

Datenbereinigung mit OpenRefine



OpenRefine

Jakob Frohmann

j.frohmann@ub.uni-frankfurt.de

Spielregeln

- Eigenes Mikrofon: **AUS**
- Eigenes Video: **AN** oder **AUS**
- Buttons:



Alles ok



Halt, ich
komme
nicht mit!



langsamer



schneller



- Fragen bitte nur im Chat

Am besten nutzen Sie zwei Bildschirme, falls dies möglich ist.

Vorbereitung | Hinweise

- Melden Sie sich für den moodle-Kurs [Praxislabor Digitale Geisteswissenschaften](#) an und schreiben Sie sich ein:

Praxislabor Digitale Geisteswissenschaften: Einführungsworkshops zu Methoden und Werkzeugen der Digital Humanities

Praxislabor Digitale Geisteswissenschaften:
Einführungsworkshops zu Methoden und Werkzeugen der Digital Humanities

Zum ersten Kennenlernen von Methoden und Werkzeugen der Digital Humanities bietet die Universitätsbibliothek JCS (im Bibliothekszentrum Geisteswissenschaften) Studierenden und Mitarbeiterinnen der Goethe-Uni im kommenden Wintersemester Workshops an. In niederschweligen Einführungen werden anhand von überschaubaren, konkreten Beispielen aus der Praxis Methoden, Tools oder Themen der digitalen Geisteswissenschaften vorgestellt und geübt und so ein erster Einblick in die Möglichkeiten gegeben, wie klassische Methoden der Geisteswissenschaften mithilfe digitaler Verfahren der Textanalyse sowie der Text- und Datenaufbereitung sinnvoll ergänzt werden können.

Die Workshopreihe besteht jeweils aus inhaltlich zusammenhängenden Zweierblöcken, in denen auf eine Präsentation eine Sitzung zur Vertiefung und Übung folgt.

Die Workshops richten sich an interessierte Einsteiger; besondere Kenntnisse werden nicht vorausgesetzt. Nähere Informationen sowie die Möglichkeit zur Anmeldung finden Sie unter: <http://www.ub.uni-frankfurt.de/digitalhumanities/workshops.html>.

Praxislabor Digitale Geisteswissenschaften: Einführungsworkshops zu Methoden und Werkzeugen der Digital Humanities

Startseite / Kurse / Verschiedenes / Praxislabor Digitale Geisteswissenschaften

Allgemeines

In dieses kollaborative Dokument können Themenvorschläge für die Hands-on-Sessions eingetragen werden. Informationen zur Anmeldung und Kurszeiten unter: <http://www.ub.uni-frankfurt.de/digitalhumanities>

Einführung in TEI / XML

Dozentin: Agnes Brauer

Der Workshop führt in die Grundlagen der Textauszeichnung mit TEI ein, einer XML-basierten und sich mittlerweile als De-facto-Standard etablierten Auszeichnungssprache speziell für die Zwecke der Geisteswissenschaften. Nach einer knappen allgemeinen Einführung werden die Teilnehmer anhand einer kleinen Übung die Praxis der Textauszeichnung mit TEI kennenlernen und sich einen ersten Überblick über die Bedeutung und die verschiedenen Module dieser Sprache verschaffen.

Link: <http://www.tei-c.org/>

Hands-on Übung zur TEI/XML-Einführung

Dozentin: Agnes Brauer

Einführungsworkshops zu Methoden und Werkzeugen der Digital Humanities

Praxislabor Digitale Geisteswissenschaften:
Einführungsworkshops zu Methoden und Werkzeugen der Digital Humanities

Zum ersten Kennenlernen von Methoden und Werkzeugen der Digital Humanities bietet die Universitätsbibliothek JCS (im Bibliothekszentrum Geisteswissenschaften) Studierenden und Mitarbeiterinnen der Goethe-Uni im kommenden Wintersemester Workshops an. In niederschweligen Einführungen werden anhand von überschaubaren, konkreten Beispielen aus der Praxis Methoden, Tools oder Themen der digitalen Geisteswissenschaften vorgestellt und geübt und so ein erster Einblick in die Möglichkeiten gegeben, wie klassische Methoden der Geisteswissenschaften mithilfe digitaler Verfahren der Textanalyse sowie der Text- und Datenaufbereitung sinnvoll ergänzt werden können.

Die Workshopreihe besteht jeweils aus inhaltlich zusammenhängenden Zweierblöcken, in denen auf eine Präsentation eine Sitzung zur Vertiefung und Übung folgt.

Die Workshops richten sich an interessierte Einsteiger; besondere Kenntnisse werden nicht vorausgesetzt. Nähere Informationen sowie die Möglichkeit zur Anmeldung finden Sie unter: <http://www.ub.uni-frankfurt.de/digitalhumanities/workshops.html>

Trainerin: Agnes Brauer
Trainerin: Jakob Frohmann

Selbsteinschreibung (Teilnehmer/in)

Kein Einschreibeschlüssel notwendig

EINSCHREIBEN

Vorbereitung | Hinweise

Laden Sie bitte **OpenRefine Version 3.4.1** herunter (93MB) und entpacken Sie die Software auf Ihrem Computer.

Sie benötigen einen **Browser** (am besten Firefox, Chrome oder Safari) sowie eine Java-Umgebung zum Ausführen des Programms. Bitte installieren Sie **Java** separat oder nutzen Sie den Download mit eingebettetem Java (nur für Windows).

