Toward a Second Revolution:
Data Citation, Altmetrics, and the Decoupled Journal

Jason Priem @jasonpriem
UNC-Chapel Hill School of Information and Library Science

Presented at DataCite's Summer Meeting: Data and the Scholarly Record, the Changing Landscape. August 24-25, 2011

What's a citation?

Precisely:

<resource> <uses> <resource>

More generally, we can think of citations as a trace of a person's use:

<person> <uses> <resource>

The traditional citation

<some scholar>

<wrote part of a peer-reviewed article with the vaguely-defined assistance of>

<some other peer-reviewed article>

This is a Good Thing.

1. Traditional citation maps closely to what we intuitively think of as "impact."

2. For a long time, traditional citation was the only kind of use/impact that left a trace.

But only part of the picture

1. Only one type of `person`: academics.

2. Only one kind of `resource`: scholarly articles.

3. Only one kind of `use`*:* using to support a scholarly article.

What about all the other uses?

Reading, annotating, bookmarking, saving, discussing, teaching, etc?

We start to confuse “the kind of use we can track” with *use* and “citation impact” with *impact*.

Enter the Web.

Suddenly, many forms of use are leaving traces

reference managers

blog citations

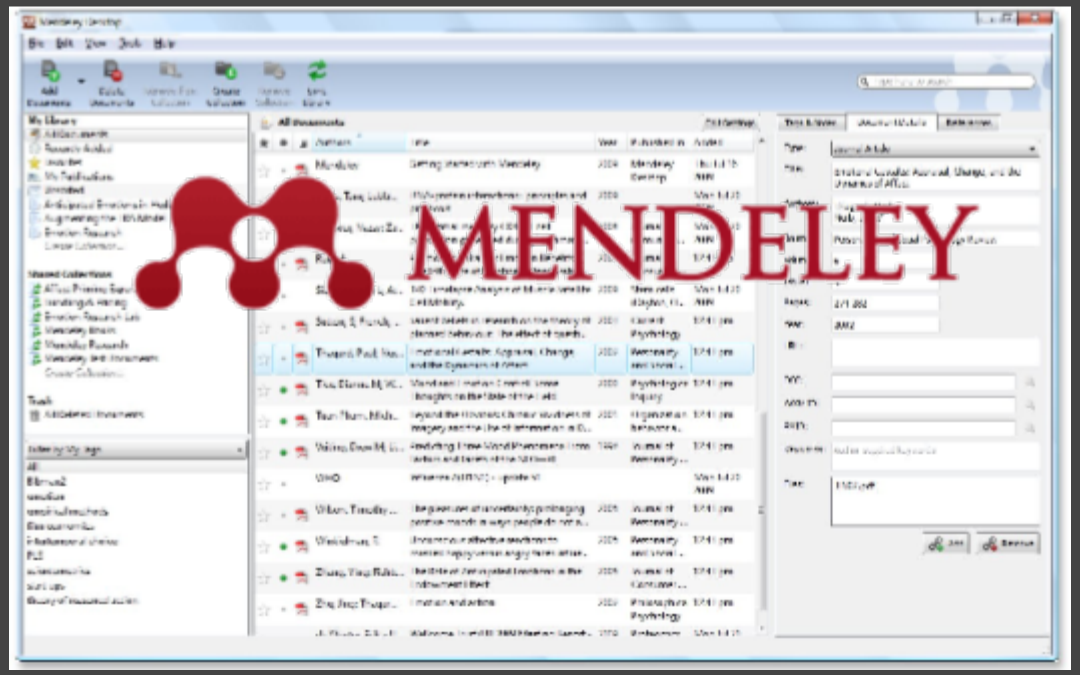
social bookmarking

social networks

Examples: Mendeley

1.1 million user libraries

106 million papers (MEDLINE has 18 million...)



