Crossref
CELEBRATING 25 YEARS

Strides towards co-creating the research nexus

Dominika Tkaczyk and Kora Korzec
Crossref

19/11/2025

DOI: https://doi.org/10.7557/5.8286

Agenda

- Introductions
- Crossref system, our metadata and ways of accessing it
- Discussion of your use cases for Crossref metadata
- Identifying gaps – opportunities for development of new metadata elements and tools
- *Optional:* strategies for using metadata for example use-cases

25 Crossref
CELEBRATING 25 YEARS

What is Crossref?

- We're a not-for-profit membership organisation that exists to make scholarly communications better
- Crossref operates the largest DOI registry for scholarly materials and content (books, journal articles, preprints, grants, and more).
- Crossref makes research objects easy to find, cite, link, assess, and reuse, thanks to the rich metadata underpinning all the records.

25 Crossref
CELEBRATING 25 YEARS

Scale of Crossref today

23,000 organisational members from **163** countries
- **50%** of members are based in Asia
- **35%** of members are universities or scholar-led

120 Sponsor orgs; **50** Ambassadors

174 million open metadata records with DOIs

1.3 billion DOI resolutions per month (**94%** of all DOI use)

2 billion calls to our API every month by **0000s** (?) systems ingesting and reusing this open metadata

25 Crossref
CELEBRATING 25 YEARS

Required, recommended and optional

crossref.org/documentation/schema-library/required-recommended-elements

Jounal article metadata

Required	
Journal (journal_metadata)	full_title, ISSN *or* title-level DOI and URL
Issue (issue_metadata)	issue, publication_date (year)
Article (article_metadata)	titles, publication_date (year), doi_data

Recommended	
Journal (journal_metadata)	abbrev_title, doi_data, coden, journal_issue, archive-locations (with one archive name), title-level DOI and URL
Issue (issue_metadata)	publication_date (month, day), journal_volume, contributors, issue, doi_data
Article (article_metadata)	contributors, ORCID, publication_date (day, month), pages (first_page, last_page), citation_list, funding, license, Crossmark metadata and JATS-formatted abstracts

Peer-review metadata

Required
title, review_date (year), relation (isReviewOf)

Recommended
contributors, institution, competing_interest_statement, running_number, license

Contributor metadata

Recommended
given_name, suffix, affiliation, ORCID

Required
Surname

25 Crossref
CELEBRATING 25 YEARS

Research Nexus - records

	Sept-2020	Sept-2021	Sept-2022	Sept-2023	Sept-2024	Sept-2025
Total records registered	117,517,271	127,859,845	138,330,747	150,418,250	162,763,867	174,111,214
Number of books	1,288,722	1,646,769	1,785,084	1,935,109	2,127,534	2,307,700
Number of peer-reviews	**158,000**	**253,805**	**348,507**	**463,171**	**634,692**	**834,147**
Number of journal-articles	83,823,982	89,897,215	96,096,724	102,347,516	109,661,283	116,683,154
Number of book-chapters	14,907,311	17,050,806	18,762,706	20,258,647	22,125,761	23,573,835
Number of grants	**353**	**25,753**	**39,181**	**86,226**	**127,542**	**175,435**
Number of preprints	**350,336**	**554,672**	**760,034**	**1,071,146**	**1,468,016**	**1,937,986**

Crossref
CELEBRATING 25 YEARS

Research Nexus - Connections

	Sept-2020	Sept-2021	Sept-2022	Sept-2023	Sept-2024	Sept-2025
Records with references	54,672,689	58,927,014	62,951,258	67,098,048	71,527,823	76,310,425
Number of citation relationships	1,181,044,739	1,312,213,008	1,447,502,991	1,587,915,525	1,740,330,734	1,909,635,516
Records with abstracts	18,613,257	21,897,035	25,226,192	29,206,046	33,463,847	38,020,681
Number of preprint-to-article links	**156,476**	**228,478**	**292,855**	**564,277**	**665,633**	**715,750**
Retractions (Crossmark + RW)	**35,084**	**43,663**	**56,179**	**60,373**	**63,299**	**64,280**
Records with ROR IDs	167,433	203,325	233,481	285,031	370,073	538,805
Records with ≥1 authors with ORCID IDs	**5,068,792**	**7,200,787**	**9,444,464**	**12,043,826**	**14,856,202**	**18,027,754**
Total unique ORCID IDs	2,968,049	4,062,108	5,102,937	6,177,065	7,360,056	8,678,223

Crossref
CELEBRATING 25 YEARS

Principles of Open Scholarly Infrastructure

GOVERNANCE

- Coverage across the research enterprise
- Stakeholder Governed
- Non-discriminatory participation
- Transparent governance
- Cannot lobby
- Living will
- **Regular review of purpose and community value**

SUSTAINABILITY

- **Transparent operations**
- Time-limited funds for time-limited activities
- Goal to generate surplus
- **Establish and maintain financial reserves**
- Mission-consistent revenue generation
- Revenue generated from services, not data
- **Volunteer labour**
- **Transition planning**

INSURANCE

- Open source
- Ensure open and secure data accessibility within legal and ethical constraints
- Available and preserved
- Patent non-assertion

https://openscholarlyinfrastructure.org/

Co-creation: collaborations and integrations

"Together we stand a greater chance of encouraging an open, fair and fully inclusive future for scholarly publishing"

