

Showcase(s): Was kann man mit XML-Daten forschen?

Ulrike Henny, ulrike.henny@uni-wuerzburg.de

8. September 2016

Zentrum für Informationsmodellierung
Austrian Centre for Digital Humanities
Elisabethstraße 59/III, SR 81.31

Workflow Edition ?

Konzeption

Forschungsinteresse

- Zielbestimmung, Requirements
- Dokumentauswahl
- Modellierung
- der Edition insgesamt
- der Inhalte / Dokumente

Umsetzung

- Digitalisierung, Reproduktion
- Äußere Beschreibung, Metadaten
- Editionswerkzeuge, Datenmanagement
- Transkription
- Textkritik
- Annotationen, Kommentare
 - Erläuterungen
 - named entities, Taxonomien, etc.
- Paratexte
- Kontextualisierung

Publikation

- Backend, technisches System
- Nutzungsumgebung, Tools
- Frontend
 - Funktionalitäten, Usability
 - Adressierbarkeit
 - Interfaces
 - Spin-Offs
- Dokumentation, Langzeitverfügbarkeit

+ sprachliche oder inhaltliche Analyse

Weitere Forschungsfragen



Was kann man mit XML-Daten forschen?

- Mögliche Wege:
 - Von der Edition und XML zurück zu „Text“ (und Sprache)
 - Von der Edition und XML aus das Codierte weiterführen
 - **Beispiel: Briefe in Zeit und Raum mit dem DARIAH Geo-Browser**

Führt oft zu weiteren digitalen Methoden und Tools, die Forschung **unterstützen**

Briefe in Zeit und Raum

- Von der Edition und XML aus das Codierte weiterführen
 - Beispiel: der DARIAH Geo-Browser

DARIAH Geo-Browser

- im DARIAH-Projekt entwickelt
- <https://de.dariah.eu/geobrowser>
- für die Analyse von Raum-Zeit-Relationen von Daten
 - Ermöglicht: vergleichende Visualisierung mehrerer Datensets
 - Korrelation von geographischen Raumverhältnissen zu entsprechenden Zeitpunkten und -abläufen

DARIAH Geo-Browser

- Features:
 - Eigene Daten können in den Geo-Browser geladen werden (Datasheet-Editor / KML / KMZ / CSV)
 - Historische Karten
 - Beschreibung der Datensätze
 - Suche
 - Interaktive Zeitleiste
- Nötig: Georeferenzierung der Datensätze

Briefe in Zeit und Raum: DARIAH Geo-Browser

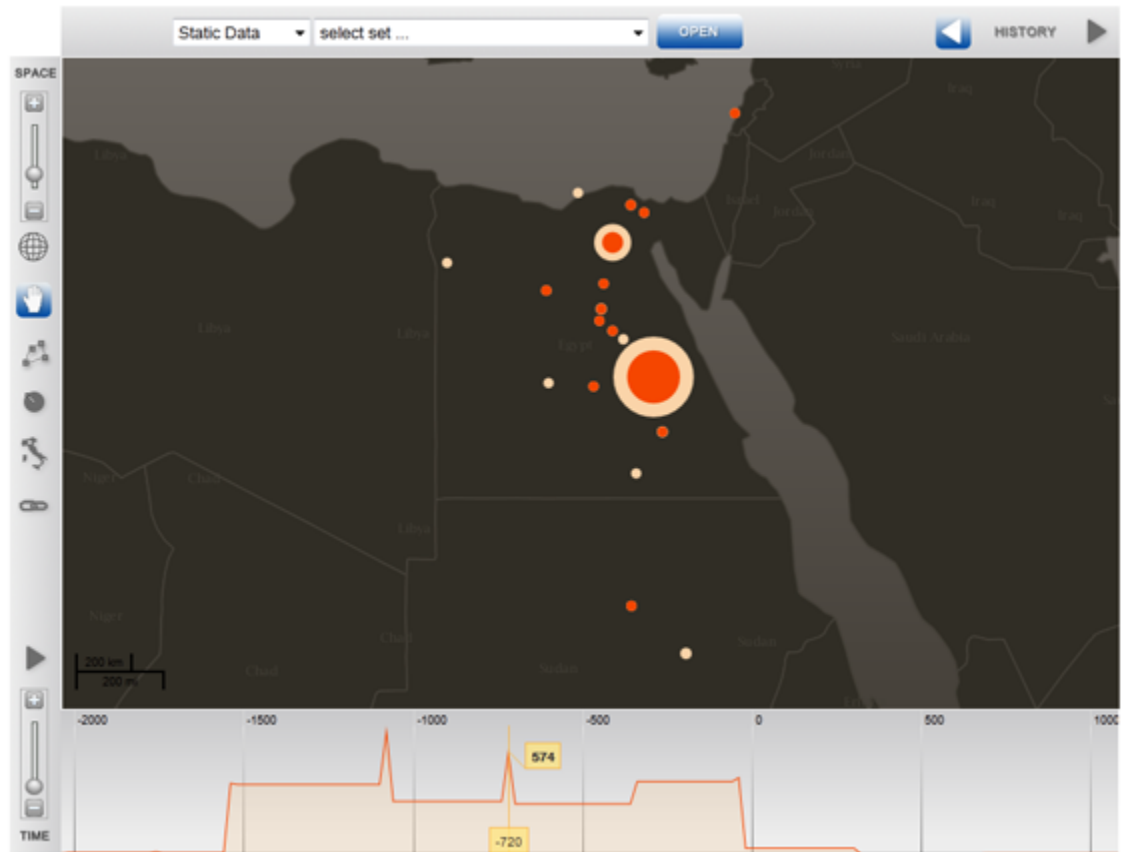
Beispielanwendung:

