
Metadaten III

Daten bereinigen und anreichern mit OpenRefine

Referentin: Anna-Lena Nowicki

digiS – Forschungs- und Kompetenzzentrum Digitalisierung Berlin |  @digiSberlin | 30.09.2021

Index



- ▶ OpenRefine – was ist das und wofür?
- ▶ Eigenes Projekt anlegen mit Beispiel-CSV
- ▶ Bedienoberfläche kennenlernen
- ▶ Tutorial vorstellen
- ▶ Ggf. tiefer in die Daten einsteigen

OpenRefine



- ... 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 3.4.1
- ▶ Quellcode und Nutzer-Dokumentation:
<https://github.com/OpenRefine/>

OpenRefine



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
- Inhalte (gruppiert) vereinheitlichen / verändern / ergänzen
- Datenstruktur verändern
- Daten mit externen Quellen abgleichen / anreichern
- Formatkonvertierungen (z.B. CSV => JSON)

OpenRefine



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

OpenRefine



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)

Datenbereinigung und -anreicherung



Refine^{OPEN}  u.a.

Fragen?