

Datenblatt & Dokumentation: Relationales Datenset des Reichsstädteprojekts

Das vorliegende Datenset basiert auf einer relationalen Geodatenbank des Projekts "**Civitas imperialis**" am Institut für vergleichende Städtegeschichte (IStG). Es stellt strukturierte historische, geografische und bibliografische Informationen zu Reichsstädten bereit. Das Datenset ist in drei separate Dateien aufgeteilt: zwei relationale Tabellen im CSV-Format und eine kombinierte Gesamtdatei im GeoJSON-Format.

Inhaltsverzeichnis

1. Allgemeine technische Spezifikationen
2. Das relationale Datenmodell (CSV-Dateien)
3. Spezifikation der CSV-Dateien
 - 3.1 Tabelle 1: Ortsbasierte Stammdaten (reichsstadt_ort_mastertabelle.csv)
 - 3.2 Tabelle 2: Ereignisbasierte Daten (reichsstadt_ereignis_mastertabelle.csv)
4. Spezifikation des GeoJSON-Gesamtdatensets (reichsstadt_gesamt.geojson)
5. Nutzungshinweise und Software-Kompatibilität
 - 5.1 Import-Richtlinien für Microsoft Excel
 - 5.2 Verwendung in Geoinformationssystemen (GIS)
6. Metadaten, Lizenzierung und Kontakt

1. Allgemeine technische Spezifikationen

- **Zeichenkodierung:** Unicode UTF-8 mit Byte Order Mark (BOM).
- **Feldtrennzeichen (CSV):** Pipe-Symbol (|). Dieses Trennzeichen wird verwendet, um Konflikte mit standardmäßig in historischen Freitexten vorkommenden Kommata oder Semikola zu vermeiden.
- **Interner Listentrenner:** Doppelkreuz (#). Mehrere Werte innerhalb einer Tabellenzelle (z. B. eine Liste von Rechtsbasen oder Quellennachweisen) werden durch dieses Zeichen voneinander abgegrenzt.
- **Georeferenzierung:** Geografische Koordinaten liegen im Textformat *Well-Known Text* (WKT, EPSG:4326 - WGS 84) sowie nativ innerhalb der GeoJSON-Geometrieobjekte vor.

2. Das relationale Datenmodell (CSV-Dateien)

Die Tabellenstruktur der CSV-Dateien folgt einer klassischen 1:n-Relation (Ein Ort zu vielen Ereignissen). Diese Struktur trennt statische Stammdaten von dynamischen chronologischen Ereignissen, um Datenredundanz zu vermeiden.

Das primäre Verknüpfungselement ist das Feld `uuid_ort` (als Universally Unique Identifier). Diese Kennung ist unveränderlich und bildet die Schnittstelle zwischen beiden Tabellen.

- **Tabelle 1 (Orte):** Enthält jeden erfassten historischen Ort exakt einmal. Sie bildet das strukturelle Fundament.
- **Tabelle 2 (Ereignisse):** Bildet die Chronik ab. Jeder Datensatz entspricht einem spezifischen historischen Ereignis. Hat ein Ort mehrere dokumentierte Ereignisse, existieren entsprechend viele Zeilen mit identischer `uuid_ort`.

3. Spezifikation der CSV-Dateien

3.1 Tabelle 1: Ortsbasierte Stammdaten

- **Dateiname:** `reichsstadt_ort_mastertabelle.csv`
- **Struktur:** Eine Zeile entspricht einem eindeutigen historischen Ort.

Spaltenname	Datentyp	Beschreibung
<code>uuid_ort</code>	Text (UUID)	Primärschlüssel. Eindeutige Kennung des Ortes zur Verknüpfung mit Tabelle 2.
<code>ort_name</code>	Text	Offizieller historischer Name des Ortes bzw. der Reichsstadt.

Spaltenname	Datentyp	Beschreibung
geom_wkt	Geometrie (WKT)	Geografische Koordinate im Format POINT(X Y).
staedtetext	Text	Umfassender historischer Beschreibungstext zur jeweiligen Stadt.
rechtsbasen	Text (Liste)	Zugeordnete Rechtsbasen, alphabetisch sortiert (getrennt durch #).
literatur	Text (Liste)	Bibliographische Angaben zum Ort (einzelne Titel getrennt durch #).

3.2 Tabelle 2: Ereignisbasierte Detaildaten

- **Dateiname:** reichsstadt_ereignis_mastertabelle.csv
- **Struktur:** Eine Zeile entspricht einem einzelnen historischen Ereignis.

Spaltenname	Datentyp	Beschreibung
uuid_ort	Text (UUID)	Fremdschlüssel. Verweist auf den zugehörigen Datensatz in Tabelle 1.
ort_name	Text	Name des Ortes zur direkten Orientierung und Filterung.
geom_wkt	Geometrie (WKT)	Geografische Koordinate des Ortes (erlaubt direkte Kartierung).
ereignis_jahr_zeitraum	Text	Bereinigtes Jahr oder Zeitraum des Ereignisses (z. B. 1348 oder 1350-1355).
ereignis_zeitpunkt_roh	Text	Unbereinigter chronologischer Originaleintrag aus der Quelldatenbank.
ereignis_anfang_roh	Ganzzahl	Strukturiertes Startjahr des Ereignisses (falls zutreffend).