<http://openrefine.org/download.html>

<https://github.com/OpenRefine/OpenRefine/releases>

<https://openjdk.java.net/>

Vorbereitung | Hinweise

- Starten Sie die Anwendung aus dem entpackten Verzeichnis, es öffnen sich eine Kommandozeile und kurz danach der Browser mit dem geladenen Programm – sollte der Browser nicht starten, benutzen Sie bitte den Link: <http://127.0.0.1:3333/>.
- Spracheinstellung für den Kurs: Deutsch
(in der linken Seitenspalte unter „Language Settings“, Achtung: Übersetzung teilweise schlecht)
- Tragen Sie bitte bei Bedarf / Interesse Themenvorschläge für die Hands-on Übung in das [kollaborative Dokument](#) ein.

„A power tool for working with messy data.“

Open Source-Software in Java

(zuvor zwischenzeitlich zu Google gehörig unter dem Namen Google Refine)

Interaktives Werkzeug zum Bearbeiten, Erkunden und Bereinigen großer Mengen von **Daten in Tabellenform**

Hat große Ähnlichkeiten zu einem Tabellenkalkulationsprogramm mit Zeilen und Spalten (z.B. Excel), aber funktioniert eher wie eine relationale **Datenbank**

→ ein „OpenRefine-Projekt“ = eine Tabelle.

Läuft **lokal** auf dem Rechner im **Browser** (keine Internetverbindung notwendig).

Anwendungsmöglichkeiten und Basisfunktionen

Reinigung und Aufbereitung von Daten

- Korrektur von häufigen Tippfehlern
- Änderungen von Formaten, Normalisierungen (z.B. Groß-/Kleinschreibung),
- Angleichen unterschiedlicher Schreibweise usw.

Daten zerlegen und neu strukturieren

- Inhalte eines Tabellenfeldes in einzelne Bestandteile zerlegen und in eigene Spalten verschieben (z.B. die Bestandteile einer Adresse)

Umstrukturierung von Daten

Anwendungsmöglichkeiten und Basisfunktionen

Einfache und kontrollierte Überführung von Daten in eine **andere Form / andere Dateiformate**

- z.B. von .tsv zu .xlsx

„Exploration“ von Daten

- Aufspüren von Inkonsistenzen oder Fehlern im Datenformat („the big picture of your data“)

Typische Bearbeitungsbeispiele

Data you have	Desired data
London	London
London]	London
London,]	London
london	London

<https://librarycarpentry.org/lc-open-refine/03-working-with-data/index.html>

Typische Bearbeitungsbeispiele

Address in single field	Institution	Library name	Address		Town/City	Region	Country	Postcode
			Address 1	2				
University of Wales, Llyfrgell Thomas Parry Library, Llanbadarn Fawr, ABERYSTWYTH, Ceredigion, SY23 3AS, United Kingdom	University of Wales	Llyfrgell Thomas Parry Library	Llanbadarn Fawr		Aberystwyth	Ceredigion	United Kingdom	SY23 3AS
University of Aberdeen, Queen Mother Library, Meston Walk, ABERDEEN, AB24 3UE, United Kingdom	University of Aberdeen	Queen Mother Library	Meston Walk		Aberdeen		United Kingdom	AB24 3UE
University of Birmingham, Barnes Library, Medical School, Edgbaston, BIRMINGHAM, West Midlands, B15 2TT, United Kingdom	University of Birmingham	Barnes Library	Medical School	Edgbaston	Birmingham	West Midlands	United Kingdom	B15 2TT
University of Warwick, Library, Gibbett Hill Road, COVENTRY, CV4 7AL, United Kingdom	University of Warwick	Library	Gibbett Hill Road		Coventry		United Kingdom	CV4 7AL

<https://librarycarpentry.org/lc-open-refine/03-working-with-data/index.html>

Typische Bearbeitungsbeispiele

Data you have	Desired data
1st January 2014	2014-01-01
01/01/2014	2014-01-01
Jan 1 2014	2014-01-01
2014-01-01	2014-01-01

<https://librarycarpentry.org/lc-open-refine/03-working-with-data/index.html>

Grundfunktionen

Werkzeuge in OpenRefine:

- verschiedene Arten von **Filtern** und **Facetten** mit vordefinierten Kriterien zum Anpassen der Anzeige

Alle Operationen geschehen i.d.R. nur auf den gewählten / **selektierten Daten**.

Alle Veränderung der Daten geschehen auf einer **Kopie des Datensets** und können wieder rückgängig gemacht werden.

- “Play with your data”

Abfolgen von **Operationen können gespeichert** werden und dann auf andere Datensätze ebenfalls angewendet werden.

Anwendungsmöglichkeiten – fortgeschrittene Funktionen

Daten abgleichen (eng. reconcile & match)

- Vergleichen / Angleichen der eigenen Daten anhand von Datenbanken (z.B. Wikidata)
- Anreicherung der eigenen Daten mit externen Quellen
(z.B. mit eindeutigen Identifikatoren)
- Verlinkung von Daten

... mit Hilfe von Webservices mit offenen Schnittstellen (APIs) über das Internet.

OpenRefine: Daten importieren/exportieren

- mögliche Dateiformate für den **Import** (Auswahl): TXT, TSV, CSV, Microsoft Excel, JSON, XML, RDF-XML, Google Data

- mögliche Dateiformate für den **Export** (Auswahl): TSV, CSV, Microsoft Excel, HTML (Tabelle)

OpenRefine_Hands_on_20190116-Hiltmann-Frohmann - Editor

Datei Bearbeiten Format Ansicht ?

