

# CrossRef

AG Discovery

Gerald Steilen

Göttingen, 10.5.2017

# Was ist Crossref?

- Offizielle „Digital Object Identifier (DOI) Registration Agency“
- Datenbank mit Millionen verknüpfter „Objekte“
  - Artikel
  - Bücher
  - Konferenzbeiträge
  - ....

# Schnittstellen

- OAI-PMH: OAI nur für zahlende Kunden

<https://support.crossref.org/hc/en-us/articles/213679866-OAI-PMH-subscriber-only>

- REST-API: kostenfrei, evtl. Rate Limits

[https://github.com/CrossRef/rest-api-doc/blob/master/rest\\_api.md](https://github.com/CrossRef/rest-api-doc/blob/master/rest_api.md)

# Rest-API

- Dokumentation:

<https://github.com/CrossRef/rest-api-doc>

- Format: JSON

[https://github.com/CrossRef/rest-api-doc/blob/master/api\\_format.md](https://github.com/CrossRef/rest-api-doc/blob/master/api_format.md)

- Lizenz: „[...] available to use for whatever purpose you may have.“

<https://www.crossref.org/services/metadata-delivery/rest-api/>

# Übersicht (Stand: 24.4.2017)

resource	description	Ergebnismenge
<b>/works</b>	returns a list of all works (journal articles, conference proceedings, books, components, etc)	87.845.084
<b>/funders</b>	returns a list of all funders in the <a href="#">FundRef Registry</a>	14.200
<b>/members</b>	returns a list of all CrossRef members (mostly publishers)	8.084

# Übersicht (Stand: 24.4.2017)

resource	description	Ergebnismenge
<b>/types</b>	returns a list of valid work types	26
<b>/licenses</b>	return a list of licenses applied to works in CrossRef metadata	329
<b>/journals</b>	return a list of journals in the CrossRef database	52.135

# Vorgehensweise

1. Journals werden als Titel und ISSN(s) in einem Array gespeichert.
2. Hat ein Journal keine ISSN(s), wird nur über den Titel gematched.

# Vorgehensweise

3. Gibt es ISSN(s), wird über Titel UND alle ISSNs gematched.
4. Duplikate mit gleichem Titel, bei dem einer der beiden Einträge keine ISSNs hat, werden als unterschiedliche Journals interpretiert.

# Variante 1

- **Query 1:** Ausgabe aller Zeitschriften  
/journals?  
– Ergebnis: ~50.000 Journals
- **Query 2:** Artikel etc. hinzufügen  
/journals/{issn}/works?

# Variante 1

<b>Gesamt</b>	<b>52.135</b>
Duplikate	42.602
einzigartig	9.598
<b>mit ISSN</b>	<b>9.176</b>
unbek. Typ	112
pISSN (ges.)	6.724
eISSN (ges.)	5.332
pISSN & eISSN	3.098

# Variante 2

- **Query 1:** Ausgabe aller Zeitschriften  
/works?filter=type:journal  
– Ergebnis: ~20.000 Journals
- **Query 2:** Artikel etc. hinzufügen  
/journals/{issn}/works?

# Variante 2

<b>Gesamt</b>	<b>20.267</b>
Duplikate	4.905
einzigartig	15.362
<b>mit ISSN</b>	<b>15.176</b>
unbek. Typ	966
pISSN (ges.)	11.130
eISSN (ges.)	11.268
pISSN & eISSN	7.970

# Vergleich

	<b>Variante 1</b>	<b>Variante 2</b>
<b>Gesamt</b>	<b>52.135</b>	<b>20.267</b>
Duplikate	42.602	4.905
einzigartig	9.598	15.362
<b>mit ISSN</b>	<b>9.176</b>	<b>15.176</b>
unbek. Typ	112	966
pISSN (ges.)	6.724	11.130
eISSN (ges.)	5.332	11.268
pISSN & eISSN	3.098	7.970