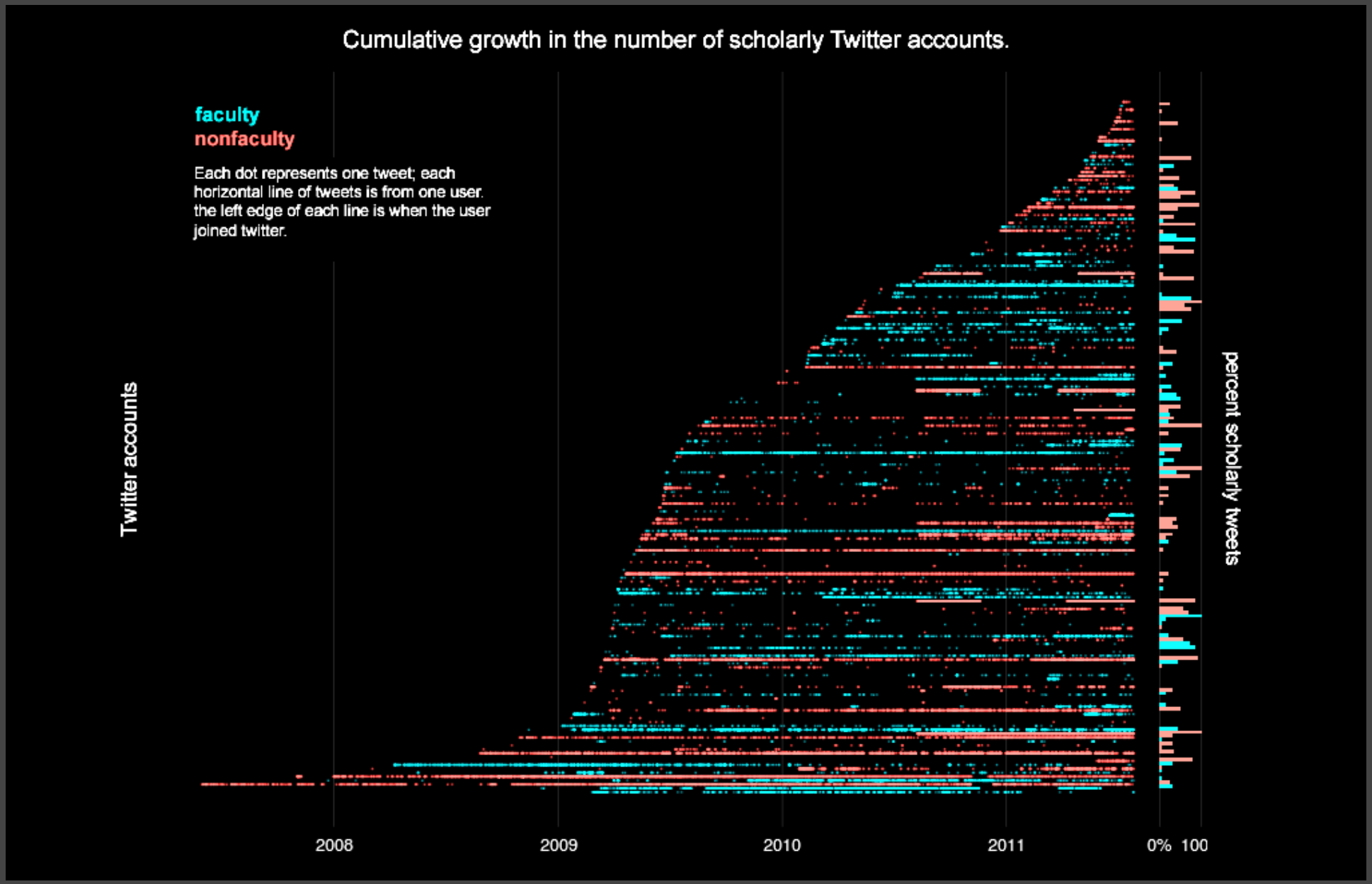
Examples: Twitter

In one month, over 58k citations from Twitter to scholarly articles (citwaitions?)

> It is like having a jury preselect what will probably interest you…. Occasionally there will be something that people will link to, and it will change what I think, or what I'm doing, or what I'm interested in.