- Lars Bjørnshauge, DOAJ Founder

PKP

DOAJ

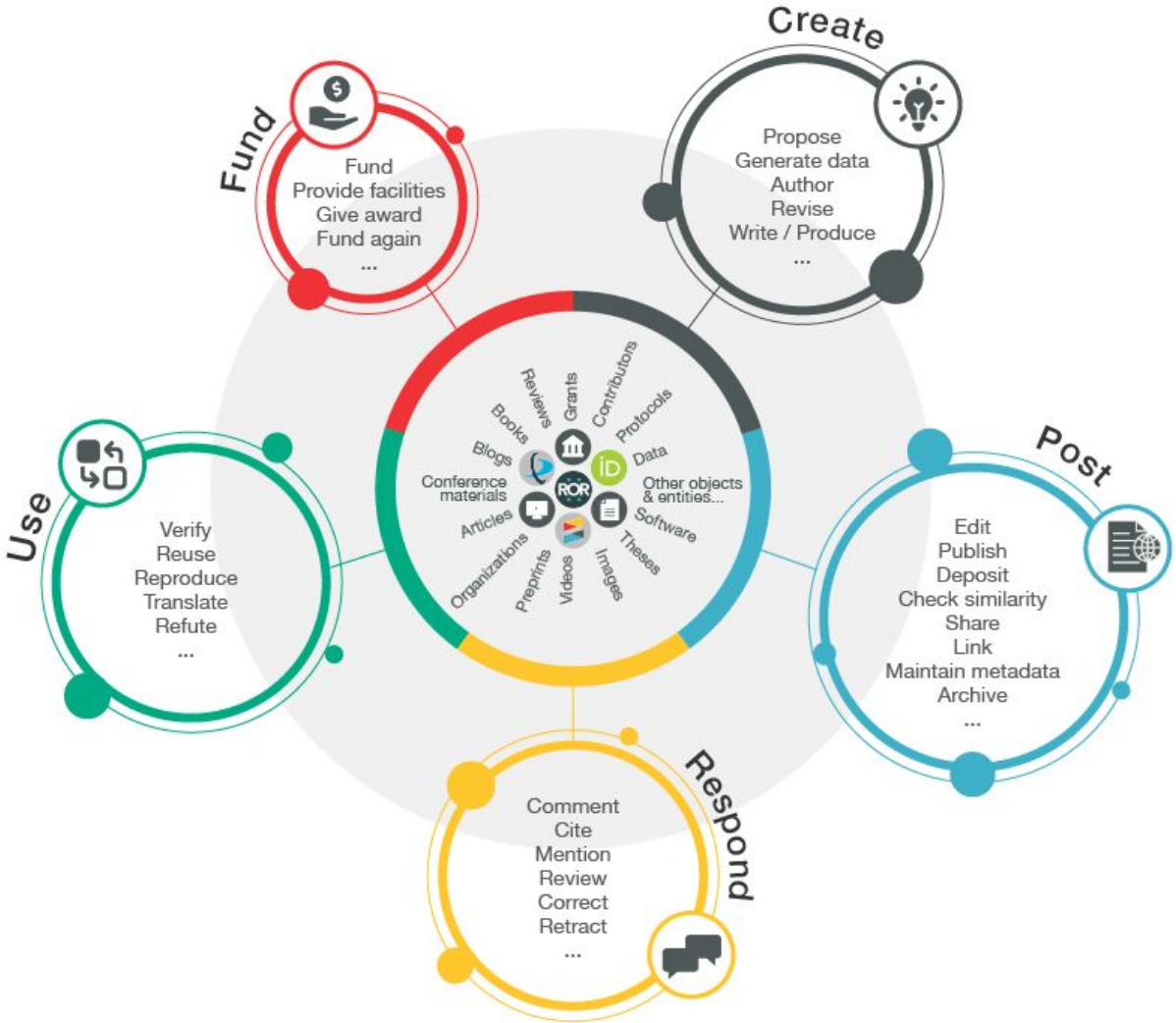
Retraction Watch

ROR

iD

BARCELONA DECLARATION ON OPEN RESEARCH INFORMATION

25 Crossref
CELEBRATING 25 YEARS

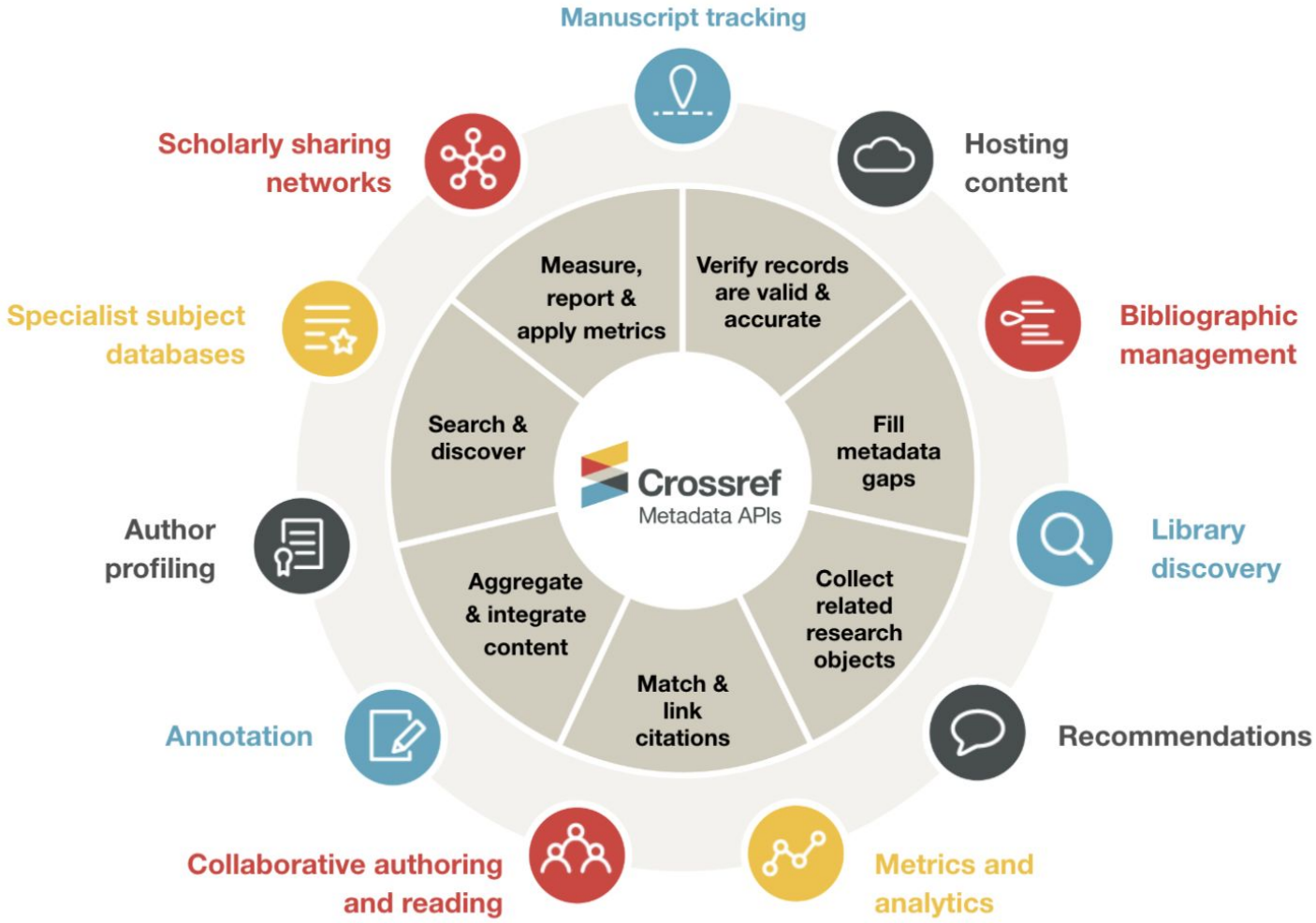


Research nexus

Like others,

“we envision a rich and reusable open network of relationships connecting research organizations, people, things, and actions;

a scholarly record that the global community can build on forever, for the benefit of society.”



