

Linked Places: A Modeling Pattern and Software for Representing Historical Movement

Karl Grossner^a, Merrick Lex Berman^b, and Rainer Simon^c

^aWorld Heritage Web; ^bHarvard University, ^cAustrian Institute of Technology

Introduction

This paper reports on work in progress aimed at facilitating the creation, sharing, linking, and analysis of data about the movement of people, ideas, cultural practices, and commodities between places, over the course of history. Products of the *Linked Places* project include: conceptual and logical models for historical routes; a temporal extension of the popular GeoJSON data format, called *GeoJSON-T*; several varied exemplar data sets converted to *GeoJSON-T* format; prototype web software for browsing and visualizing that data; and Python scripts to convert data between CSV, GeoJSON-T, and RDF compatible with the Pelagios Gazetteer Interconnection Format¹. Substantial interim work products are shared in the Linked Places and GeoJSON-T GitHub repositories² and have been reported in some detail in two blog posts³.

Motivation

A growing number of historical gazetteers are being developed in the course of digital humanities research projects (Berman, Mostern & Southall, 2016). Their spatial-temporal coverage is typically limited to a particular area and period due to factors of scholarly quality, cost, and relevance to a given project. Coverage extents do vary considerably, from a single city for a few generations to a region for several centuries. With few exceptions, these gazetteers are unpublished as such; instead they are spatial tables contained within, and integral to, the larger project data store.

Because historical gazetteers are difficult and time-consuming to produce, it is vital they be published, when possible, in a way that permits linking them—activity the *Pelagios* project⁴ has

¹ <https://github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format>

² <https://github.com/kgeographer/linked-places>; ² <https://github.com/kgeographer/geojson-t>

³ <http://commons.pelagios.org/2016/10/linking-linked-places-project-update/>;
<http://commons.pelagios.org/2017/01/linked-places-dec-2016/>

⁴ <http://commons.pelagios.org/>

made great strides in facilitating. An emergent network of specialized gazetteers holds terrific promise, not only for re-use, but ultimately as a distributed, increasingly comprehensive geographical (i.e. spatial-temporal) index to linked data from numerous domains, including history, archaeology, literary studies, philology, and several of the social sciences. The focus of such an index, and encyclopedic applications it enables, will be on *individual places*, typically at the scales of cities and points of interest.

Such systems are highly desirable, but given a large volume of data about individual places we can also begin harvesting, creating, and sharing data about the *connections between them*. We should be able to ask of historical gazetteers: What journeys and historical routes has a given place been a waypoint on? And, what flows of people, ideas, and commodities has it been a source or sink for?

But the *Linked Places* and *GeoJSON-T* projects have been undertaken with an even larger, “moonshot” vision in mind: a system allowing scholars and the general public to visualize and analyze the emergence, growth and spread of human settlements, their changing attributes, and the dynamic connections between them, including the diffusion of technologies and cultural practices.

To realize these ideas, we need a) lots of data, and b) methods and means for merging or linking them. In some respects, we are starting from scratch; data about historical movement is sparse and stored in disparate forms. Much of it will be newly generated, for example by parsing texts, transforming tabular records, or digitally tracing lines on historical maps. Merging and linking operations will require that the form of data from different sources (or abbreviated catalogues thereof) be either standardized (in the case of merging), or similar enough that automated alignment is feasible.

The majority of works on geographic networks concerns physical media like roads and rail, whereas movement data is eventive. Geographers have modeled migration flows and disease diffusion for several decades, providing theoretical bases for their analysis that are outside our present scope. An overview of that work is found in (Lowe & Moryadas 1975). An excellent and more recent work on mobility and geographic movement is Tim Cresswell’s “On the Move” (2006). We are not aware of any efforts to model data for historical routes computationally,

however the core abstraction we build upon is the traditional graph/network model of nodes and edges credited to 18th century work of Euler (Biggs, et al 1986).

A Modeling Pattern

Data modeling is as much an art as a science (Simsion & Witt 2004), but some core best practices are well-known. A usual first step is establishing what entities are to be represented, what their essential attributes are, and what relationships obtain between them (cf. Chen, 1976). This step is often best accomplished collaboratively, in an iterative process undertaken by domain experts. Our results were immediately published to blog posts and relevant listservs, and the resulting input was useful in refining the model.