-study participant

Examples: Twitter



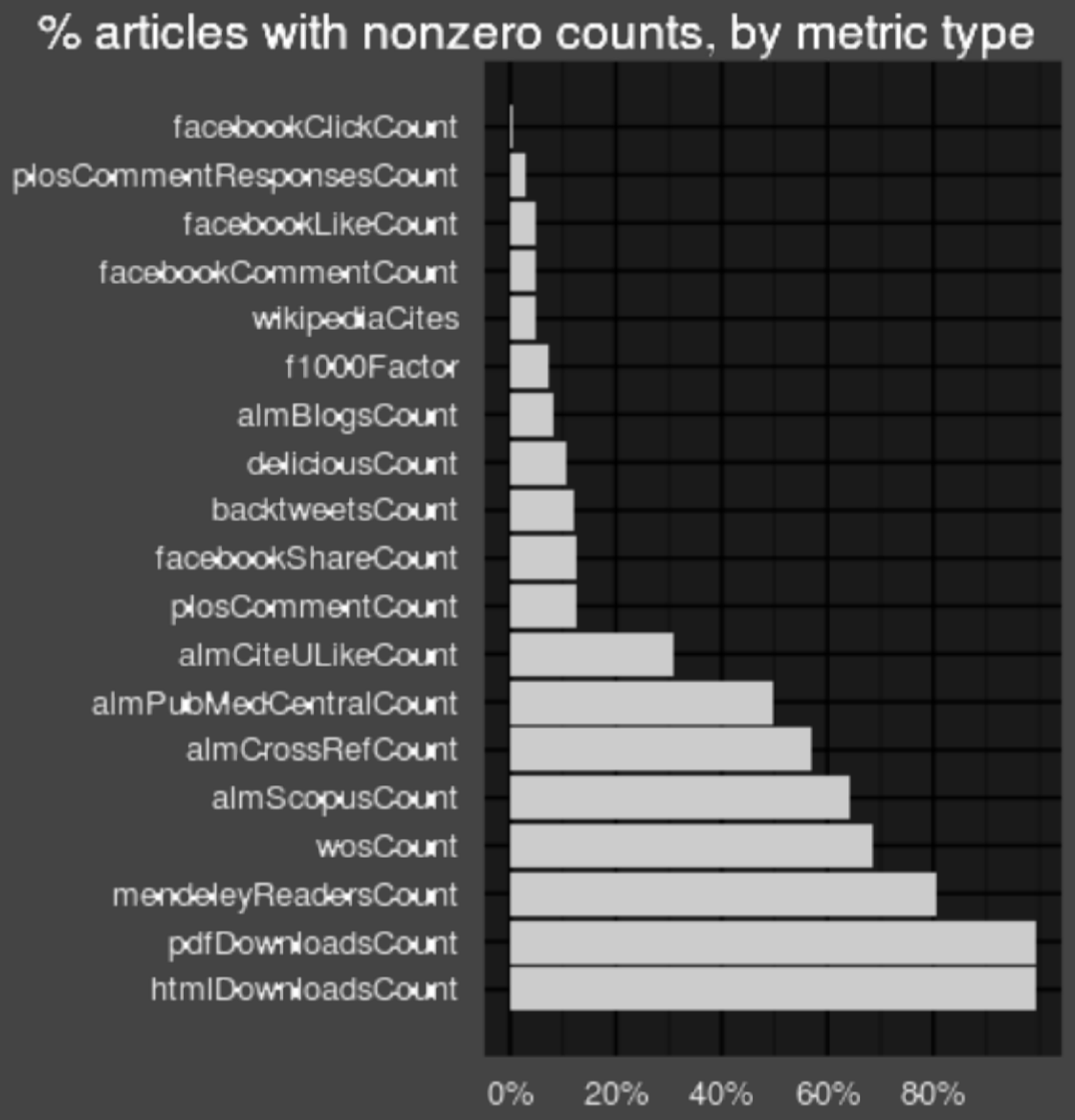
Altmetrics

Measuring use other than traditional citations.

- The Altmetrics Manifesto
- altmetrics11 workshop at ACM Web Science Conference 2011
- altmetrics groups on Mendeley and Zotero

Altmetrics: What do we know?

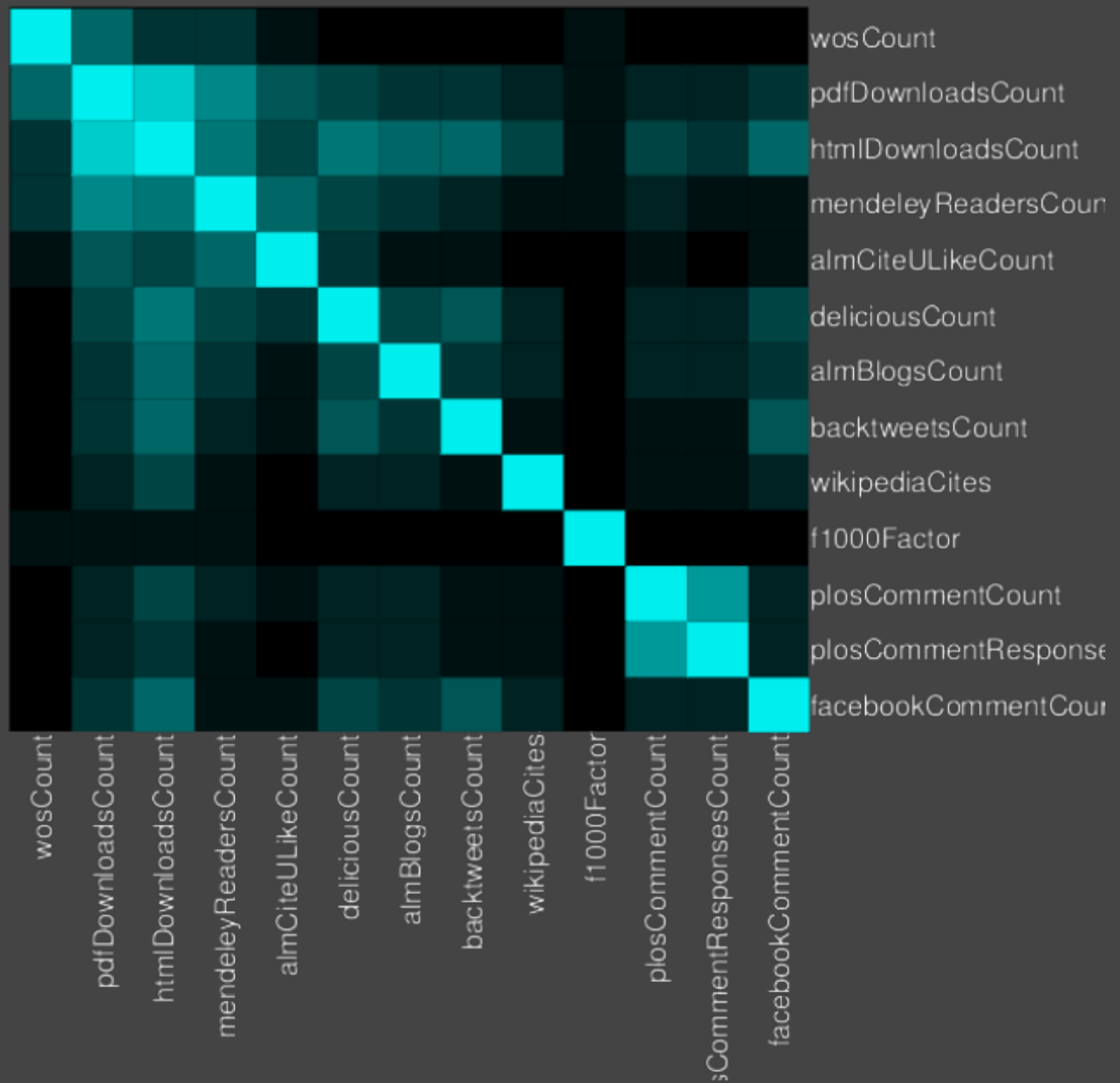
There's lots of altmetrics data out there already.