Crossref
CELEBRATING 25 YEARS

Why does metadata matter for research?

Research integrity

Metadata as signals of trustworthiness including provenance information

Reproducibility

Metadata as relationships between literature, data, software, protocols and more

Assessment

Metadata can be used to analyze the outcomes of research and demonstrate compliance

Discoverability

More metadata means more systems have more pathways to find your work

Crossref
CELEBRATING 25 YEARS

Who's in the room?

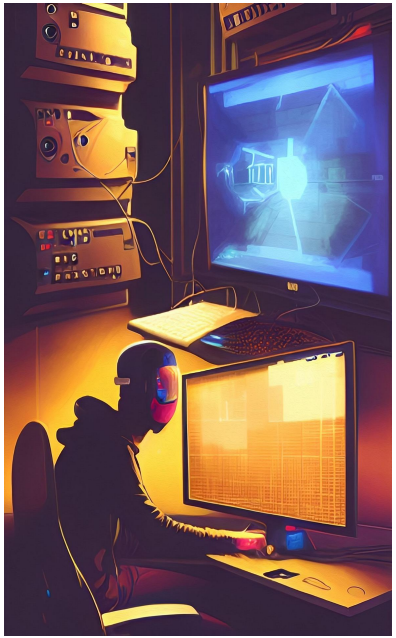
Crossref
CELEBRATING 25 YEARS

25
Crossref
CELEBRATING 25 YEARS

Crossref system overview

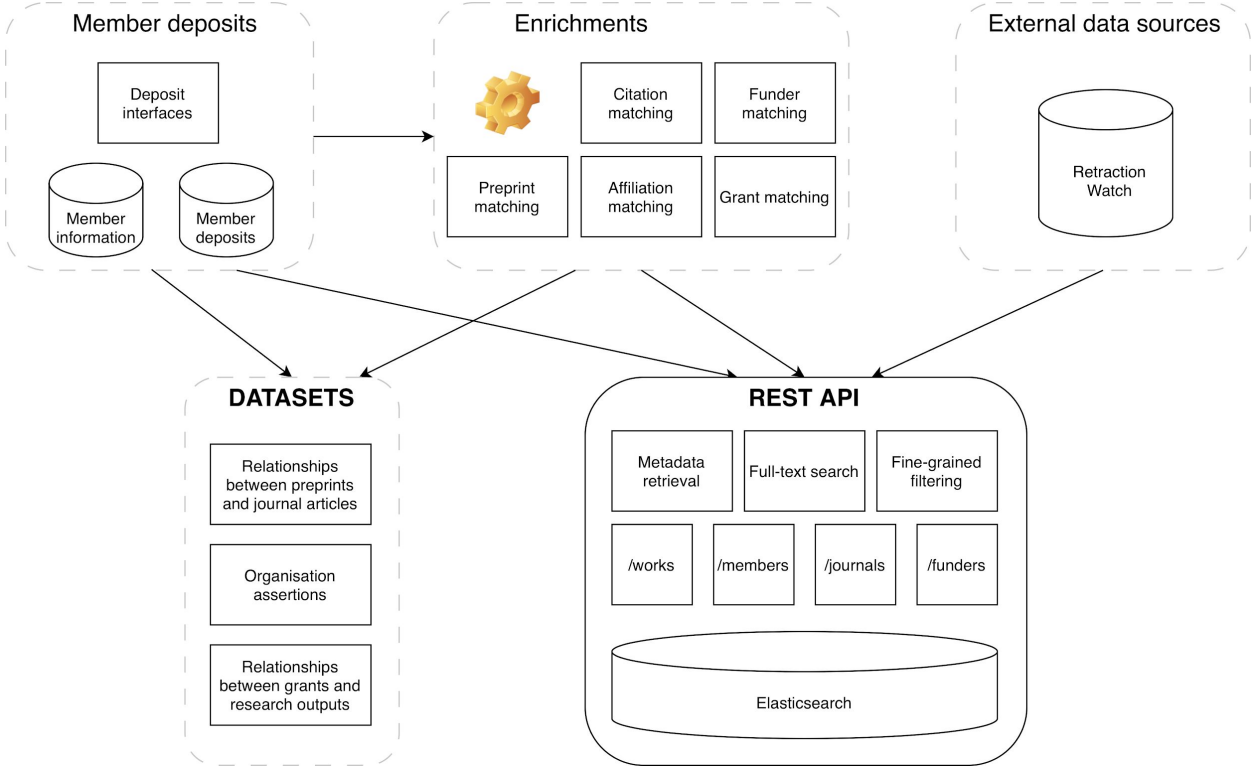
The role of technology

The technology team combines infrastructure, software development, and data science expertise to **take care of the Crossref system**, modernise it, and prepare it to support global scholarly communication **in the long term**.



25 Crossref
CELEBRATING 25 YEARS

Crossref system



Crossref
CELEBRATING 25 YEARS

Crossref system: inputs

- Crossref member deposits
  - **metadata records** of journal articles, conference proceedings, books and book chapters, datasets, preprints, reports, software, grants, and more
  - registered by publishers, societies, institutions, universities, funders, museums, government organisations, libraries, data and subject repositories, conference providers, standards bodies, individual scholars, news outlets
- Additional data sources
  - retractions and updates from **Retraction Watch**
- Enrichment workflows
  - **metadata matching** strategies