When the modeling context is an individual research project, it hardly matters what names are given those entities and relationships—only that the data store’s internal logic be sound and well understood by project members. But if, as in this case, the system will accommodate data from many sources or be accessed by others, we need to find broad agreement on a conceptual model and a vocabulary for its constituents between as many prospective participants as possible—that is, to describe the ontology of the research domain. Although much *ontology engineering* of this sort has involved comprehensive high-level ontologies such as the CIDOC-CRM⁵, the development and implementation of small *ontology design patterns* (ODP) has been gaining favor since the introduction of that paradigm by Aldo Gangemi (2005). Such patterns, by any name, are “reusable successful solutions to a recurrent modeling problem”⁶ which can be used alone or assembled in modular fashion for larger requirements. Examples include patterns for “Place,” “Event,” “Participation,” and “Region.”

And so the first step taken in the Linked Places project has been to develop an ontology design pattern for the *historical movement* of something between two or more places over some physical channel, either for some time during or throughout a timespan. The pattern, visualized in *Figure 1*, comprises the following conceptual understandings:

⁵ The Conceptual Reference Model initially developed for the International Council of Museums

⁶ This definition provided by the Association for Ontology Design & Patterns (ODPA), <http://ontologydesignpatterns.org/wiki/Odp:WhatIsAPattern>

A **route**⁷ describes an attestation of one or more occurrences of the movement of something (e.g. people, commodities, information) between two or more **places**, either for some time during or throughout a **time_period**. Routes are composed of one or more **segment**, each of which is composed of two places and a path (corresponding to nodes and edges in network parlance), the locations and temporal attributes for which may be unknown or unspecified. Movement between places occurred upon **ways**⁸—physical channels such as roads, rivers, canals, railways, footpaths, and sea lanes—and may have been directional.

The three types of routes considered here are **journeys**, **flows**, and **historical_routes**:

A **journey** is the record of a specific instance of travel by one or more individuals. Examples include: the 7th century pilgrimage of the Buddhist monk Xuanzang across China and India; the first voyage of Captain James Cook, between 1768 and 1771.

A **flow** is the record of the movement of something (commodities, people, ideas) between two places, aggregated as a magnitude, throughout some period. Examples include: the transport of captive Africans between West Africa and Bahia in the 17th century; letters between certain correspondents in Paris and Prague in the 18th century; a source network of late Neolithic obsidian artifacts and known source locations on the Anatolian Plateau.

A **historical_route** asserts a single or composite named course of travel between places, taken repeatedly by unspecified individuals over time, usually for purposes of commerce. Examples include *the Silk Road* and *the Amber Routes*. Some correspond with named roads, for instance the *Via Salaria* in Italy is both a *way* and a *historical_route*.

Additional axioms indicated by the relations and cardinality expressions (e.g. 0...*) in *Figure 1* include:

- All routes are sourced, normally to textual or cartographic documents
- The *way* for a *segment* (its physical path described by a geometry) may be known and represented, unknown, or ignored⁹.
- Each *segment* has one or more temporal attribute (“when”), which can be a *time_period*, (possibly named) or a sequence (e.g. after segment *n*)

⁷ In the sense of the course taken between *places*. “Route” is arguably both more general and familiar than alternatives like *path*, *itinerary*, or *trajectory*.

⁸ This sense of “way” corresponds to one OED definition, and is used by OpenStreetMap

⁹ Segments with unspecified ways will typically be visualized as a line or arc

- Routes and their component segments can have any number of attributes (properties), dependent upon data sources and project requirements