Altmetrics: What do we know?

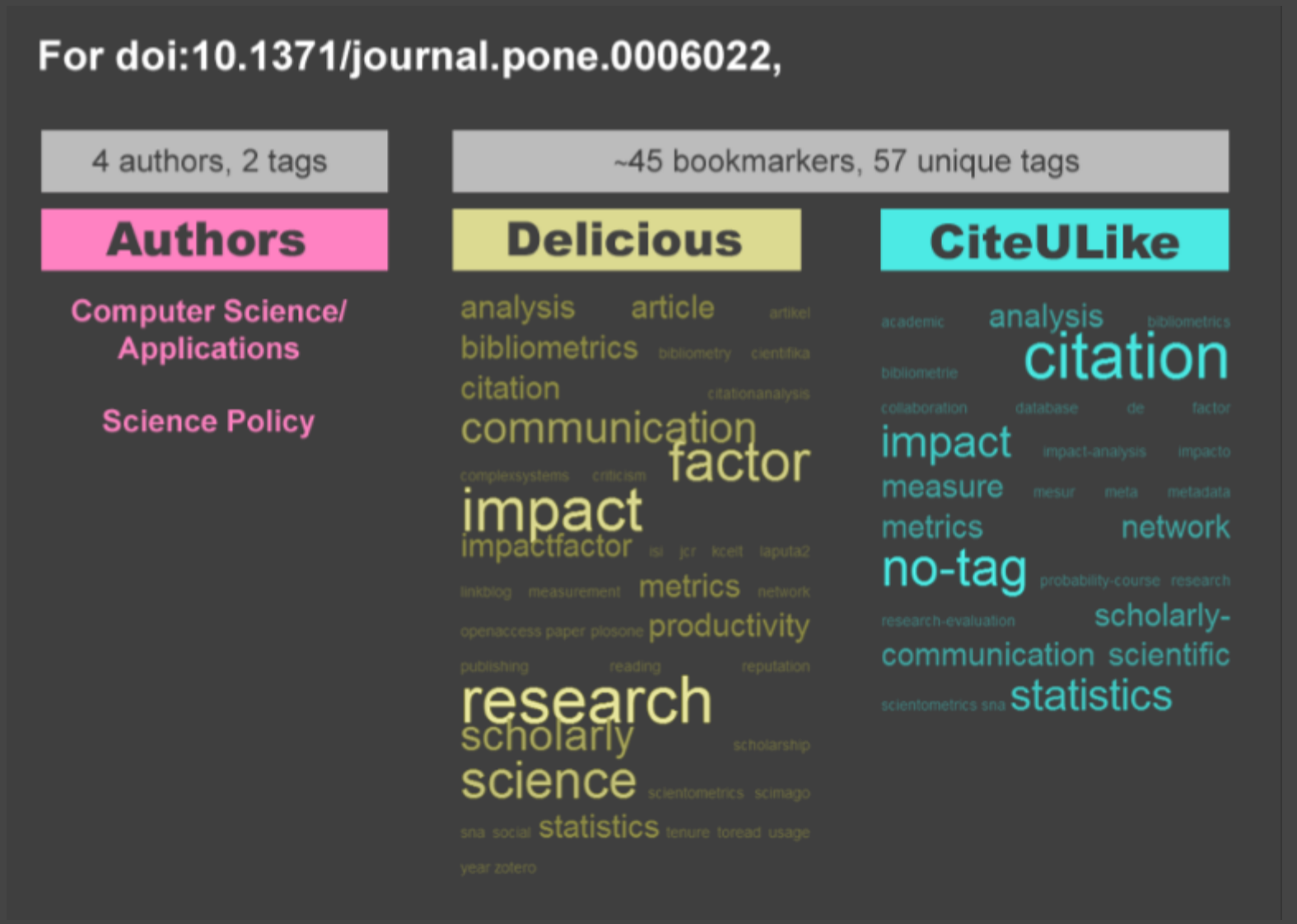
Altmetrics impact is mostly orthogonal to traditional citation impact.