25 Crossref
CELEBRATING 25 YEARS

Crossref system: outputs

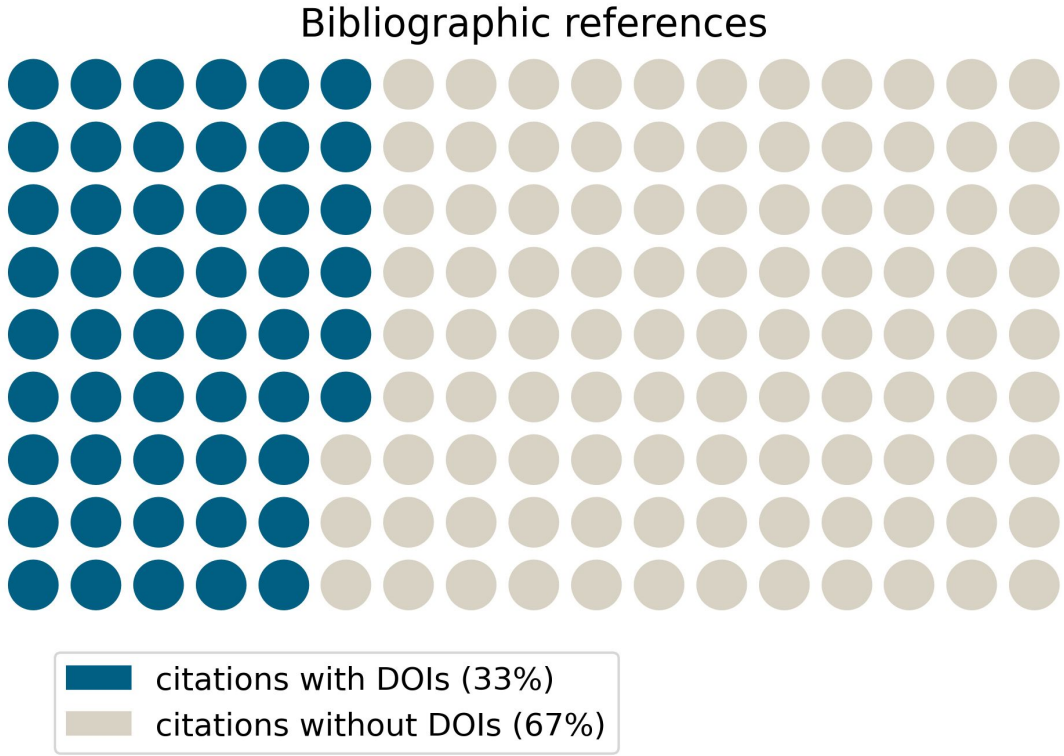
- For people
  - Metadata Search
  - Participation Reports
- For machines
  - REST API
  - Matching datasets