Spaltenname	Datentyp	Beschreibung
ereignis_ende_roh	Ganzzahl	Strukturiertes Endjahr des Ereignisses (falls zutreffend).
kategorie	Text	Thematische Zuordnung des Ereignisses (z. B. "Privilegierung", "Verpfändung").
kategorie_ist_reichsstadt	Text	Indikator, ob das Ereignis den Status als Reichsstadt begründet oder beendet.
ereignis_beschreibung	Text	Detaillierte inhaltliche Beschreibung des historischen Vorfalles.
ereignis_spezifische_belege	Text (Liste)	Exakte Quellennachweise und Belegstellen für dieses Ereignis (getrennt durch #).

4. Spezifikation des GeoJSON-Gesamtdatensets

- **Dateiname:** reichsstadt_gesamt.geojson
- **Struktur:** Standardisiertes hierarchisches Geodaten-Format gemäß RFC 7946.

Das GeoJSON-Format führt die relational getrennten Daten der CSV-Dateien in einem einzigen, geschachtelten Dokument zusammen. Die Datei ist als FeatureCollection aufgebaut, bei der jedes Objekt (Feature) eine Stadt repräsentiert.

Hierarchischer Aufbau der Property-Attribute:

- **type:** "Feature"
- **geometry:** Punktgeometrie (POINT)
- **properties:**
 - **uuid_ort** (*Text*): Eindeutige Kennung.
 - **ort_name** (*Text*): Historischer Name.
 - **staedtetext** (*Text*): Beschreibungstext der Stadt.
 - **rechtsbasen** (*JSON-Array*): Liste der zugeordneten Rechtsbasen als eigenständige Strings.
 - **literatur** (*JSON-Array*): Bibliographie als Liste einzelner Strings.
 - **ereignisse** (*JSON-Array von Objekten*): Chronologische Liste aller Ereignisse. Jedes eingebettete Ereignis-Objekt enthält die Attribute:

- **beschreibung** (*Text*), **kategorie** (*Text*), **jahr_zeitraum** (*Text*), **zeitpunkt_roh** (*Text*), **anfang_roh** (*Ganzzahl*), **ende_roh** (*Ganzzahl*), **ist_reichsstadt** (*Boolean/Text*), **belege** (*Text*).

5. Nutzungshinweise und Software-Kompatibilität

5.1 Import-Richtlinien für Microsoft Excel

Aufgrund der enthaltenen Sonderzeichen, Zitate und Trennzeichen innerhalb der Freitexte führt das Öffnen per Doppelklick zu Formatierungsfehlern. Folgender Import-Pfad ist anzuwenden:

1. Ein leeres Excel-Dokument öffnen.
2. Im Menüband den Reiter **Daten** auswählen und auf **Aus Text/CSV** klicken.
3. Die entsprechende CSV-Datei auswählen.
4. Im Konfigurationsfenster folgende Parameter zwingend einstellen:
 - **Dateiursprung:** 65001: Unicode (UTF-8)
 - **Trennzeichen:** Benutzerdefiniert wählen und das Pipe-Symbol (|) manuell eintragen.
5. Auf **Laden** klicken.

5.2 Verwendung in Geoinformationssystemen (GIS)

- **CSV-Dateien:** Beim Import in GIS-Software (z. B. QGIS oder ArcGIS) müssen Geometrie-Importe über die Spalte **geom_wkt** definiert werden. Zur gemeinsamen Analyse der Ereignisse muss ein Tabellen-Join über das gemeinsame Feld **uuid_ort** durchgeführt werden.
- **GeoJSON-Datei:** Die Datei kann ohne Transformation oder Join-Operationen direkt per Drag-and-Drop in ein GIS geladen werden. Alle Sachdaten, inklusive der verschachtelten Ereignis-Objekte, sind direkt über die Geometrien abrufbar.

6. Metadaten, Lizenzierung und Kontakt

- **Metadaten-Standard:** Die GeoJSON-Datei enthält auf der obersten Ebene das Objekt "metadata", welches vollständig dem **DataCite-Metadaten-Schema (Version 4)** entspricht. Es enthält maschinenlesbare Angaben zum Projekt, den beteiligten Personen des Instituts für vergleichende Städtegeschichte (IStG) sowie institutionelle Identifikatoren (ROR).
- **Nutzungslizenz:** Die Daten stehen unter der Lizenz **Creative Commons Attribution Share Alike 4.0 International (CC-BY-SA-4.0)**.
- **Kontakt:** Tobias Runkel, IStG [tobias.runkel@uni-muenster.de]