=f.occur_id
WHERE e.type_id = 'E7-5-4-1' AND (e.yr > 1757 AND e.yr < 1835)
```

Some other questions that should be generalizable to other large event datasets involving interactions include:

- What activity and events drove, or were concurrent with, landings (purchases and sales) in the various regions, and did the magnitude of trade ebb and flow in response to them?
- What are the current politics and demographic characteristics of the historical slave exporting and importing regions?
- Broadly, does the present nature of slave trading places (exporters and

importers) reflect those events?

- Has this history been memorialized in any of those places, and if so how?

7.2.5 The Pilgrimage of Xuanzang, 629-645 CE (XUANZANG)

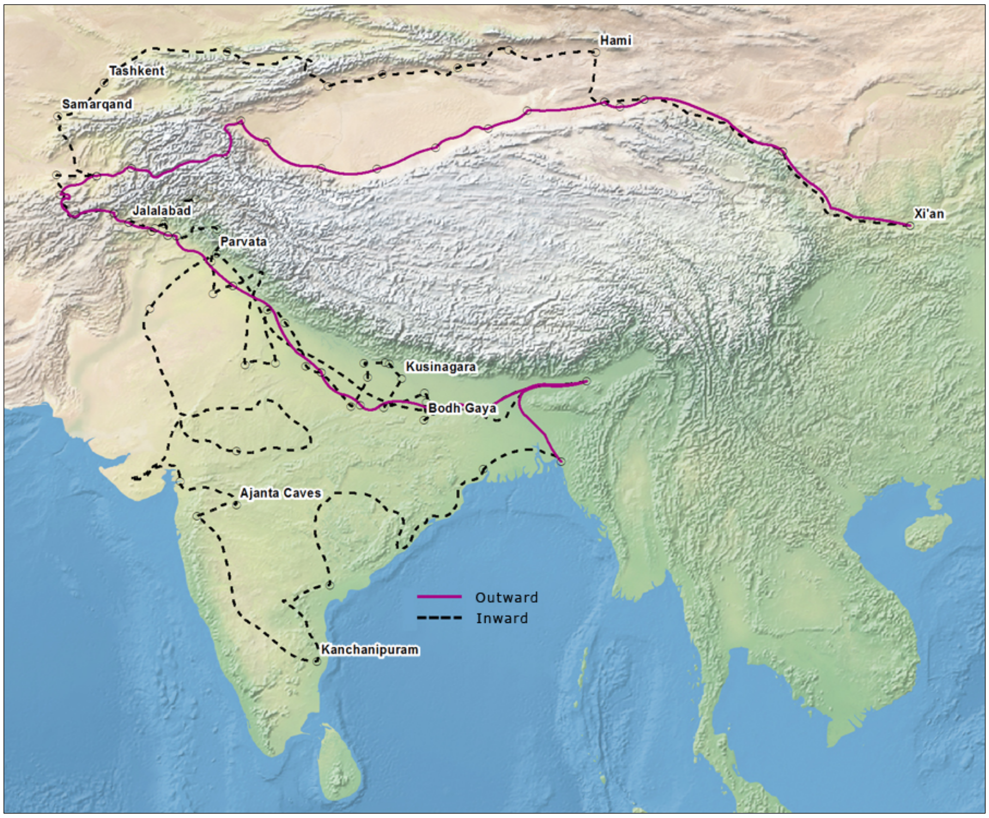


Figure 7-10. Route of Xuanzang, 629-645

An important class of event in print historical atlases is the journey. Print atlases commonly represent routes taken by exploratory, scientific and trading expeditions, as well as some extraordinary individual journeys. One of these is the well-chronicled 16-year pilgrimage of the Chinese Buddhist monk and scholar, Xuanzang. An encoding of certain events in one such chronicle (Wriggins, 2004)[74] was recently undertaken by a University of California, Merced history student (Thorpe, 2010), for a Google Earth narrative visualization. Only arrival, stay and departure events were encoded (162 in number), and their importation into DRUMLIN raised several useful challenges:

- Place names in 7th century China are difficult to align with modern gazetteer entries. Annotation will be essential in historical atlas applications.
- Dates and date sequences are often vague and may be absent altogether; we may know he arrived somewhere in early spring of 633 and somewhere else in late fall, but not the dates of the journey to get between the two places. This pressed the design and implementation of the Period datatype (§7.1.6), to allow interval boundaries to be specified by intervals instead of instants.
- The digitized paths between locations represent routes that cannot be known with the precision a mapped polyline implies.

I would argue the Xuanzang data is an example of one benefit of digital humanities methods in pressing researchers to explicitly confront the degree of uncertainty in their data. A plausible counter-argument could be made about losing the forest for the trees (or angels on pin-heads, etc.); that is, do we really need computers to help us reason about such imprecisely described spatial and temporal boundaries?

The most obvious purpose for digitizing multiple accounts of Xuanzang's journey would be to spatially index text passages of first-hand accounts and interpretive

---

[74] Many other works about Xuanzang, aimed at both scholarly and popular audiences, are outside the present scope.

materials in a digital historical atlas. Such atlases would certainly encourage exploration, e.g. investigating what is known about the origins of Buddhism in what is now Burma (bypassed by Xuanzang); comparing Xuanzang's path with the travels of Gautama Buddha himself centuries earlier, or with the Silk Road as described by various scholars; learning what sites along the route are currently protected, and why; also whether they would be interesting to visit.

As to whether there are analytical questions such digitized paths might help answer, with the caveat that I am not a historian, these few come to mind:

- How did elevation change vary between his outward and inward journeys across the Himalayas?
- Was there a plausible southerly route to India?

Leaving aside such truly extraordinary journeys, it would be interesting to have a very large set of life-paths of historical individuals, in whatever limited detail can be managed. In modern social science research, where data for ordinary individuals is increasingly available, very interesting research using time-geographic methods is on the rise (*cf.* §2.5.3).

7.2.6 UNESCO 2009 World Heritage List (WHS2009)

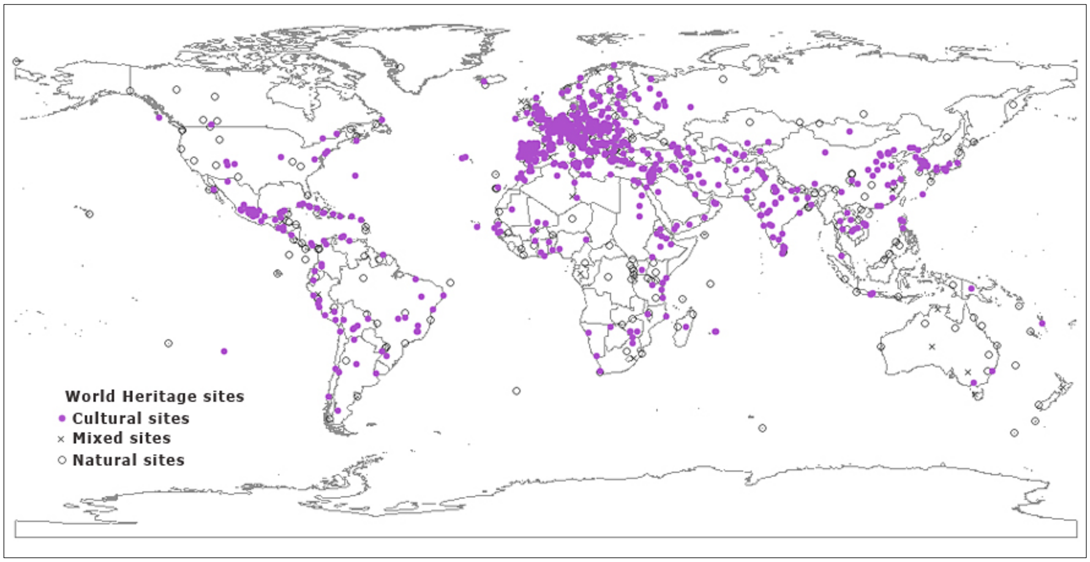


Figure 7-11. 2009 UNESCO World Heritage sites

As of 2009, 890 places worldwide had been designated as World Heritage sites by UNESCO[75]. Data about these, including locations and designation criteria, have been added to DRUMLIN to illustrate their potential as a contextualizing framework for the *Cultural Heritage Web* digital historical atlas project discussed in §3.3. Of the 890 sites, 714 are designated as either 'cultural' or 'mixed cultural/natural' (the remaining 176 are 'natural'). A dataset containing 50 events for each, well-described according to the data model presented in this work—totaling 35,700—would be quite a good start for a digital historical atlas, and eminently doable by a global community of interest.

---

[75] United Nations Educational , Scientific and Cultural Organization

7.3 Discussion of physical implementation

The DRUMLIN database is a work-in-progress. It holds five event-centered datasets constituting a range of activity types and mapping geometry requirements. The database schema implements the SHO ontology incompletely, and some elements, like check constraints for domain and range are not yet written. That said, as a proof-of-concept it does successfully demonstrate that the underlying ontology and physical data model can support digital historical atlases for a number of varied themes, including: social protest movements, military activity, expeditions and pilgrimages, global trade or electoral politics. Further discussion of results appears in Chapter 8.

8 Conclusions

This work has taken steps towards a general model for representing knowledge about historical human activity. It has produced an extended GIS data model based on a novel spatial history ontology (SHO), to support both an emerging genre of digital historical atlases and individual analytical applications. The integration of spatial, temporal and thematic perspectives on dynamic geographic phenomena is achieved by modeling the semantic structure of a spectrum of information about occurrences—from raw observational data (*what* or *who*, *where*, *when* and *how much*) to complex knowledge constructs asserting historical processes as causal and telic theories (the *why*). Record-level documentation of sources—an essential element of historical scholarship—is integral to the model.

8.1 Results

In Chapter 1 these questions were posed: (i) "What may digital historical atlases be?" (ii) "What are their representational requirements?" and (iii) "What extensions to existing GIS data models will effectively meet those requirements?" In this dissertation I have described the research and engineering tasks undertaken to answer those questions, and the results achieved. In the course of this work I have:

1. Developed a theoretical framework for the digital historical atlas genre and described a development scenario in the context of two visionary systems—the digital geolibrary and Digital Earth—and an actual project in development, Cultural Heritage Web.

2. Enumerated the entities to be represented in such systems, including physical and conceptual things (endurants) and happenings (perdurants), and developed a set of competency questions to assess the effectiveness of proposed representational models.
3. Identified and defined the elements of important *geo-historical information constructs*.
4. Developed a logical model to represent those constructs, subsumed as extensions and modifications (the SHO) to an existing upper ontology in widespread use, DOLCE.
5. Built an exemplar database whose schema implements the SHO logic insofar as possible.
6. Developed parameterized query functions and materialized views to answer the competency questions and demonstrate the effectiveness of the SHO and its implementation.

8.2 Discussion

The theoretical framework for the digital historical atlas genre proposed here will ultimately be judged by the community of practice, evidenced by whether future atlases built on event-centered data models along these lines are developed, and whether any of those have successful outcomes.

A few omissions and intentional shortcomings in this work should be mentioned here, with some explanation:

1. The expression of the description logic of the SHO in the DRUMLIN database schema is incomplete, as discussed in §7.1.1. The immediate goals in this implementation step were pragmatic, amounting to doing what can be managed in order to meet application requirements—as opposed to achieving a complete logical mapping. Whether to complete such a mapping should depend upon whether further requirements are revealed as an actual atlas is built. What has been accomplished here is sufficient to get started with that project. It appears likely that the Spatial History Ontology can be made considerably simpler, and many constructs in the current and most complete version of DOLCE (DLP397) can and should be abandoned.
2. The OWL-DL version of SHO has not been verified as sound and complete with reasoning software such as Pellet, Fact++, or Jena. In the course of fitting new classes and relations to DOLCE, some existing classes considered extraneous were removed or ignored. For example, (i) DOLCE's Process and Figure in the Descriptions and Situations (DnS) extension have been ignored; (ii) in DOLCE the class Non-physical Object is the lone sub-class of Non-physical Endurant and in turn has a single sub-class, Social Object; Non-physical Object was removed. Such excisions are likely to result in inconsistencies, however for the reasons cited in item 1 above, repairs at this interim stage are not judged to be productive.
3. This work has not examined material artifacts, and no relevant taxonomies appear in the Type framework, as has been done for Events, Activity and

Roles for example. Such vocabularies (built environment, implements, etc.) will be essential for digital historical atlases going forward.

Two aspects of the novel capabilities provided by the SHO and resulting relational data model are noted in the context of the Election 2008 dataset. First, occurrences modelled in this way make a powerful organizing framework for *faceted search* and navigation in encyclopaedic information systems amounting to a potentially endless graph. Beginning with any node (an event, person, place, etc.) one may navigate in turn to any related node ad infinitum. Persons participate in multiple events performing activities, in various roles. Places are settings for multiple occurrences. An initial query—say, for a given politician's speeches (products of speech-giving events)—can return the locations of each, the persons and/or groups present and their attributes, as well as aggregate or comparative measures of the speeches' semantic content. Each of those entities may be browsed or queried directly, returning for example the characteristics of attendees, their own activities of any type at that location or elsewhere, speeches with similar content, or the demographics of places involved.

Secondly, this modeling approach enables representing assertions of purpose-driven causation—as well as their documented factual bases. This has been demonstrated in two software applications supported by the DRUMLIN database at an early phase of this work. The *IssueBrowser* project (Grossner, 2009) had two realizations[76]. The first was an art installation displaying a force-directed graph visualization depicting political distance between Presidential candidates and campaign issues as screen distances between nodes (Figure 8-1).

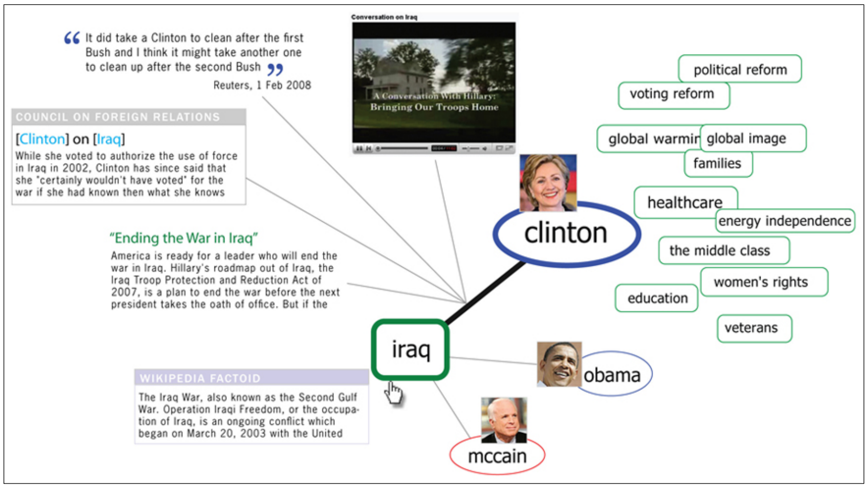


Figure 8-1. IssueBrowser: candidate distance graph visualization

Distances were a function of a semantic similarity measure, handled in the database as discussed in §7.2.2. The second was an interactive web-based mapping and graphing application illustrating variation in campaign issue focus by geographic region, political party and time period (Figure 8-2). Due to the data model, those values could be aggregated across space (US region), time, and theme (issue, party, or candidate). They could also be compared with corresponding values one might generate for other elections, other kinds of complex events or other spatiotemporal

---

[76]In the first case, a development graphic is shown; in the second, a screen shot from a running prototype. The functionality described and depicted for each was realized in prototype software that is not publicly available at this time. (Grossner, 2009; Ventura, et al., 2009) Additional material is available (Grossner, 2009b).

periods.

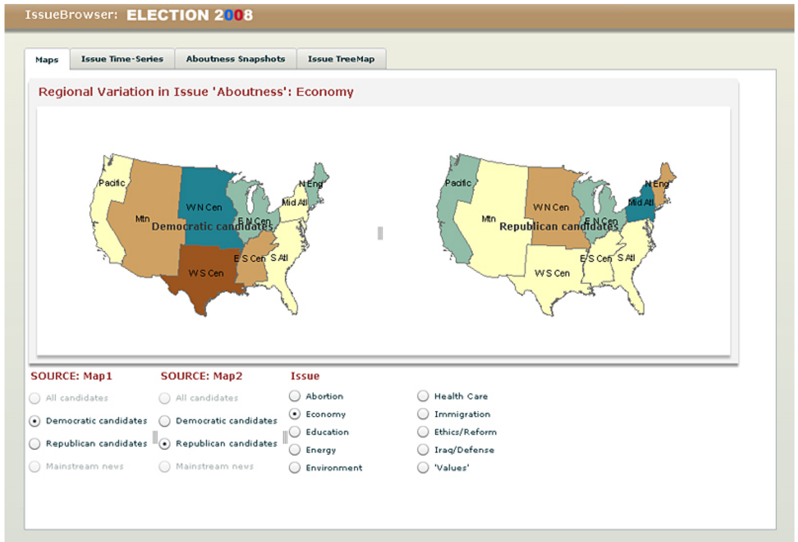


Figure 8-2. IssueBrowser: geographic issue variation

In another analysis mentioned in §7.2.2.1, the dynamic changes in candidates' issue space positions (Figure 7-6) are the *factual substrate* for assertions of causal and telic relationships with other occurrences, in this case both "internal" to the election cycle (polls, primary results, candidate withdrawals) and external to it (stock market trends or major overseas events).

Finally, it is worth noting that this research process of analysis, knowledge production and documentation is itself represented and attributed in the database as a Measurement event, with this author as Participant in the role of Analyst. The particular measurement, in this case, semantic similarity, is only one of an unlimited variety that could be applied to stored textual data. Results could be readily filtered by any property of the analysts and of any groups and organizations they belong to.

8.3 Future work

Several tracks of future research are suggested by the results of this work, concerning (i) event dimensions and classification, (ii) measurement, and (iii) flow and interaction.

8.3.1 Event dimensions and classification

In the SHO, complex event objects having spatial-temporal locations are generated by describing the participation or involvement of actors, witnesses, artifact products, etc. At this stage, formal event and activity subclass hierarchies have not been developed. Instead a system of extensible controlled vocabularies has been implemented. The Couclelis framework for ontologies of geographic information (2010; also §2.4.3) has not yet formalized the progression for temporal *geographic information constructs*. The possibilities discussed in §6.3.8.1 seem compatible with that project, and it would be interesting to pursue them. If an event is an information construct (and I think it is), what kinds are there, structurally speaking? Taken together, the characteristics of participants, the activity they perform, and aggregate spatial-temporal attributes, comprise a large set of event dimensions, relevant for a wide range of historical inquiries. These descriptors are the event object 'facets' used in the description and exploratory browsing functionality described in §8.1. We can also inventory, display and visually compare the life-paths of individuals.

Two related and entirely inductive approaches might be worth pursuing. Both are based on the notion of event dimensions:

8.3.1.1 Discovering dimensions

Given a collection of events for a project domain, we might wish to measure their similarity, classify them to facilitate analysis and consequent discovery of patterns and relationships, and to represent findings and interpretations. Extending the event-as-object metaphor, we can ask: What is it for? What is it made of? How is it made, i.e. what is its internal structure? What binds its parts, gives it unity?

Events might be classified along dimensions of

- composition (e.g. activity types, participant types);
- purpose and effect, as asserted or inferred from observed and measured change (e.g. of identity, position or possession); and
- spatiotemporal attributes, including locational patterns (e.g. distribution, density, diffusion), duration, periodicity and topology (e.g. adjacency, containment, overlap).

In a subsequent phase of this research I will examine a broader range of event datasets and develop a comprehensive set of "competency questions" we might ask of them, perhaps exposing additional dimensions. Given a general set of activity and event dimensions, it should be possible to create corresponding functions for measurement and set membership. These should enable the discovery, classification and analysis of spatiotemporal and thematic *structures* in historical event data. Such structures may be internal—patterns and clusters within complex composite events (e.g. a war)—or external, in the sense of historical *processes*, which are defined here as asserted meaningful sequences of events and their settings (e.g. a social movement). A working application to iteratively test, improve and demonstrate them can follow that. For many applications the measurement of similarity between events and processes along those dimensions will be important, perhaps as distance in *n*-dimensional conceptual spaces.