Totenbuchprojekt

[totenbuch.kml](#)

ca. 3000 Metadaten-
sätze zu Textträgern
mit Sprüchen des
altägyptischen
Totenbuchs

Visualisiert:
Herkunftsorte und -zeit



DARIAH Geo-Browser

- Laden Sie die Datei [totenbuch.kml](https://geobrowser.de.dariah.eu/totenbuch.kml) im Geo-Browser (<https://geobrowser.de.dariah.eu/>) hoch
 - Links bei Load Data „local KML file“ auswählen
 - Datei suchen und hochladen
- Wählen Sie eine historische Karte
 - Historical map of 1000 BC
- Testen Sie die Animation
 - Zeitpunkt auf der Leiste setzen + Play
- Schauen Sie sich die Beschreibung der Datensätze im unteren Bereich an

KML = XML!

Briefe in Zeit und Raum: DARIAH Geo-Browser

Beispielanwendung:

Hispanoamerikanische Unabhängigkeitskriege

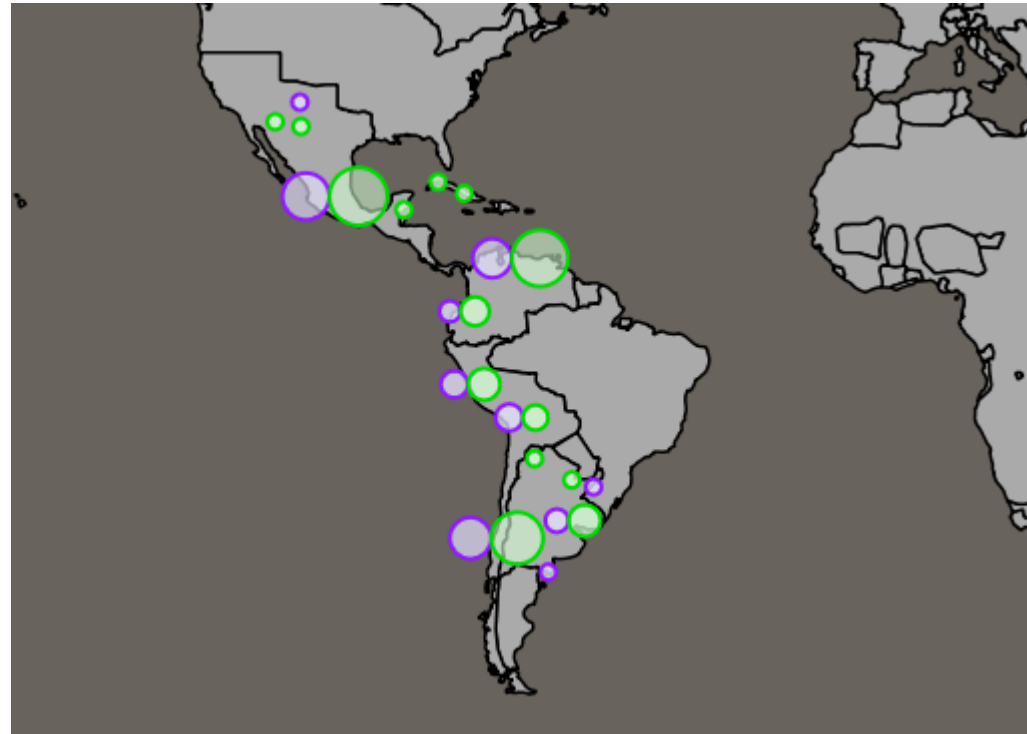
[batallas-1.kml](#)

