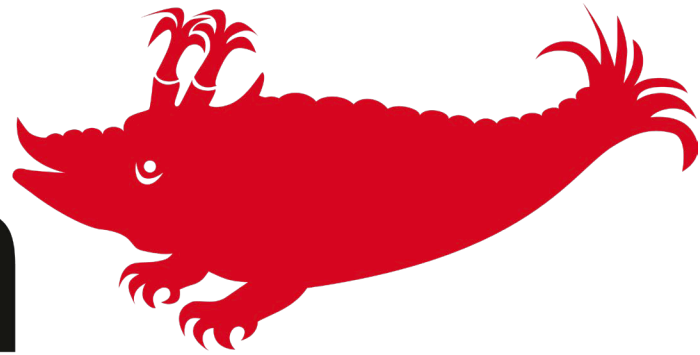


film
museum





HOW I STOPPED WORRYING AND LEARNED TO LOVE MESSY DATA.

OpenRefine in Datenanalyse, Datenbereinigung
und Datenabgleich

Anna Högner, Österreichisches Filmmuseum

Kulturpool Stakeholder Forum 2026

5.-6. Mai 2026, kärnten.museum, Klagenfurt

Noch messy oder schon dirty data? Datenqualitätsprobleme

- Inkonsistente Werte
- Dubletten
- Mangelnde Validierung und Pflege (“Data Janitor”)
- Datensilos

OpenRefine

“OpenRefine ist ein Power Tool, das es ermöglicht, Daten zu laden, zu verstehen, zu bereinigen, abzugleichen und mit Daten aus dem Internet anzureichern.”

OpenRefine

- Open Source
- Läuft lokal im eigenen Browser
- Nutzt den eigenen Arbeitsspeicher –
effizient für große Datensätze

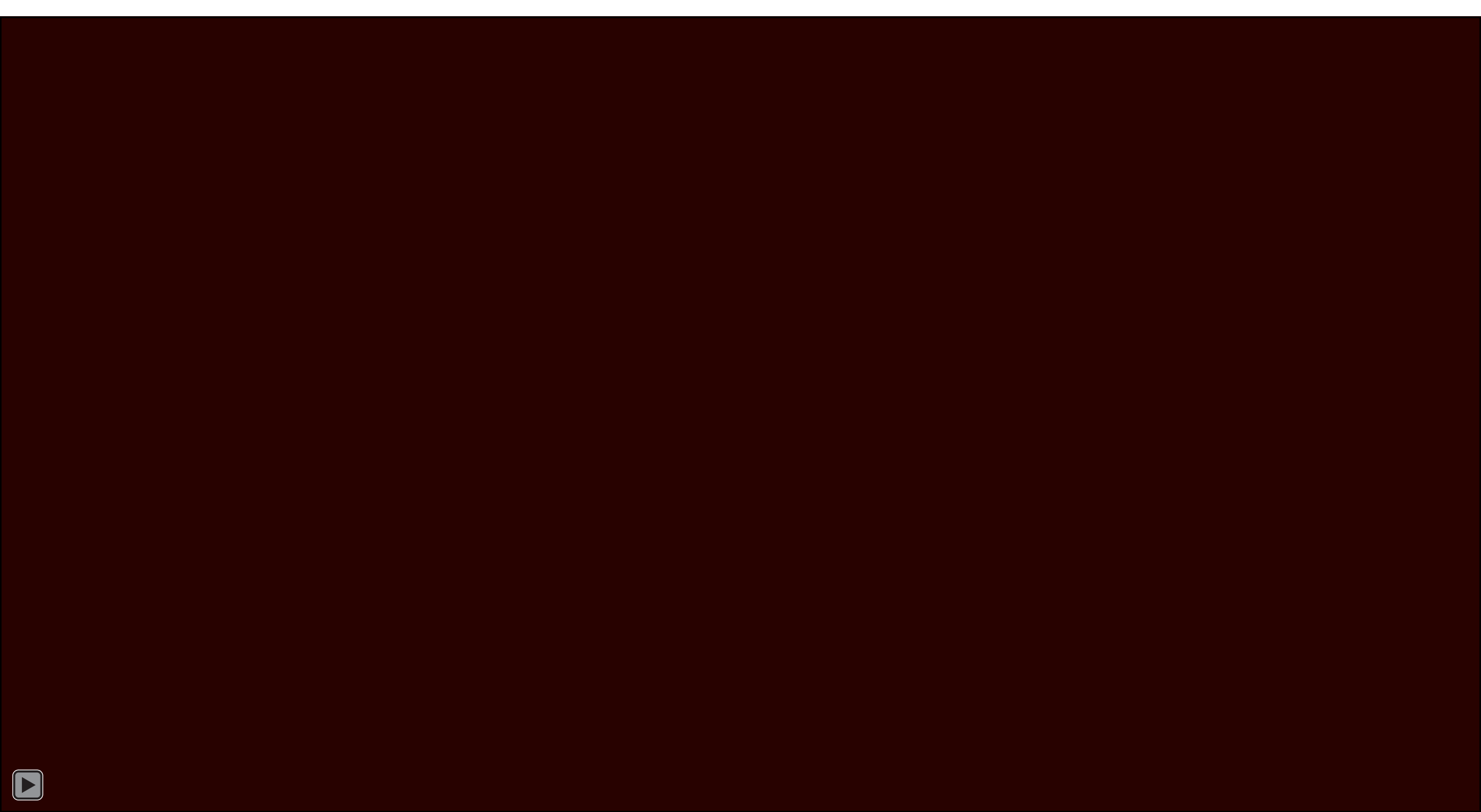
Daten laden...

- Unterstützt die meisten gängigen Austauschformate: CSV, TSV, TXT, .xlsx, JSON, XML...



Daten verstehen und bereinigen

- Facettierung von Werten
- Unterscheidet zwischen Reihe und Datensatz
- “Expressions” erlauben komplexe Such-, Modellierungs und Transformationsoperationen





Daten anreichern

- Anreicherung mit Identifiern aus öffentlichen Datenrepositorien (GND, wikidata, Getty)
- Anreicherung mit Daten aus anderen öffentlichen Repositorien
- Validierung von Daten



Vielen Dank für Ihre
Aufmerksamkeit!

Anna Högner
a.hoegner@filmmuseum.at