8.3.1.2 Metric conceptual spaces

In this dissertation I contribute a description model for the temporal portion of Berry's matrix but not its third leg—theme. I have characterized activity as *temporal substance*—that which discrete temporal entities such as events are composed of. It has been useful to analogize it to *material substance*—that which physical things are composed of. It seems to me the same formulation can be applied usefully to conceptual constructs, in the sense of DOLCE's Non-physical Endurants (NPED). In CIDOC-CRM these are termed Conceptual Objects. Such entities include information objects, descriptions, plans and theories. We might consider whether conceptualizations of physical, temporal and mental objects are *conceptual substance*—components of information objects stitched together in asserted relations.

An elaboration of the SHO along these lines would contribute to a *semantic reference system* per Kuhn (2003), able to reference entities within metric conceptual spaces for navigation and analysis, as suggested by Gärdenfors (2004), Raubal (2004) and Adams and Raubal (2009). DOLCE's Abstract Regions (*AR*) appear to be an appropriate ontological slot corresponding to metric conceptual spaces. Formalizations for conceptual spaces would permit a new, perhaps better approach to answering queries such as, 'where and when has anything *like this* happened?'

8.3.2 Measurement

Two issues regarding measurement addressed briefly in this work will be further pursued: probabilistic intervals and spatiotemporal boundaries.

- I have used a Period array datatype to define indeterminate time-periods bounded not by instants but by intervals. Those bounding intervals need to permit weighting by confidence and/or probability. Containment and intersection operations on such intervals should return probability values that can be symbolized in maps and timelines. I plan to seek collaboration with a computer scientist for a fuller implementation of Period [ ].
- Fuzziness aside, the bounds of complex events and processes can be computed as unions of their temporal and spatial locations. If these are represented as single footprints or hulls, we can readily calculate topological relations between them such as adjacency, overlap and containment. However, if spatial locations are non-contiguous, or there are temporal gaps, such calculation is expensive. It would be interesting to experiment with *n*-dimensional cubes over the union of an event's spatiotemporal locations to reduce computational cost.

It should also be mentioned that the PostgreSQL Date datatype is limited to the Gregorian calendar. Historical applications will require conversion and mapping capability between multiple calendars.

8.3.3 Interaction and flow

This dissertation research has made it clear that an ontology for spatial history must account for the practices of aggregation, abstraction and generalization routinely undertaken by researchers in their descriptions of temporal entities—and the SHO does, in part. A formalization of flows and trajectories should join those of events and historical processes proposed herein.

The trans-Atlantic slave trade data were modeled both as individual voyages having trajectory locations, and as aggregated Commercial Exchange events having flow (point-pair) locations. The concept of flow in this case points to some conceptual and terminological inconsistencies, and a gap in the models developed so far. It seems intuitively incorrect to call 300 voyages made under the British flag during a given year an "exchange event." The temporal bounds are arbitrary, albeit conventional. In fact a year's slave trade flow is a collection of voyage events having results that include the movement of humans as a commodity; each voyage is associated with multiple commercial transaction events having monetary values. The Voyages dataset does include detail on those transactions.

The *path*, *trajectory* and *flow* sub-classes of *place* introduced in this work are minimally defined and considered placeholders. I plan to continue development of trajectory and flow datatypes, to complement the choice of point, polyline and polygon primitives available to historical GIS modelers. Many historical phenomena one might represent in a digital historical atlas can be abstracted to directional paths. Some require time-stamped begin and end control points (e.g. expeditions and military maneuvers); others correspond to directional flows valid for an interval (e.g. trade, migration, cultural diffusion,). Many kinds of flow can be somewhat clumsily modeled in commercial GIS, but temporal aspects have been largely ignored.

Another important step is determining whether database references to *flow* activity and events can be usefully linked to the mathematical models created and manipulated in external software. One example of work that might benefit from such an inquiry is found in the ENFOLD-ing project[77] at University College London's Center for Advanced Spatial Analysis (CASA). Researchers there are developing models of global flows in trade, migration, development aid and security-related activity to explain and predict abrupt economic and political change events.

There will necessarily be several stages of refining the SHO and its derived databases, best undertaken collaboratively by a group of interested historians, geographical knowledge modelers and atlas authors.

[77] http://www.casa.ucl.ac.uk/

9 Appendices

9.1 Spatial History Ontology (SHO) classes and relations

An OWL-DL version of SHO that can be browsed in an ontology editor such as Protégé[78] will be maintained at http://www.geog.ucsb.edu/~grossner/sho_latest.owl. The upper-level classes and relations are listed here. SKOS-style taxonomies that specialize several upper categories appear in §9.2.

Table 9-1. SHO Perdurant, Quality and Abstract categories (new are underlined)

CODE	CLASS	PARENT
PT	particular	
PD	**PERDURANT**	PT
EV	**event**	PD
ACC	accomplishment	EV
ACT	action	ACC
ACH	achievement	EV
A	**activity**	PD
AGA	agentive-activity	A
NAGA	non-agentive-activity	A
ST	**state**	PD
PRO	**process**	PD
Q	**QUALITY**	PT
TQ	temporal-quality	Q
TL	temporal-location	TQ
PQ	physical-quality	Q
SL	spatial-location	PQ
AQ	abstract-quality	Q
AB	**ABSTRACT**	PT
R	**region**	AB
AR	abstract-region	R
PR	physical-region	R
S	space-region	PR
TR	temporal-region	R
T	time-interval	TR
SET	set	AB
PROP	proposition	AB

[78] http://protege.stanford.edu/

Table 9-2. SHO Endurant categories (new are underlined; * = not yet defined)

CODE	CLASS	CODE	CLASS
PT	particular		*non-agentive-social-object, contd.*
ED	**ENDURANT**	CPT	**concept**
PED	**physical-endurant**	TYP	type
M	**amount-of-matter**	COU	course
FEAT	**feature**	FR	role
EFEA	earth-feature	SROL	social-role
PO	**physical-object**	CROL	causal-role
APO	**agentive-physical-object**	QROL	qualitative-role
RPO	rational-physical-object	AROL	agent-driven-role
PER	person	PAR	parameter
PGRP	group-of-persons	D	**description**
NAPO	**non-agentive-physical-object**	DCON	constitutive-description
MART	material-artifact	HPRO	historical-process
PBOD	physical-body	HPRD	historical-period
BOBJ	biological-object	HEV	historical-event *
PER	*person*	PL	place *
PLPH	physical-place	TH	theory
GOBJ	geographical-object	METH	method
EFEA	*earth-feature*	PLAN	plan
MART	*material-artifact*	NARR	narrative
PPLU	**physical-plurality**	SUBJ	subject
PGRP	*group-of-persons*	ISYS	inf-encoding-system
NPED	**non-physical-endurant**	CSYS	classification-system
SOB	**social-object**	IOBJ	**information-object**
ASO	**agentive-social-object**	DOC	document
ORG	organization	ADOC	authority-document
SPER	social-person	LING	linguistic-object
NASO	**non-agentive-social-object**	TXT	text
PLNP	**non-physical-place**	APPL	appellation
PLG	geographical-place	APAC	actor-appelation
PLPG	political-geog.-place	APPL	place-appellation
RGN	geographic-region	APTI	time-appellation
PLIM	imaginary-place	ID	identifier
COLN	**collection**	VIS	visual-item
COLL	collective	S	**situation**
FGRP	functional-group		
TCOL	type-collection		

Table 9-3. SHO relations (additions to DOLCE are underlined; inverses not listed)

REL_ID	REL_NAME	DOMAIN	RANGE
ref	**references**	NPO	PT
ref-a	about	IOBJ	PT
ref-i	identifies	APPL	PT
ref-d	<u>documents</u>	DOC	ED or PD or Q
ref-c	classifies	CPT	PT
play	**played-by**	FR	ED
seq	sequences	COU	PD
val	valued-by	PAR	R
ref-l	<u>lists</u>	ADOC	TYP
ref-r	<u>represents</u>	VIS	PT
ref-t	<u>has-type</u>	PT	TYP
infl	**<u>influenced</u>**	PD or ED	PD or HP
infl-m	<u>motivated</u>	PD or ED	PD or HP
infl-sp	<u>specific-purpose-of</u>	EV	PD or HP
infl-gp	<u>general-purpose-of</u>	TYP	PD or HP
infl-c	<u>caused</u>	PD or ED	PD or HP
infl-f	<u>facilitated</u>	PD or ED	PD or HP
infl-i	<u>initiated</u>	PD or ED	PD or HP
infl-p	<u>perpetuated</u>	PD or ED	PD or HP
infl-h	<u>hindered</u>	PD or ED	PD or HP
infl-t	<u>terminated</u>	PD or ED	PD or HP
infl-s	<u>setting-for</u>	HP	ST
gk	**generic-constituent**	PT	PT
memb	**member**	COLN	ED
rmemb	<u>r-member</u>	ORG or PGRP	PER
sk	**specific-constant-constituent**	PT	PT
p	**part**	PT	PT
pp	proper-part	PT	PT
pp-t	temporary-proper-part	ED	ED
p-t	temporary-part	ED	ED
pp-t	*temporary-proper-part*	ED	ED
partic	**participant**	PD	ED
perf	carried-out (performed)	ACT	RPO or SAG
prod	**product**	ACT	ED
loc-g	**generic-location**	PT	PT
loc-x	exact-location	PT	R

REL_ID	REL_NAME	DOMAIN	RANGE
loc-sd	d-spatial-location	NPED	S
loc-sp	p-spatial-location	PD	S
loc-p	physical-location	PED	S
loc-t	temporal-location	PD	TR
dur	duration	PD	T
loc-te	e-temporal location	ED	TR
loc-ap	approximate-location	ED or PD or Q	ED or PD or Q
loc-dpl	descriptive-place	ED	NPED
loc-ppl	participant-place	PD	ED
t-rel	**temporal-relation**	PD	PD
t-prec	precedes	PD	PD
t-rslt	**result**	AC	PD
t-coin	temporally-coincides	PD	PD
t-conn	temporally-connected	PD	PD
t-meet	meets	PD	PD
t-met	met-by	PD	PD
t-incl	temporally-includes	PD	PD
t-conc	concluded-by	PD	PD
t-start	started-by	PD	PD
t-olap	temporally-overlaps	PD	PD
gd	**generic-dependent**	PT	PT
sd	**specific-constant-dependent**	PT	PT

9.2 Type taxonomies in DRUMLIN

Table 9-4. Event types

TYPE_ID	EVENT TYPE NAME	NAMESPACE
EV	**Event**	sho
E7-1	**Political event**	sho
E7-1-1	Candidacy	elec08
E7-1-2	Election	elec08
E7-1-2-1	Primary Election	elec08
E7-1-2-2	Caucus	elec08
E7-1-2-3	Electoral Vote	elec08
E7-1-3	Electoral Cycle	elec08
E7-1-4	Inauguration	elec08

TYPE_ID	EVENT TYPE NAME	NAMESPACE
E7-1-5	Nomination Race	elec08
E7-1-6	Political Debate	elec08
E7-1-7	Political Gathering	elec08
E55-1-1	Campaign Speech	elec08
E55-1-2	Ceremony	elec08
E55-1-3	Webcast	elec08
E55-1-4	Town Hall Meeting	elec08
E55-1-6	Fundraiser	elec08
E55-1-8	Meeting	elec08
E55-1-9	House Party	elec08
E55-1-10	Rally	elec08
E55-1-11	Party Event	elec08
E55-1-12	Contentious Gathering	brit
E55-1-12-1	Poachers Vs. Gamekeepers	brit
E55-1-12-2	Smugglers Vs. Customs	brit
E55-1-12-3	Brawls In Drinking Places	brit
E55-1-12-4	Other Violent Gatherings	brit
E55-1-12-5	Attacks On Blacklegs	brit
E55-1-12-6	Market Conflicts	brit
E55-1-12-7	Other Unplanned Gatherings	brit
E55-1-12-8	Authorized Celebrations	brit
E55-1-12-9	Delegations	brit
E55-1-12-10	Parades, Demonstrations, Rallies	brit
E55-1-12-11	Strikes, Turnouts	brit
E55-1-12-12	Pre-Planned Meetings Of Named Associations	brit
E55-1-12-13	Pre-Planned Meetings Of Public Assemblies	brit
E55-1-12-14	Other Pre-Planned Meetings	brit
E7-1-8	Nomination Convention	elec08
E7-1-9	General election race	elec08
E7-1-10	Tenure	elec08
E7-1-11	Demonstration	elec08
E7-1-13	Boycott	elec08
E7-1-14	Enactment	elec08
E7-1-15	Judicial Decision	elec08
E7-1-16	Political Speech	elec08
E7-1-18	Political Movement	elec08
E7-2	**Military event**	sho
E7-2-2	Military Campaign	nap1812
E7-2-3	Battle	nap1812

TYPE_ID	EVENT TYPE NAME	NAMESPACE
E7-2-3-1	Ground Battle	nap1812
E7-2-4	Military Maneuver	nap1812
E7-2-4-1	Military Advance	nap1812
E7-2-4-2	Military Retreat	nap1812
E7-2-6	Military Occupation	nap1812
E7-2-7	Bivouac	nap1812
E7-11	**Journey**	sho
E7-11-2	Voyage	sho
E7-11-2-1	Trade Voyage	sho
E7-11-4	Journey Segment	sho
E7-11-5	Pilgrimage	sho
E7-3	**Crime**	sho
E7-3-1	Assassination	elec08
E7-3-2	Civil Disobedience	elec08
E7-3-3	Assault	elec08
E7-4	**Law enforcement event**	elec08
E7-4-2	**Guard**	elec08
E7-4-2-1	Escort	elec08
E7-5-4	**Commercial Event**	sho
E7-5-4-1	Sale (Purchase)	sho
E7-5-4	Barter	sho
E7-6	**Fire**	sho
E7-7	**Stay**	sho
E7-8	**Visit**	sho
E7-9	**Depart**	sho
E7-10	**Arrive**	sho

Table 9-5. Role types

TYPE_ID	ROLE TYPE NAME	NAMESPACE
FR	***Functional role***	***sho***
E55-2-2	Leader	sho
E55-2-5	Executive	sho
E55-2-5-1	Chief Executive	sho
E55-2-10	Candidate	sho
E55-2-19	Debater	sho
E55-2-20	Observer	sho
E55-2-21	Speaker	sho

TYPE_ID	ROLE TYPE NAME	NAMESPACE
E55-2-22	Guest	sho
E55-2-23	Oath Administrator	sho
E55-2-24	Principle	sho
E55-2-25	Victim	sho
E55-2-26	Perpetrator	sho
E55-2-27	Signatory	sho
E55-2-28	Commander	sho
E55-2-29	Combatant	sho
E55-2-30	Sub-group	sho
E55-2-31	Captain	sho
E55-2-32	Captive	sho

Table 9-6. Organization and group types

TYPE_ID	ORG/GRP TYPE NAME	NAMESPACE
ORG	***Organization***	*sho*
E40-1	**Advocacy organization**	sho
E40-1-1	Political Campaign	elec08
E40-1-2	Political Party	elec08
E40-2	**Commercial Business**	elec08
E40-2-1	**News Organization**	elec08
E40-2-1-1	Newspaper	elec08
E40-2-1-4	News Agency	elec08
E40-2-1-5	Political Blog	elec08
E40-2-1-5-1	Liberal Blog	elec08
E40-2-1-5-2	Conservative Blog	elec08
E40-2-1-7	News Magazine	elec08
E40-2-2	Publisher	elec08
E40-3	**Educational Institution**	sho
E40-3-2	University	sho
E40-5	**Government Agency**	sho
E40-7	**Inter-governmental Organization**	sho
E40-8	**Religious Organization**	sho
E40-9	**Research Organization**	sho
E40-10	**Charitable Organization**	sho
E40-11	**Open Source Software Project**	sho
E40-12	**Military Organization**	sho
E40-12-1	Army	sho

TYPE_ID	ORG/GRP TYPE NAME	NAMESPACE
E40-12-4	Military Unit	sho
E40-12-4-1	Infantry Unit	sho
E40-12-4-2	Cavalry Unit	sho
GRP	***Group***	*sho*
FGRP	***Functional group***	*sho*
E74-3	Political Office	sho
E74-4	Informal political affiliation	sho
E74-5	Contributing Authors	sho
E74-6	Expedition Party	sho

Table 9-7. Information object types

TYPE_ID	INF OBJECT TYPE NAME	NAMESPACE
DOC	***Document***	*sho*
E31-1	Transcript	sho
E31-2	Database	sho
E31-2-1	Spatial Database	sho
E31-2-1-1	Shapefile	sho
E31-3	Encyclopedia	sho
E31-4	Spreadsheet	sho
LING	***Linguistic Object***	*sho*
E55-3-1	Book	sho
E55-3-1-1	Atlas	sho
E55-3-2	Blog Entry	sho
E55-3-3	News Report	sho
E55-3-6	Speech Text	sho
E55-3-9	Essay	sho

Table 9-8. Activity types

TYPE_ID	ACTIVITY TYPE NAME	NAMESPACE
AGA	***Agentive Activity***	*sho*
E7-5	**Economic activity**	sho
E7-5-1	Agriculture	sho
E7-5-1-1	Crop Production	sho
E7-5-1-2	Animal Husbandry	sho
E7-5-1-3	Logging	sho

TYPE_ID	ACTIVITY TYPE NAME	NAMESPACE
E7-5-2	Manufacturing	sho
E7-5-3	Resource Extraction	sho
E7-5-3-1	Mining	sho
E7-5-3-2	Oil Extraction	sho
E7-5-3-3	Natural Gas Extraction	sho
E55-5	**Activity verbs**	**brit**
E55-5-1	Adjourn	brit
E55-5-2	Assemble	brit
E55-5-3	Attack	brit
E55-5-4	Block	brit
E55-5-5	Bracket	brit
E55-5-6	Celebrate	brit
E55-5-7	Cheer	brit
E55-5-8	Communicate	brit
E55-5-9	Control	brit
E55-5-10	Decry	brit
E55-5-11	Delegate	brit
E55-5-12	Deliberate	brit
E55-5-13	Demonstrate	brit
E55-5-14	Die	brit
E55-5-15	Dine	brit
E55-5-16	Donkey	brit
E55-5-17	End	brit
E55-5-18	Enter	brit
E55-5-19	Fight	brit
E55-5-20	Gather	brit
E55-5-21	Hear Petition	brit
E55-5-22	Hunt	brit
E55-5-23	March	brit
E55-5-24	Meet	brit
E55-5-25	Move	brit
E55-5-26	Negotiate	brit
E55-5-27	Observe	brit
E55-5-28	Oppose	brit
E55-5-29	Other	brit
E55-5-30	Petition	brit
E55-5-31	Proceed	brit
E55-5-32	Request	brit
E55-5-33	Resist	brit

TYPE_ID	ACTIVITY TYPE NAME	NAMESPACE
E55-5-34	Resolve	brit
E55-5-35	Smuggle	brit
E55-5-36	Support	brit
E55-5-37	Thank	brit
E55-5-38	Turnout	brit
E55-5-39	Vote	brit
E55-5-41	Attempt	brit
E55-5-42	Chair	brit
E55-5-43	Disperse	brit
E55-5-44	Receive	brit
E55-5-45	Address	brit
E55-5-46	Give	brit
E55-5-47	Transport	brit
E55-5-48	Travel	brit

9.3 Code for database functions

9.3.1 f_occurnear(integer, double precision, integer)