[batallas-2.kml](#)

Informationen zu Schlachten
in Unabhängigkeitskriegen
Aus Wikipedia-Infoboxen

Visualisiert:

Ort und Zeit der Schlacht

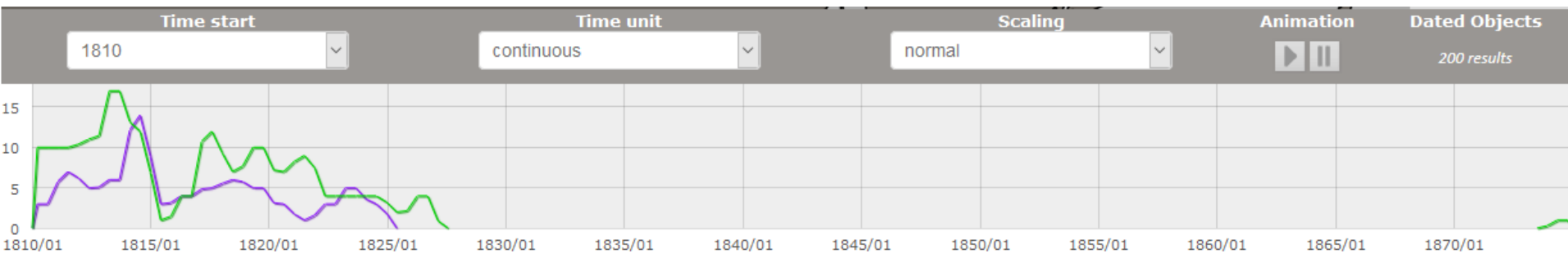


grün: Sieg

lila: Verlust

Briefe in Zeit und Raum: DARIAH Geo-Browser

- Laden Sie die Dateien [batallas-1.kml](#) und [batallas-2.kml](#) nacheinander in den Geo-Browser (<https://geobrowser.de.dariah.eu/>) hoch
- Wählen Sie eine historische Karte
 - Historical map of 1815 BC
- Testen Sie die Animation & schauen Sie sich die Beschreibungen an



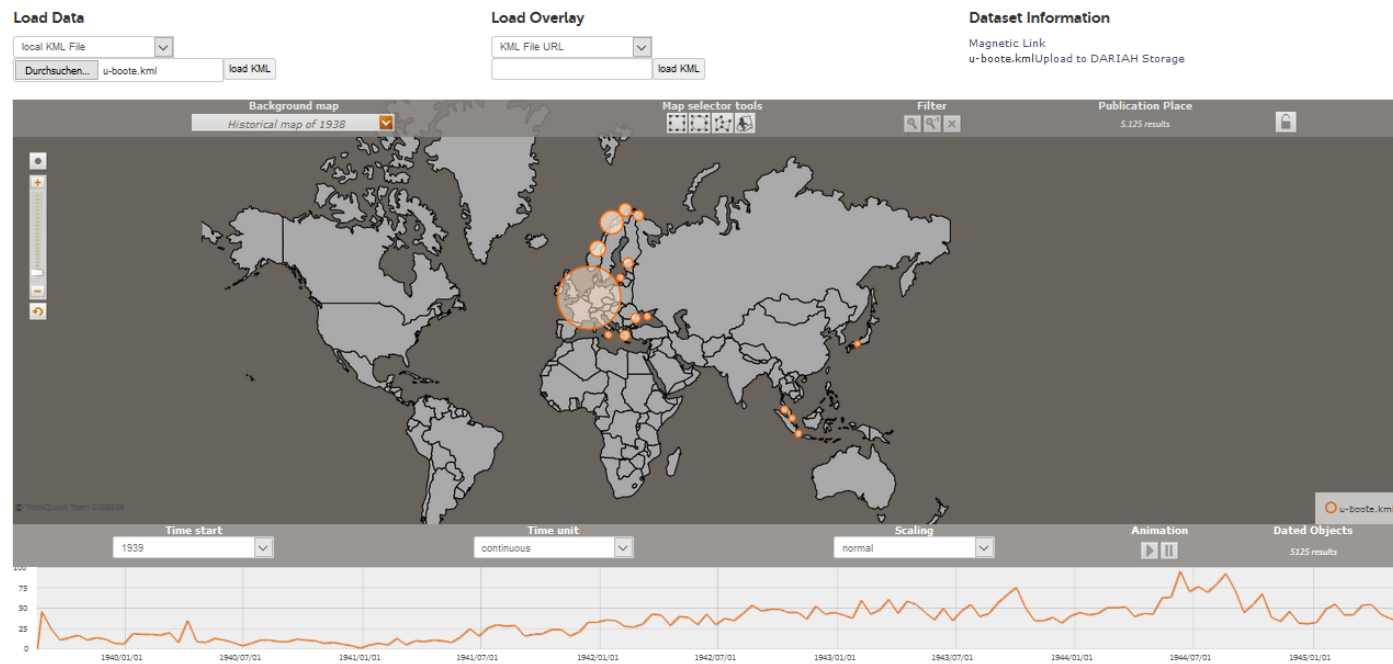
Briefe in Zeit und Raum: DARIAH Geo-Browser

