

GeoBib



GEOREFERENZIERTE
ONLINE-BIBLIOGRAPHIE FRÜHER
HOLOCAUST- UND
LAGERLITERATUR

DER GEOBIB- PROTODATENSATZ

M4.2S

Bastian Entrup, Sandra Schaarschmidt

Juni 2013

INHALT

1	Übersicht: Daten	2
2	Protodatenätze	2
2.1.	Primärliteratur und Scans	2
2.2.	XML-Daten	3
2.3.	Wiki-Daten	3
2.4.	Geo-Daten und Kartenmaterial	3
3	Zusammenfassung	4
4	Quellenverzeichnis	4

1 ÜBERSICHT: DATEN

In diesem Dokument wird ein Protodatensatz des GeoBib-Projekts beschrieben. Der Datensatz, bestehend aus literarischen Metadaten, bibliographischen Daten, Orts- und Personendaten im Wiki sowie geographischem Kartenmaterial, wurde zur Verwendung in geplanten Prototypen der Verarbeitungspipeline, der Datenbank sowie der Website zusammengestellt.

Die folgenden Dateien umfassen sowohl Primärtexte als auch zugehörige XML-Dateien. Bei einigen Texten sind auch Scans oder Annotationsammlungen im Word-Format beigelegt. Diese sind allerdings für die weitere Verarbeitung nicht relevant.

1939_Rudolph_Albert (aka Rudolf Frank).pdf

1939_Rudolph_Albert.doc

1939_Rudolph_Albert.xml

1945_Kozlik_Francois (Annotation).doc

1945_Kozlik_Francois.pdf

1945_Kozlik_Francois.xml

1945_o.A._Extermination of Polish Jews.pdf

1946_Kopp_Guido.doc

1946_Kopp_Guido.pdf

1946_Kopp_Guido.xml

Test_1945_o.A._Extermination of Polish Jews.xml

Test_1947_Pankiewicz_Tadeusz.xml

Hinzu kommen die Geographischen Datensätze und die Informationen aus dem Wiki.

2 PROTODATENSÄTZE

2.1.PRIMÄRLITERATUR UND SCANS

Die Primärdaten, also die literarischen Texte in Form von Scans, werden dem hier beschriebenen Datensatz nur der Vollständigkeit halber beigelegt. Die Scans sind mittels OCR (*optical character recognition*) um ihren Textinhalt in auslesbarer Form erweitert worden. Diese Daten werden in der Verarbeitung nur von den Mitarbeitern, die für die Annotation und Zusammenstellung der diversen (Meta-) Daten verantwortlich sind, genutzt und spielen für die automatische Verarbeitung in späteren Arbeitsschritten keine Rolle. Teilweise werden für die Sammlung dieser Daten Word-Dokumente angelegt. Auch diese werden nicht automatisch verarbeitet.

2.2.XML-DATEN

Die XML-Dateien repräsentieren jeweils einen Primärtext. Der Inhalt dieser XML-Dateien ist im Meilenstein M4.1S genauer beschrieben.

2.3.WIKI-DATEN

Jede XML-Datei verweist auf mehrere Instanzen, die im Wiki durch eine Seite repräsentiert werden. Diese Trennung von Daten, die den Text selbst bzw. die Personen und Orte, die im Text erwähnt werden, betrifft, ist ebenfalls im Meilenstein M4.1S beschrieben. Der Protodatensatz umfasst ein Backup des Wikis, in dem sich auch die referenzierten Daten befinden. Insgesamt sind zurzeit 3511 Personen-Seiten sowie 664 Orts-Seiten vorhanden.