```
-- everything that's happened within x dec. deg. of place
CREATE OR REPLACE FUNCTION f_occurnear(IN placeid integer, IN buffer
double precision, IN world integer DEFAULT 0)
  RETURNS TABLE(place_id integer, occur_id integer, class_id text,
   occur_name text, period period[], placename text,
   geom_point geometry) AS
$BODY$
SELECT el.place_id, e.occur_id, e.class_id, e.occur_name, e.period,
p.placename, p.geom_point
  FROM evloc el join e on el.occur_id=e.occur_id join place p on
el.place_id=p.place_id
 WHERE ST_Contains(
    ST_Buffer((select geom_point from place p where
p.place_id=$1),$2::double precision),p.geom_point) -- e.g. 163782
(Chicago); 258454 (Denver)
    and el.dataset_id=$3 order by class_id, first(period[1])
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.2 *f_concurrent(integer)*

```
-- all events concurrent with (occurred during) a given event
CREATE OR REPLACE FUNCTION f_concurrent(IN event integer)
  RETURNS TABLE(occur_id integer, occur_name text, where text,
   when period[], geom_point geometry) AS
$BODY$
SELECT e.occur_id, e.occur_name, p.placename || coalesce(', '
||p.admin1, '') AS "where", e.period AS "when", p.geom_point
FROM e
JOIN evloc el on e.occur_id=el.occur_id
JOIN place p on el.place_id=p.place_id
WHERE contains(period,(select period[1] FROM e WHERE occur_id=$1))
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.3 *f_particip(integer)*