# Beispiele Metadaten

- Springer Nature: ISSN 1432-2072

<http://api.crossref.org/journals/%7B7B1432-2072%7D/works?filter=type:journal-article>

- Walter de Gruyter: ISSN 2191-4664

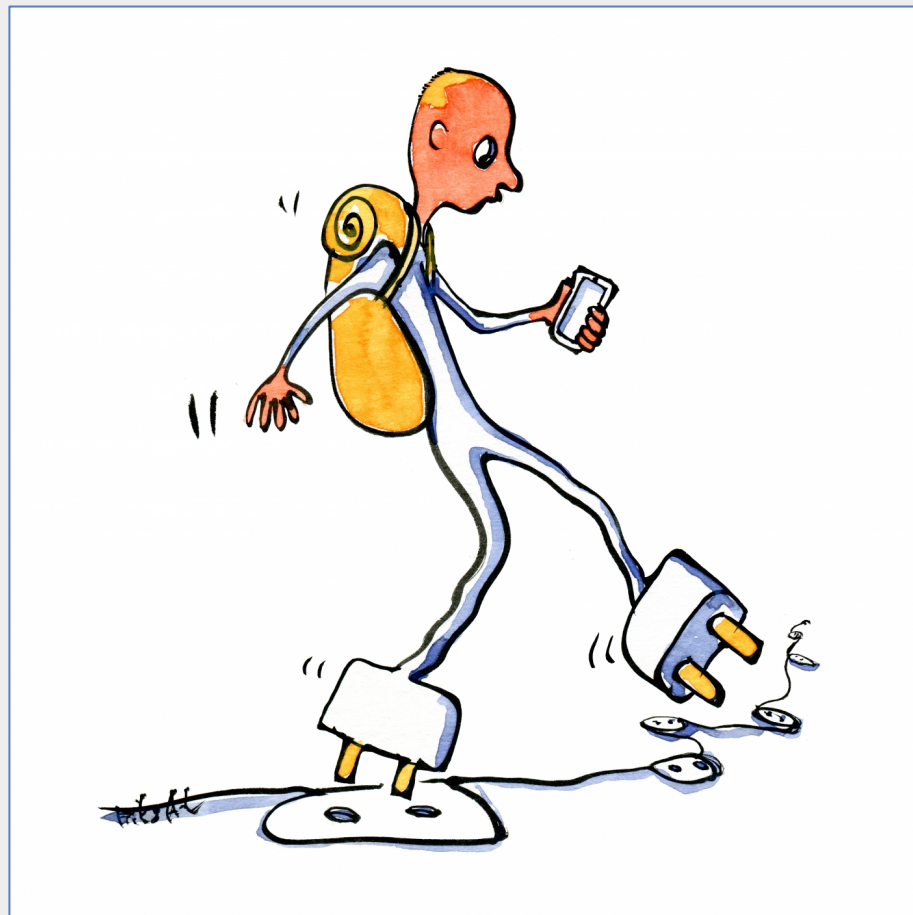
<http://api.crossref.org/journals/%7B2191-4664%7D/works?filter=type:journal-article>

**Hinweis:** URLs am besten mit Browser Firefox Version 53.0 oder größer öffnen.

# Fazit

- keine 1:1 Zuordnungen möglich
- Qualität der Metadaten unterschiedlich:
  - Umfang (Anzahl der Objektemerkmale)
  - Homogenität (Merkmalverteilung zwischen Objekten)
- Einzelfallprüfung erforderlich
- Direkter Vergleich mit anderen Quellen für Bewertung notwendig.

# 5. Kontakt



Frits Ahlefeldt:  
<https://hikingorg.files.wordpress.com/2015/12/electrical-hiking-color-illustration-by-frits-ahlefeldt.jpg>



Gerald Steilen  
[steilen@gbv.de](mailto:steilen@gbv.de)  
0551 39-91272

Maßgeblich für diese Präsentation ist das  
gesprochenen Wort.

