

HistoGIS: Vom Punkt zur Fläche in Raum und Zeit

Schlögl, Matthias

matthias.schloegl@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

Andorfer, Peter

peter.andorfer@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

Geographische Angaben können in historischen Kontexten nicht als simple 2-dimensionale Daten (Längen- und Breitengrade) verstanden werden. Punkte die auf der Karte nur wenige Kilometer voneinander entfernt sind waren vor 500 Jahren wegen geographischer Hürden (Berge, Schluchten, Flüsse etc.) vielleicht ewig weit voneinander entfernt. Ähnliches gilt für politische Grenzen: Vor 30 Jahren waren Orte in Deutschland die heute wenige Autominuten voneinander entfernt liegen durch den eisernen Vorhang voneinander getrennt. Kamzelak (2018) hat es so formuliert: “Orte haben eine historisch-politische Dimension, die bei einer übergreifenden Registererfassung erst sichtbar zu einem Problem wird. [...] Für die Visualisierung von Briefen etwa sind historische Karten ein Desiderat; generell auch Geodaten für Flächen. Und alle mit Geodaten versehenen Einträge müssen mit einem Zeitstempel kombiniert sein, denn beispielsweise die Altstadt von Jerusalem ist eben heute nicht am selben Ort wie vor 2.000 Jahren.”.

Trotzdem arbeitet eine Vielzahl an Digital Humanities Projekten auch heute noch mit 2-dimensionalen geographischen Angaben. Unserer eigenen Erfahrung nach liegt das hauptsächlich an der Verfügbarkeit von Daten und Services. Moderne Punkt-Daten können einfach als geonames, openstreetmap etc. dump heruntergeladen werden bzw. über API Schnittstellen abgefragt werden. Für historische Daten existieren diese Services noch nicht vollumfänglich. Mit Pelagios gibt es ein Ökosystem an Services für Ortsdaten in der antiken Welt (<http://commons.pelagios.org/>), das im Zuge des “World-Historical Gazetteer” Projektes (<http://whgazetteer.org/>) auch in jüngere Zeiten ausgedehnt wird. Abgesehen von diesen notwendigen Initiativen und sehr wertvollen Datensätzen fehlen den Digital Humanities immer noch Polygondaten zu politischen Entitäten im Verlauf der Zeit.

HistoGIS hat sich zum Ziel gesetzt genau diese Lücke zu füllen in dem:

- ein Repository historischer Polygondaten mit einheitlichen Metadaten aufgebaut und zum Download angeboten wird. Für dieses Repository werden zunächst schon existierende Polygondaten eingesammelt, aufbereitet und erst in einem zweiten Schritt für historisch wichtige Zeitspannen neue

Polygone erstellt. Dabei setzt sich HistoGIS zunächst zum Ziel die Periode zwischen dem ausgehenden 18. Jhdt. und 1918 für das Gebiet der KuK Monarchie und des deutschen Bundes abzudecken.

- und RestAPI Services zur einfachen Anreicherung historischer Daten mit Hilfe des Repositories angeboten werden. Z.B.: Wo war Punkt X/Y zum Zeitpunkt Z¹? Die API antwortet mit den Metadaten zu den einzelnen überlappenden Polygonen die für den Zeitraum (oder Zeitpunkt) Gültigkeit haben. Nehmen wir an: Ein Projekt hat Reiseberichte in seiner Datenbank. Eine Station war Bolzano 1910. Die Geokoordinaten (46.49067, 11.33982) wurden über Geonames gefunden. Schickt man diese mit dem Jahr an die API bekommt man die Metadaten für Bozen Stadt, An der Etsch, Tirol und Österreich-Ungarn zurück².

Eine Fokussierung auf die politischen Verwaltungseinheiten und ihre Grenzen erlaubt es in vielen Bereichen - zumindest für jüngere Zeiten - einen Mix aus modernen Daten/Methoden und historischen zu verwenden ohne historisch falsche Daten zu generieren. So können bei oben angeführten Beispiel Bozen die Google Maps API, Geonames oder Openstreetmap zur Geolokalisation verwendet werden (die Geokoordinaten von Bozen haben sich ja nicht geändert) und die HistoGIS API um die Eingliederung in die Verwaltungshierarchien zu verbessern (Bozen war 1910 Teil der KuK Monarchie). Dieses Vorgehen reduziert nicht nur den Aufwand für die Erstellung/Kuratierung der Daten drastisch, es ermöglicht es auch leichter Punkte denen schon Längen- und Breitengrade zugewiesen wurden politisch/historisch zu verorten.