Beispielanwendung: **Der U-Boot-Krieg im Zweiten Weltkrieg**

Ausfahrten von U-Booten (von <http://uboat.net>)

Visualisiert: Datum und Ort der Ausfahrt

uboat.net



DARIAH Geo-Browser

- Wie kommen die eigenen Daten in den Geo-Browser?
 - Ortsnamen, Zeitpunkte/-räume und Ereignisbeschreibungen aus den Editionsdaten extrahieren
 - Falls nicht vorhanden: Anreichern mit Geokoordinaten
 - Format produzieren, das der Geo-Browser versteht

DARIAH Geo-Browser

- Mögliches Zielformat: KML
 - Keyhole Markup Language
 - Auszeichnungssprache zur Beschreibung von Geodaten
 - = XML
 - <https://developers.google.com/kml/>
(Referenz und Beispiele)

Briefe in Zeit und Raum: Beispiel-KML

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Document>
    <Placemark>
      <name>Ausfahrt Nr. 1 von U 1</name>
      <address>Kiel</address>
      <description><![CDATA[<ul><li><a
href="http://de.wikipedia.org/wiki//wiki/U_1_(Schiff,_1935)">W
ikipedia-Artikel zu U 1</a></li><li>Klasse: II
A</li></ul>]]></description>
      <TimeStamp>
        <when>1940-03-15</when>
      </TimeStamp>
      <Point>
<coordinates>10.1405555555556,54.3252777777778</coordinates>
      </Point>
    </Placemark>
```

DARIAH Geo-Browser

- KML
 - Wurzelement mit KML-Namensraum:
`<kml
xmlns="http://www.opengis.net/kml/2.2">`
 - Dokument-Element: enthält das gesamte KML-Dokument
`<Document>`

DARIAH Geo-Browser

- KML
 - Placemark: ein einzelner Ort-Zeit-Datensatz.
Enthält weitere Kindelemente.
`<Placemark>...`
`</Placemark>`

DARIAH Geo-Browser

- KML: Kindelemente von Placemark
 - Titel für den Datensatz
<name>
 - Ortsangabe als Ortsname
<address>
 - Zusätzliche Beschreibung zum Datensatz
<description>

DARIAH Geo-Browser

- KML: Kindelemente von Placemark
 - Zusätzliche Beschreibung zum Datensatz