We need to think about n-dimensional "impact space"



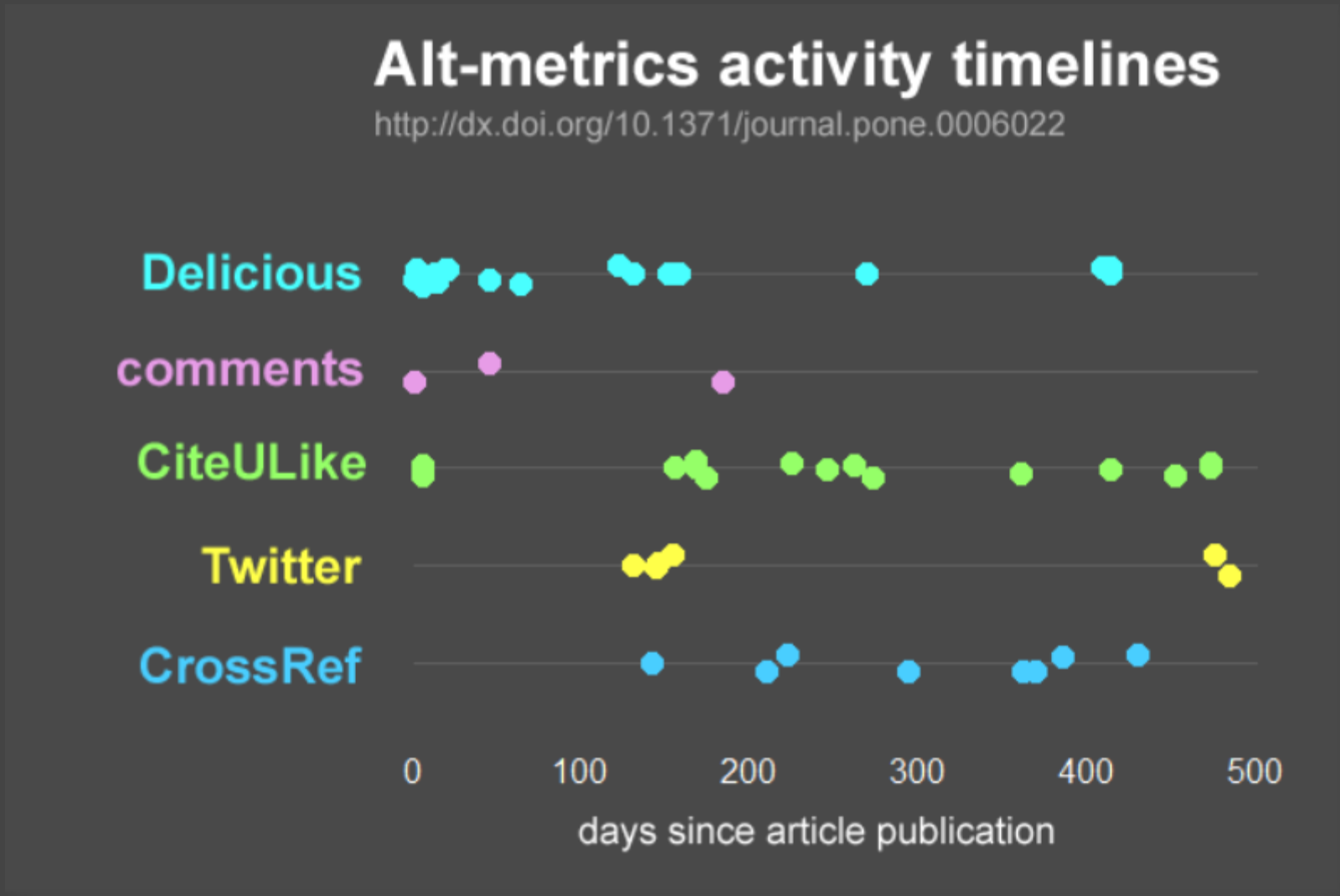
Altmetrics: What do we know?

Altmetrics can provide rich metadata, not just counts.



Altmetrics: What do we know?

Altmetrics tell a story (or sing a song?)



Altmetrics tools

- PLoS Article-level metrics
  - Designed for for journals, lots of features
- CitedIn
  - For individuals, dozens of metrics, focused on life sciences
- Total-impact: http://total-impact.org/report.php?id=hljHeI
  - For individuals, calculates impact for all sorts of products including datasets. Doesn't exactly, um, work.

Why altmetrics?

- We can more fairly evaluate researchers, publications, and institutions if we look at the whole impact picture.
- We can assess impact faster
- Etc, etc.
- Much more important: quantifying impact means we can teach machines what’**s** important**.** This changes the game. And the game needs to be changed.

Journals need an upgrade.

- First journals went hand in hand with the Scientific Revolution. Applied the most advanced technology available to the problem of spreading scholarship.
- Today's journals are still the best scholarly communication system possible using 17th-century technology.
- They've got some problems:
  - Slow
  - Restrictive format: function follows form
  - Closed
  - Inconsistent quality control
  - Hard to innovate

We don’t use the Web.