```
-- all participants in an event and any sub-events
CREATE OR REPLACE FUNCTION f_particip(IN occur_id integer)
  RETURNS TABLE(actor_id integer, participant text) AS
$BODY$
WITH eventparts(occur_id,occur_name) AS (select occur_id, occur_name
from f_subevents($1))
SELECT distinct on (p.actor_subj) p.actor_subj AS actor_id,
a1.prefname AS subject
FROM particip p
    JOIN e on p.occur_id=e.occur_id
    LEFT JOIN actor a1 ON p.actor_subj=a1.actor_id
    LEFT JOIN actor a2 ON p.actor_obj=a2.actor_id
WHERE p.occur_id IN(SELECT occur_id FROM eventparts)
UNION
SELECT distinct on (p.actor_obj) actor_obj, a2.prefname
FROM particip p
    JOIN e on p.occur_id=e.occur_id
    LEFT JOIN actor a1 ON p.actor_subj=a1.actor_id
    LEFT JOIN actor a2 ON p.actor_obj=a2.actor_id
WHERE p.occur_id IN(SELECT occur_id FROM eventparts)
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.4 *f_particip(integer, integer, integer)*

```
-- structure of an event according to a dataset/source combo;
-- subj/obj participants, roles, activity
-- e.g. select * from f_particip(20441,3,26009) returns the
assertions of Esposito, et al about Napoleon's advance
CREATE OR REPLACE FUNCTION f_particip(IN occur_id integer, IN world
integer, IN source integer)
```

```
  RETURNS TABLE(actor_id integer, participant text, relation text,
   occur_id integer, period period[], event text, role text,
   subj_class text, activity text, obj_class text, obj_actorid
integer,
   obj_name text, world integer, source integer) AS
$BODY$
WITH eventparts(occur_id,occur_name) AS (select occur_id, occur_name
from f_subevents($1))
SELECT p.actor_subj AS actor_id, a1.prefname AS subject, r.rel_name
AS relation, p.occur_id,
    e.period,e.occur_name, t1.type_name AS role, a1.class_id AS
subj_class, t2.type_name AS activity,
    a2.class_id AS obj_class, p.actor_obj, a2.prefname AS object,
p.dataset_id, p.source_id
FROM particip p
    JOIN e on p.occur_id=e.occur_id
    LEFT JOIN actor a1 ON p.actor_subj=a1.actor_id
    LEFT JOIN actor a2 ON p.actor_obj=a2.actor_id
    LEFT JOIN type t1 ON p.role_id=t1.type_id
    LEFT JOIN type t2 ON p.activity=t2.type_id
    LEFT JOIN relation r ON p.rel_id=r.rel_id
WHERE p.occur_id IN(SELECT occur_id FROM eventparts)
AND p.dataset_id = $2 AND p.source_id = $3
ORDER BY first(e.period[1])
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.5 f_activity(integer)