Crossref
CELEBRATING 25 YEARS

25
Crossref
CELEBRATING 25 YEARS

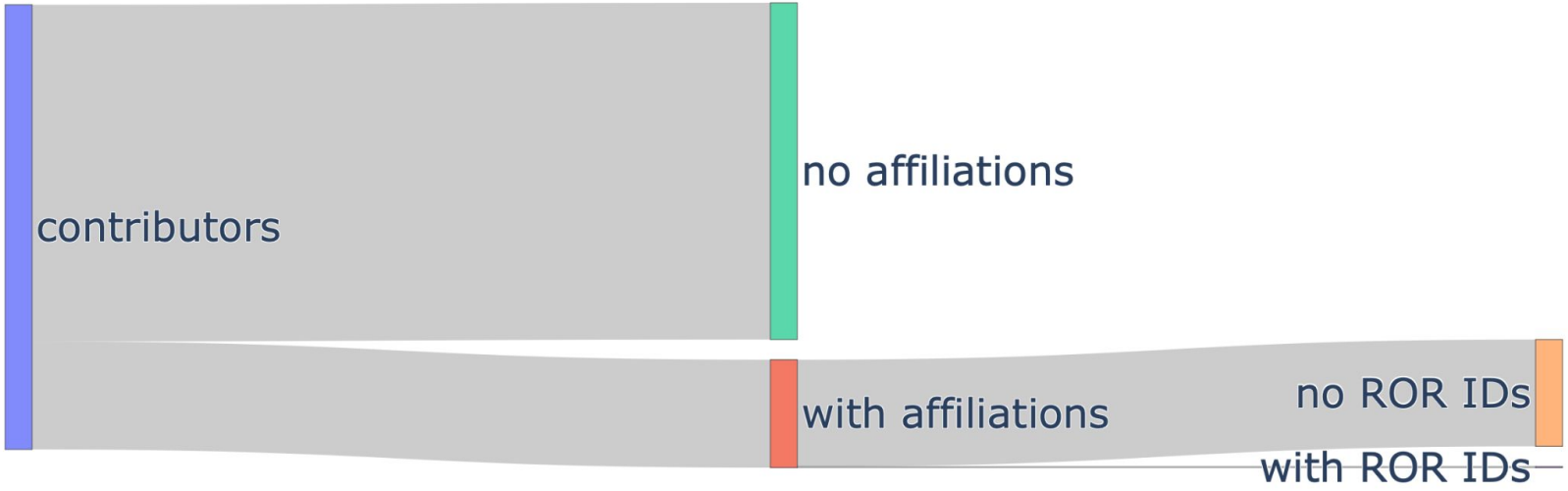
Metadata enrichment with matching

Metadata gap: bibliographic references



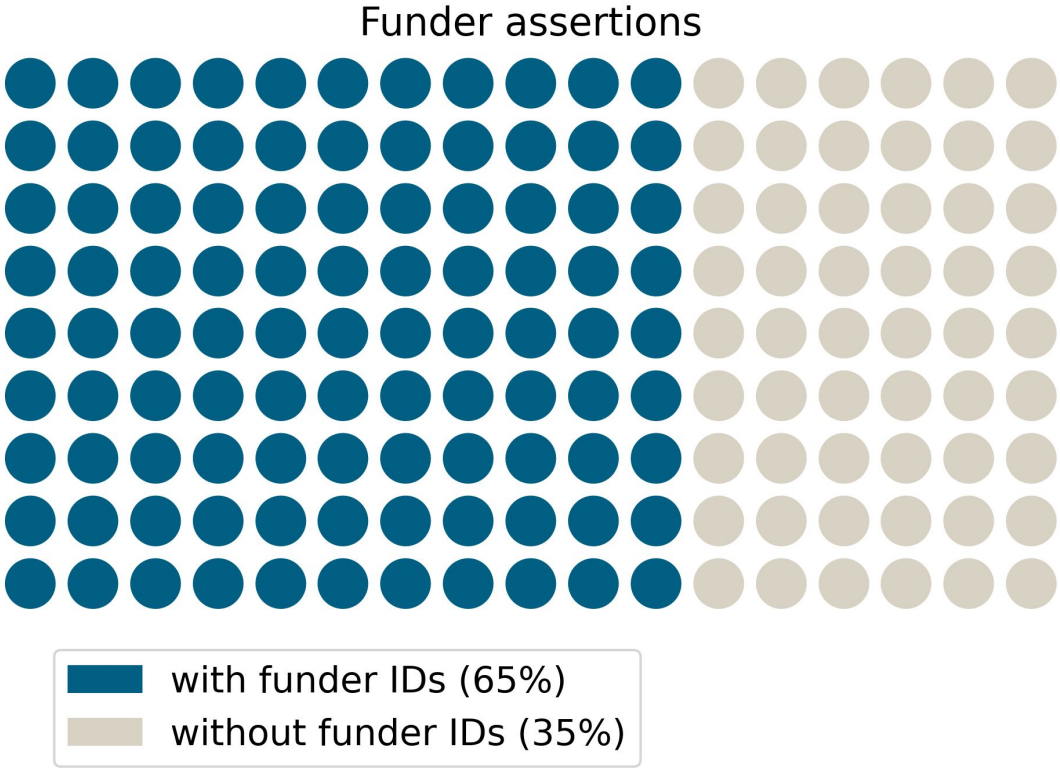
Crossref
CELEBRATING 25 YEARS

Metadata gap: affiliations



Crossref
CELEBRATING 25 YEARS

Metadata gap: funders



Crossref
CELEBRATING 25 YEARS

Metadata matching

some information about an item

that item's identifier

Crossref
CELEBRATING 25 YEARS

Bibliographic reference matching

1. Boucher RC (2004) New concepts of the pathogenesis of cystic fibrosis lung disease. Eur Resp J 23: 146–158.

New concepts of the pathogenesis of cystic fibrosis lung disease

R.C. Boucher

European Respiratory Journal 2003 23(1): 146-158; DOI: https://doi.org/10.1183/09031936.03.00057003

Crossref
CELEBRATING 25 YEARS

Funder matching

Emberi Eroforrások Minisztériuma

https://ror.org/00rb16m44

Ministry of Human Capacities

ORGANIZATION TYPES

Funder, Government

OTHER NAMES

Labels

Emberi Erőforrások Minisztérium (hu)

Aliases

Ministry of National Resources (en)

Acronyms

Emmi

LOCATIONS

Budapest (GeoNames ID 3054643), Hungary

WEBSITE

http://www.kormany.hu/en/ministry-of-human-resources

OTHER IDENTIFIERS

GRID grid.467820.f

Crossref Funder ID 501100005881

Wikidata Q989266

Affiliation matching

Department of Molecular Medicine, Sapporo Medical University, Sapporo 060-8556, Japan

https://ror.org/01h7cca57

Sapporo Medical University

ORGANIZATION TYPES

Education, Funder

OTHER NAMES

Labels

札幌医科大学 (ja)

Aliases

Sapporo ika daigaku

Acronyms

SMU

LOCATIONS

Sapporo (GeoNames ID 2128295), Japan

WEBSITE

http://web.sapmed.ac.jp/e/

OTHER IDENTIFIERS

GRID grid.263171.0

ISNI 0000 0001 0691 0855

Crossref Funder ID 100018376

Wikidata Q835726

25 Crossref CELEBRATING 25 YEARS

Preprint matching

Open Access Review

High-Entropy Materials: Features for Lithium–Sulfur Battery Applications

by Yikun Yao, Jiajun Chen, Rong Niu, Zhenxin Zhao and Xiaomin Wang *

College of Materials Science and Engineering, Taiyuan University of Technology, Taiyuan 030024, China

* Author to whom correspondence should be addressed.

Metals **2023**, *13*(5), 833; https://doi.org/10.3390/met13050833

Submission received: 28 February 2023 / Revised: 17 March 2023 / Accepted: 28 March 2023 / Published: 24 April 2023

Preprint Review

This version is not peer-reviewed.

Status and Prospects of High-Entropy Materials for Lithium–Sulfur Batteries

Yikun Yao, Jiajun Chen, Rong Niu, Zhenxin Zhao, Xiaomin Wang *

A peer-reviewed article of this preprint also exists.

Version 1

Submitted: 28 February 2023

Posted: 28 February 2023

Crossref
CELEBRATING 25 YEARS

Grant matching

```
funder ID:
10.13039/100004440
funder name:
Wellcome
award:
088858/Z/09/Z
```

Funded by
Wellcome Trust

£ 627,828

Duration
14 Sep 2009 - 13 Oct 2013

Internal grant ID
088858

Grant DOI
https://doi.org/10.35802/088858

Funding stream
Immune System in Health and Disease

Grant type
Biomedical Resources Grant

25 Crossref
CELEBRATING 25 YEARS

Current state of matching

Bibliographic reference matching	In production, results in the REST API
Funder matching	In production, results in the REST API
Affiliation matching	Prototype, available as a separate dataset
Preprint matching	Prototype, available as a separate dataset
Grant matching	Prototype, available as a separate dataset

Crossref
CELEBRATING 25 YEARS

Crossref
CELEBRATING 25 YEARS

Our metadata sources

The REST API

https://api.crossref.org/

- 175M metadata records of research outputs (**/works**)
- 30 record types
- 79 metadata fields
- functionality:
  - filtering
  - faceting
  - full-text search
  - iteration
  - sampling

