



You are
Crosser

Joe Wass
Head of Software Development
jwass@crossref.org
@joewass



The role of open scholarly infrastructure in metrics



The role of open scholarly infrastructure in metrics



The role of open scholarly infrastructure in metrics





Scholarly publications



Datasets



Research organisations



Researchers and Authors



Scholarly infrastructure organisations

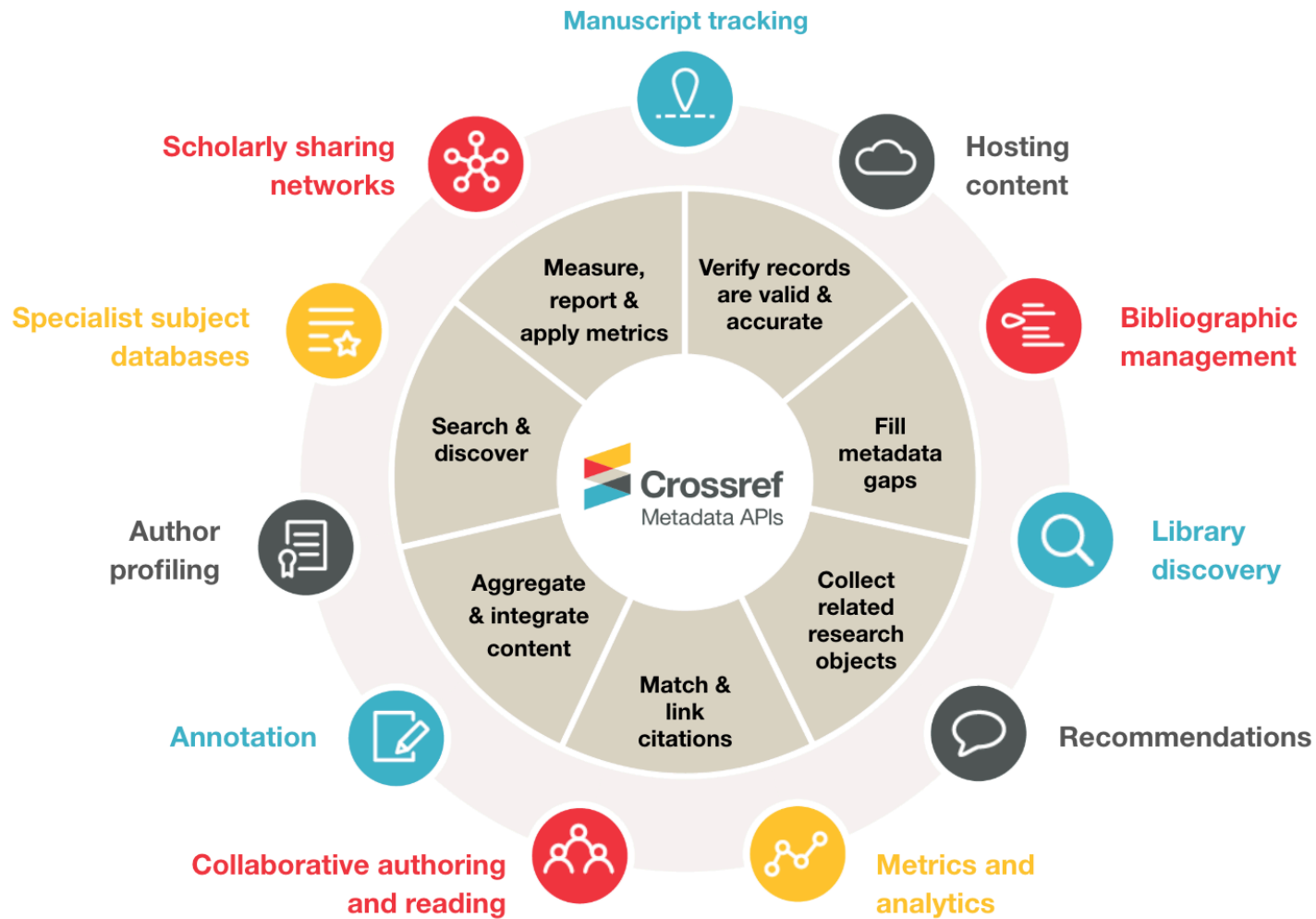
- Focus on a specific area.
- Serve and work with different communities.
- Work together to improve scholarship.
- Are nearly invisible.



Crossref

- Focus on scholarly publications.
- Serve the publishing, science, funding, authoring communities and anyone who wants information about publications.
- Work together with other organisations like ORCID and DataCite.
- Is *nearly* invisible.





