

# Modeling and Measuring ‘Campaign 2008’ to Support Visualization and Analysis

Karl Grossner  
Department of Geography  
University of California, Santa Barbara  
karlg@geog.ucsb.edu

## ABSTRACT

The two-year long complex social event that is a US Presidential election cycle is modeled using a modified and extended version of the CIDOC Conceptual Reference Model (ISO 21127:2006). That high-level ontology was originally developed for cultural heritage information in digital applications for the domains of museums, libraries and archives. The data model is instantiated in a spatiotemporal database driving a novel interactive exhibit entitled *IssueBrowser 2008* and supporting geographic and political science analyses. It is argued the challenges of modeling this ongoing composite event are similar in many respects to numerous knowledge domains in historical social science research, and successful application here is presented as a first step in the development of an ontological framework for geohistorical computing.

## Keywords

Ontology, database, data model, GIS, historical GIS

## 1. INTRODUCTION

Complex social events like the 2008 US Presidential election cycle are dynamic geographic phenomena: they comprise human activities associated with particular locations on the earth surface, and location over time is integral to their analysis. As such, they represent a relatively unexamined domain for evaluating recent theoretical work in spatiotemporal data models for geographic information systems (GISs), as well as exploring new approaches to geo-historical knowledge representation. The term *event* is used broadly here to refer to an occurrence, *complex* in the sense of being composed of many parts (sub-events), and having significant if discontinuous spatial extension and numerous participants. For GIS, granting ‘first-class’ object status to temporal entities at all scales is one requisite for the “third stage in development of spatiotemporal information systems...a full-blooded treatment of change, in terms of events and actions”[1].

This election offers a data-rich exemplar for many kinds of complex events one might seek to represent at various levels of granularity in historical applications. The prototype data model discussed here is instantiated in a relational database using the open-source PostgreSQL software, and drives both a novel large-scale interactive public information exhibit entitled *IssueBrowser* and a web site. The initial database used a straightforward relational model to meet the requirements of the first visualization (Fig. 1), a force-directed “endless graph” of candidates, issues and media content about them. Candidates and issues are nodes; node size and edge distances are derived from automated analyses of text content, principally a measure of similarity to 34 issue-specific reference documents. User selection of a candidate-issue pair generates display of text and video content specific to that

pair. Query results persist across four large screens as results of new queries join previous ones, revealing contrasts and similarities. The new data model introduced here continues to support that exhibit as well as web visualizations emphasizing geographic variation in issue emphasis by candidates, news media and analysts.

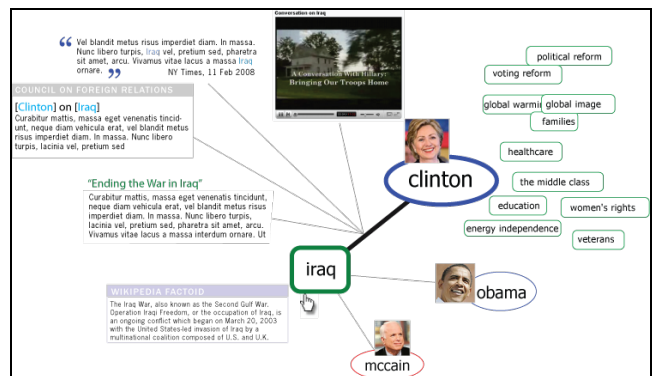


Figure 1. *Issue Browser* visualization (schematic), a navigable force-directed “endless graph”

## 2. EVENT-CENTRIC, ONTOLOGY-DRIVEN GIS DEVELOPMENT

A more general model is the goal of this work, one capable of representing complex geopolitical events and associated public discourse in *any* context, e.g. the US Civil Rights Movement or “The Troubles” of 20<sup>th</sup> Century Ireland. Furthermore, it will be general enough to also represent human activities as diverse as intellectual and technological diffusion, commercial flows, wars and the rise and demise of states. The natural organizing concept for such a model—at once atomic and unifying—is the *event*, broadly defined. To describe an event well means accounting for its participants (individuals and groups), its products (e.g. ideas and artifacts) and its locale in terms of place and/or location. One must also represent its temporal, spatial and thematic extents, and critically, place it in a chronology. Further, to assert purpose and causality for constructing narratives, events must be associated with one another not only chronologically but thematically and analytically. A relatively simple and highly usable formalization of these facets can form the basis for a new class of HGIS—and by extension, digital historical atlases. Such representations must begin from a high-level ontological framework, extensible and flexible enough to support domain ontologies of increasing detail beneath.

A promising candidate for seeding this development is the CIDOC Conceptual Reference Model (CIDOC-CRM) developed over a 10-year period by a working group of the International

Council of Museums and now ISO standard 21127:2006 [2]. A community of interest is working to align the CIDOC-CRM with various emergent interoperability standards, including the DOLCE ontology [3], which has received attention for its potential applicability in GIS [4]; [5].

