



Crossref



Crossref Event Data A New, Open Underlying Data Set for Altmetrics

Joe Wass
Principal R&D Engineer
Crossref
@joewass

Hello again!

**What is an
“underlying dataset”?**

**Problems to tackle
when building a metric
or doing research.**

Identifying a data set.

Identifying a data set.

**Links from
where?**

Identifying a data set.

Links to where?

Identifying a data set.

**What kind of
links?**

**Non-traditional
publications, non-
traditional rules.**

**Collecting data
from the corpus.**

**Identifying &
canonicalising links.**

Link rot in *your* dataset.

Recording data for reproducibility.

“The data we got are not transparent””

Who's human?

Who's a machine?

Two categories.

Category A

**Everyone has to
solve these.**

Category B

**Everyone has to
different issues.**

Category B

Particular corpus of articles.

Links from a particular website.

Blogs from known authors.

Links via DOIs or not.

Building different things?

**Category A: We can
help you solve.**

Category B: We can do.

Crossref

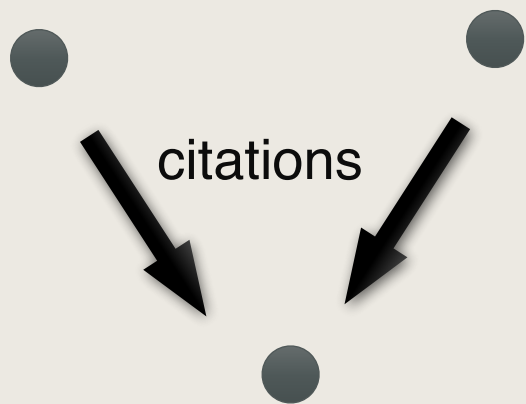
~100 million items.

Metadata.

**Infrastructure to make it
work.**

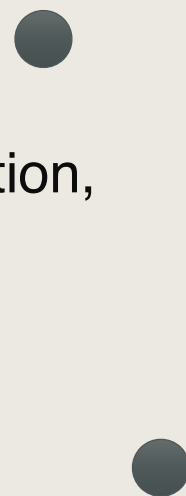
**Metadata typically
provided by
publishers.**

Crossref Registered Content



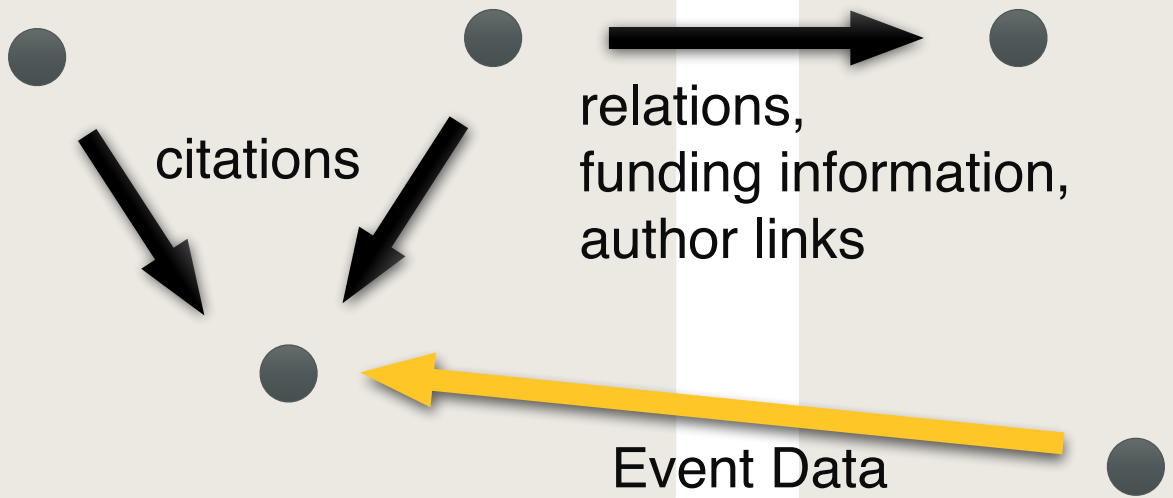
relations,
funding information,
author links

The Rest of the world



Crossref Registered Content

The Rest of the world



**No longer supplied
by publishers.**

An underlying data set.

