Metadaten

Datenbereinigung und –anreicherung mit OpenRefine

Index

Datenbereinigung und -anreicherung

- OpenRefine – Kurze Vorstellung von Oberfläche und Bedienung
- Praxisteil mit Beispiel-Datensatz

- Optional:
  Datenvorbereitung für Wikidata
  Arbeiten mit eigenen Daten

WS-Evaluierung

Datenbereinigung und -anreicherung

Daten-erschließungs-system

Datenbereinigung / -anreicherung

OpenRefine u.a.

standardisiertes Austauschformat

Zielsystem, Nutzer etc.

OpenRefine

„... working with messy data“

Open Refine

... ist eine Desktop-Anwendung, die Ihren Browser zur Darstellung nutzt
=> Sie arbeiten lokal, Ihre Daten bleiben lokal

- ursprünglich von Google entwickelt, mittlerweile Open Source
- im Workshop verwendete Version: OpenRefine v2.8 Release
- Quellcode und Nutzer-Dokumentation: https://github.com/OpenRefine/

Open Refine

Wofür:

„OpenRefine is a power tool that allows you to load data, understand it, clean it up, reconcile it, and augment it with data coming from the web.“

- inhaltlichen Datenüberblick verschaffen
- Inkonsistenzen erkennen / Schreibweisen vereinheitlichen
- Inhalte gruppiert verändern / ergänzen
- Datenstruktur verändern
- Daten mittels externer Quellen abgleichen / anreichern
- Formatkonvertierungen (z.B. CSV => JSON)

Open Refine

Input-Formate:

- „TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.“

- **Empfehlung**: zur Bearbeitung tabellarischer Datenstruktur verwenden um unkomplizierten, verlustfreien Export zu gewährleisten

Open Refine

Unterschiede zu Excel etc.:

- dient Datenbearbeitung, nicht Datenverwaltung
- Funktionen nicht in Tabellenzellen gespeichert => d.h. Funktionen werden nach Nutzereingabe **einmalig** auf Daten ausgeführt
- ausführlichere (selbsterklärende) grafische Oberfläche zur Benutzerführung
- breites Spektrum an Exportformaten mittels Templating (für Fortgeschrittene)

Open Refine

Fahrplan:

- eigenes Projekt anlegen mit Beispiel-CSV
- Grundfunktionalitäten kennenlernen
- Datenset kennenlernen
- gemeinsame Überarbeitung ausgewählter Datenelemente
- CSV-Export zur Weiterverwendung