```
-- all events + activity performed by one actor;
-- calls f_membership() as subroutine
CREATE OR REPLACE FUNCTION f_activity(IN actor integer)
  RETURNS TABLE(occur_id integer, occur_name text, period period[],
   begun date, class_id text,event_role text, group_role text,
   activity text, actor_obj integer) AS
$BODY$
-- join id's of person/group and groups they've belonged to
WITH member(group_actorid,group_role) AS (SELECT group_id, role_id
FROM f_membership($1) f
    UNION SELECT $1 AS group_id, 'grp participant')
-- then get activity of all actors in that list
SELECT distinct(p.occur_id), e.occur_name, e.period,
first(period[1])::date AS begun, e.class_id,
        p.role_id AS event_role, m.group_role, p.activity,
p.actor_obj
    FROM member m, particip p
    JOIN e ON p.occur_id=e.occur_id
    WHERE p.actor_subj IN (SELECT group_actorid FROM member)
    AND m.group_role NOT LIKE 'grp %' -- filter avoids duplication
    ORDER BY begun
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.6 *f_actorpath(integer)*

```
-- a person’s life-path
CREATE OR REPLACE FUNCTION f_actorpath(IN actor_id integer)
  RETURNS TABLE(actor_id integer, occur_id integer, occur_name text,
   begun date, role text, place_id integer, placename text, admin1
   text, ccode text, geom_point geometry) AS
$BODY$
select $1::int AS actor, fa.occur_id, fa.occur_name, fa.begun,
t.type_name::text, el.place_id, p.placename::text,
    p.admin1::text,p.ccode::text,p.geom_point
from f_activ($1) fa
join evloc el on fa.occur_id=el.occur_id
join type t on fa.event_role=t.type_id
join place p on el.place_id=p.place_id
where p.featclass = 'P' -- cities and towns, not regions
order by begun
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;

-- a support function for f_actorpath
CREATE OR REPLACE FUNCTION f_activ(IN actor integer)
  RETURNS TABLE(occur_id integer, occur_name text, period period[],
   begun date, class_id text, event_role text, activity text,
   actor_obj integer) AS
$BODY$
SELECT distinct(p.occur_id), e.occur_name, e.period,
first(period[1])::date AS begun, e.class_id,
        p.role_id AS event_role, p.activity, p.actor_obj
    FROM particip p
    JOIN e ON p.occur_id=e.occur_id
    WHERE p.actor_subj = $1
    ORDER BY begun
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
ALTER FUNCTION f_activity(integer) OWNER TO postgres;
```

9.3.7 *f_subevents(integer)*

```
-- event-subevents with locations
CREATE OR REPLACE FUNCTION f_subevents(IN integer)
  RETURNS TABLE(occur_id integer, occur_name character varying,
   period period[], place_id integer, path_id integer) AS
$BODY$
WITH RECURSIVE
included_events(subj,occur_id,occur_name,period,place_id,path_id) AS
(
 SELECT subj, obj, e.occur_name::text, e.period, el.place_id,
el.path_id FROM relev r
    JOIN e on r.obj=e.occur_id
    LEFT JOIN evloc el on e.occur_id=el.occur_id
    WHERE subj = $1
```

```
 UNION ALL
    SELECT r.subj, r.obj, r.obj_label, ie.period, el.place_id,
el.path_id
    FROM included_events ie, relev r
    JOIN evloc el on r.obj=el.occur_id
    WHERE r.subj = ie.occur_id
  )
SELECT occur_id,occur_name,period,place_id,path_id
FROM included_events
ORDER BY place_id, path_id
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.8 f_subev(numeric)

```
-- parameterizes view v_relev; full breakdown of an event's
-- sub-events
CREATE OR REPLACE FUNCTION f_subev(numeric)
  RETURNS SETOF v_relev AS
$BODY$
    SELECT * FROM v_relev WHERE subj = $1 $BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.9 f_particip_a(integer)

```
-- simple list of activity performed by all participants in an
-- event and sub-events
CREATE OR REPLACE FUNCTION f_particip_a(IN occur_id integer)
  RETURNS TABLE(actor_id integer, participant text, activity text,
occur_id integer) AS
$BODY$
WITH eventparts(occur_id,occur_name) AS (select occur_id, occur_name
from f_subevents($1)) -- 20279 $1
SELECT p.actor_subj AS actor_id, a1.prefname AS subject,
t.type_name, p.occur_id
FROM particip p
    JOIN e on p.occur_id=e.occur_id
    LEFT JOIN actor a1 ON p.actor_subj=a1.actor_id
    LEFT JOIN actor a2 ON p.actor_obj=a2.actor_id
    LEFT JOIN type t ON p.activity=t.type_id
WHERE p.occur_id IN(SELECT occur_id FROM eventparts)
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.10 f_everhappened(text)

```
-- returns all instances of an event type and any sub-types of it
CREATE OR REPLACE FUNCTION f_everhappened(IN type_id text)
  RETURNS TABLE(when period[], event_type text, place text) AS
$BODY$
```

```
WITH incl_types AS (SELECT * FROM f_subtypes($1))
SELECT e.period, t.type_name, p.placename FROM e
    JOIN type t ON e.type_id=t.type_id
    LEFT JOIN evloc el on e.occur_id=el.occur_id
    LEFT JOIN place p ON el.place_id = p.place_id
    WHERE e.type_id IN (select type_id from incl_types)
    ORDER BY first(period[1]) DESC
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.11 f_evmeasure(integer, character varying, character varying)

```
-- returns results of any analysis results of macro-event and sub-
event -- products e.g. semantic analysis for speech text: [select *
from
-- f_evmeasure(4000, 'semsig34', 'E55-3-6') =
-- all Dem speeches during 2008 election cycle
CREATE OR REPLACE FUNCTION f_evmeasure(IN event integer, IN measure
character varying, IN "type" character varying)
  RETURNS TABLE(occur_id integer, infobj_id integer, class_id text,
   type_id text, valarray numeric[], place_id integer, admin1 text)
AS
$BODY$
SELECT
ref.occur_id,m.infobj_id,i.class_id::text,i.type_id::text,m.valarray
,ev.place_id,p.admin1::text
  FROM refer ref
  JOIN m_infobj m on ref.infobj_id=m.infobj_id
  JOIN inf i ON m.infobj_id=i.infobj_id
  JOIN e on ref.occur_id=e.occur_id
  JOIN evloc ev on e.occur_id=ev.occur_id
  JOIN place p on ev.place_id=p.place_id
 WHERE ref.occur_id in (select occur_id from f_subevents($1)) --
3998, e.g.
   AND ref.rel_id = 'ref-d' AND m.measure=$2 AND i.type_id=$3
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.12 f_members_date(integer, date)