(Category A)

Problems

A

**Which data set
to use?**

A

**Which link
source corpus?**

A

- **Reddit ~ 100,000**
- **Twitter ~ 3,000,000**
- **Wikipedia ~ 30,000,000**
- **hypothes.is ~ 20,000**
- **wordpress.com ~ 30,000**
- **Blogs ~ 200,000**
- **Stack Exchange ~ 6,000**

A

**What kind of
links?**

A

Canonicalisation and link rot.

A

**Reproducibility of
data collection.**

A

**Reproducibility of
state.**

A

**Reproducibility of
data processing.**

A

**Representing an
evolving set of links.**

B

**Telling humans
from machines.**

**External APIs &
Websites**

**External APIs &
Websites**



**Agents monitor
external services**



**Percolator consults
external services**

Artifacts



Agents



Evidence



Percolator



Event

**Agents
consume
Artifacts**



**Agents
produce
Evidence
Records**

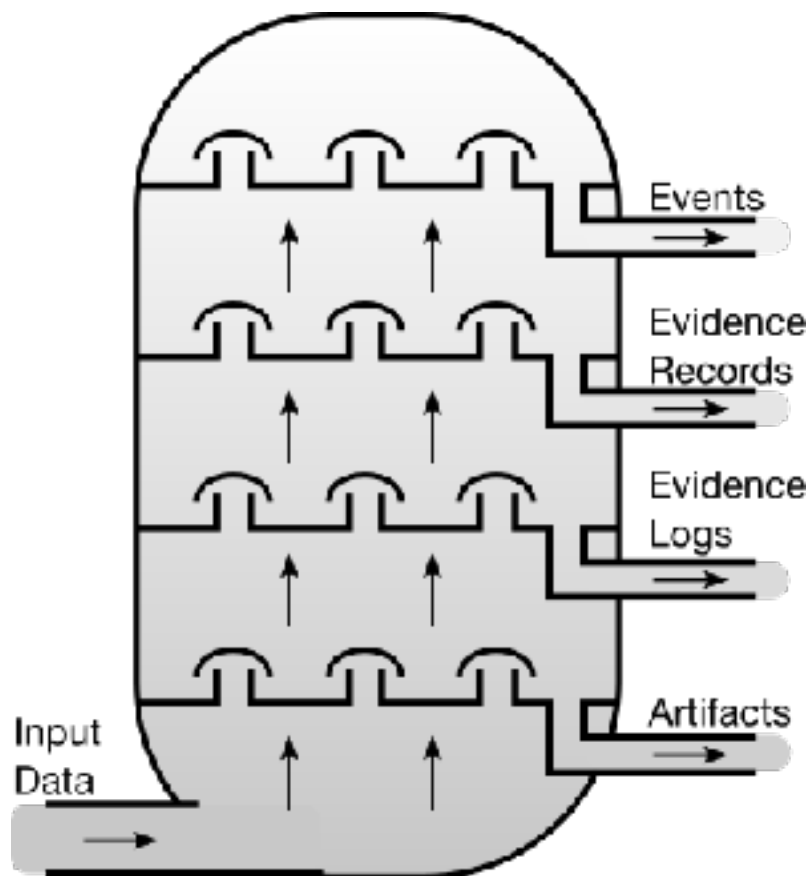


**Percolator
consumes
Evidence
Records**



**Percolator
produces
Events**

Evidence Logs



All data is free.

CC0,

CC-BY-4.0,

etc

Open for Beta

<https://www.eventdata.crossref.org/guide/>

<https://www.crossref.org/blog>

4:AM Talk & Hackathon!

eventdata@crossref.org

[@joewass](#)

