

Altmetrics, altdocuments, altlanguages...
Are your altmetrics alt- enough?



Luc Boruta, Thunken Inc.
luc@thunken.com — @thunkenizer
5:AM Altmetrics Conference, London, 2018/09/26

Diversity in altmetrics



Altmetric, ALM, Impactstory, Plum, et al. are great projects.

But **diversity is good**, and we think we can do even better.



Are your altmetrics alt- enough?

NO.



Are your altmetrics alt- enough?

Not everything that counts can be counted,
and not everything that can be counted counts.

Metrics are a sampling game.

Sampling/selection biases are an issue.



Selection biases in altmetrics: document types

Too much focus on traditional peer-reviewed publications.

Preprints are still treated as **second-class documents**.

What about **patents**, **clinical trials**, **law articles**, etc.?

What about **non-textual objects**, e.g. datasets or software?



Selection biases in altmetrics: identifiers

The ideal identifier should be **persistent**,
findable, accessible, interoperable, and reusable...

We all **copy-paste from the address bar** of our browser.

doi:**10.1351/goldbook.A00046** = <http://doi.org/ftz65g> = <http://bit.ly/2DZINQT>



Selection biases in altmetrics: languages

<1% of the world population
speak English as their first language

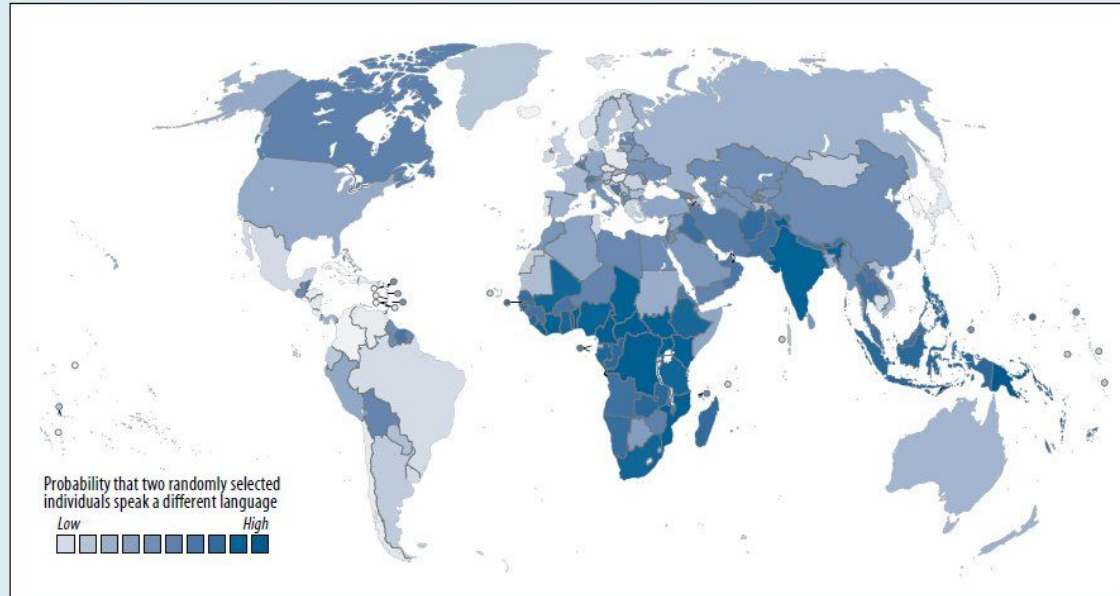
Wikipedia: **298 languages, 47.8M articles**

English Wikipedia: 5.6M articles (<12%)



Linguistic diversity is not optional

Map B6.3.1 Linguistic diversity around the world



Source: WDR 2018 team, using data from Ethnologue (2015). Data at http://bit.do/WDR2018-Map_B6-3-1.