```
-- returns all members of a group/org on a given date;
-- e.g. select * from f_members_date(1019, '1862-01-01') =
-- US President on that date, Abe Lincoln
CREATE OR REPLACE FUNCTION f_members_date(IN grp_id integer, IN
valid_day date)
  RETURNS TABLE(member_id integer, name text, period period[],
   place_id integer, placename text) AS
$BODY$
SELECT m.member_id AS actor_id,a.prefname, m.period, p.birthplace,
pl.placename AS birthplace
FROM member m
JOIN actor a on m.member_id=a.actor_id
```

```
JOIN pers p on m.member_id=p.actor_id
LEFT JOIN place pl ON p.birthplace=pl.place_id
WHERE group_id=$1
AND contains(range(period),$2::timestamptz);
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.13 f_membership(integer)

```
-- returns groups a person has belonged to, when and in what role
CREATE OR REPLACE FUNCTION f_membership(IN integer)
  RETURNS TABLE(group_id integer, prefname character varying,
   role_id character varying, rolename character varying,
   period period[], notes text) AS
$BODY$
SELECT group_id, g.prefname, role_id, t.type_name AS rolename,
period, m.notes
FROM member m
JOIN grp g ON m.group_id=g.actor_id
JOIN type t ON m.role_id=t.type_id
WHERE member_id = $1
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.14 f_members(integer)

```
-- returns all members of a group/org;
-- e.g. select * from f_members(1019) is all US Presidents
CREATE OR REPLACE FUNCTION f_members(IN grp_id integer)
  RETURNS TABLE(member_id integer, name text, period period[],
place_id integer, placename text) AS
$BODY$
SELECT m.member_id AS actor_id,a.prefname, m.period, p.birthplace,
pl.placename AS birthplace
FROM member m
JOIN actor a on m.member_id=a.actor_id
JOIN pers p on m.member_id=p.actor_id
LEFT JOIN place pl ON p.birthplace=pl.place_id
WHERE group_id=$1
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.15 f_referenced(integer, double precision, integer);

```
-- Inf objects referencing event at or within n dec. deg of a place
CREATE OR REPLACE FUNCTION f_referenced(IN occurid integer, IN
buffer double precision, IN world integer DEFAULT 0)
  RETURNS TABLE(title text, author integer, event text, where text,
   when period[]) AS
$BODY$
SELECT i.title, i.auth_id AS auth, e.occur_name AS event,
```

```
    p.placename || coalesce(', ' ||p.admin1, '') AS where, e.period
AS when
    FROM refer r
    JOIN inf i ON r.infobj_id=i.infobj_id
    JOIN e ON r.occur_id=e.occur_id
    JOIN evloc el ON e.occur_id=el.occur_id
    JOIN place p ON el.place_id=p.place_id
   WHERE r.occur_id IN (
    SELECT occur_id FROM f_occurnear($1,$2,$3))
     ORDER BY first(period[1]) DESC
$BODY$
  LANGUAGE 'sql' VOLATILE   COST 100  ROWS 1000;
```

9.3.16 f_region(integer)

```
-- returns geometry of regions described as collections of places,
e.g. census regions, Appalachia
CREATE OR REPLACE FUNCTION f_region(IN integer)
  RETURNS TABLE(parent_id integer, place_id integer, placename text,
   admin1 text, the_geom geometry) AS
$BODY$
WITH RECURSIVE included_places(parent_id,
place_id,placename,admin1,the_geom) AS (
 SELECT r.parent_id, r.place_id, p.placename, p.admin1, p.geom_mpoly
FROM region r
    JOIN place p on r.place_id=p.place_id
    WHERE parent_id = $1
 UNION ALL
    SELECT r.parent_id,r.place_id, p.placename, p.admin1,
p.geom_mpoly
    FROM included_places ip, region r
    JOIN place p on r.place_id=p.place_id
    WHERE r.parent_id = ip.place_id
  )
SELECT parent_id,place_id,placename,admin1,the_geom
FROM included_places
ORDER BY placename
$BODY$
  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.3.17 f_subtypes(text)

```
--utility function called elsewhere; returns all subtypes
CREATE OR REPLACE FUNCTION f_subtypes(IN type_id text)
  RETURNS TABLE(parent_type text, type_id text, type_name text) AS
$BODY$
WITH RECURSIVE included_types (parent_type, type_id, type_name) AS (
 SELECT t.parent_type, t.type_id, t.type_name FROM type t
    WHERE t.parent_type = $1 --'E7-1-2'
 UNION ALL
    SELECT it.parent_type, it.type_id, t.type_name
```

```
    FROM included_types it, type t
    WHERE it.type_id = t.parent_type )
SELECT t.parent_type, t.type_id, t.type_name
FROM type t where type_id = $1
  UNION ALL
SELECT * FROM included_types
ORDER BY type_id
$BODY$  LANGUAGE 'sql' VOLATILE  COST 100  ROWS 1000;
```

9.4 Existing digital history projects

Digital historical projects

The following digital projects exemplify the considerable interest and activity in digital atlases, HGIS and historical mapping.

1. Historical and Cultural Atlases

    1.1. CultureSampo: Finnish Culture on the Semantic Web 2.0 is an ambitious project developed by the Semantic Computing Research Group at Helsinki University of Technology (2003-2010), calling itself "a national communal publishing conduit for both institutional memory organizations as well as private citizens." The system uses Semantic Web technologies to present linked information ~128,000 cultural artifacts and a great deal of contextualizing material about places and events. (http://www.kulttuurisampo.fi/?lang=en)

    1.2. Pleiades; Ancient World Mapping Center atthe University of North Carolina at Chapel Hill; the aim of this project is "to provide on-line access to all information about Greek and Roman geography assembled by the Classical Atlas Project for the Barrington Atlas of the Greek and Roman World." It will also "enable large-scale collaboration in order to maintain and diversify this dataset." (http://www.unc.edu/awmc/pleiades.html)

    1.3. Atlas of Switzerland; the 3rd edition of this high-quality digital atlas is available on DVD. It represents perhaps the finest example of "GIS in multimedia," the preferred approach to digital atlas development as described in (Craglia & Raper, 1995). Project directors (Sieber & Huber, 2007) have suggested a "Culture and History" edition is forthcoming. (http://www.atlasderschweiz.ch/atlas/en/)

    1.4. ECAI Silk Road and Iraq Atlases; examples of temporal web-GIS driven cultural and historical atlases; typically comprise political boundaries, point locations of historic sites with short articles, images and bibliographic references; well described in (Buckland & Lancaster, 2004). (http://www.ecai.org)

    1.5. TimeMap is a mapping application implemented by ECAI as an interactive web-based interface to a clearinghouse of geo-historical and cultural datasets. (http://ecaimaps.berkeley.edu/clearinghouse/; http://www.timemap.net)

1.6. Theban Mapping Project; "Since its inception in 1978, the Theban Mapping Project (TMP, now based at the AmericanUniversity in Cairo) has been working to prepare a comprehensive archaeological database of Thebes." This web presentation has exceptional production values. (http://www.thebanmappingproject.com)

1.7. Historical Atlas of the 20th Century; this web-based collection of GIS-generated static maps essentially replicates digitally what would be a fairly poor quality print historical atlas. To be fair, it was first published in 1998 and would have been novel at that time. Produced by librarian and amateur historian Matthew White; (http://users.erols.com/mwhite28/20centry.htm)

1.8. Periodical Historical Atlas of Europe "shows the evolution of this continent through a sequence of 21 historical maps, every map depicting the political situation at the end of each century." As above, a collection of GIS-generated static maps. (http://www.euratlas.com/summary.htm)

1.9. HyperHistory is a digital version of Andreas Nothinger's paper "World History Chart," a detailed timeline of world history. It is self-described as "an expanding scientific project presenting 3,000 years of world history with an interactive combination of synchronoptic lifelines, timelines, and maps." It is the most advanced digital atlas of world history I'm aware of. However, while events, people and places are represented, the relationships between them are not as well developed or formalized as they might be. Maps are static and not GIS-driven. (http://www.hyperhistory.com)

2. Historical GIS

2.1. Vision of Britain through Time is the web-based interface to the Great Britain Historical GIS Project based in the Department of Geography, University of Portsmouth, UK. (http://www.visionofbritain.org.uk)

2.2. The China Historical GIS project (CHGIS) is a large-scale effort of several research units at HarvardUniversity to produce "a documented database of places and administrative units or the period of Chinese history between 222 BCE and 1911 CE;"(http://www.fas.harvard.edu/~chgis/)

2.3. National Historical Geographic Information System (NHGIS) provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2000; (http://www.nhgis.org/)

2.4. AfricaMap is "at its core a digital base map of the continent, viewable dynamically at a range of scales, and composed of the best cartographic mapping available." It includes both contemporary maps and "scholarly maps focused on Africa in various historical periods."

(http://africamap.harvard.edu)

2.5. The German Historical GIS is focused on the myriad boundary changes in Germany from 1820 to 1914. (http://hgis-germany.de)

2.6. Chinese Civilization in Time and Space; *From the web site*: “This system consists of three major components: basic geospatial materials, WebGIS integrated application environment, and thematic information. The fundamental base maps are based on Dr. Tan's ‘The Historical Atlas of China.’ "The Historical Atlas of China" provides users with Chinese historical features, covering Chinese history over the past 2000 years, from the ancient time to Qing dynasty.” (http://ccts.sinica.edu.tw/index.php?lang=en)

3. Geolibraries

3.1. Perseus Digital Library; TuftsUniversity, Gregory Crane, Editor; “Perseus is an evolving digital library (DL), engineering interactions through time, space, and language. Our primary goal is to bring a wide range of source materials to as large an audience as possible.” The project has for several years been at the forefront of DL research efforts seeking to move beyond delivering objects to opening and processing them, delivering views on their contents. Perseus researchers have digitized the contents of a large number of classical and other source texts and have georeferenced and mapped the place names therein. (Crane, 2002). (http://www.perseus.tufts.edu)

3.2. Alexandria Digital Library (& gazetteer); provides access to 15,000+ maps, images and GIS datasets in the UC Santa Barbara Map and Image Library; produced the ADL Gazetteer Content Standard, most recent version: 3.2 (2004-02-26). (http://www.alexandria.ucsb.edu)

4. Museums and museum-like projects

4.1. Becoming Human: Paleoanthropology, Anthropology and Human Origins; not a museum, but what digital museums should aspire to; an interactive multimedia documentary and interpretive exhibit, remarkable for its scholarly content and high production values; produced by the Institute for Human Origins, Arizona State University; (http://www.becominghuman.org)

4.2. Smithsonian American History; timeline and other visual navigation of the museum’s collections and exhibits; no maps; (http://www.americaslibrary.gov)

4.3. National Palace Museum of Taipei; browse collections, eLearning multimedia exhibits—non-interactive presentations—with some hand-drawn maps; (http://www.npm.gov.tw/en/home.htm)

4.4. American Museum of Natural History; promotional previews of exhibits and collections; no maps; (http://www.amnh.org)

4.5. America's Story; the US Library of Congress; a single US Map indexes short historical narratives about events involving each of the 50 states; (http://www.americaslibrary.gov)

4.6. BritishMuseum; incredibly, although exhibits and collections are arranged geographically by country, continent or region, the museum's digital exhibits, e.g. for World Cultures have not a single map; (http://www.britishmuseum.org)

9.5 Semi-interval operations for Period [ ] datatype

```
/*
* Author: Karl Grossner; Scott Bailey
* License: BSD
* Purpose: Adds period type boolean functions corresponding to the
* semi-intervals of C. Freksa (1992). Temporal Reasoning Based on
* Semi-Intervals. Artificial Intelligence 54: 199-227
*/
-- ol  older (<, m, o, fi, di)
CREATE OR REPLACE FUNCTION ol(period, period)
RETURNS boolean AS
$$
    SELECT first($1) < first($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION ol(period, period)
IS 'True if p1 starts before p2';