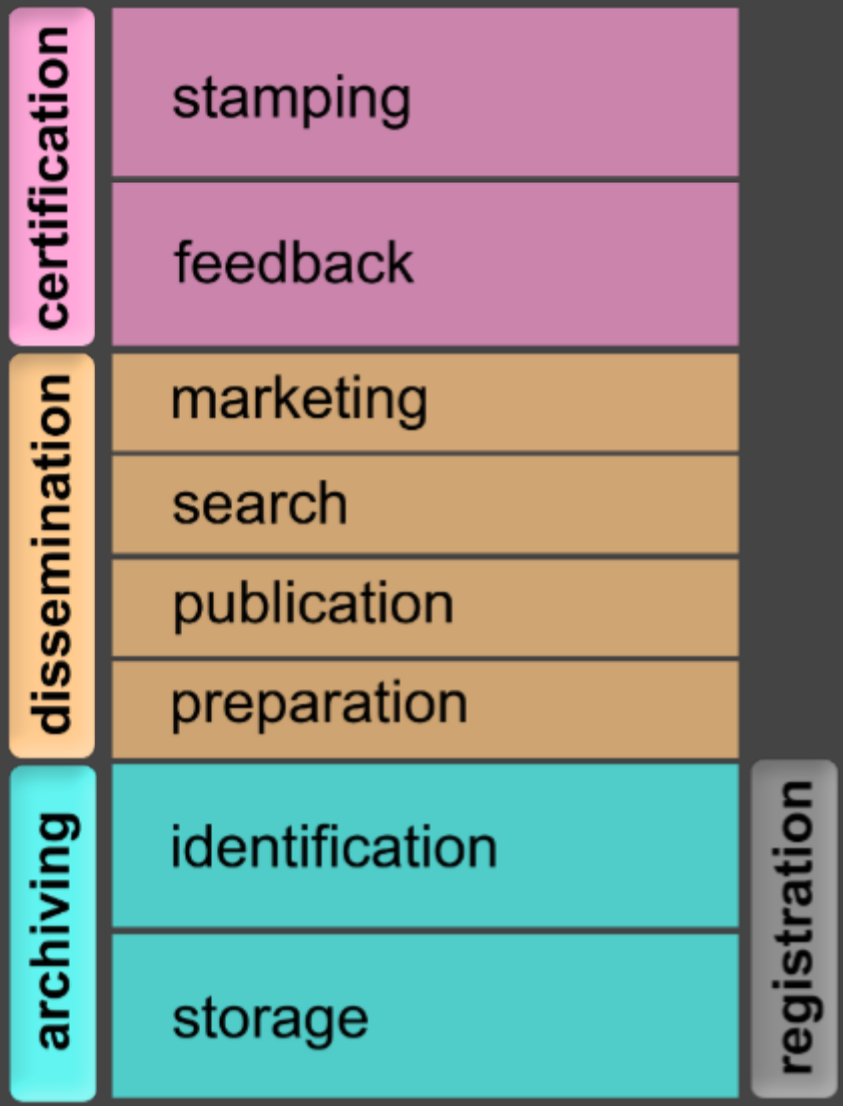
- Berners-Lee created the Web as a scholarly communication tool.
- Today the Web has revolutionized *everything but* scholarly communication.
- Online journals are essentially paper journals, delivered by faster horses.

  In the early days, CERN maintained a *list* of all the world’s Web servers. Haha, CERN ur so krazy.
- But today, we can fit *every single meaningful outlet for scholarship (*25k journals + some conferences) on one list, too.

What does the system do now?

Any modernization will have to do everything the current system does, better.