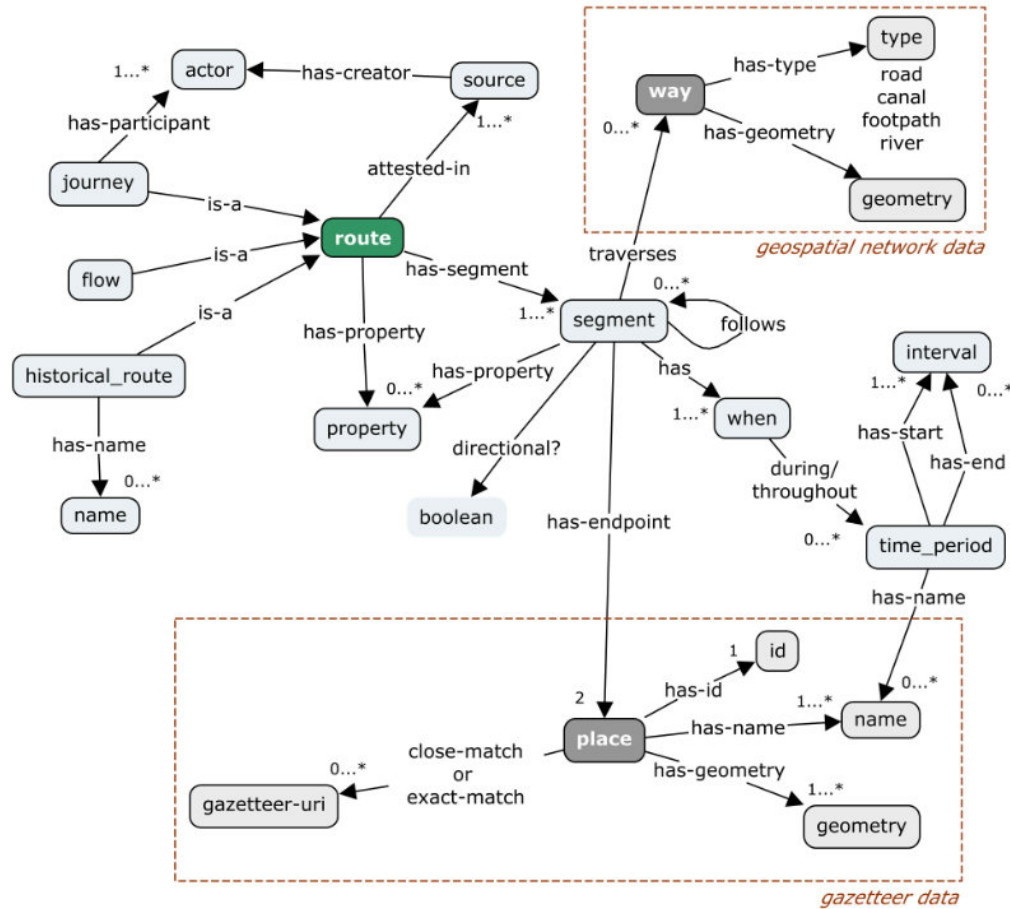


Figure 1 - A conceptual model for historical movement (routes)

The ontology pattern we introduce here is specialized, as compared to high level ontologies like CIDOC-CRM. We have not yet mapped our distinctive entities (route, journey, flow, historical_route, segment, when) to existing ontologies. The term place is commonly found, but usually is synonymous with location; the sense we are adopting is that of the Pleiades gazetteer, but is not in a published ontology that we're aware of. In any case, we feel it is best to first lay down a logically coherent set of terms and at a later date attempt to align them with other ontologies.

Formats

The *route* ODP has informed our development and implementation of recommended standard data formats. It turns out all three types of routes can be effectively described in GeoJSON-T, an extended version of GeoJSON, the widely-used format for representing geographic

*FeatureCollections*¹⁰. A *FeatureCollection* of routes will include both *Place* and *Route* features. *Route* segments are articulated as an array of one or more geometries in a route's *GeometryCollection*. GeoJSON-T allows optional “when” objects, both for each feature at the same level as its *geometry* object and for segment geometries (*Figure 2*). Features and segments have certain required properties as shown, and can have unlimited project-specific properties.

```

FeatureCollection {
  Features [
    PLACE FEATURES
    { @featureType: Place
      id: f1234
      type: Feature,
      geometry: {
        type: Point,
        coordinates: [ <lng>, <lat> ]
      }
    },
    { . . . },

    ROUTE FEATURE
    { @featureType: [ Journey | Flow | hRoute ]
      type: Feature,
      id: f9876,
      when: { ... },
      geometry: {
        type: GeometryCollection
        geometries: [
          SEGMENTS
          { type: LineString,
            coordinates: [[ ], [ ]]
            when: { timespan: [0651,,,0651,"3 months in 649"],
              duration: "3m", follows: ""},
            properties: {
              source: f1234,
              target: f2345,
              directional: 1,
              . . .
            }
          },
          { . . . }
        ]
      }
    }
  ]
}

```

Figure 2 - GeoJSON-T applied to route data

Data

To date, seven exemplar datasets have been converted from a typical CSV format to GeoJSON-T, using a newly developed Python program. Three are for journeys: two by individuals (a 7th century pilgrimage and a modern circumnavigation), the third by 840 Venetian ship convoys in the 13-15th centuries. Another dataset aggregates those ship journeys as flows having

magnitudes of journeys and ships. The last three are historical_routes: the Roman era itinerary of the Vicarello Beakers, the route system between courier stations in Ming Dynasty China, and a large set of “Old World” trade and pilgrimage routes¹¹.