Diese Daten wurden dem Interaktiven Katalog des Münzkabinetts der Staatlichen Museen zu Berlin entnommen (<https://ikmk.smb.museum/>) und für die Hands-on Session noch etwas verändert (verunreinigt). Es sind jene Münzen, die mit den Stichworten "Mittelalter", "Deutschland" und "Heraldik / Wappen" versehen sind. (Torsten Hiltmann)

Brandenburg: Johann I. und Otto III.Brandenburg: Johann I. und Otto III.

Brandenburg: Johann I. und Otto III.

(Germany, Brandenburg)

Denar (MA), um 1250

18206726

Waldeck-Pyrmont: GrafschaftWaldeck-Pyrmont: Grafschaft

Waldeck-Pyrmont: Grafschaft

Lügde (Germany, Westfalen)

Denar (MA), 1250-1300

18213795

Brandenburg: MarkgrafschaftBrandenburg: Markgrafschaft

Brandenburg: Markgrafschaft

(Germany, Brandenburg)

Denar (MA), um 1275

18206724

Schlotheim: Herren

Schlotheim: Herren

Schlotheim (Germany, Thüringen)

Brakteat, 1280-1290

18215954

Brandenburg: MarkgrafschaftBrandenburg: Markgrafschaft

Brandenburg: Markgrafschaft

(Germany, Brandenburg)

Denar (MA), um 1290

18206725

Köln: Erzbistum und Waldeck-Pyrmont: GrafschaftKöln: Erzbistum und Waldeck-Pyrmont: Grafschaft

Köln: Erzbistum und Waldeck-Pyrmont: Grafschaft

Lügde (Germany, Westfalen)

Denar (MA), 1297-1304

18213796

Köln: Erzbistum und Waldeck-Pyrmont: GrafschaftKöln: Erzbistum und Waldeck-Pyrmont: Grafschaft

Köln: Erzbistum und Waldeck-Pyrmont: Grafschaft

Lügde (Germany, Westfalen)

Denar (MA), 1297-1304

18213798

Köln: Erzbistum und Waldeck-Pyrmont: GrafschaftKöln: Erzbistum und Waldeck-Pyrmont: Grafschaft

Kön: Erzbistum und Waldeck-Pyrmont: Grafschaft

Lügde (Germany, Westfalen)

Denar (MA), 1297-1304

18213801

Montabaur: StadtMontabaur: Stadt

Grundsätzliche Arbeitsweise mit Open Refine

Grundsätzlich: Die Daten herausfiltern, welche man verändern möchte, dann die selektierten Daten in einer Operation bearbeiten.

Mögliche Vorgehensweise:

- Quelldaten importieren
- Daten analysieren
- Daten selektieren (z.B. spaltenweise/zeilenweise oder nach bestimmten Kriterien)
- Daten bearbeiten Daten anreichern
- Daten im gewünschten Zielformat wieder exportieren

Siehe auch: <https://www.tanner.de/blog/systeme-technologien/datenanalyse-und-optimierung-mit-openrefine-1/>

Ein Projekt erstellen...

← → ↻ ⓘ 127.0.0.1:3333 ☆ ⌵ 🏠 👤 ...

OpenRefine Ein leistungsstarkes Werkzeug für die Bearbeitung von ungeordneten Daten.

- Projekt erstellen
- Projekt öffnen
- Projekt importieren
- Spracheinstellungen

Erstellen Sie ein Projekt, um Daten zu importieren. Welche Art von Daten können Sie importieren?
TSV, CSV, *SV, Excel (.xls und .xlsx), JSON, XML, RDF als XML, und Google Data Dokumente werden unterstützt. Die Unterstützung für andere Formate kann mit OpenRefine-Erweiterungen hinzugefügt werden.

Daten abrufen von

Suchen Sie eine oder mehrere Dateien auf Ihrem Computer, die Sie hochladen möchten:

Dieser Computer OpenRefin..._DNB.csv

Nächste »

Webadressen (URLs)

