Five principles for community altmetrics data

Joe Wass
Principal R&D Engineer
@joewass
Mission: build a community data source for altmetrics and other links.
What kind of ecosystem am I trying to support?
How do I design a service that supports it?
Five* principles

* five-and-a bit?
1. Metrics should be **decoupled** from the underlying data that they are derived from.
2. Underlying data should be **open**, available data sets.
3a. Underlying data sets should reflect the **diversity** of the behaviour being **observed**.
3b. Metrics should be as diverse and individual as their audiences, and reflect that diversity.
4. The **software** that collects the underlying data should be **open source**. Both the **intended behaviour** and the **actual behaviour** of those systems should also be open.
5. Community contributions and feedback, from platforms, researchers and authors, should be integral to the operation and collection of the data.
Background
Altmetrics is circular. Feedback loop. Some of the same users participate in creation and analysis. Engagement helps us keep up to date.
No new barriers. Open dialogue. DIY benefits the whole community.
We had citations before we had metrics. Metrics aren’t the only use.
In depth
1. Metrics should be decoupled from the underlying data that they are derived from.
Metrics can be provided as black boxes. Researchers can’t always tell how they were created.
Clearly separate outputs (metrics) from inputs (data).
Decoupled

Metrics

Metrics

Underlying Data

{ you could remove the metrics without affecting the data }
2. Underlying data should be open, available data sets.
How do you make your research reproducible?
Publish the dataset, give it an open data license.
How do you choose to trust in a metric when making big decisions?
Publish the dataset and method so anyone can inspect it.
Other people are doing similar things.
Open data gives the best chance at compatibility and comparability.
Open Data

Underlying Data

Other uses

No restriction
3a. Underlying data sets should reflect the diversity of the behaviour being observed.
How do you capture the diversity of all online conversations?
Engage those communities.
How do you avoid cultural and selection bias?
Broaden the inputs to as wide an audience as possible. Enable the community to collect and submit data.
How do you enable novel combinations and uses?
Reduce reuse restrictions.
Observing diverse data sources

Underlying Data

Websites
Twitter
Specialist platform
3b. Metrics should be as diverse and individual as their audiences, and reflect that diversity.
How do you enable the widest variety of uses?
The right level of processing. Enough detail to capture diversity.
How do you enable researchers of mixed levels of experience to use it?
Get the model right. Make data easy to use, but include enough detail and provenance.
Serving diverse outputs

- Metrics
- Recommendation Systems
- Specialist research

Underlying Data
4. The software that collects the underlying data should be open source. Both the *intended* behaviour and the *actual* behaviour of those systems should also be open.
It’s not clear how data was collected.
Make the software open source.
Software and processes changes over time.
Include versioned software and right in the dataset.
Sometimes there are bugs.
Publish the activity logs to allow inspection of what happened.
The Internet is unreliable and heterogenous. Weird things happen.
Data should be backed up by provenance information and evidence.
Serving diverse outputs

Underlying Data

Source code

Agent collects data

Documentation

Record of activity
5. Community contributions and feedback, from platforms, researchers and authors, should be integral to the operation and collection of the data.
One organisation can never keep up to date with all the evolving communities.
Include the community in the whole process.
One organisation can never track everything.
Accept actual data from community sources.
Feedback loop

Users

Data

Feedback

Underlying Data
Altogether now
1) Metrics should be **decoupled** from the underlying data that they are derived from.

2) Underlying data should be **open**, available data sets.

3a) Underlying data sets should reflect the **diversity of the behaviour** being observed.

3b) Metrics (and other derivatives) should reflect the **diversity of audiences**.

4) The software that collects the underlying data should be **open source**. Both the intended behaviour and the **actual behaviour** of those systems should also be open.

5) **Community contributions** and feedback should be integral to the operation and collection of the data.
Underlying Data

1) Outputs are decoupled from underlying data.

2) Open data

3a) Data should reflect diversity of sources.

3b) Data should be suitable for diverse uses and audiences.

4) Open source

5) Community can feed back to underlying data & collection.

Agents

Use A

Use B

Use C

Community layers

Websites

Twitter

Specialist platform

Record of activity

Source

Docs
Who’s with me?

Full story on these principles:
- https://altmetrics.org/altmetrics18/ - Paper
- https://blog.afandian.com/2018/05/five-principles-altmetrics/ - Blog

Crossref Event Data:
- https://www.crossref.org/services/event-data/

jwass@crossref.org
@joewass
Bonus questions!