Software

The widespread adoption of GeoJSON has demonstrated that for a data format to be useful, there must be software with visualization and analysis capabilities that supports it. Accordingly, an essential element of the Linked Places project is development of proof-of-concept web software to render GeoJSON-T data, for both routes and places alone, to a map and timeline together. The development of that software is ongoing, and publicly available¹². (Figure 3).

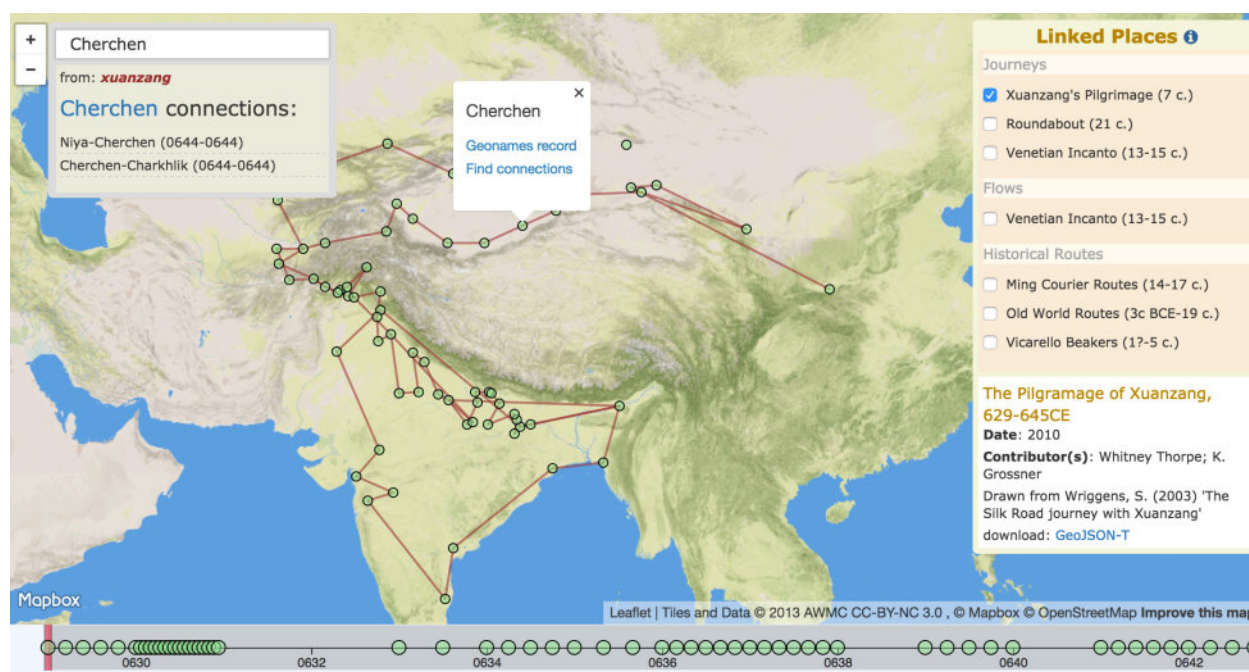


Figure 3 - Linked Places interface (partial view as of March 2017)

References

- Berman, M.L., Mostern, R., & Southall, H. (2016). *Placing Names: Enriching and Integrating Gazetteers*. Bloomington: Indiana University Press.
- Biggs, N.; Lloyd, E.; Wilson, R. (1986), *Graph Theory, 1736-1936*, Oxford: Oxford University Press

¹¹ T. Matthew Ciolek’s OWTRAD dataset, <http://www.ciolek.com/owtrad.html>

¹² <http://topotime.org/linkedplaces>

- Chen, P. P. S. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, 1(1), 9-36.
- Cresswell, T. (2006). *On the move: Mobility in the modern western world*. New York: Routledge.
- Gangemi, A. (2005). Ontology design patterns for semantic web content. In *International semantic web conference* (pp. 262-276). Springer Berlin Heidelberg.
- Lowe, J. C., & Moryadas, S. (1975). *The geography of movement*. Boston: Houghton Mifflin
- Simsion, G., & Witt, G. (2004). *Data modeling essentials*. San Francisco: Morgan Kaufmann.