Zwischenablage

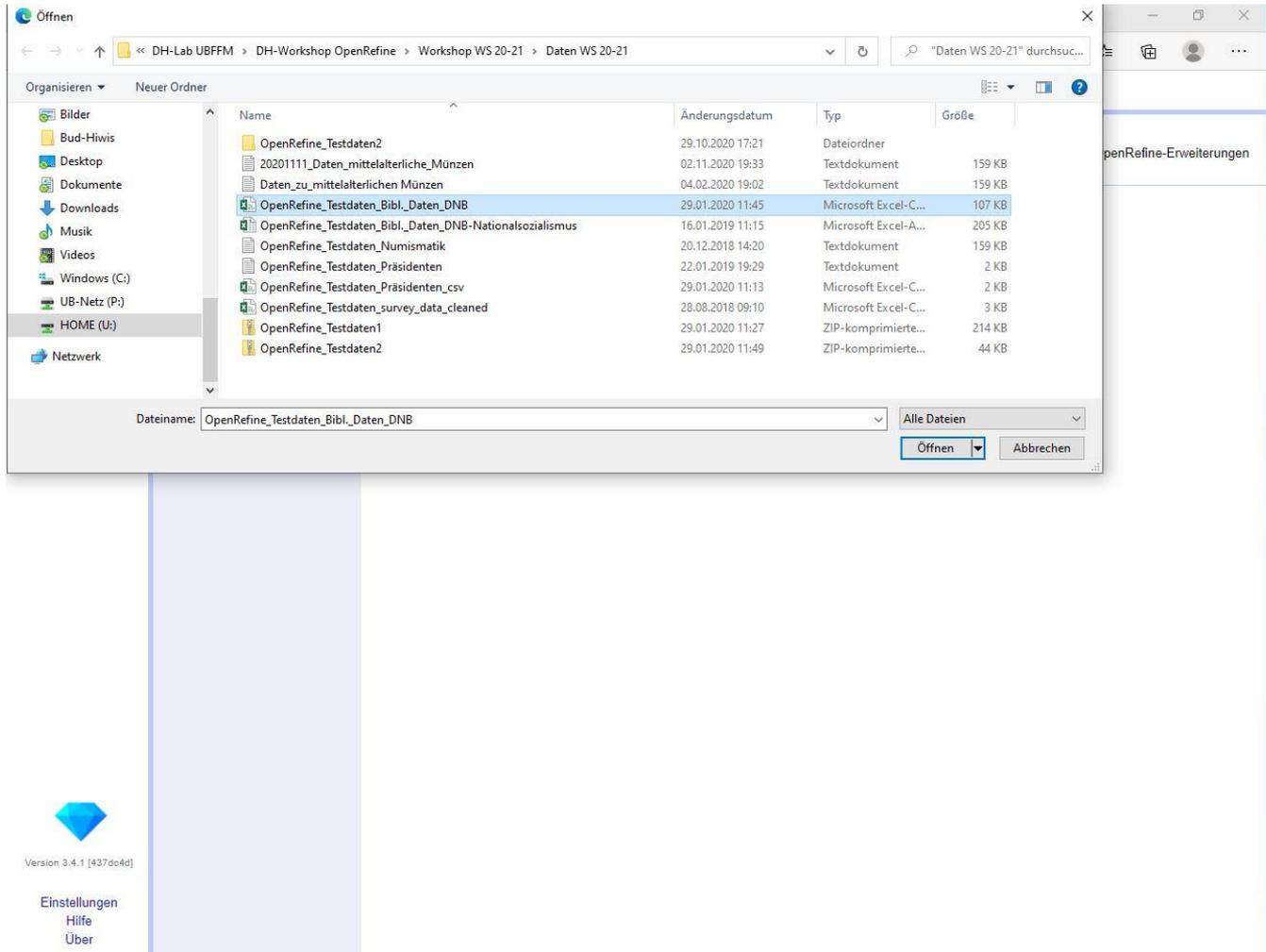
Database

Google Data


Version 3.4.1 [437dc4d]

Einstellungen
Hilfe
Über

Ein Projekt erstellen...



Ein Projekt erstellen...

OpenRefine Ein leistungsstarkes Werkzeug für die Bearbeitung von ungeordneten Daten.

Projekt erstellen «Von vorne anfangen Parsing-Optionen konfigurieren Projekt name OpenRefine_Testdaten_Bibl_Date Tags Projekt erstellen »

	subject	identifier	type	creator	title	volume	edition	publisher	year	format	ISBN	ISSN	ISMN	EAN/UPC
1.		IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.:] 1851			Frankfurt a. M. : Krebs-Schmitt	1850	58 Bl. ; gr. 8				
2.		IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J[ohann] B[aptist] Scholl. Mit begleit. Text v. Dornroder			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2				
3.		IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkblümlein / v. Heinrich Zschokke			Frankfurt, M. : J. D. Sauerländer's Vert.	1850	VIII, 142 S., Front., [2] Bl. Verfl. Anz. ; kl. 8				
4.		IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Dänemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frankfurt a. M. : Brönnner	1851	72 S. ; kl. 8				
5.		IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frankfurt a. M. : [s. n.]	1851	LXIV, 338 S. ; 8				
6.	59 Belletristik	ISBN-978-3-628-39627-4, 3-628-39627-1,		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche-Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3-628-39627-			

Daten analysieren als

Zeichen kodierung Aktualisieren Vorschau

Spalten werden getrennt durch

Kommas (CSV)

Tabs (TSV)

benutzerdefiniert: ;

Führende & nachgestellte Leerzeichen aus Zeichenfolgen entfernen

Markiere Sonderzeichen mit \

Spaltennamen (durch Kommata getrennt):

Erste ignorieren 0 Zeile(n) am Dateianfang

Nächste analysieren 1 Zeile(n) als Spaltenüberschriften

Anfängliche verwerfen 0 Datenzeile(n)

Maximale Last 0 Datenzeile(n)

Zeichen verwenden " zum Einschließen von Zellen mit Spaltentrennern

Analysiere Zelltext in Zahlen, Datumsangaben, ...

Leere Zeilen speichern

Speichern von leeren Zellen als null-Werte

Dateiquelle in jeder Zeile speichern

Version 3.4.1 [437do4d]

Einstellungen Hilfe Über

Ein Projekt erstellen...

OpenRefine Ein leistungsstarkes Werkzeug für die Bearbeitung von ungeordneten Daten.