Top-level ontologies like DOLCE and CIDOC-CRM differentiate at the highest level between *Persistent Items*, as things that *are* in time, and *Temporal Entities*, as things that *happen* in time. The CRM specifies 85 such entity classes (35 temporal, 50 persistent), and 134 *property declarations* that define inheritable relationships necessary and/or possible between those classes. By making *when* co-equal with *what* and *where*, a framework like CIDOC-CRM becomes a practical starting point for ontology-driven GIS development [6].

Once expressed and instantiated in a formal language, such as the RDF/OWL standard of the emerging Semantic Web<sup>1</sup>, a basic level of inferential reasoning is possible. Each property holds for a given *domain* and *range*. For example, the property *had Participant* holds for the domain of *Event*, with a range of *Actor*. The OWL language also permits value and cardinality constraints. Both entity classes and properties may be specialized, into sub-classes and sub-properties inheriting their parents' attributes and properties, and/or generalized, into super-classes and super-properties with the inverse relationship. For example, 27 of the 35 temporal entity classes are sub-classes of *activity*, i.e. agentive occurrences. In this election exemplar, a sub-class of *political activity* has been added, itself with eight subclasses (Table 2). This presumptive hierarchy of event types enables explorations of relatedness and similarity among and between all political activities (and by extension their myriad participant entities) represented in a given database, and potentially across multiple federated HGIS systems.

### 3. MODELING AN ELECTION

In the first stage of this process the entities to be modeled were identified (Table 1) and mapped to appropriate entity classes within the CRM, adding sub-classes as required (Table 2). The new entities were added to an RDF/OWL instantiation of the CIDOC-CRM<sup>2</sup>, along with enough instance data for a preliminary evaluation of the model's logical consistency, using the Protégé<sup>3</sup> ontology editor. Because it is impractical to maintain large-scale data stores in RDF format, the second stage entails instantiating the data model and inferencing capability in a relational database system.

The conceptual UML diagram in Figure 2 illustrates a schema sufficient to represent the election data according to the conceptual model of the extended CRM. Note that object-relational systems like PostgreSQL support table inheritance, which corresponds nicely to sub-classes in CRM. For example, *Person* and *Group* are sub-classes of *Actor*.

E2	TEMPORAL_ENTITY
E3	Condition State
E4	Period
E5	Event
E7	Activity
	<i>Political_Activity: [Election_Cycle, Political_Campaign, Political_Debate, Primary_Election, Caucus, Nominating_Convention, Election, Inauguration]</i>
	<i>Public Appearance</i>
	<i>Public Statement</i>
E63	Beginning of Existence
E66	Formation
E64	End of Existence
E68	Dissolution
E77	PERSISTENT_ITEM
E39	Actor
E74	Group
E40	Legal_Body
	<i>Advocacy_Organization: [Political_Campaign, Political_Party]</i>
	<i>Commercial_Enterprise: [News_organization]</i>
	<i>Educational_Institution</i>
	<i>Government: [Country_government, First-level_admin]</i>
	<i>Inter-governmental_Organization</i>
	<i>Non-governmental_Organization</i>
	<i>Religious_Organization</i>
	<i>Research_Organization</i>
	<i>Ethnic_Group</i>
E21	Person
E70	Thing
E71	Man-Made Thing
E24	Physical_Man-Made_Thing
E28	Conceptual Object
	<i>Issue</i>
	<i>Role</i>
	<i>Occupation</i>
	<i>Religion</i>
E73	Information_Object
E33	Linguistic_Object: [News_report, Essay, Speech, Tag_set, Book]
E36	Visual_Item
E55	Type
E41	Appellation
E52	TIME-SPAN
E53	PLACE
E54	DIMENSION

Table 1. Partial listing of CRM classes with additions (Enn Existing in CRM, active, new, [subclasses])

The text documents which are the subject of analysis in this application are instances of the *infobject* class, products of *PublicAppearance* events, typed as *SpeechTranscript* per column 3 in Table 1. The similarity measures used in the analysis are stored in a separate association table as <document, *hasMeasure*, value> corresponding to the <subject, predicate, object> form of the RDF language. The values—measures of issue “aboutness” in election-related speech—can be aggregated by candidate, party, region, time-period, etc. As mentioned earlier, this offers the possibility for analyses across election cycles and other arenas of political discourse at any time in history.

### 4. FUTURE WORK

The CIDOC-CRM does not elaborate time-span and place entity classes, and its gaps and shortcomings with respect to the requirements of historical geographers and 'spatial historians,' are being addressed. While the current database supports queries on data for a US Presidential election in the *IssueBrowser* application, its flexibility for other kinds of complex events must be tested by extending it to support other exemplar applications. That work constitutes a significant methodological portion of the author's ongoing doctoral work.