2.4.GEO-DATEN UND KARTENMATERIAL

Für das Projekt GeoBib können verschiedene digitale geographische Grundlagenkarten verwendet werden. Ein Teil der voraussichtlich zu nutzenden Datensätze stammt aus dem Projekt ‚GIS-Kollektion Bevölkerungsgeschichte‘ vom Max-Planck-Institut für demographische Forschung in Rostock (MPIDR). Das Projekt unterstützt die „Forschung, in dem es bestehende Lücken in der europäischen GIS-Dateninfrastruktur hinsichtlich historischer [...] Verwaltungsgrenzen sowie historischer Ortsnamen schließt“ [1]. Die in diesem Rahmen entstandenen Geodaten liegen in der Lambertschen Azimutalprojektion vor (ETRS89 / LAEA Europe; EPSG 3035) und können für nichtkommerzielle, wissenschaftliche Zwecke frei verwendet werden.

Für das Projekt GeoBib werden nicht sämtliche in den Datenpaketen enthaltene Shape-Files benötigt. Von Interesse sind die Daten für Europa 1930, Polen 1931, das Deutsche Reich 1933-45 sowie Deutschland 1949. Für Europa sind die administrativen Einheiten der 2. Ebene vorhanden, die z.B. in Deutschland den Ländern entspricht. Die 1. Ebene, die die Staaten abbildet, muss also nachdefiniert werden. Die einzelnen Datensätze für Polen und das Deutsche Reich/Deutschland beinhalten dagegen die administrativen Einheiten der 3. Ebene, welche in Deutschland den Kreisen entspricht und in Polen den Powiaty. Dort müssen also die 2. und 1. Ebene nachdefiniert werden. Für einige Länder, vor allem in Westeuropa, sind aufgrund territorialer Beständigkeit Grenzverläufe aus den Daten für Europa 1930 auch auf weitere Jahre übertragbar. Genauer ist dem Meilenstein M5.2R zu entnehmen.

Die jeweiligen Vektordaten liegen als Shape-Files vor. Ein großer Vorteil ist, dass sie im Gegensatz zu vielen anderen verwendbaren Daten bereits georeferenziert sind, was eine zeitintensive Nachbearbeitung erspart. Jedoch sind die Attributdaten nicht in allen Shape-Files vollständig, sodass sie teilweise überarbeitet werden müssen. Auch stimmen bestimmte Grenzverläufe in verschiedenen Shape-Files an Stellen, an denen sie deckungsgleich sein müssten, nicht überein. Daher muss auch diesbezüglich eine nachträgliche Anpassung erfolgen.

Auch wenn schon jetzt abzusehen ist, dass aufgrund der bestehenden Lücken bezüglich historisch-administrativer Geodaten [2] auch für das Projekt GeoBib kein vollständiger Datensatz für den Zeitraum zwischen 1933 und 1949 angelegt werden kann, so ist dennoch angestrebt, aus den verfügbaren Quellen für den Map-Server des Projekts einen einheitlichen Datensatz mit verschiedenen Jahres-Layer ,wie aus einem Guss‘ zu fertigen. Damit wird das aktuelle Kartenmaterial um historisch-administrative Grenzen ergänzt.

3 ZUSAMMENFASSUNG

Die hier vorgestellten Daten wurden im ersten Projektlaufjahr des GeoBib-Projekts für Zwecke der Softwareentwicklung, besonders für die Entwicklung einer Verarbeitungspipeline, die Entwicklung der Datenbank, sowie zur Erstellung der Web-Plattform zusammengestellt. Die Daten liegen in Form einer kommentierten Zip-Datei auf den internen Projektservern vor und sind dort für Projektpartner einsehbar und, sofern nötig, verwendbar.

4 QUELLENVERZEICHNIS

[1] MPIDR (MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH) (2013): Population History GIS Collection.

http://www.demogr.mpg.de/de/abteilungen/historische_demografie_892/projekte/gis_kollektion_bevoelkerungsgeschichte_1556.htm (18.06.2013).

[2] PIEPER, J./SCHWEIKART, J./MARTI-HENNEBERG, J. (2005): Grenzen im Wandel – GIS-gestützte Erfassung und Visualisierung der historischen, regionalen Strukturen Europas (1870-2000).

https://labor.beuth-hochschule.de/fileadmin/labor/geovis/Publikationen/Grenzen_im_Wandel_-_GIS-gestuetzte_Erfassung_und_Visualisierung.pdf (06.05.2013).