Projekt erstellen «Von vorne anfangen Parsing-Optionen konfigurieren Projekt name OpenRefine_Testdaten_Bibl_Date Tags Projekt erstellen »

	subject	identifier	type	creator	title	volume	edition	publisher	year	format	ISBN	ISSN	ISMN	EAN/UPC
1.		IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.:] 1851			Frankfurt a. M. : Krebs-Schmitt	1850	58 Bl. ; gr. 8				
2.		IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J[ohann] B[aptist] Scholl. Mit begleit. Text v. Dornroder			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2				
3.		IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkblümlein / v. Heinrich Zschokke			Frankfurt, M. : J. D. Sauerländer's Vert.	1850	VIII, 142 S., Front., [2] Bl. Verfl. Anz. ; kl. 8				
4.		IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Dänemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frankfurt a. M. : Brönnner	1851	72 S. ; kl. 8				
5.		IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frankfurt a. M. : [s. n.]	1851	LXIV, 338 S. ; 8				
6.	59 Belletristik	ISBN-978-3-628-39627-4, 3-628-39627-1,		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche-Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3-628-39627-			

Daten analysieren als **Zeichen kodierung** UTF-8 Aktualisieren Vorschau

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

JSON-LD files

RDF/N3 files

RDF/N-Triples files

RDF/Turtle files

Spalten werden getrennt durch

Kommas (CSV)

Tabs (TSV)

benutzerdefiniert: ;

Führende & nachgestellte Leerzeichen aus Zeichenfolgen entfernen

Markiere Sonderzeichen mit \

Spaltennamen (durch Kommata getrennt):

Erste ignorieren 0 Zeile(n) am Dateianfang

Nächste analysieren 1 Zeile(n) als Spaltenüberschriften

Anfängliche verwerfen 0 Datenzeile(n)

Maximale Last 0 Datenzeile(n)

Zeichen verwenden " zum Einschließen von Zellen mit Spaltentrennern

Analysiere Zelltext in Zahlen, Datumsangaben, ...

Leere Zeilen speichern

Speichern von leeren Zellen als null-Werte

Dateiquelle in jeder Zeile speichern

Ein Projekt erstellen...

OpenRefine Ein leistungsstarkes Werkzeug für die Bearbeitung von ungeordneten Daten.

Projekt erstellen «Von vorne anfangen Parsing-Optionen konfigurieren Projekt name OpenRefine_Testdaten_Bibl_Date Tags Projekt erstellen »

	subject	identifier	type	creator	title	volume	edition	publisher	year	format	ISBN	ISSN	ISMN	EAN/UPC
1.		IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.:] 1851			Frankfurt a. M. : Krebs-Schmitt	1850	58 Bl. ; gr. 8				
2.		IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J[ohann] B[aptist] Scholl. Mit begleit. Text v. Dornroder			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2				
3.		IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkblümlein / v. Heinrich Zschokke			Frankfurt, M. : J. D. Sauerländer's Vert.	1850	VIII, 142 S., Front., [2] Bl. Verfl. Anz. ; kl. 8				
4.		IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Dänemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frankfurt a. M. : Brönnner	1851	72 S. ; kl. 8				
5.		IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frankfurt a. M. : [s. n.]	1851	LXIV, 338 S. ; 8				
6.	59 Belletristik	ISBN-978-3-628-39627-4, 3-628-39627-1,		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche-Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3-628-39627-			

Daten analysieren als

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

JSON-LD files

RDF/N3 files

RDF/N-Triples files

RDF/Turtle files

Zeichen kodierung Aktualisieren Vorschau

Spalten werden getrennt durch

Kommas (CSV)

Tabs (TSV)

benutzerdefiniert: ;

Führende & nachgestellte Leerzeichen aus Zeichenfolgen entfernen

Markiere Sonderzeichen mit \

Spaltennamen (durch Kommata getrennt):

Erste ignorieren 0 Zeile(n) am Dateianfang

Nächste analysieren 1 Zeile(n) als Spaltenüberschriften

Anfängliche verwerfen 0 Datenzeile(n)

Maximale Last 0 Datenzeile(n)

Zeichen verwenden " zum Einschließen von Zellen mit Spaltentrennern

Analysiere Zelltext in Zahlen, Datumsangaben, ...

Leere Zeilen speichern

Speichern von leeren Zellen als null-Werte

Dateiquelle in jeder Zeile speichern

Ein Projekt erstellen...

OpenRefine Ein leistungsstarkes Werkzeug für die Bearbeitung von ungeordneten Daten.

Projekt erstellen «Von vorne anfangen Parsing-Optionen konfigurieren Projekt name OpenRefine_Testdaten_Bibl_Date Tags **Projekt erstellen »**

	subject	identifier	type	creator	title	volume	edition	publisher	year	format	ISBN	ISSN	ISMN	LAN	OPC
1.		IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.:] 1851			Frankfurt a. M. : Krebs-Schmitt	1850	58 Bl. ; gr. 8					
2.		IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J[ohann] B[aptist] Scholl. Mit begleit. Text v. Dornroder			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2					
3.		IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkblümlein / v. Heinrich Zschokke			Frankfurt, M. : J. D. Sauerländer's Vert.	1850	VIII, 142 S., Front., [2] Bl. Verfl. Anz. ; kl. 8					
4.		IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Dänemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frankfurt a. M. : Brönnner	1851	72 S. ; kl. 8					
5.		IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frankfurt a. M. : [s. n.]	1851	LXIV, 338 S. ; 8					
6.	59 Belletristik	ISBN-978-3-628-39627-4, 3-628-39627-1,		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche-Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3-628-39627-				

Daten analysieren als

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

JSON-LD files

RDF/N3 files

RDF/N-Triples files

RDF/Turtle files

Zeichen kodierung

Aktualisieren Vorschau

Spalten werden getrennt durch

Kommas (CSV)

Tabs (TSV)

benutzerdefiniert: ;

Führende & nachgestellte Leerzeichen aus Zeichenfolgen entfernen

Markiere Sonderzeichen mit \

Spaltennamen (durch Kommata getrennt):

Erste ignorieren 0 Zeile(n) am Dateianfang

Nächste analysieren 1 Zeile(n) als Spaltenüberschriften

Anfängliche verwerfen 0 Datenzeile(n)

Maximale Last 0 Datenzeile(n)

Zeichen verwenden " zum Einschließen von Zellen mit Spaltentrennern

Analysiere Zelltext in Zahlen, Datumsangaben, ...