Datenmodell, Technische Grundlage und Workflow

Die Modellierung historischer Verwaltungsräume hinsichtlich ihrer räumlichen und zeitlichen Ausdehnungen erfolgte bewusst in einer äußerst vereinfachten Art und Weise. Das mittels Python (bzw. GeoDjango) definierte und als Postgresql implementierte Datenmodell besteht in seinem Kern aus den drei Hauptklassen bzw. Tabellen “TempSpatial”, “Source” und “TempSpatialRel”

Ein historischer Verwaltungsraum (Temporalized Spatial Entity oder eben “TempSpatial”) definiert sich über die im gesamten Datenset einzigartige Kombination der Eigenschaften zeitliche Ausdehnungen (“start_date” und “end_date”, Datumsfelder), räumliche Ausdehnung (“geom”; Multipolygon) sowie einem Feld “date_accuracy”, welches Auskunft über den Grad der Genauigkeit der angegebenen Datumswerte gibt. Ergänzt wird diese Klasse um die Eigenschaften “name” für, entsprechend den Projektkonventionen einen zeitgenössischen Namen der Entität, “alt_names” für

alternative Namen sowie einem Feld “additional_data”, welches das Speichern arbiträrer weiterer Daten im JSON Format ermöglicht.

Das Feld “administrativ_unit” zeigt auf eine Hilfsklasse (“SkosConcept”) für kontrolliertes Vokabular, welche in weiten Teilen das SKOS Datenmodell implementiert.

Die Quelle jeder Instanz der Klasse TempSpatial bzw. jeder im Datenset erfassten historischen Verwaltungseinheit wird mit Hilfe der Klasse “Source” beschrieben. Darin werden URLs zu verwendeten Daten anderer Projekte gespeichert wie auch eine Beschreibung der (weiteren) Datenerhebung und Kuration im Rahmen des Projektes sowie ein Zitationsvorschlag. Jedes Source Objekt ist außerdem auch mit einem ESRI-Shapefile verbunden welches Projektintern als primäres Datenformat dient. Mehr dazu im Abschnitt Workflow.

Die Modellierung beliebiger Relationen zwischen beliebigen historischen Verwaltungsräumen ist im Datenmodell durch die Klasse TempSpatialRelation grundgelegt. Hier können jeweils zwei TempSpatial Objekte (“instance_a” und “instance_b”) für eine Zeitspanne (“start_date” und “end_date”) in eine typisierte (Verweis auf die bereits erwähnte Hilfsklasse “SkosConcept”) Relation gebracht werden. Hierbei ist jedoch weniger an die in Gazetteern üblichen “part of” Beziehungen gedacht, sondern an Relationen wie beispielsweise “ist Vorgänger von” oder “wurde zusammengelegt mit”. Allerdings muss darauf hingewiesen werden, dass im derzeitigen Status des Projektes noch keine derartigen Beziehungen erfasst werden.

Die Art und Weise wie eben erwähnte “part of” Beziehungen erfasst werden sollen, wurde im Projekt ausgiebig diskutierte, wobei hier neben formal-konzeptionellen Argumenten vor allem auch die konkreten Arbeitsvoraussetzungen und -bedingungen im Projekt berücksichtigt werden mussten.

Schlussendlich wurde eine explizite Modellierung hierarchischer Strukturen der erfassten Verwaltungseinheiten verzichtet, sprich im Datenmodell wird nicht ausdrücklich festgehalten, dass z.B. die Verwaltungseinheit A für einen gegebenen Zeitraum, Teil der Verwaltungseinheit B und diese Teil der Verwaltungseinheit D war. Dies erscheint deshalb als zulässig, weil im Projekt von der Prämisse ausgegangen wird, dass, zumindest für den für das Projekt primär interessante (Zeit)Raum, die Fläche der übergeordnete Einheit stets die Summe aller ihr untergeordneten Einheiten bildet. Dies ermöglicht es, dass part-of Beziehungen zwischen TempSpatial Objekten ‘on the fly’ mit Hilfe von spatial queries und unter Berücksichtigung der jeweiligen Start- und Enddatumswerten berechnet werden können. Und dies wiederum erleichtert einerseits die Arbeit der DatenkuratorInnen im Projekt, da diese keine expliziten Verbindungen pflegen müssen, andererseits ermöglicht dieses Flexibilität die Integration anderer Datensatz mit relativ geringem Aufwand.