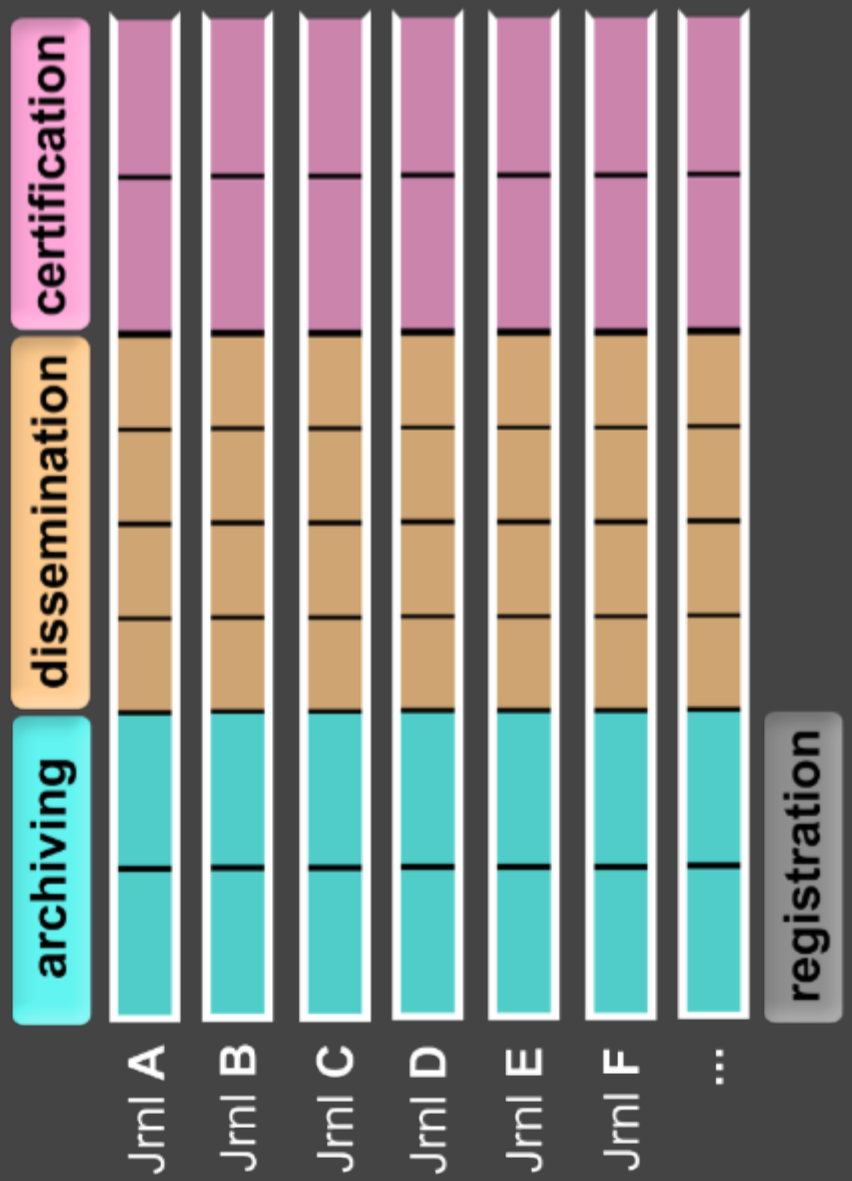
Journals have four "traditional functions."



How does this system work?

Every journal does every function itself. Each produces the same product. Little variety, little choice:

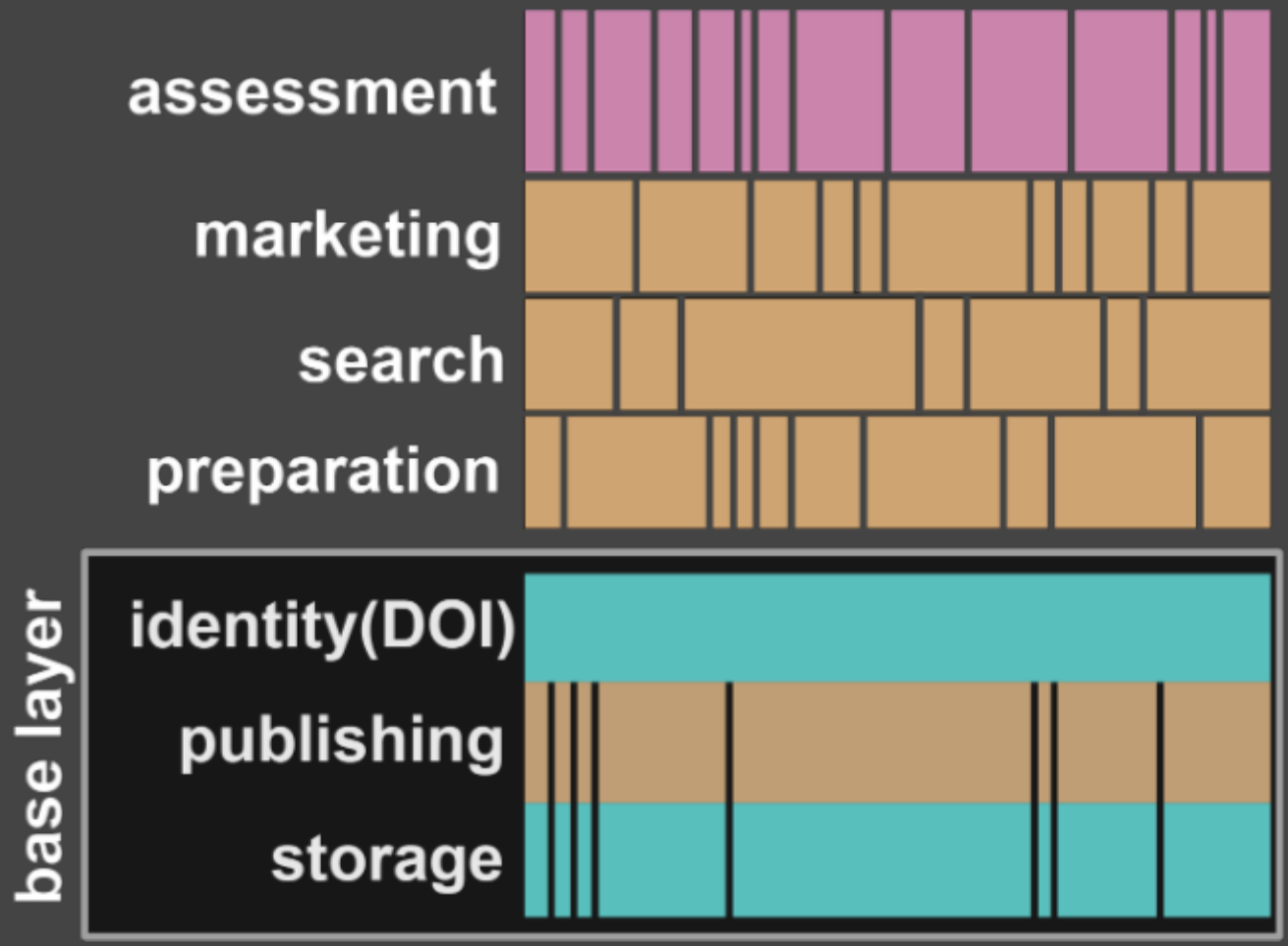
publishing as a fixed-price menu.