The role of open **scholarly** infrastructure in metrics





What is “scholarship”?

*The scholarly method or scholarship is the body of principles and practices used by scholars to make their **claims about the subject** as valid and trustworthy as possible, and to make them known to the scholarly public.*

https://en.wikipedia.org/wiki/Scholarly_method



What are “claims about the subject”?

- Basic research.
- Research about research.
- Bibliometrics, altmetrics and ... *metrics.
 - The impact of research or funding.
 - How communities interact with the research.

The role of open scholarly infrastructure in metrics

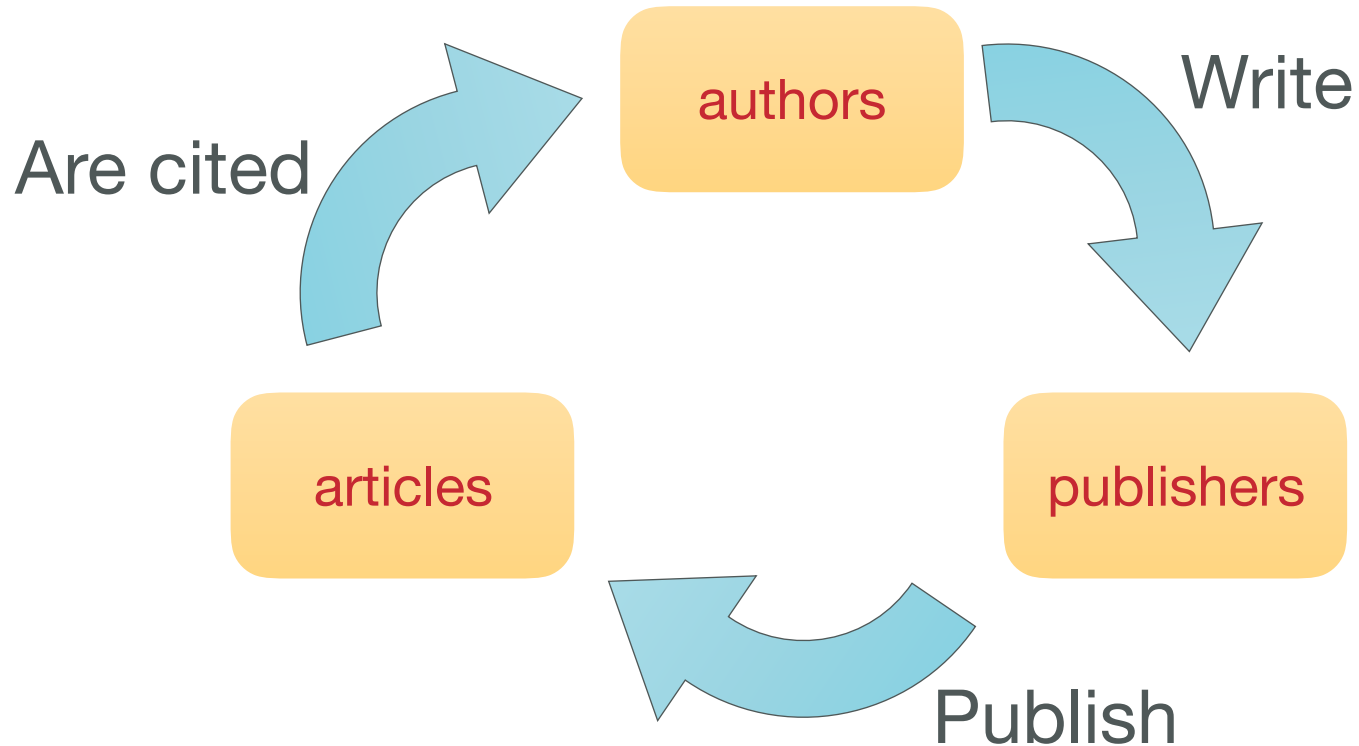


The landscape
is changing



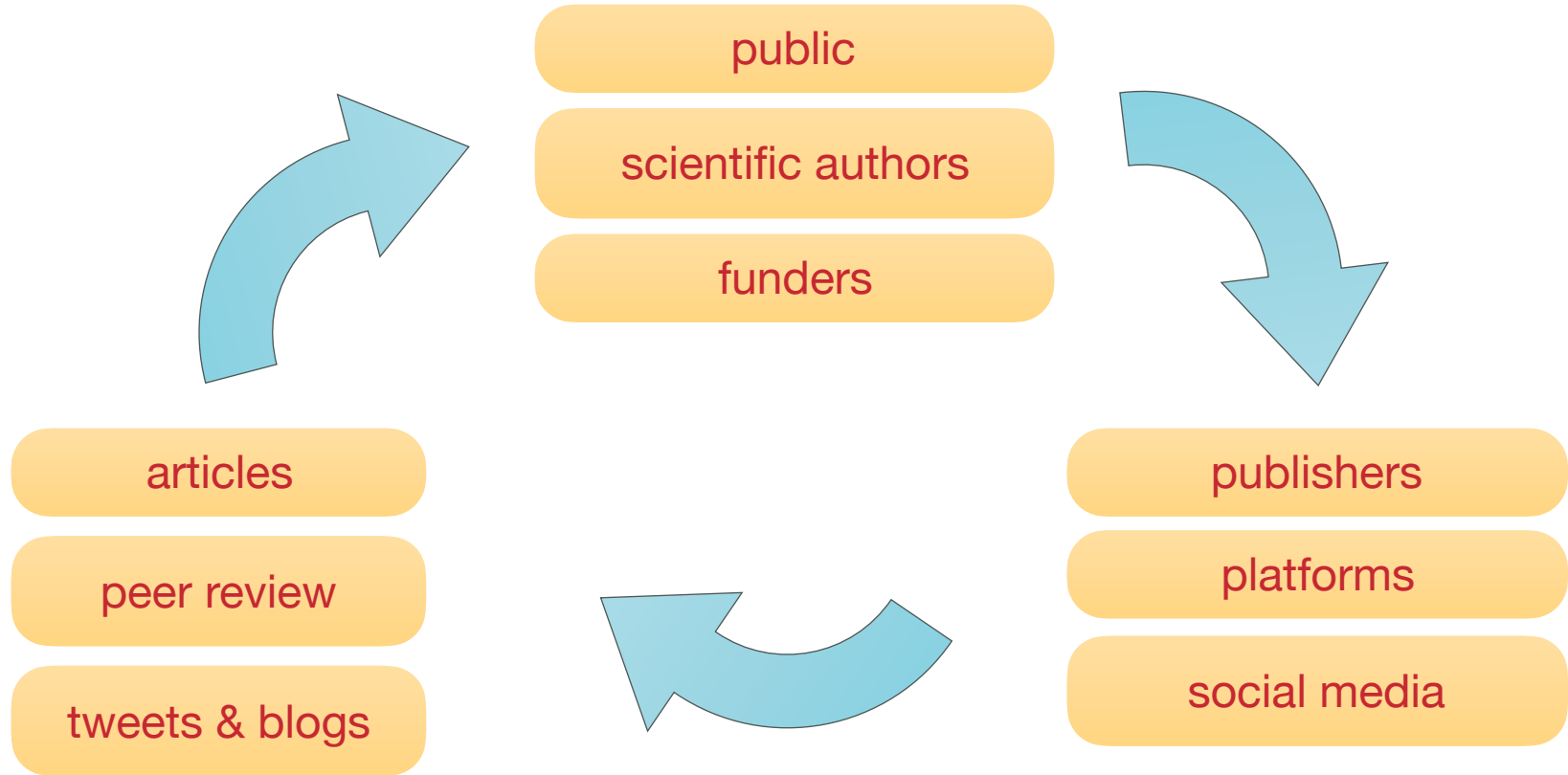


Metadata flow for “Traditional” scholarship





“New” scholarship





Where does metadata come from?

Old world:



New world:



The **role** of open scholarly infrastructure in metrics





Why do we exist?

- We exist to do those things that all of our members need to do, or that happens across all of our members.
- e.g. They all need identifiers so we give DOIs. They all need to exchange metadata, so we accept and distribute that.
- We know that DOI Registered Content is being mentioned out there on the web, so we can go and collect those links.
- There are common problems we can solve.



What role can we play?

- Provide open platform for exchanging data, Event Data, for exchanging data.
- Collect web mention data.
- Combining traditional publisher-originated data with new fetched metadata.
- Allowing sharing of data usage metrics.



Crossref collects data from:

- Twitter
- Wikipedia
- Reddit
- StackExchange
- Blogs
- Crossref - DataCite data citations
- Hypothes.is annotations



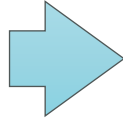
What happens if we don't?

- Our members will do it.
 - But what if one member does it, and another does not?
 - What if they do it differently?
- Researchers will do it.
 - Long-term sustainability? Open data? Re-use of data? Focus of study?
- Private companies will do it.
 - Data is not open.
 - Collected for specific purpose, which forces you to make choices.