An dieser Stelle muss jedoch betont werden, dass die bis dato kuratierte Menge an Daten noch nicht ausreicht um

konkrete Aussagen hinsichtlich der Belastbarkeit des hier vorgestellten Datenmodells treffen zu können. Erste Tests diesbezüglich fielen jedoch durchwegs positiv aus, sowohl was die Genauigkeit der spatial queries, vor allem aber auch was deren Performanz betrifft.

Die technischen Komponenten des Projektes sind überschaubar. Als Storage Layer fungiert eine Postgresql Datenbank mit PostGIS Erweiterung. Die Interaktion damit erfolgt über einen mittels dem Python basierten Webframework (Geo)Django implementierten Applikation Layer, wobei mit (Geo)Django, respektive django-rest-framework sowohl die HistoGIS-Webapplikation als auch ein entsprechender REST Webservice implementiert wurde bzw. wird.

Die eigentlich Datenkuration erfolgt davon völlig unabhängig und unter Verwendung der open source Software Qgis. Bis dato wurden damit vorwiegend bereits existierende Daten harmonisiert und als ESRI shapefiles gespeichert. Das Schema dieser Dateien entspricht dabei weitgehend dem oben skizzierten Datenmodell. Als ‘fertig’ erachtete Datensätze (gezippte Shapefiles) werden dann über ein Webformular in HistoGIS als sogenannte “Source” objekte hochgeladen, entpackt, die Features der Shapfiles als TempSpatial Objekte gespeichert und mit dem Source Objekt verknüpft.

Neben der Kuration und Harmonisierung bereits bestehender “Vektordaten” werden im Projekt aber auch selbst Daten erzeugt. Dazu wählt ein Historiker im Team verwertbare (historische) Karten aus, welche idealerweise bereits digitalisiert (gescannt) sind. Die Datakuratorinnen georeferenziert diese Scans (geotiffs) und extrahieren die darin auffindbaren Informationen zu historischen Verwaltungsgrenzen als Vektordaten.

Kartenmaterial bis dato und Ausblick

Bis dato befinden sich knapp 4000 Polygone in der Production Instanz des Systems. Das Anpassen und Einspielen schon vorhandener Polygone wurde mit Daten aus dem Census mosaic Projekt (<https://censusmosaic.demog.berkeley.edu/>) und dem HGis Archiv (<http://www.hgis-germany.de/>) begonnen. Damit können große Teile des 19. Jhdts für die Gebiete Österreich-Ungarns und des Deutschen Bundes (inkl. der jeweiligen Nachfolgeentitäten) bereits abgedeckt werden. Die Daten werden nicht nur technisch aufbereitet, sondern auch inhaltlich von einem Verwaltungshistoriker überprüft. HistoGIS implementiert dafür ein Ampelsystem. Vom HistoGIS-Team technisch wie inhaltlich überprüfte Daten werden mit Grün markiert, vom HistoGIS-Team ausgewählte Daten die noch nicht überprüft wurden mit Gelb und von Usern zur Verfügung gestellte, nicht überprüfte Daten mit Rot (momentan befinden sich lediglich gelbe und grüne Daten im System).

Die Aufbereitung der Daten, wie auch die Entwicklung des technischen Systems schreitet erfreulicher Weise schneller voran als geplant. Es wurde deshalb erst kürzlich beschlossen in HistoGIS schon während der Projektlaufzeit auch Daten außerhalb der geplanten räumlich-zeitlichen Grenzen aufzunehmen.

In unserer Präsentation werden wir vor allem das Datenmodell und die RestAPI Schnittstellen des Systems diskutieren und vorstellen.

Fußnoten

1. Eine auf diese API aufbauende Abfragemaske findet sich hier: <https://histogis.acdh.oeaw.ac.at/shapes/where-was/>
2. Beispielhaft das Polygon für Tirol: <https://histogis.acdh.oeaw.ac.at/shapes/shape/detail/3352>

Bibliographie

Kamzelak, Roland S. (2018): *“Von der Raupe zum Schmetterling oder Wie fliegen lernen – Editionsphilologie zwischen Infrastruktur und Semantic Web.”* In: **Kamzelak, Roland S / Steyer, Timo (eds.):** *Digitale Metamorphose: Digital Humanities und Editionswissenschaft.* (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 2). text/html Format. DOI: 10.17175/sb002_004 [letzter Zugriff 28. September 2018]

Nüssli, Marc-Antoine / Nüssli, Christos (2017): *“A formal model for historical atlases and historical knowledge”*, <http://www.academia.edu/35853762> [letzter Zugriff 28. September 2018]