Leere Zeilen speichern

Speichern von leeren Zellen als null-Werte

Dateiquelle in jeder Zeile speichern

Ein Projekt erstellen...

← → ↻ 127.0.0.1:3333/project?project=2450891248065 ☆ ⚙️ 📄 👤 ...

OpenRefine OpenRefine_Testdaten_Bibl_Daten_DNB csv [Permalink](#) Öffnen... Export Hilfe

Facette / Filter Rückgängig / Wiederholen 0 / 0

Über Facetten und Filter 

Verwenden Sie Facetten und Filter, um Teilmengen Ihrer Daten auszuwählen. Wählen Sie Facetten- und Filtermethoden aus den Menüs am oberen Rand jeder Datenspalte.

Sie sind sich nicht sicher, wie Sie anfangen sollen? [Sehen Sie sich diese Screencasts an](#)

32 Datensätze Erweiterungen: Wikidata

Anzeigen als: Zeilen **Datensätze** Anzeigen: 5 10 25 50 Datensätze « Erste < Vorherige 1 - 10 nächste > letzte »

▼ Alle	▼ subject	▼ identifier	▼ type	▼ creator	▼ title	▼ volume	▼ edition	▼ publisher	▼ year	▼ format	▼ ISBN	▼ ISSN
☆	1.	IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.] 1851			Frankfurt a. M. : Krebs-Schmitt	1850	58 Bl. ; gr. 8		
☆	2.	IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J[ohann] B[aptist] Scholl. Mit begleit. Text v. Dornroder			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2		
☆	3.	IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkblümlein / v. Heinrich Zschokke			Frankfurt, M. : J. D. Sauerländer's Verl.	1850	VIII, 142 S., Front., [2] Bl. Verl. Anz. ; kl. 8		
☆	4.	IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Dänemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frankfurt a. M. : Bröner	1851	72 S. ; kl. 8		
☆	5.	IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frankfurt a. M. : [s. n.]	1851	LXIV, 338 S. ; 8		
☆	6.	59 Belletristik ISBN-978-3-628-39627-4, 3-628-39627-1, IDN:932277861		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche-Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3-628-39627-4	
☆		IDN:361226438		Meyer, Hermann von [Verfasser]	Über die Reptilien und Säugetiere der verschiedenen Zeiten der Erde : 2 Reden / Hermann von Meyer			Frankfurt a. M. : Schmerber	1852	150 S. ; 8		
☆		IDN:361008236		Irmischer, Johann Konrad [Verfasser]	Handschriften-Katalog der Königlichen Universitäts-Bibliothek zu Erlangen / bearb. von Johann Konrad Irmischer			Frankfurt a. M. : Erlangen : Heyder & Zimmer	1852	469 S. : Mit 2 Taf. Schriftproben ; 8		
☆	7.	59 Belletristik ISBN-978-3-628-39876-6, 3-628-39876-2, IDN:932318401		Trautmann, Franz [Verfasser]	Eppelein von Gailingen, und was sich seiner Zeit mit diesem ritterlichen Friesenpiel und		[Mikrofiche-Ausg.]	Frankfurt am Main : Sauerländer	1852	IV, 145 S. : III.	978-3-628-39876-6	

Ein Projekt erstellen...

← → ↻ 127.0.0.1:3333/project?project=2450891248065 ☆ ⚙️ 🏠 👤 ...

OpenRefine OpenRefine_Testdaten_Bibl_Daten_DNB.csv [Permalink](#) Öffnen... Export Hilfe

Facette / Filter Rückgängig / Wiederholen 0 / 0

32 Datensätze Erweiterungen: Wikidata

Anzeigen als: Zeilen Datensätze Anzeigen: 5 10 25 50 Datensätze « Erste < Vorherige 1 - 10 nächste > letzte »

▼ Alle	▼ subject	▼ identifier	▼ type	▼ creator	▼ title	▼ volume	▼ edition	▼ publisher	▼ year	▼ format	▼ ISBN	▼ ISSN
☆		IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.] 1851			Frankfurt a. M. : Krebs-Schmitt	1850	58 Bl. ; gr. 8		
☆	2.	IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J(ohann) B(aptist) Scholl. Mit begleit. Text v. Dornroder			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2		
☆	3.	IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkblümlein / v. Heinrich Zschokke			Frankfurt, M. : J. D. Sauerländer's Verl.	1850	VIII, 142 S., Front., [2] Bl. Verl. Anz. ; kl. 8		
☆	4.	IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Danemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frankfurt a. M. : Bröner	1851	72 S. ; kl. 8		
☆	5.	IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frankfurt a. M. : [s. n.]	1851	LXIV, 338 S. ; 8		
☆	6.	59 Belletristik ISBN-978-3-628-39627-4, 3-628-39627-1, IDN:932277861		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche-Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3-628-39627-4	
☆		IDN:361226438		Meyer, Hermann von [Verfasser]	Über die Reptilien und Säugetiere der verschiedenen Zeiten der Erde : 2 Reden / Hermann von Meyer			Frankfurt a. M. : Schmerber	1852	150 S. ; 8		
☆		IDN:361008236		Irmischer, Johann Konrad [Verfasser]	Handschriften-Katalog der Königlichen Universitäts-Bibliothek zu Erlangen / bearb. von Johann Konrad Irmischer			Frankfurt a. M. : Erlangen : Heyder & Zimmer	1852	469 S. : Mit 2 Taf. Schriftproben ; 8		
☆	7.	59 Belletristik ISBN-978-3-628-39876-6, 3-628-39876-2, IDN:932318401		Trautmann, Franz [Verfasser]	Eppelein von Gailingen, und was sich seiner Zeit mit diesem ritterlichen Friesenpiel und		[Mikrofiche-Ausg.]	Frankfurt am Main : Sauerländer	1852	IV, 145 S. : III.	978-3-628-39876-6	