-- hh  head to head with (si, =, s)
CREATE OR REPLACE FUNCTION hh(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT first($1) = first($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION hh(period, period)
IS 'True if p1 has the same start time as p2';

-- yo  younger (d, f, oi, mi, >)
CREATE OR REPLACE FUNCTION yo(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT first($1) > first($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION yo(period, period)
IS 'True if p1 starts after p2';

-- sb  survived by (<, m, o, s, d)
CREATE OR REPLACE FUNCTION sb(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT last($1) < last($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION sb(period, period)
IS 'True if p1 ends before p2 ends';

-- tt  tail to tail with (fi, =, f)
CREATE OR REPLACE FUNCTION tt(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT last($1) = last($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION tt(period, period)
```

```
IS 'True if p1 has same end as p2';

-- sv  survives (di, si, oi, mi, >)
CREATE OR REPLACE FUNCTION sv(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT last($1) > last($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION sv(period, period)
IS 'True if p1 ends after p2 ends';

-- pr  precedes (<, m)
CREATE OR REPLACE FUNCTION pr(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT last($1) <= first($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION pr(period, period)
IS 'True if p1 ends before or at p2 start';

-- bd  born before death of (<, m, o, fi, di, si, =, s, d, f, oi)
CREATE OR REPLACE FUNCTION bd(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT first($1) < last($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION bd(period, period)
IS 'True if p1 begins before p2 ends';

-- ct  (this is already in Scott Baily's chronos functions)
CREATE OR REPLACE FUNCTION ct(period, period)
RETURNS boolean AS
$$
    SELECT first($1) < last($2)
    AND last($1) > first($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;

-- db  died after birth of (o, fi, di, si, =, s, d, f, oi, mi, >)
CREATE OR REPLACE FUNCTION db(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT last($1) > first($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION db(period, period)
IS 'True if p1 ends after p2 starts';

-- sd  succeeds (mi, >)
CREATE OR REPLACE FUNCTION sd(p1 period, p2 period)
RETURNS boolean AS
$$
    SELECT first($1) >= last($2)
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION sd(period, period)
```

```
IS 'True if p1 starts after p2 ends';

-- ob  older and survived by (<, m, o)
CREATE OR REPLACE FUNCTION ob(period, period)
RETURNS boolean AS
$$
    SELECT first($1) < first($2)
    AND last($1) < last($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION ob(period, period)
IS 'True if p1 starts before p2 starts and ends before p2 ends';

-- oc  older contemporary of (o, fi, di)
CREATE OR REPLACE FUNCTION oc(period, period)
RETURNS boolean AS
$$
    SELECT first($1) < first($2)
    AND last($1) > first($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION oc(period, period)
IS 'True if p1 starts before p2 starts and ends after p2 starts';

-- sc  surviving contemporary of (di, si, oi)
CREATE OR REPLACE FUNCTION sc(period, period)
RETURNS boolean AS
$$
    SELECT first($1) < last($2)
    AND last($1) > last($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION sc(period, period)
IS 'True if p1 starts before p2 ends and ends after p2 ends';

-- bc  survived by contemporary of (o, s, d)
CREATE OR REPLACE FUNCTION bc(period, period)
RETURNS boolean AS
$$
    SELECT last($1) > first($2)
    AND last($1) < last($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION bc(period, period)
IS 'True if p1 ends after p2 starts and ends before p2 ends';

-- yc  younger contemporary of (d, f, oi)
CREATE OR REPLACE FUNCTION yc(period, period)
RETURNS boolean AS
$$
    SELECT first($1) > first($2)
    AND first($1) < last($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION yc(period, period)
IS 'True if p1 starts after p2 starts and before p2 ends';