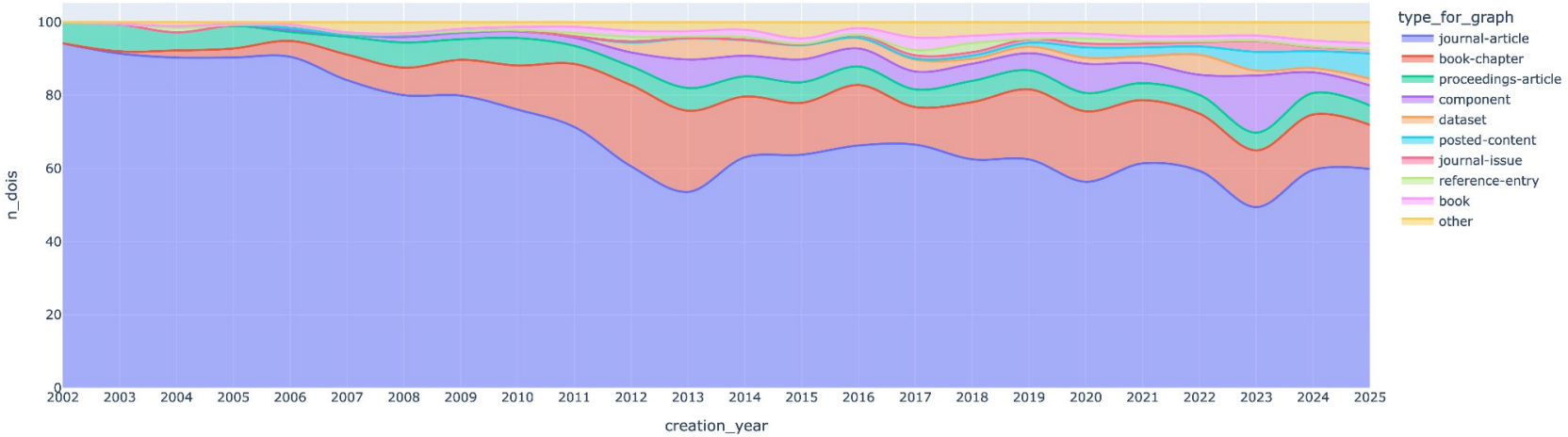
Crossref
CELEBRATING 25 YEARS

The REST API

https://api.crossref.org/

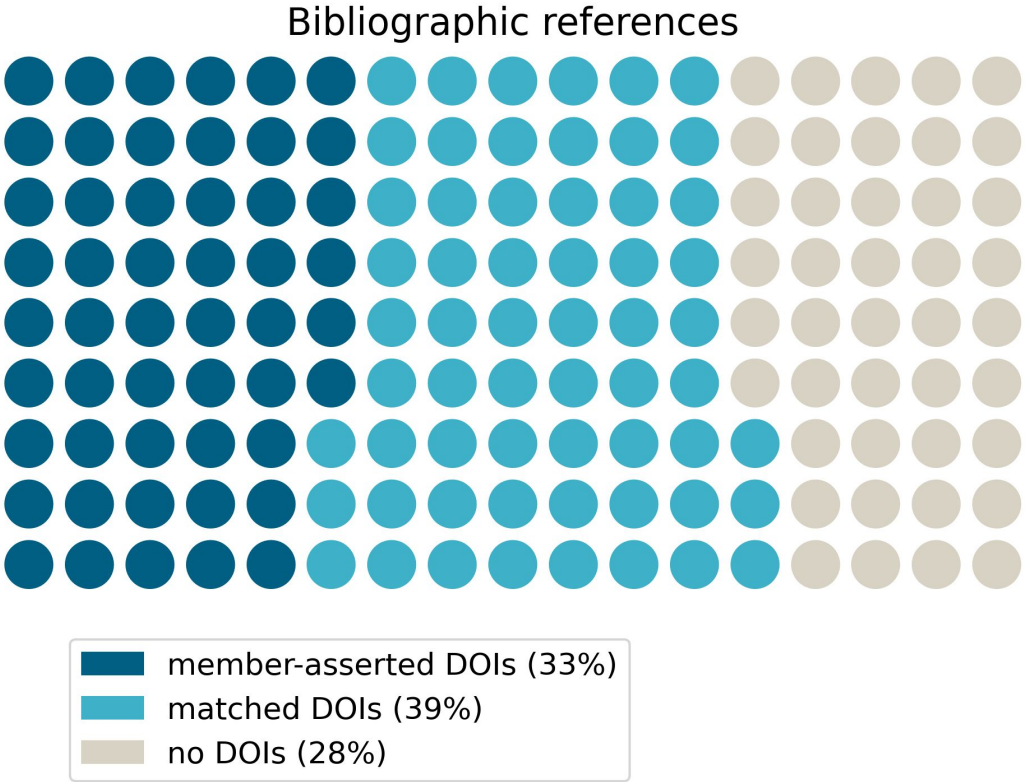
- 30k records of Crossref members (**/members**)
- 156k records of journals (**/journals**)
- 45k records of funders (**/funders**)

The REST API



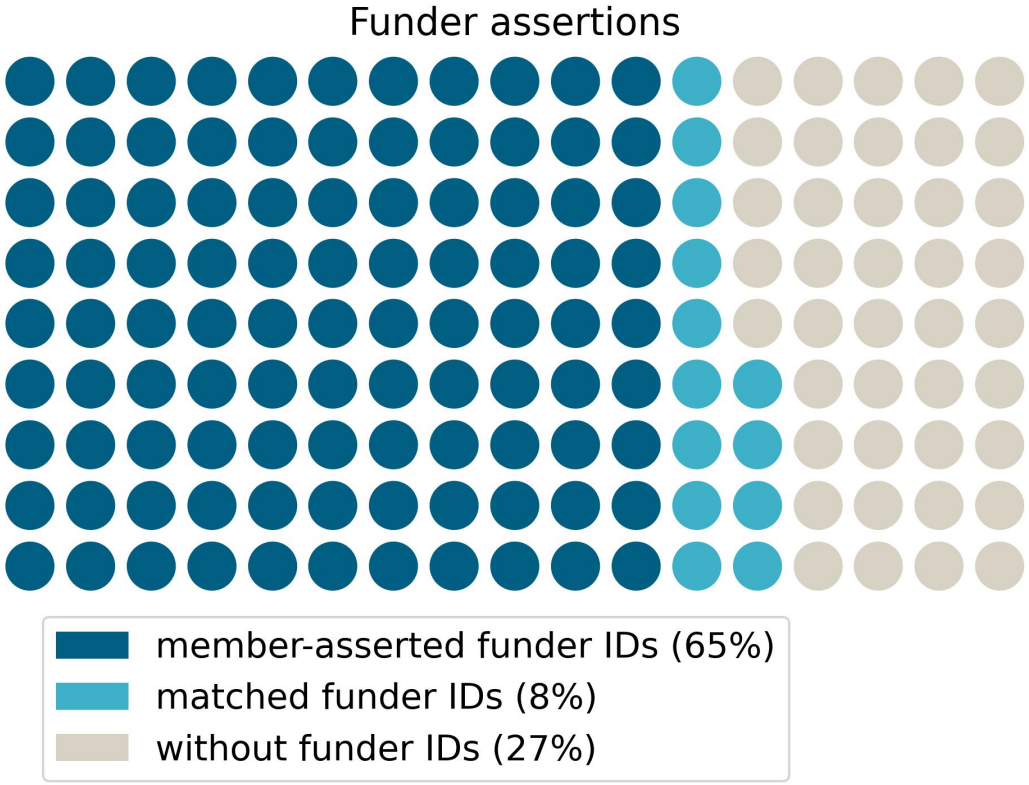
Crossref
CELEBRATING 25 YEARS

The REST API



Crossref
CELEBRATING 25 YEARS

The REST API



Crossref
CELEBRATING 25 YEARS

Matching datasets

- relationships between grants and research outputs https://doi.org/10.13003/waej1een
- relationships involving research organisations https://doi.org/10.5281/zenodo.15254993
- relationships between preprint and research outputs https://doi.org/10.5281/zenodo.15124417

Crossref
CELEBRATING 25 YEARS

Grant matching dataset

Relationships between **grants** and **research outputs**
https://doi.org/10.13003/waej1een