<sup>1</sup> <http://www.w3.org/TR/owl-ref/>

<sup>2</sup> developed at Universität Erlangen-Nürnberg; retrieved 18 May 2008 from [http://www8.informatik.unilangen.de/IMMD8/Services/cidoc\\_crm/cidoc\\_crm\\_4.2.4\\_owl\\_dl.owl](http://www8.informatik.unilangen.de/IMMD8/Services/cidoc_crm/cidoc_crm_4.2.4_owl_dl.owl)

<sup>3</sup> <http://protege.stanford.edu/>

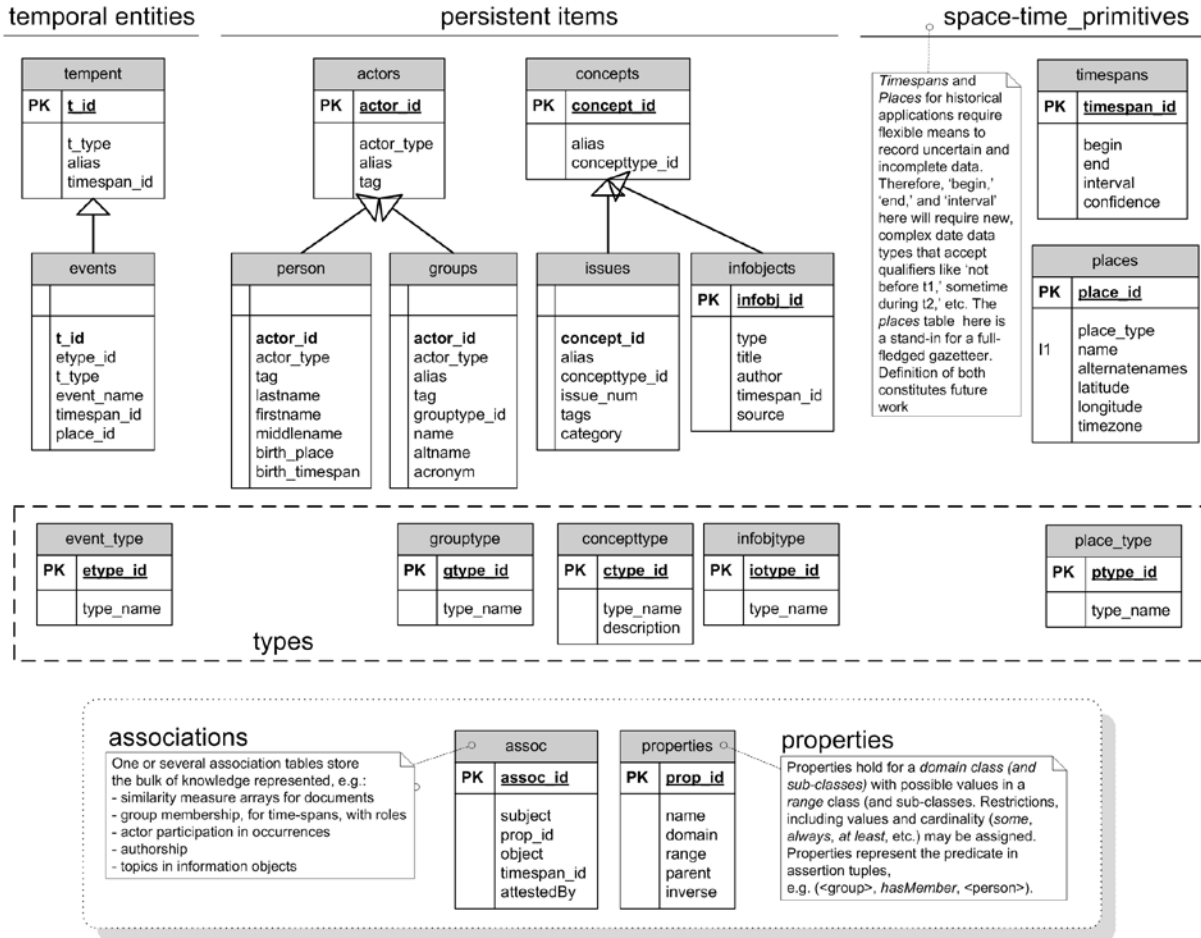


Figure 2. Partial listing of CRM classes with additions (**Enn** Existing in CRM, active, new, [subclasses])

## 5. ACKNOWLEDGMENTS

The author's research has been generously supported by NSF IGERT grant #

## 6. REFERENCES

- [1] Worboys, M. (2005). Event-Oriented Approaches to Geographic Phenomena. *International Journal of Geographical Information Science*, 19(1): 1-28
- [2] Crofts, N., M. Doerr, T. Gill, S. Stead, M. Stiff (Eds.) (2008). *Definition of the CIDOC Conceptual Reference Model*. Retrieved from [http://cidoc.ics.forth.gr/docs/cidoc\\_crm\\_version\\_4.2.4\\_Marc\\_h2008\\_.pdf](http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.4_Marc_h2008_.pdf), 4 April 2007.
- [3] Gangemi, A., N. Guarino, C. Masolo, A. Oltramari and L. Schneider (2002). Sweetening Ontologies with DOLCE. *Lecture Notes in Artificial Intelligence*, 2473: 166-181.
- [4] Kuhn, W. (2007). An Image-Schematic Account of Spatial Categories. In S. Winter et al. (Eds.): *COST 2007*, LNCS 4736, pp. 152-168
- [5] Agarwal, P. (2005). Ontological Considerations in GIScience. *International Journal of Geographical Information Science*, 19(5): 501-536
- [6] Fonseca, F., M.J. Egenhofer, P. Agouris & G. Câmara (2002). Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, 6(3): 231-2