Über Facetten und Filter

Verwenden Sie Facetten und Filter, um Teilmengen Ihrer Daten auszuwählen. Wählen Sie Facetten- und Filtermethoden aus den Menüs am oberen Rand jeder Datenspalte.

Sie sind sich nicht sicher, wie Sie anfangen sollen?
[Sehen Sie sich diese Screencasts an](#)

Daten bearbeiten

← → ↻ ⓘ 127.0.0.1:3333/project?project=2450891248065 ☆ ☆ 🗑️ 👤 ⋮

OpenRefine OpenRefine_Testdaten_Bibl_Daten_DNB.csv Permalink Öffnen... Export Hilfe

Facette / Filter Rückgängig / Wiederholen 0/0

32 Datensätze Erweiterungen: Wikidata

Anzeigen als: Zeilen Datensätze Anzeigen: 5 10 25 50 Datensätze « Erste < Vorherige 1 - 10 nächste > letzte »

Alle	subject	identifizier	type	creator	title	volume	edition	publisher	year	format	ISBN	ISSN
1.		Facette Textfilter			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [Hs.] 1851			Frankfurt a. M. : Febs- Schmidt	1850	58 Bl. ; gr. 8		
2.		Zellen bearbeiten Spalte bearbeiten Austauschen Sortieren... Ansicht Abgleichen	Umwandeln... Gemeinsame Umwandlungen Entfernen von führenden und nachstehenden Leerzeichen Aufeinanderfolgende Leerstellen zusammenfassen Leer nach unten HTML-Entitäten unescapen Typografische Anführungszeichen durch ASCII ersetzen In titlecase In Großbuchstaben In Kleinbuchstaben In Nummer In Datum In Text In Null In Leerzeichenkette		... eine für			Frankfurt, M. : Schmerber	1850	[1] Bl. : 6 Taf. m. Abb. ; qu. 2		
3.									1850	VIII, 142 S., Front., [2] Bl. Verl. Anz. ; kl. 8		
4.		IDN:361197144							1851	72 S. ; kl. 8		
5.		IDN:579178315			Bericht des Ausschusses öffentliche ur- mündliche Ve- in Strafsache d. gesetzgeb Versammlung freien Stadt Frankfurt ; N Gesetz - Ent- Apl...				1851	LXIV, 338 S. ; 8		
6.	59 Belletristik	ISBN-978-3-628- 39627-4, 3-628- 39627-1, IDN:932277861 IDN:361226438		Gutzkow, Karl [Verfasser]	Vergangene Tage		[Mikrofiche- Ausg.]	Frankfurt am Main : Literar. Anst.	1852	XXVIII, 211 S.	978-3- 628- 39627-4	
				Meyer, Hermann von [Verfasser]	Über die Reptilien und Säugetiere der verschiedenen Zeiten der Erde ; 2 Reden / Hermann von Meyer			Frankfurt a. M. : Schmerber	1852	150 S. ; 8		
		IDN:361008236		Irmischer, Johann Konrad [Verfasser]	Handschriften- Katalog der Königlichen Universitäts- Bibliothek zu Erlangen / bearb. von Johann Conrad Irmischer			Frankfurt a. M. ; Erlangen : Heyder & Zimmer	1852	469 S. : Mit 2 Taf. Schriftproben ; 8		
7.	59 Belletristik	ISBN-978-3-628- 39876-6, 3-628- 39876-2, IDN:932318401		Trautmann, Franz [Verfasser]	Eppelein von Gailingen, und was sich seiner Zeit mit diesem ritterlichen Fulenspiegel und		[Mikrofiche- Ausg.]	Frankfurt am Main : Sauerländer	1852	IV, 145 S. : Ill.	978-3- 628- 39876-6	

Über Facetten und Filter

Verwenden Sie Facetten und Filter, um Teilmenen Ihrer Daten auszuwählen. Wählen Sie Facetten- und Filtermethoden aus den Menüs am oberen Rand jeder Datenspalte.

Sie sind sich nicht sicher, wie Sie anfangen sollen?
[Sehen Sie sich diese Screencasts an](#)

Daten speichern / exportieren

127.0.0.1:3333/project?project=2450891248065

OpenRefine OpenRefine_Testdaten_Bibl_Daten_DNB csv Permalink

Facette / Filter

Rückgängig / Wiederholen 0 / 0

Über Facetten und Filter

Verwenden Sie Facetten und Filter, um Teilmengen Ihrer Daten auszuwählen. Wählen Sie Facetten- und Filtermethoden aus den Menüs am oberen Rand jeder Datenspalte.