- 250,163 total relationships
- 2,491 (1%) were deposited by a Crossref member
- 247,672 (99%) were matched through award number and funder information

Crossref
CELEBRATING 25 YEARS

Organisation matching dataset

Relationships involving **research organisations**
https://doi.org/10.5281/zenodo.15254993

- relationships:
  - contributor's affiliations
  - institutional contributors
  - work's institutions
  - grant investigator's affiliations
- 140,906,929 total assertions
- 1,014,325 (0.7%) assertions contain a ROR ID deposited by Crossref members
- 94,988,729 (67%) assertions contain a matched ROR ID

Crossref
CELEBRATING 25 YEARS

Preprint matching dataset

Relationships between **preprints** and **journal articles**
https://doi.org/10.5281/zenodo.15124417

- 1,060,572 relationships in total
  - 954,782 preprints
  - 953,453 journal articles
- 462,092 (44%) relationships deposited by the Crossref members
- 598,480 (56%) matched relationships

Crossref
CELEBRATING 25 YEARS

25
Crossref
CELEBRATING 25 YEARS

Your metadata use cases

Generate ideas with these questions in mind

- **As** [your role] **I need to** [process or result you need to deliver]
- **I need evidence about** [phenomenon in the scholarly ecosystem]
- **Can metadata help me automate** [process that you currently do manually]?
- **How can Crossref metadata be used for** [process you undertake] **instead of proprietary data?**

Crossref
CELEBRATING 25 YEARS

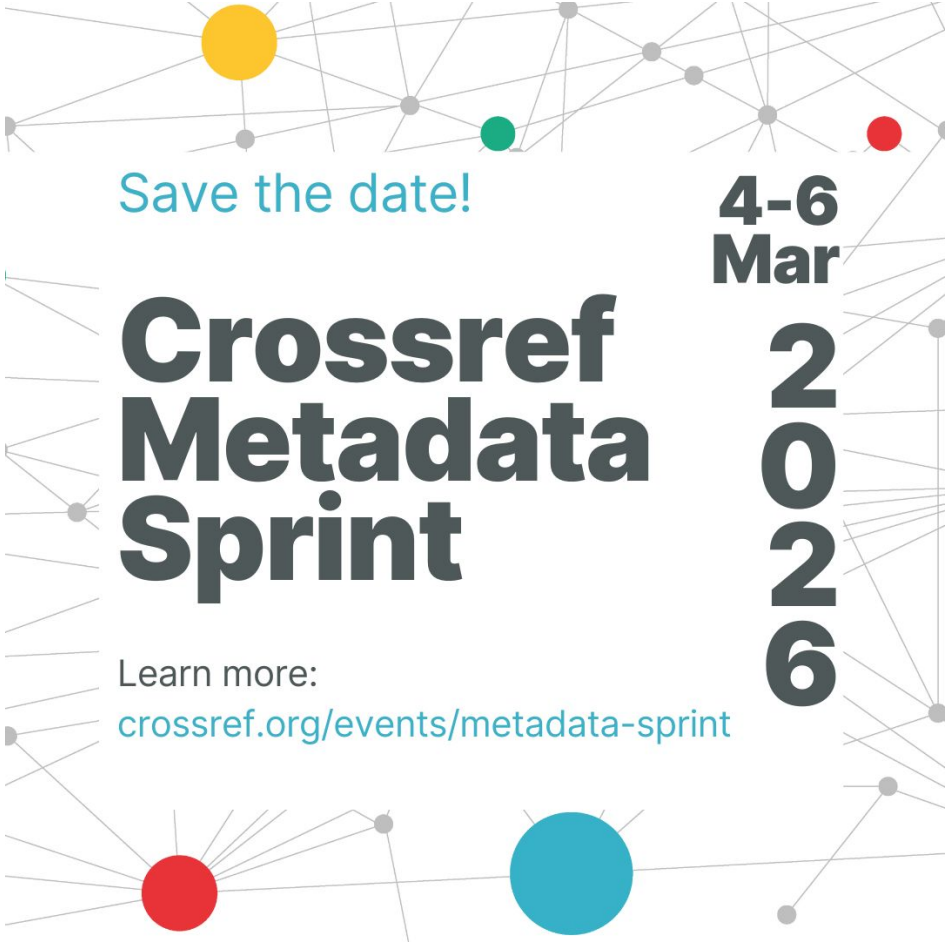
25
Crossref
CELEBRATING 25 YEARS

Metadata letter to Santa

What’s missing? Think of...

- Tools that would help you make the most of metadata
- Metadata elements that are insufficiently covered
- Metadata elements not currently included in Crossref schema

25 Crossref
CELEBRATING 25 YEARS



Crossref
CELEBRATING 25 YEARS

25 Crossref
CELEBRATING 25 YEARS

How to interact with Crossref metadata?

How to use: REST API requests

Example: explore open access status of works funded by the Spanish National Research Council

Crossref
CELEBRATING 25 YEARS

How to use: REST API requests

Example: explore open access status of works funded by the Spanish National Research Council

Step #1: Get the Funder ID

/funders?query=Spanish+National+Research+Council

↳ id 501100003339

25 Crossref
CELEBRATING 25 YEARS

How to use: REST API requests

Example: explore open access status of works funded by the Spanish National Research Council

Step #2: Get the works associated with the funder

/funders/501100003339/works?rows=1000
/funders/501100003339/works?rows=1000&offset=1000
/funders/501100003339/works?rows=1000&offset=2000
...

Crossref
CELEBRATING 25 YEARS

How to use: REST API requests

Example: explore open access status of works funded by the Spanish National Research Council

Step #3: Check the licenses

/funders/501100003339/works?facet=license:*&rows=0
/funders/501100003339/works?facet=license:5&rows=0

Crossref
CELEBRATING 25 YEARS

How to use: REST API requests

Example: explore open access status of works funded by the Spanish National Research Council

Licence	Work count
https://www.elsevier.com/tdm/userlicense/1.0/	2,221
https://www.elsevier.com/legal/tdmrep-license	1,909
https://creativecommons.org/licenses/by/4.0/	868
https://creativecommons.org/licenses/by/4.0	769
http://creativecommons.org/licenses/by/4.0/	611

How to use: scripting

Example: who are the most cited authors from UiT?

How to use: scripting

Example: who are the most cited authors from UiT?