The purpose defines the processing

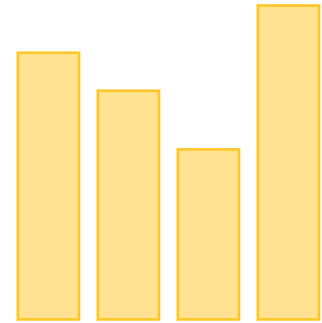
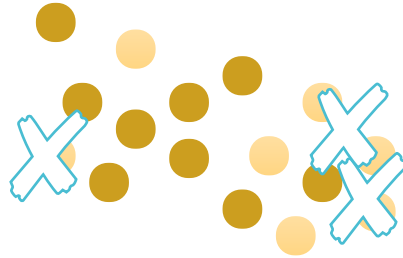
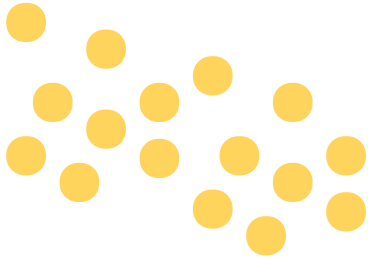
input



process

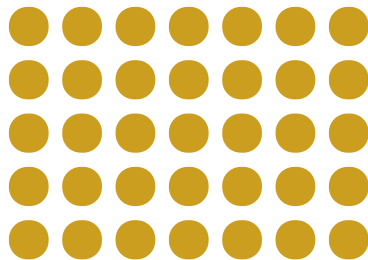
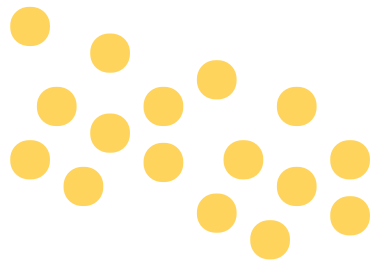