```
<description><![CDATA[<ul><li><a href="http://de.wikipedia.org/wiki//wiki/U_1_(Schiff,_1935)">Wikipedia-Artikel zu U 1</a></li><li>Klasse: IIA</li></ul>]]></description>
```

DARIAH Geo-Browser

- XML Character Data
 - CDATA
 - Zeichendaten, die nicht als Markup verstanden werden sollen
 - Z. B., wenn XML verarbeitet wird und darin anderes Markup steht (wie HTML), das einfach so weitergegeben werden soll, wie es ist
 - Syntax:
`<!CDATA[...<p>Dies soll nicht geparst werden</p> ...]]>`

DARIAH Geo-Browser

- KML: Kindelemente von Placemark
 - Zeitpunkt
`<TimeStamp>`
 `<when>2007-01-03T00:00:00Z</when>`
`</TimeStamp>`
 - Zeitraum
`<TimeSpan>`
 `<begin>2004-03</begin>`
 `<end>2004-04</end>`
`</TimeSpan>`

Briefe in Zeit und Raum:

DARIAH Geo-Browser

Anwendungsfall Weber-Briefe

DARIAH Geo-Browser

Anwendungsfall Weber-Briefe

- Wann wurden wo Briefe von Carl Maria von Weber verschickt?
- Wann haben wo Empfänger Briefe von ihm erhalten?
- (Welche Orte wurden wann in Briefen erwähnt?)

DARIAH Geo-Browser

Anwendungsfall Weber-Briefe

- Ausgangsdaten: [briefcorpus-weber.zip](#)
- Enthält 100 TEI-Dateien mit Briefen
- Herunterladen und die Zip-Datei entpacken

DARIAH Geo-Browser

- Welche Daten haben wir? Welche fehlen noch?
 - Wer an wen →
`//correspAction[@type='sent']/persName`
`//correspAction[@type='received']/persName`
 - Wann →
`//correspAction/date/@when`
 - Von wo nach wo →
`//correspAction[@type='sent']/placeName`
`//correspAction[@type='received']/placeName`

DARIAH Geo-Browser