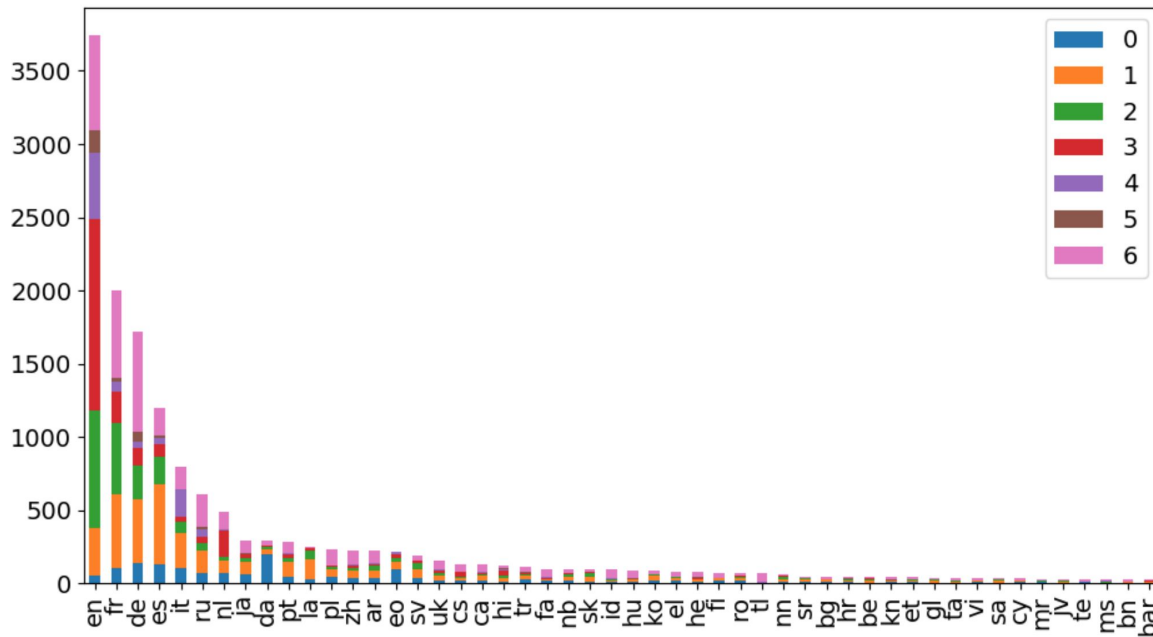
IBRD 43166 | SEPTEMBER 2017

http://bit.do/WDR2018-Map_6-3-1



Linguistic diversity is not optional

Top 50 languages of BabelBox users on Wikidata



Selection biases in altmetrics: wiki languages

Altmetric: 3 languages (en, fi, sv)

PlumX Metrics: 3 languages (en, es, pt)

ALM: 25 most popular languages

Cobaltmetrics: 180+ languages!



Natural language processing to the rescue!

Anglo-centrism is prejudicial to science.

Algorithmic complexity cannot be used as an excuse.

Citation data is mostly **machine-readable**,
and/or can be described using **local grammars**.

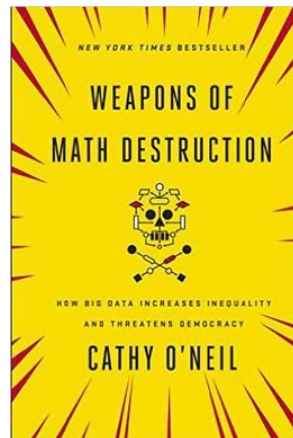


Altmetrics of math destruction

Altmetrics are a sampling game.

Imbalanced datasets reinforce discrimination.

Diversity and **fairness** must be integral parts
of how we drive our projects.



isbn:0553418815





Cobaltmetrics

cobaltmetrics.com [#altmetricsforall](https://twitter.com/thunkenizer)

Join the demo session tomorrow over lunch!

luc@thunken.com — [@thunkenizer](https://twitter.com/thunkenizer)