output



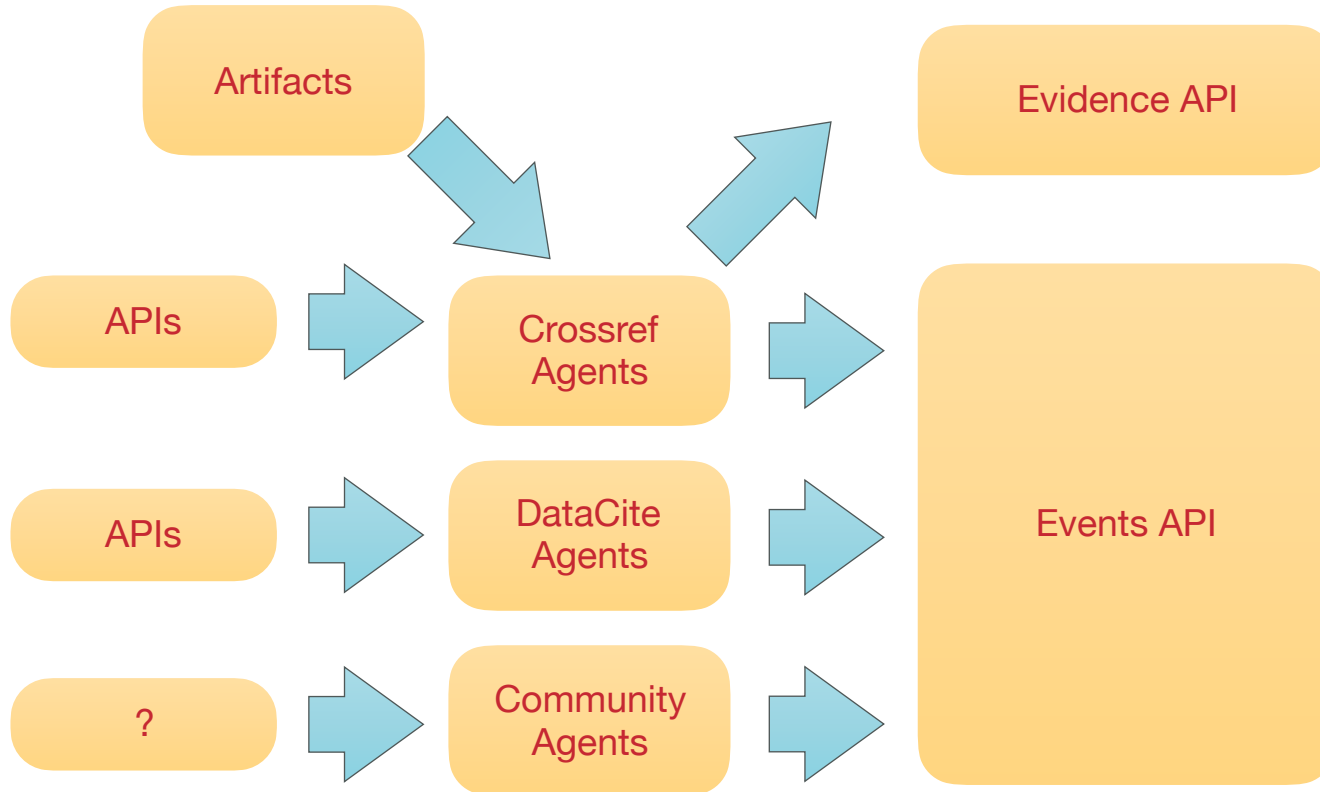


Wider purpose, less processing





How does Event Data work?



The role of open scholarly infrastructure in metrics

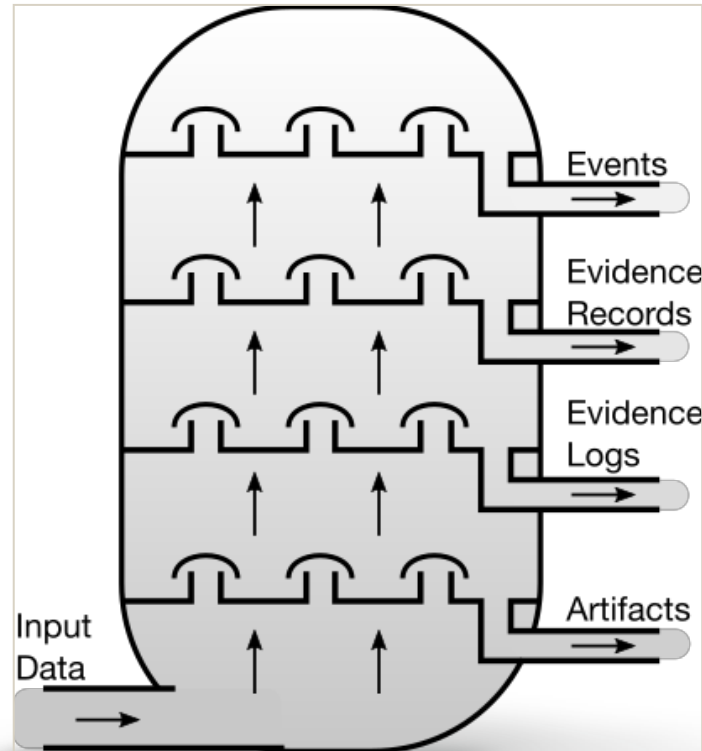


More than one
role





Intermediary data



<https://www.crossref.org/blog/you-do-want-to-see-how-its-made-seeing-what-goes-into-altmetrics/>

The role of **open** scholarly infrastructure in metrics



Open
data = Open
opportunities





Open

- **Open Data**

CC0 and CC BY 4.0 allows the widest range of uses.

- **Open Source**

Users can understand where the data comes from.

- **Open Evidence**

Users can look at what our agents did.

- **Open Contributions**

Community can contribute data.

- **Open Ended**

We collect the underlying data, it can be used for anything.



Open Data

- CC0 and CC BY 4.0 are unrestrictive.
- Each Event has a license so you always know how you can use the data.
- You don't need to ask us first. We have terms and conditions, but they are as unrestrictive as possible.



Open Source

- All the software is available on GitHub.
- You can see how the Agent gathered data.
- If you have questions about the data, you can look to see how it was made at a given point in time.
- The community can suggest improvements to our code.



Open Evidence

- “What did the Agent do”
- The activity is provided in open data. Events link to the Evidence Record and Logs which describes the software version and input version.
- Nothing is hidden. We have the same data as you!



Open Contributions

- We can't collect everything. We allow community data sources to send us data.
- There is no risk to them, because the license is open.
- New tools, like Plaudit.pub, can send data directly in.





Open Ended

- We don't make metrics. We just collect links.
- We have some good ideas about how you can use it.
- But if you have a new idea for using data, just do it!

The role of open scholarly infrastructure in metrics



That's up to you!





More information

- Event Data User Guide:
<https://www.eventdata.crossref.org/guide/>
- Ideas in this presentation:
<https://www.crossref.org/authors/joe-wass/>

Thanks

Joe Wass

Head of Software Development

@joewass jwass@crossref.org