- Aber: Wo liegen die Orte genau?
- Wir brauchen Geokoordinaten!
 - Liste aller Absende- und Empfangsorte extrahieren
 - XSLT-Stylesheet: [Übungsdateien/ortsliste-erstellen.xsl](#)
 - Führt zu: [Übungsdateien/ortsliste-ohne-koordinaten.xml](#)
 - Automatisch oder händisch mit Geokoordinaten anreichern
 - Ergebnis: [Übungsdateien/ortsliste.xml](#)

DARIAH Geo-Browser

- **Gemeinsame Übung:**
- Transformationsszenario in oXygen einrichten
 - Als Ausgangsdatei: eine beliebige Briefdatei, z. B. [briefcorpus-weber/A040382.xml](#)
 - Als Transformationsdatei: [Übungsdateien/ortsliste-erstellen.xsl](#)
 - Als Zieldatei: [Übungsdateien/ortsliste-ohne-kordinaten.xml](#)

Briefe in Zeit und Raum:

DARIAH Geo-Browser

- Der nächste Schritt:
 - Händisch oder automatisch die Geokoordinaten ergänzen
 - Ergebnis: [Übungsdateien/ortsliste.xml](#)

DARIAH Geo-Browser

- Wie haben jetzt ein Korpus mit Briefen. Und eine Ortsliste mit Geokoordinaten.
- Wie geht es weiter? Was ist zu tun?
 - Editionsdaten (=TEI) in das Eingabeformat für den Geo-Browser (=KML) transformieren
 - Pro Korrespondenzhandlung (`//correspAction`) ein KML-Datensatz
 - Dabei die Geokoordinaten (aus `orte-koordinaten.xml`) einbinden
 - Womit? → XSLT!

DARIAH Geo-Browser

- Nützliches für den Weg:
 - XPath-Funktion `collection(Pfad)`
 - Zum Einlesen mehrerer XML-Dokumente zugleich
 - XPath-Funktion `document(Pfad)`
 - Zum Einlesen eines einzelnen XML-Dokuments
- Achtung: Namensräume! **TEI != KML**

Auf geht's!