Sie sind sich nicht sicher, wie Sie anfangen sollen?
Sehen Sie sich diese Screenscasts an

32 Datensätze

Anzeigen als: Zeilen **Datensätze** Anzeigen: 5 10 25 50 Datensätze

	Alle	subject	identifier	type	creator	title	volume	edition	p	
1.			IDN:368449920			Tagebuch für das Geschäftsleben : Jg. 8. 1850 [hs.] 1851			Frank M. : K Schm	
2.			IDN:1001287584		Scholl, Johann Baptist [Verfasser] ; Dornroder, ... [Mitwirkender]	Lebens-Bilder : eine Frühlingsgabe für das deutsche Volk / Entworfen u. auf Stein gezeichnet v. J[ohann] B[aptist] Scholl. Mit begleit. Text v. Dornroder			Frank M. : B	
3.			IDN:1003764746		Zschokke, Heinrich [Verfasser]	Feldblumen : eine andere Selbstschau in poet. Gedenkbüchlein / v. Heinrich Zschokke			Frank J. D. Sauerl. Verl.	
4.			IDN:361197144		Marggraff, Hermann [Verfasser]	Trutz Dänemark! : Dt. Lieder f. Schleswig-Holstein / Ges. u. hrsg. H. Marggraff			Frank M. : B	
5.			IDN:579178315			Bericht des Ausschusses für das öffentliche und mündliche Verfahren in Strafsachen : An d. gesetzgebende Versammlung d. freien Stadt Frankfurt ; Nebst Gesetz - Entwurf u. Anlagen A. bis E. desselben			Frank M. : [S...]	
6.	59 Belletristik		ISBN:978-3-628-39627-4, 3-628-39627-1, IDN:932277861		Gutzkow, Karl [Verfasser]	Vergangene Tage	[Mikrofiche-Ausg.]		Frankfurt am Main : Literar. Anst.	1852 XXVIII, 211 S. 978-3-628-39627-4
			IDN:361226438		Meyer, Hermann von [Verfasser]	Über die Reptilien und Säugetiere der verschiedenen Zeiten der Erde : 2 Reden / Hermann von Meyer			Frankfurt a. M. : Schmerber	1852 150 S. ; 8
			IDN:361008236		Irmischer, Johann Konrad [Verfasser]	Handschriften-Katalog der Königlichen Universitäts-Bibliothek zu Erlangen / bearb. von Johann Konrad Irmischer			Frankfurt a. M. : Erlangen : Heyder & Zimmer	1852 469 S. : Mit 2 Taf. Schriftproben : 8
7.	59 Belletristik		ISBN:978-3-628-39876-6, 3-628-39876-2, IDN:932318401		Trautmann, Franz [Verfasser]	Eppelein von Gallingen, und was sich seiner Zeit mit diesem ritterlichen Filensienel und	[Mikrofiche-Ausg.]		Frankfurt am Main : Sauerländer	1852 IV, 145 S. : Ill. 978-3-628-39876-6

Öffnen... Export

Exportieren nach lokal

Tabulatorgetrennter Wert

Kommagetrennter Wert

HTML-Tabelle

Excel (.xls)

>=Excel 2007 (.xlsx)

ODF-Tabelle

Benutzerdefinierter tabellarischer Exporteur...

SQL-Exporteur...

Vorlagen...

OpenRefine project archive to Google Drive...

Google Sheets

Wikidata edits...

QuickStatements file

Wikidata schema

Achtung:
Ihre Daten gehen verloren, wenn der Browser geschlossen wird, ohne dass das Projekt exportiert wurde!

Nächste Woche gemeinsam...

- Zellen teilen und (wieder) vereinigen mit Hilfe von Separatoren
- Facetieren und Filtern
- Clustern von Daten
- Anreichern von Daten (Beispiele)

Bitte versuchen Sie, die Software vor dem Kurs zum Laufen bringen.

<http://openrefine.org/download.html>

Tipps & Tricks / Links / Literatur

OpenRefine

<http://openrefine.org/>

Ressourcen

<https://www.wikidata.org/> (Daten aller Art, default in OpenRefine)

<https://www.geonames.org/> (Geografika)

<http://swb.bsz-bw.de/DB=2.104/> (Personen, Körperschaften, Konferenzen, Geografika, Sachschlagwörter, Werktitel) [GND - Gemeinsame Normdatei der Deutschen Nationalbibliothek, mehr Infos [hier](#)]

Visualisierung:

<https://geobrowser.de.dariah.eu/>

<http://hdlab.stanford.edu/palladio/>

Tipps & Tricks / Links / Literatur

Blogs

<https://histhub.ch/histhub-lab-tutorials-zu-openrefine/>

(Blog-Serie zur Arbeit an historischen Daten mit OpenRefine)

<http://blog.lobid.org/2018/08/27/openrefine.html> (einfache Anreicherung von Daten in OpenRefine mit [Personen-] Daten aus der GND via lobid.org)

Literatur

Ruben Verborgh/Max de Wilde, Using OpenRefine. The essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the Web (Community experience distilled), Birmingham, Mumbai 2013. [[Online-Ressource über UB FFM](#)]

Danke für Ihre Aufmerksamkeit!

Vielen Dank für die Zurverfügungstellung von Materialien und Daten an Agnes Brauer (Universitätsbibliothek Johann Christian Senckenberg) und Prof. Dr. Torsten Hiltmann (Zentrum für Digitale Geschichtswissenschaft, Universität Münster).

Workshop konzipiert in Anlehnung an "Library Carpentry: OpenRefine"

<https://librarycarpentry.org/lc-open-refine/>

(Licensed under CC-BY 4.0 2016–2020 by [Library Carpentry](https://librarycarpentry.org/))