-- ys  younger and survives (oi, mi, >)
```

```
CREATE OR REPLACE FUNCTION ys(period, period)
RETURNS boolean AS
$$
    SELECT first($1) > first($2)
    AND last($1) > last($2);
$$ LANGUAGE 'sql' IMMUTABLE STRICT COST 1;
COMMENT ON FUNCTION ys(period, period)
IS 'True if p1 starts after p2 starts and ends after p2 ends';
```

10 References cited

Activity (n.d.). In *Oxford English Dictionary*. Retrieved from http://dictionary.oed.com/cgi/entry/ 50002218

Adams, B. & Raubal, M. (2009). A Metric Conceptual Space Algebra. In K. Hornsby, G. Ligozat, C. Claramunt, & M. Denis (Eds.), *Spatial Information Theory - 9th International Conference, COSIT 2009, Aber Wrac'h, France*. Lecture Notes in Computer Science 5756, pp. 51-68, Springer, Berlin.

Agarwal, P. (2005). Ontological considerations in GIScience. *International Journal of Geographical Information Science*, 19(5): 501–536

Allemang, D. & Hendler, J. (2007). *Semantic web for the working ontologist*. Burlington, MA: Morgan Kaufmann

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26: 832-843

Allen, J. F. & Ferguson, G. (1994). Actions and events in interval temporal logic. *Journal of Logic and Computation*, 4(5): 531-571

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language: Towns, buildings, construction*. New York: Oxford University Press.

Arctur, D. & Zeiler, M. (2004). *Designing geodatabases*. Redlands, CA: ESRI Press

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy,* 9: 5-16.

Baker, A.R. (2003). *Geography and history: Bridging the divide.* Cambridge: Cambridge University Press

Barnard, W. (2001). *Conceptions of geography*. Matieland, South Africa: Centre Geographical Analysis, University of Stellenbosch.

Berners-Lee, T. (2006). *Linked Data*. Web article, retrieved 5 October 2010, from http://www.w3.org/DesignIssues/LinkedData.html

Berry, B. (1964) Approaches to regional analysis: A synthesis. *Annals of the Association of American Geographers* 54 (1): 2–11

Bittner, T. (2007). From top-level to domain ontologies: ecosystem classifications as a case study. In S. Winter, et al. (Eds.): *COSIT 2007, LNCS 4736*, pp. 61-77

Black, J. (1997). *Maps and history: Constructing images of the past*. New Haven: Yale University Press.

Bodenhammer, D. (2008). History and GIS: Implications for the discipline. In A. K. Knowles (Ed.) *Placing History: How maps, spatial data, and GIS are changing historical scholarship.* Redlands, CA: ESRI Press, pp. 219-234.

Bol P.K. (2007) *The China historical geographic information system (CHGIS): Choices faced, lessons learned*. Presented at Conference on Historical Maps and GIS, Nagoya University, August 23-24, 2007

Brachman R. J. & Levesque, H. J. (2004). *Knowledge representation and reasoning*. San Francisco: Elsevier

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

Brewster, C., Alani, H., Dasmahapatra, S. & Wilks, Y. (2004). Data driven ontology evaluation. *Proceedings of International Conference on Language Resources and Evaluation*, Lisbon, 2004.

Buckland, M. & Lancaster, L. (2004). Combining place, time and topic: The Electronic Cultural Atlas Initiative. *D-Lib Magazine*, (10):5, doi:10.1045

Carter, P. (1987). *The road to Botany Bay: An essay in spatial history*. London: Faber and Faber

Clarke, K. C. & Cloud, J. G. (2000). On the origins of analytical cartography. *Cartography and Geographic Information Science*, 27(3), pp.195-204.

Codd, E.F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), pp. 377-387

Cook, W. A. (1989). *Case grammar theory*. Washington, DC: Georgetown University Press.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In Frank, A., U., Campari, I. and Formentini, U. (eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Berlin: Springer-Verlag, 65–77.

Couclelis, H. Personal communication, 27 Jan 2009

Couclelis, H. (2010). Ontologies of geographic information. *International Journal of Geographical Information Science*, *iFirst*, pp. 1-25. Retrieved October 20, 2010 from http://www.informaworld.com/smpp/ftinterface~content=a928352549 ~fulltext=71324093

Craglia, M., & Raper, J. (1995). GIS and multimedia. *Environment and Planning B*, *22*(6), 634-636.

Craglia, M., Goodchild, M. F., Annoni, A., Camara, G., Gould, M. Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S. & Parsons, E. (2008). Next-Generation Digital Earth. *International Journal of Spatial Data Infrastructures Research*, (3):146-167

Crane, G. (2002). Cultural heritage digital libraries: Needs and components. In *Research and Advances Technology for Digital Technology : 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002. Proceedings* (pp. 51-60).

Crofts, N., Doerr, M., Gill, T., Stead, S., & Stiff M. (Eds.) (2008). *Definition of the CIDOC conceptual reference model*. Retrieved from http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.4_March2008_.pdf, 4 April 2007.

Cunfer G. (2008) Scaling the Dust Bowl. In Knowles, A. (Ed.) *Placing history: How maps, spatial data, and GIS are changing historical scholarship*. Redlands, CA: ESRI Press 95-121

Date, C. J. (2003). *Introduction to database systems*. Boston:Addison-Wesley

de Fezensac, M. (1970). *The Russian campaign, 1812*. (L. Kennett, Trans.) Athens, GA: University of Georgia Press. (Original work published 1849).

Doerr, M. & Iorizzo, D. (2008). The dream of a global knowledge network—A new approach. *ACM Journal on Computing and Cultural Heritage* 1(1*)*: 5.1-5.23.

DLP397 (2006). *DOLCE2.1-Lite-Plus, version 3.9, Build 3971*. Retrieved 22 Jan 2010 from Laboratory for Applied Ontology, Institute of Cognitive Science and Technology, http://www.loa-cnr.it/ontologies/DLP_397.owl

Drucker, J. & Nowviskie, B. (n.d.). *Temporal modeling: conceptualization and visualization of temporal relations for humanities scholarship*. Retrieved 10 October 2010 from: http://www3.iath.virginia.edu/time/reports

Elmasri, R., & Navathe, S. B., (2007). *Fundamentals of database systems*. Boston: Addison-Wesley

Eschenbach, C. (2001). On changes and diachronic identity of spatial socio-economic units. In A. Frank, J. Raper and J. Cheylan (Eds.) *Life and Motion of Socio-economic Units* (pp. 63-77).

Fonseca, F. (2006). The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology*, 58(6), pp. 786-793

Fonseca, F., Egenhofer, M.J., Agouris, P. & Câmara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3): 231-257

Ford, A., Clarke, K. C., & Raines, G. (2009). Modeling Settlement Patterns of the Late Classic Maya Civilization with Bayesian Methods and Geographic Information Systems. *Annals of the Association of American Geographers*, 99(3): 496-520

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.

Frank, A. (2003). Ontology for spatio-temporal databases. In M. Koubarakis (Ed.), *Spatiotemporal Databases: The Chorochronos Approach*. Lecture Notes in Computer Science 2520, pp. 9-77, Springer, Berlin.

Frank, A. (2007). Data quality ontology: An ontology for imperfect knowledge. In S. Winter et al. (Eds.): *COSIT 2007*, LNCS 4736, 406-420, Berlin: Springer-Verlag

Frank, A. & Raubal, M. (1999). Formal specifications of image schemata - A step to interoperability in geographic information systems. *Spatial Cognition and Computation* 1(1): 67-101.

Frank A., Raper, J. and Cheylan, J. (Eds.) (2001). *Life and motion of socio-economic units*. London: Taylor & Francis.

Franzosi, R. (2004). *From words to numbers: Narrative, data and social science.* Cambridge, UK: Cambridege University Press.

Franzosi, R. (2010). *Quantitative narrative analysis*. Los Angeles: Sage Publications.

Freksa, C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence, 54*: 199-227

Freundschuh, S.M. & Egenhofer, M.J. (1997). Human conceptions of spaces: Implications for geographic information systems. *Transactions in GIS*, 2(4): 361-375.

Frye, C. (2008). *The ArcGIS citation data model and method*. Conference presentation, 2008 Social Science History Association, Miami, FL, 25 Oct 2008.

Gaddis, J. L. (2002). *The Landscapes of history: How historians map the past*. New York: Oxford University Press.

Gahegan, M. & Pike, W. (2006). A situated knowledge representation of geographical information. *Transactions in GIS*, 10(5): 727-749

Galton, A. (1992). Logic as a formal method. *The Computer Journal*, 35(5). Cambridge, UK: Cambridge University Press, pp. 431-440

Galton, A. (2003). Desiderata for a spatio-temporal geo-ontology. In W. Kuhn, M.F. Worboys and S. Timpf (Eds.): *COSIT 2003*, LNCS 2825, pp 1-12

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.

Galton, A. (2006). On what goes on: The ontology of processes and events. In B. Bennett & C. Fellbaum (Eds.), *Formal Ontology in Information Systems: Proceedings of the fourth international conference (FOIS2006)*, IOS Press, 2006, pages 4-11

Galton, A. (2009). *Processes and events in geographical space*. Unpublished manuscript supporting tutorial given at Conference on Spatial Information Theory (COSIT), Aber Wrac'h, France, 21 September, 2009.

Galton, A. (2010). How is a collection related to its members? In B. Smith, R. Mizoguchi, & S. Nakagawa (Eds.), *Interdisciplinary Ontology*, *Vol. 3, Proceedings of the Third Interdisciplinary Ontology Meetings*, February 27-28, 2010, Tokyo. Keio University, pp. 9-17.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. M. (1994). *Design patterns: Elements of reusable object-oriented software*. Boston: Addison-Wesley.

Gangemi, A. (2005). Ontology design patterns for the semantic web. In Y. Gil et al (Eds.) *The Semantic Web: International Semantic Web Conference 2005, LNCS 3729*, pp. 262-276.

Gangemi, A. & Mika, P. (2003). Understanding the semantic web through descriptions and situations. In R. Meersman et al. (Eds.): *CoopIS/DOA/ODBASE 2003, LNCS 2888*, pp. 689–706. Berlin:Springer-Verlag.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. & Schneider, L. (2002). Sweetening ontologies with DOLCE. *Lecture Notes in Artificial Intelligence*, 2473: 166-181.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Garland, K. J. & Noyes, J. M. (2004) CRT monitors: Do they interfere with learning? *Behaviour & Information Technology*, 23 (1), 43-52.

Gelernter, D. (1991). *Mirror worlds: Or, the day software puts the universe in a shoebox...how it will happen and what it will mean*. New York: Oxford University Press

Glennon, A. (2010). Creating and validating object-oriented geographic data models: Modeling flow within GIS. *Transactions in GIS*, 14(1), 23-42

Gnoli, C. (2008). Facets: A fruitful notion in many domains. *Axiomathes*, 18:127–130

Goffart W. (2003). *Historical atlases: The first three hundred years, 1570-1870*. Chicago: University of Chicago Press.

Goodchild, M. F. (1992). Geographic information science. *International Journal of Geographical Information Science*, 6(1): 31-45.

Goodchild, M.F. (1998). The geolibrary. In S. Carver (editor) *Innovations in GIS 5*. London: Taylor and Francis, pp. 59-68.

Goodchild, M.F. (2000). Cartographic futures on a digital earth. *Cartographic Perspectives* 36: 7-11

Goodchild, M. F. (2004). GIScience, geography, form and process. *Annals of the Association of American Geographers*. 94(4), pp.709-714.

Goodchild, M.F., Yuan, M. & Cova, T.J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science* 21(3), 239-260.

Goodchild, M. F. & Glennon, A. (2008). Representation and computation of geographic dynamics. In K.S. Hornsby and M. Yuan, editors, *Understanding Dynamics of Geographic Domains***.** Boca Raton: CRC Press, pp. 13–30. [448]

Gore, A. (1998). *The Digital Earth: Understanding our planet in the 21st century*. Retrieved fromhttp://www.digitalearth.gov/VP19980131.html, 18 February 2006.

Gregory, I. N. & Ell, P. S. (2007). *Historical GIS: Technologies, methodology and scholarship.* Cambridge, UK: Cambridge University Press

Grenon, P. (2003). BFO in a nutshell: A bi-categorical axiomatization for BFO and comparison with DOLCE. *IFOMIS Technical Report Series*, 06/2003.

Grenon, P. & Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition and Computation* 4(1): 69-104.

Grossner, K. (2009). *Modeling and measuring 'Campaign 2008' to support visualization and analysis*. Unpublished manuscript. Retrieved October 20, 2010 from http://www.geog.ucsb.edu/~grossner/assets/Grossner_MAST09.pdf

Grossner, K. (2009b). *The first spatial history of the 2008 US Presidential campaign: Modeling and measuring election discourse*. (Invited talk and slides presented at ThinkSpatial series, spatial@ucsb). Retrieved November 6, 2010 from http://spatial.ucsb.edu/events/brownbags/docs/2008-2009/Grossner-brownbag-presentation.pdf

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

Guarino, N. (1998). Formal ontology and information systems. In *Formal Ontology in Information Systems. Proceedings of the 1st International Conference*, edited by N. Guarino. Amsterdam: IOS Press. pp. 3-15.

Guarino, N. & Welty, C. (2002). Evaluating ontological decisions with OntoClean. *Comm. of the ACM*, 45(2):61–65.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association* 24: 7-21.

Hall, S., & Hornsby, K. (2005). Ordering events for dynamic geospatial domains. In A.G. Cohn and D.M. Mark (Eds.): *COSIT 2005*, LNCS 3693, pp. 330-346

Hartshorne, R. (1939). The character of regional geography. In Agnew, J., D.N. Livingstone, and A. Rogers, (Eds.). (1996). *Human geography: an essential anthology*. Cambridge, MA: Blackwell Publishers.

Heise, D. (1991) Event structure analysis: A qualitative model of quantitative research. In *Using Computers in Qualitative Research*, N. Fielding and R. Lee (Eds), pp. 136–163(Newbury Park, CA: Sage).

History (n.d.). In *Oxford English Dictionary*. Retrieved from http://dictionary.oed.com/cgi/entry/ 50106603

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

Hornsby, K. S. & Cole, S. (2007). Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS*, 11(4): 555-573.

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

Howarth, J. (2007). *Landscape and purpose: modeling the functional and spatial organization of the land*. Unpublished doctoral dissertation, University of California, Santa Barbara. Retrieved from ProQuest Dissertations and Theses, http://gateway.proquest.com/openurl?rft_dat=xri:pqdiss:3330472

Janowicz, K. (2006). Towards a similarity-based identity assumption service for historical places. In M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild, (Eds.), *GIScience, 4th International Conference* (Muenster, Germany, September 2006), LNCS 4197, 199-216, Berlin: Springer-Verlag

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

Janowicz, K., and Keßler, C. (2008). The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science* (10)22, 1129-1157.

Jensen, M. (2006). Visualizing complex semantic timelines. In *Proceedings of the ACH-ALLC, July 2006*, Paris.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination and reason*. Chicago: University of Chicago Press.

Kaupinnen, T., Paakkarinen, P., Mäkelä, E., Kuittinen, H., Väätäinen, J., & Hyvönen, E. (in press). Geospatio-temporal semantic web for cultural heritage. In M. Lytras, et al. (Eds). *Proceedings of Digital Culture and E-Tourism: Technologies, Applications and Management Approaches, 2010*, Hershey, PA US: IGI Global.

Kavouras, M. & Kokla, M. (2008). *Theories of geographic concepts*. Boca Raton, FL: CRC Press.

Klippel, A., Worboys, M., & Duckham, M. (2008). Identifying factors of geographic event conceptualization. *International Journal of Geographical Information Science*, 22(2), 183-204

Knowles, A. K. (Ed.) (2002) *Past time, past place: GIS for history*. Redlands, CA: ESRI Press.

Knowles, A. K. (2005) Emerging trends in historical GIS. *Historical Geography*, 33: 7-13

Knowles, A. K. (2008a). What could Lee see at Gettysburg? In A. K. Knowles, (Ed.) *Placing History: How maps, spatial data, and GIS are changing historical scholarship.* Redlands, CA: ESRI Press, pp. 235-266.

Knowles A. K. (Ed.) (2008b) *Placing History: How maps, spatial data, and GIS are changing historical scholarship.* Redlands, CA: ESRI Press.

Kowalski, R. & Sergot, M. (1986). A logic-based calculus of events. In I. Mani, J. Pustejovsky, and R. Gaizauskas (Eds.) (2005). *The Language of Time: A Reader*, Oxford: Oxford University Press

Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographic Information Science*, (15)7: 613-631.

Kuhn, W. (2003). Semantic reference systems. *International Journal of Geographic Information Science*, 17(5): 405-409

Kuhn, W. (2005). Geospatial semantics: Why, of what, and how? In S. Spaccapietra and E. Zimányi (Eds.): *Journal on Data Semantics III, LNCS 3534*, pp. 1 – 24, 2005.

Kuhn, W. (2007a). An image-schematic account of spatial categories. In S. Winter et al. (Eds.): *COSIT 2007*, LNCS 4736, pp. 152–168

Kuhn, W. (2007b). *Spatio-temporal primitives*. Dangermond Lecture, delivered at U.C. Santa Barbara, 12 April 2007.

Kurata, Y., & Egenhofer, M. J. (2005). Structure and semantics of arrow diagrams. In A.G. Cohn and D.M. Mark (Eds.): *COSIT 2005, LNCS 3693*, pp. 232 – 250

Lakoff, G. (1987). *Women, fire and dangerous Things*. Chicago: University of Chicago Press.

Lakoff, G. and M. Johnson (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langran, G. (1992). *Time in geographic information systems*. London: Taylor and Francis.

Larman, C. (2002). *Applying UML and patterns*. Upper Saddle River NJ: Prentice-Hall

Leibniz, G. W. (1996). *New essays on human understanding*. P. Remnant and J. Bennett, Eds. Cambridge, UK: Cambridge University Press. (Original work published 1764).

Lilley K.D., Lloyd, C., Trick, S. and Graham C. (2005) Analysing and mapping medieval urban forms using GPS and GIS. *Urban Morphology* 9:5-15.

Liu, Y., Gou, Q. H., Wieczorek, J., & Goodchild, M. (2009). Positioning localities based on spatial assertions. *International Journal of Geographical Information Science*, 23(11), pp. 1471-1501

Livingstone, D.N. (1992). *The geographical tradition: episodes in the history of a contested enterprise*. Oxford, UK: Blackwell Publishers.

Mark, D.M., Smith, B., Egenhofer, M.J. & Hirtle, S. (2005). Ontological foundations for geographic information science. In R.B. McMaster and E.I. Usery (Eds.). *A Research Agenda for Geographic Information Science* (pp. 335-350). Boca Raton, FL: CRC Press.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N. & Oltramari, A. (2003). *WonderWeb Deliverable D18 Ontology Library*. Trento, Italy: Laboratory for Applied Ontology.

Massey, D. (1997). A global sense of place. In T. Barnes and D. Gregory (Eds.) *Reading human geography: the poetics and politics of inquiry*. London: Arnold, 315-323.

Massey, D. (2005). *For space*. London: Sage

McCarty W. (2004). Modeling: A study in words and meanings. In S. Schreibman, R. Seimens, J. Unsworth (Eds.) *A companion to digital humanities*, pp. 254-270. Malden, MA: Blackwell

McMaster, R. B. & Usery, E. I. (Eds.). (2005). *A research agenda for geographic information science*. Boca Raton, FL: CRC Press.

Mennis, J. & Peuquet, D. (2003). The role of knowledge representation in geographic knowledge discovery: A case study. *Transactions in GIS*, 7(3): 371-391

Mennis, J. L., Peuquet, D. J.& Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science*, 14(6), 501–520.

Miller, H. (2004). A measurement theory for time geography. *Geographic Analysis*, (37): 17-45

Miller, H. (2007). Place-based versus people-based geographic information science. *Geography Compass*. 1/3: 503-535.

Mostern, R. (2008) Historical gazetteers: An experiential perspective, with examples from Chinese history. *Historical Methods*, 41.1

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

Mostern, R., Grossner, K., & Johnson, I. (2009) *Models, dating and relationships: "Eventifying" history and historicizing events*. Unpublished manuscript.

Mostern, R. & Meeks, E. (2010). *Digital Gazetteer of Song Dynasty China v.1.1* (Web database). Accessed 6 November, 2010.

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

National Research Council, (Ed.) (2006). *Learning to think spatially*. Washington, DC: National Academies Press.

Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In C. Welty and B. Smith (Eds.), *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, Ogunquit, Maine, October 17-19, 2001.

Novick, P. (1988). *That noble dream: The objectivity question and the American historical profession*. Cambridge, UK: Cambridge University Press

Peuquet, D. J. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3): 441-461

Peuquet, D. J. (2002). *Representations of space and time*. New York: The Guilford Press.

Plewe, B.S. (2002). The nature of uncertainty in historical geographic information. *Transactions in GIS*, 6(4): 431-456

Plewe, B. S. (2003). Representing datum-level uncertainty in historical GIS. *Cartography and Geographic Information Science*, 30(4): 319-334

Pred, A. (1977). The choreography of existence: Comments on Hägerstrand's time-geography and its usefulness. *Economic Geography*, 33(2): 207-221

Pred, A. (1981). Production, family and free-time projects: a time-geographic perspective on the individual and societal change in nineteenth-century U. S. cities. *Journal of Historical Geography*, 7(1): 3-36

Pred, A. (1984). Place as historically contingent process: Structuration and the time-geography of becoming places. *Annals of the Association of American Geographers*, 74(2): 279-297

Process (n.d.). In *Oxford English Dictionary*. Retrieved from http://dictionary.oed.com/cgi/entry/50189148

Pustejovsky, J., Ingria, R., Soser, R., Castaño, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G., & Mani, I. (2005). *The specification language TimeML*. In I. Mani, J. Pustejovsky, R. Gaizauskas (Eds.). *The language of time: A reader*, pp. 545-558. Oxford: Oxford University Press

Raubal, M. (2004). Formalizing conceptual spaces. In A. Varzi and L. Vieu (Eds.), *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004).* Frontiers in Artificial Intelligence and Applications 114, pp. 153-164, IOS Press, Amsterdam, NL.

Rey, S.J. & Anselin, L.. (2007) PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies*. 37(1): 5-27.

Rey, S.J. & Janikas, M.V. (2006) STARS: Space–time analysis of regional systems, *Geographical Analysis* 38: 67-84.

Robertson, B. (2005). *The historic event markup and linking project*. Retrieved 4 September 2008, from http://heml.mta.ca/heml-cocoon/

Ruotsalo, T., & Hyvönen, E. (2007). An event-based approach for semantic metadata interoperability. In Abere, K. et al. (Eds.), *The Semantic Web (ISWC 2007 Proceedings, LNCS 4825*), Berlin:Springer-Verlag, pp. 409-422

Sauer, C. (1925). The morphology of landscape, In J. Agnew, D.N. Livingstone, and A. Rogers, (Eds.). (1996). *Human geography: an essential anthology*. Cambridge, MA: Blackwell Publishers.

Sauer, C. (1941). Forward to historical geography. In J. Leighly (Ed.) (1965). *Land and Life*. Berkeley, CA: University of California Press.

Schrodt, P. A. (2006). *Twenty Years of the Kansas Event Data System Project*. Retrieved August 1, 2007, from http://www.ku.edu/~keds/KEDS.history.html

Seibt, J. (2004). Free process theory: Towards a typology of occurrings. *Axiomathes* (14): 23-55.

Semy, S. K., Pulvermacher, M. K., & Obrst, L. J. (2004). Toward the use of an upper ontology for U.S. government and U.S. military domains: An evaluation. *MITRE Technical Report MTR-04B0000063*. Bedford, MA: MITRE Corp.

Sewell, W. H. (2005). *Logics of history*. Chicago: University of Chicago Press.

Shakespeare, W. (1999). *The tempest*. (P. Holland, Ed.). New York: Penguin. (Original work published 1604).

Shaw, R., Troncy, R., & Hardman, L. (2009). LODE: Linking open descriptions of events. In A. G´omez-P´erez, Y. Yu, and Y. Ding (Eds.): *The Semantic Web, Asian Semantic Web Conference 2009*, LNCS 5926, pp. 153–167

Shaw, S., L. & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space. *Journal of Transport Geography*, (17): 141-149

Shaw, S., Bombom, L. S., & Yu, H. (2008). A Space-time GIS approach to exploring large individual-based spatiotemporal datasets. Transactions in GIS, 12(4): 425-441

Sieber, R. & Huber, S. (2007). Atlas of Switzerland 2 - A highly interactive thematic national atlas, in *Multimedia Cartography, 2nd ed.* (eds. W. Cartwright, M. Peterson, G. Gartner), Springer: Berlin, 2007, pp. 161-182. Retrieved from http://www.atlasderschweiz.ch/pdf_publi/MC07_SieberHuber.pdf, 2 October, 2007.

Siebert, L. (2000) Using GIS to document, visualize and interpret Tokyo's spatial history. *Social Science History*, 24(3):537-574

Simons, P. (2000). *Parts: A study in ontology*. Oxford: Oxford University Press.

Simsion, G. C. & Witt, G. C. (2005). *Data modeling essentials*. San Francisco: Morgan Kaufmann

Skupin, A., & Hagelman, R. (2003). Attribute space visualization of demographic change. *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems, New Orleans, Louisiana, November 7-8*. New York: ACM Press. 56-62.

Smith, B. (2001). Fiat objects. *Topoi* (20**):** 131–148

Smith, B. & Grenon, P. (2004). A cornucopia of formal-ontological relations. *Dialectica* Vol. 58, No 3 , pp. 279-296

Smith, B. & Mark, D.M. (2001). Geographical categories: an ontological investigation. *International Journal of Geographic Information Science*, 15:7, pp. 591-612.

Smith, M.L. (2005). Networks, territories, and the cartography of ancient states. *Annals of the Association of American Geographers*. 95(4), pp.832-849.

Snodgrass, R. T. (1992). Temporal databases. In A. U. Frank, I. Campari, and U. Formentini (Eds.), *Theories and methods of spatio-temporal reasoning in geographic space*. Lecture Notes in Computer Science 639, pp. 22-64, Heidelberg-Berlin: Springer-Verlag.

Snodgrass, R. T. (2000). *Developing time-oriented database applications in SQL*. San Francisco: Morgan Kaufmann

Southall, H. (2007) *A place for everything, and everything in its place: The architecture behind Vision of Britain*. Presented at ECAI/PNC Joint Meetings: Area Studies, Then and Now, Berkeley, CA 18 October 2007.

Southall, H., & White, B. (n.d.). Mapping the life course: Visualising migrations, transitions and trajectories. Unpublished manuscript.

Sowa, J.F. (2000). *Knowledge representation: Logical, philosophical and computational foundations*. Pacific Grove, CA: Brooks/Cole.

Staley, D. J. (2003). *Computers, visualization, and history: How new technology will transform our understanding of the past*. Armonk, N.Y.: M. E. Sharpe

Stuff (n.d.). In *Oxford English Dictionary*. Retrieved from http://dictionary.oed.com/cgi/entry/ 50239988

Taylor, H. A. (2005). Mapping the understanding of understanding maps. In Shah, P. and A. Miyake (Eds.). *The Cambridge handbook of visuospatial thinking*. Cambridge, UK: Cambridge University Press

Thorpe, W. (2010). *A mandala and a monk*. [Data file].

Tilly, C. (1995). *Popular contention in Great Britain, 1758-1934*. Cambridge MA: Harvard University Press

Tilly, C. & Schweitzer, R. A. (1980). *Enumerating and coding contentious gatherings in nineteenth-century Great Britain*. Center for Research on Social Organization (CRSO), Working paper #210. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/50984, October 21, 2009.

Tuan, Y. F. (1977). *Space and place: The perspective of experience.* Minneapolis: University of Minnesota Press

Uschold, M. and M. Gruninger (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review* 11(2): 93-136.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, 66(2): 143-160.

Ventura, J., Forbes, A., Adams, B., Joshi, S., Grossner, K., Bulger, M., Höllerer, T., Manjunath, B. J. (2009). *IssueBrowser: Knowledge acquisition via multimedia data*. *Proceedings of The Future of Interactive Media: Workshop on Media Arts, Science, and Technology (MAST)*. Jan 29-30, 2009, Santa Barbara, CA. Retrieved November 6, 2010 from http://mast.mat.ucsb.edu/docs/paper_13.pdf

Vision of Britain (2005). *A vision of Britain through time* (www.visionofbritain.org.uk) Retrieved 8 March 2008.

Voyages (2008). *Voyages: The trans-Atlantic slave trade database*. [Data file]. Retrieved September 4, 2010 from http://slavevoyages.org/downloads/tastdb-2008.dbf

Wada, T. (2002). *BRIT in Access, Version 1.0.* [Data file]. Unpublished, contributed by the author November 1, 2009.

Weaver, C., Fyle, D., Robinson, A., Holdsworth, D., Peuquet, D. & MacEachren, A.M. (2006). Visual analysis of historic hotel visitation patterns. *IEEE Transactions on Visualization and Computer Graphic*s, 12(5)

Westermann, U. & Jain, R. (2007). Toward a common event model for multimedia applications. IEEE Multimedia, (14)1: 19-29

Wood, Z. & Galton, A. (2008a). A new classification of collectives. In C. Eschenbach and M. Grüninger (Eds.), *Formal ontology in information systems*, IOS Press, 2008, pages 109-120.

Wood, Z. & Galton, A. (2008b). Collectives and how they move: A tale of two classifications. In B. Gottfried and H. Aghajan (eds.), *Proceedings of the 2nd Workshop on Behaviour Monitoring and Interpretation (BMI'08)* at the 31st Annual German Conference on Artificial Intelligence (KI2008), Kaiserslautern, Germany, September 23-26, 2008, pages 57-71.

Worboys, M. (2001). Modelling changes and events in dynamic spatial systems with reference to socio-economic units. In A. Frank, J. Raper and J. Cheylan (Eds.) *Life and motion of socio-economic units* (pp. 129-37). London: Taylor & Francis.

Worboys, M. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19(1): 1-28

Worboys, M. & Duckham, M. (2004). *GIS: A computing perspective (2nd Ed.)*. Boca Raton, FL: CRC Press

Worboys, M. & Hornsby, K. (2004). From objects to events: GEM, the geospatial event model. In Egenhofer, M. J., C. Freksa, H. J. Miller (Eds.) *Geographic Information Science: Third International Conference*, *LNCS*3234: 327–344.

Wriggins, S. H. (2004). *The Silk Road journey with Xuanzang*. Boulder, Colorado: Westview Press.

Yee, K., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. *CHI Letters*, (5)1: 401-408

Yuan, M. (1997). Use of knowledge acquisition to build wildfire representation in geographical information systems. *International Journal of Geographic Information Science*, 11(8): 723-745

Yuan, M. (1999). Use of a three-domain representation to enhance GIS support for complex spatiotemporal queries. *Transactions in GIS*, 3(2): 137-159

Yuan, M. (2001). Representing geographic information to support queries about life and motion of socio-economic units. In A. Frank, J. Raper and J. Cheylan

(Eds.) *Life and motion of socio-economic units* (pp. 217-234). London: Taylor & Francis.

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

10.1 Atlases Reviewed

Atlas Histórico-biográfico José Martí. (1983). La Habana: Instituto Cubano de Geodesia y Cartografía y Centro de Estudios Martianos.

Ardagh, J. (1991). *Cultural atlas of France*. New York: Facts on File.

Barraclough, G. (Ed.) (1993). *The Times atlas of world history*. London: Times Books.

Barnavi, E. (Ed.) (1992). *A historical atlas of the Jewish people: From the time of the patriarchs to the present*. New York: Random House.

Black, J. (Ed.) (2000). *Atlas of world history*: *Mapping the human journey*. DK Publishing, Inc.

Bradbury, M. (Ed.) (1995). *The atlas of literature*. London: De Agostini Editions.Carnes, M. C. (2003) *Historical Atlas of the United States*. New York: Routledge.

Carnes, M. C. (2003). *Historical atlas of the United States*. New York: Routledge.

Charlesworth, A. (1983). (Ed.). *An atlas of rural protest in Britain 1548-1900*. Philadelphia: University of Pennsylvania Press.

Esposito, V. J. and Elting, J. R. (1999). *A military history and atlas of the Napoleonic wars*. London: Greenhill Books.

Gilbert, M. (2002) *The Routledge atlas of Russian history* (3rd ed ed.) London: Routledge.

Goins, C. R. (2006) *Historical atlas of Oklahoma*. Norman, OK: University of Oklahoma Press.

Haywood, J. (1997). *Atlas of World History*. New York: Metro Books.

Moore, R. I. (Ed.) (1992). *Atlas of world history*. Chicago: Rand McNally.

McEvedy, C. (1992). *The new Penguin atlas of medieval history*. New York, USA: Penguin Books.

LaRosa, M., & Mejía, G. R. (2007). *An atlas and survey of Latin American history*: M. E. Sharpe.

Laboa, J. M. (2003). *The historical atlas of eastern and western Christian monasticism*. Collegeville, MN: Liturgical Press.

Man, J. (1999). *Atlas of the year 1000*. Cambridge, MA: Harvard University Press.

Neave, S., & Ellis, S. (Eds.) (1996). *An historical atlas of East Yorkshire*. Hull, UK: University of Hull Press.

O'Brien, P. K. (Ed.). (1999). *Atlas of world history* . New York: Oxford University Press

Pluvier, J. M. (1995). *Historical atlas of Southeast Asia*. London, UK: Brill.

CODA

Figure 10-1. *Landschaft bei E* (Landscape at E), Paul Klee, 1921