Briefe in Zeit und Raum:

DARIAH Geo-Browser

- Gemeinsame Übung:
 - XSLT-Transformation durchführen, um aus den Briefdateien und der Ortsliste eine KML-Datei zu erstellen

DARIAH Geo-Browser

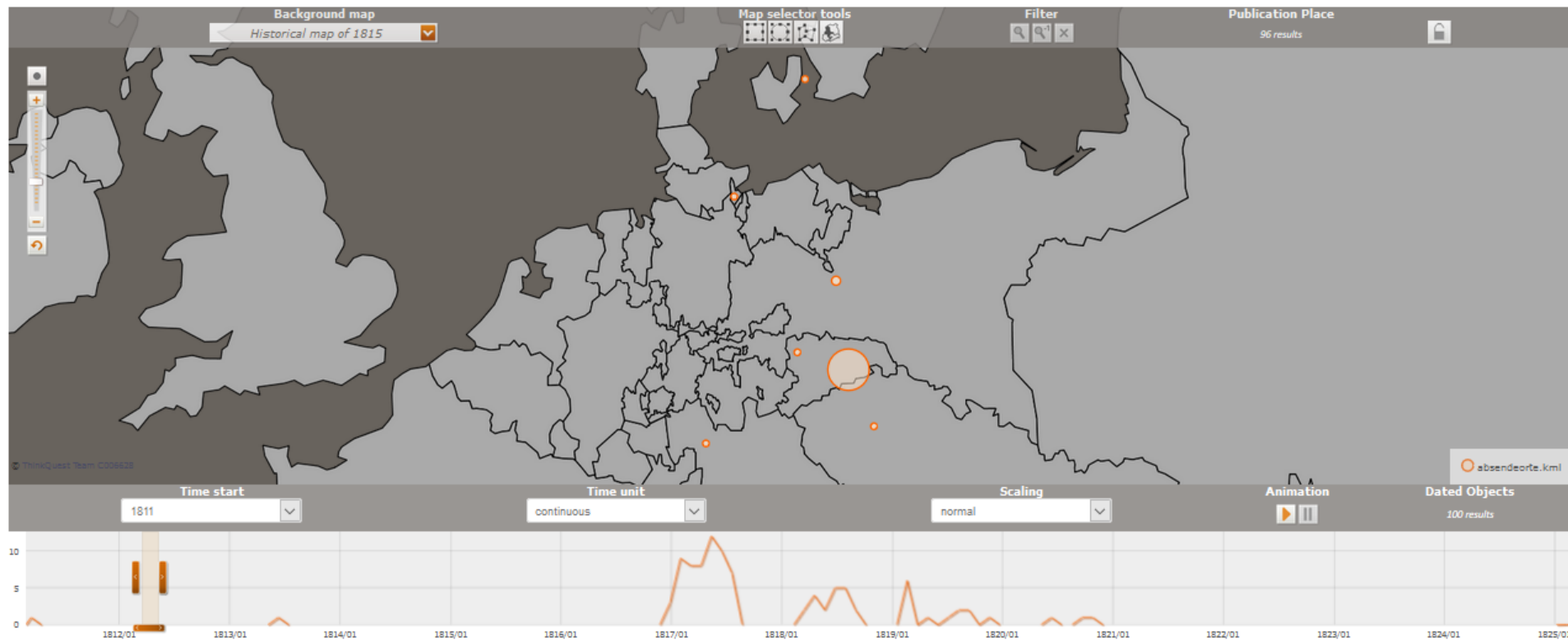
- **Gemeinsame Übung:**
- Transformationsszenario in oXygen einrichten
 - Als Ausgangsdatei: eine beliebige Briefdatei, z. B. [briefcorpus-weber/A040382.xml](#)
 - Als Transformationsdatei: [Übungsdateien/KML-absendeorte-generieren.xsl](#)
 - Als Zieldatei: [Übungsdateien/absendeorte.kml](#)

DARIAH Geo-Browser

- 2. Transformation
 - Kann entsprechend für die Empfangsorte gemacht werden (statt Absendeorte)
- Letzter Schritt:
 - Fertige KML-Dateien in den Geo-Browser laden
 - [Übungsdateien/absendeorte.kml](#)
 - Evtl. zusätzlich auch [Übungsdateien/empfangsorte.kml](#)

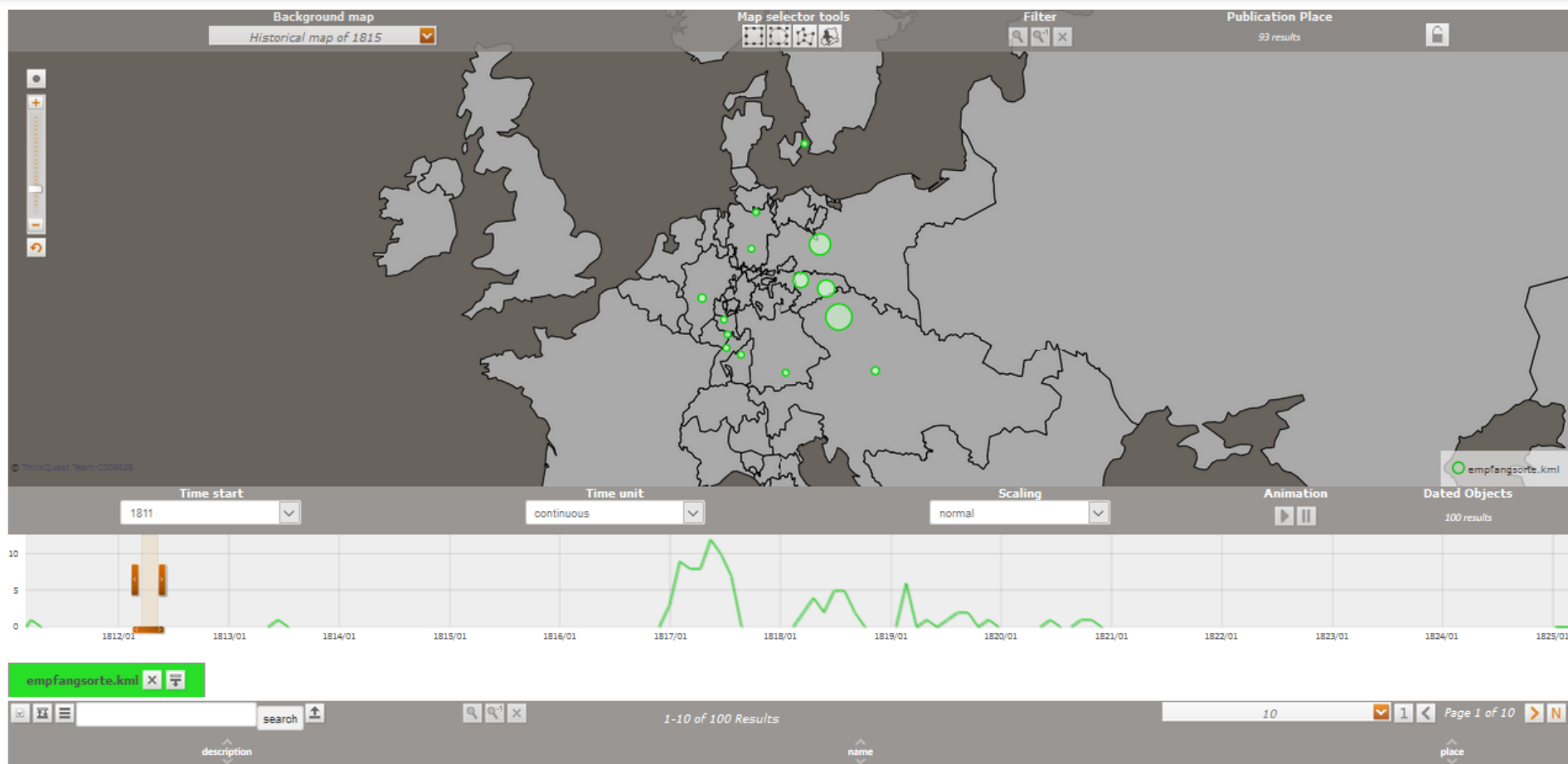
Briefe in Zeit und Raum: Briefe im Geo-Browser

Absendeorte (Karte 1815):



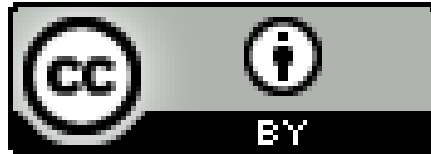
Briefe in Zeit und Raum: Briefe im Geo-Browser

Empfangsorte (Karte 1815):



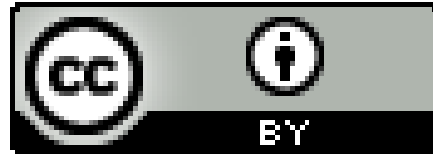
Vielen Dank!

Dieses Werk ist lizenziert unter einer **Creative Commons Namensnennung 4.0 International** Lizenz.



Alle darin verwendeten Werke anderer Urheber sind Zitate zu wissenschaftlichem Gebrauch.

This work is licensed under a **Creative Commons
Namensnennung 4.0 International** License.



All works of other authors cited here are their intellectual property and are used for academic teaching purpose only.