Step #1: Get the relevant research outputs from the dataset

```
research_outputs = []

with open("crossref-organisation-assertions-Mar-2025.jsonl", "r") as f:
    for l in f:
        record = json.loads(l)
        if record["ror-id"] == "https://ror.org/00wge5k78" \
            and record["relationship"] == "contributor-affiliation" \
            and record["contributor-role"] == "author":
            research_outputs.append(record)
```

Crossref
CELEBRATING 25 YEARS

How to use: scripting

Example: who are the most cited authors from UiT?

Step #2: Get the citation counts from the REST API

```
REST_API = "https://api.crossref.org/works/"

citation_counts = {ro["DOI"]: 0 for ro in research_outputs}

for doi in citation_counts:
    record = requests.get(f"{REST_API}{doi}").json()["message"]
    citation_counts[doi] = record["is-referenced-by-count"]
```

25 Crossref
CELEBRATING 25 YEARS

How to use: scripting

Example: who are the most cited authors from UiT?

Step #3: Get top-5 most cited works

```
for ro in research_outputs:
    ro["is-referenced-by-count"] = citation_counts[ro["DOI"]]

research_outputs.sort(key=lambda p: p["is-referenced-by-count"],
                      reverse=True)

for ro in research_outputs[:5]:
    print(ro["is-referenced-by-count"], ro["contributor"])
```

25 Crossref
CELEBRATING 25 YEARS

How to use: scripting

Example: who are the most cited authors from UiT?

DOI	Authors	Citation count
10.1001/jamaoncol.2016.5688	Elisabete Weiderpass	2,953
10.1080/15548627.2020.1797280	Yakubu Princely Abudu, Terje Johansen, Trond Lamark, Jakob Mejlvang	2,110
10.1001/jamaoncol.2017.3055	Elisabete Weiderpass	1,568
10.15252/embj.201796697	Terje Johansen	1,392
10.15252/embj.2021108863	Terje Johansen	1,232

How to use: large-scale processing

Example: are data citations growing in numbers?

Crossref
CELEBRATING 25 YEARS

How to use: large-scale processing

Example: are data citations growing in numbers?

Step #1: Get Crossref and DataCite dataset DOIs

```
cr_works = … # Crossref data dump
cr_datasets = (
    cr_works.filter(col("type") == "dataset")
    .select(lower(col("DOI")).alias("target_DOI")
)

dc_works = … # DataCite data dump
dc_datasets = (
    dc_works.filter(col("attributes.types.citeproc") == "dataset")
    .select(lower(col("id")).alias("target_DOI"))
)

datasets = cr_datasets.union(dc_datasets)
```

Crossref
CELEBRATING 25 YEARS

How to use: large-scale processing

Example: are data citations growing in numbers?

Step #2: Get citations to datasets

```
references = (
    cr_works.withColumn("ref", explode("reference"))
    .select(
        lower(col("DOI")).alias("source_DOI"),
        lower(col("ref.DOI")).alias("ref_DOI"),
        col("created.date_time").alias("created")
    )
)
dataset_references = references.join(
    datasets,
    refs.ref_DOI == datasets.target_DOI,
    "inner"
)
```

Crossref
CELEBRATING 25 YEARS

How to use: large-scale processing

Example: are data citations growing in numbers?

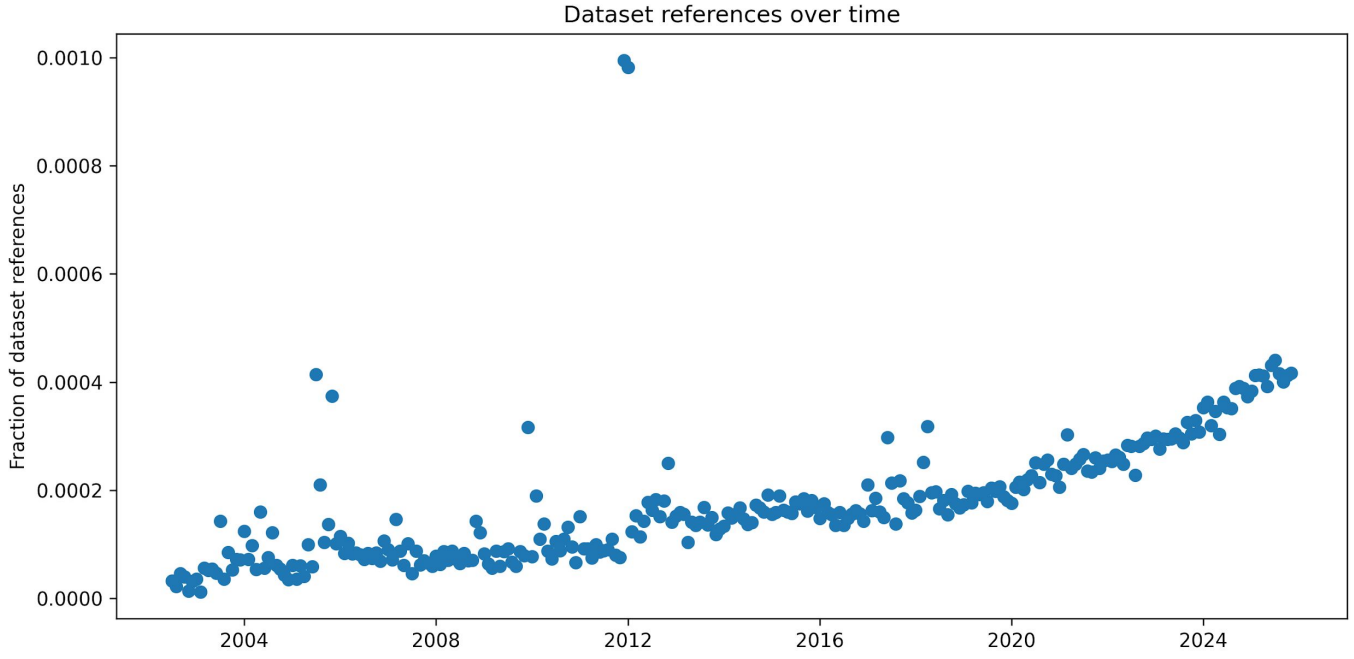
Step #3: Group and normalize

```
dataset_references = dataset_references.groupBy("created").agg(
    F.count("*").alias("dataset_refs"))

by_month = (
    references.groupBy("created")
    .agg(count("*").alias("total_refs")))

result = (
    dataset_references.join(by_month, "created", "left")
    .withColumn(
        "fraction_refs",
        col("dataset_refs") / col("total_refs")
    ))
```

Crossref
CELEBRATING 25 YEARS

How to use: large-scale processing

Example: are data citations growing in numbers?



Crossref
CELEBRATING 25 YEARS

Thank you

25
Crossref
CELEBRATING 25 YEARS