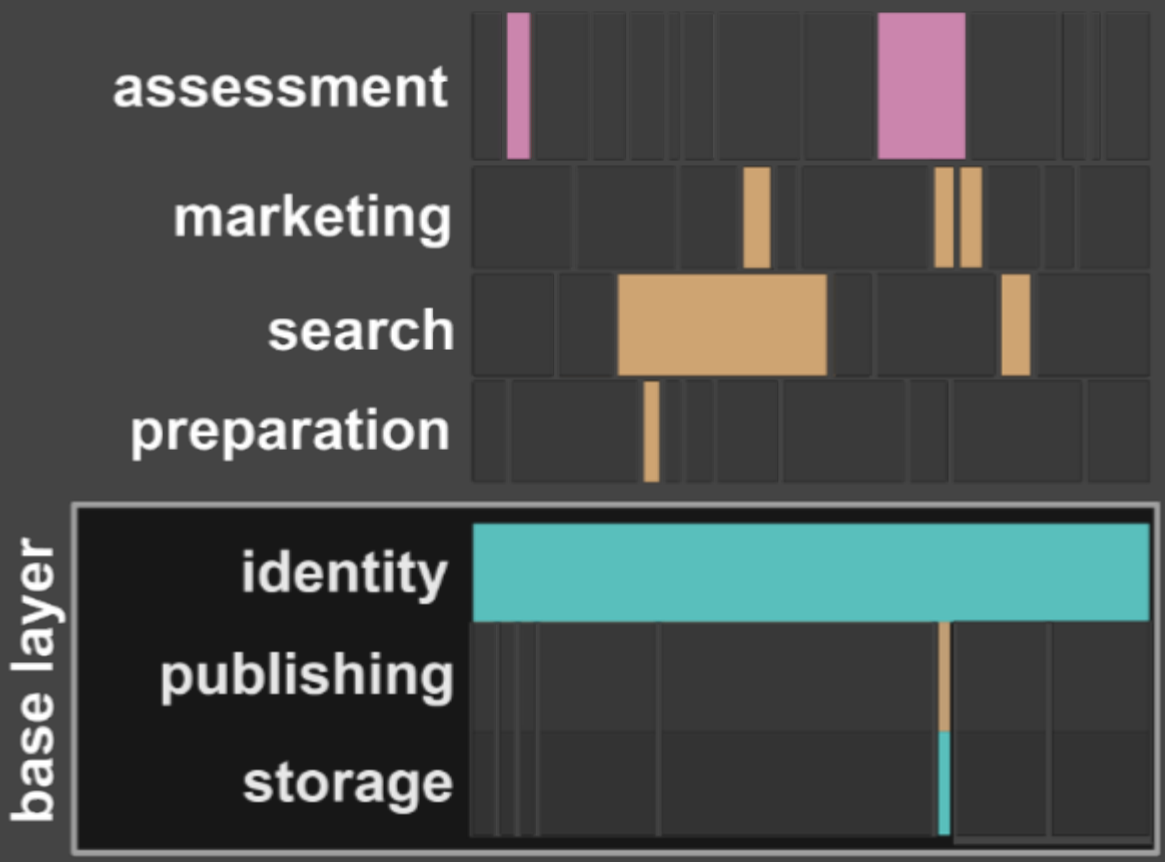
The decoupled journal (DcJ)

Functions are offered as individual services; authors pick which ones they want:

Publishing a la carte.



A DcJ example



Peer-review stamp + aggregated comments

Author's Twitter feed + mailing lists

Google + disciplinary search service

Typesetting

The product is stored and published from an institutional repository, and given a DOI as a unique ID.

DcJ properties

- Creators compete for attention in an efficient market.
- More brokers compete to deliver reputation information on more diverse content.
  - Qualitative quality control: peer review as a service
  - Quantitative quality control: aggregated review
- Layered services: form follows function
  - Three kind of items: data, annotations, and links.
  - Every scholarly product is saved, indexed, ranked, filtered

(For more on the DcJ, see prereview gDoc)

Information overload FTW

- Wait, you're going to publish *more*? We're already drowning.
- But overload is only a problem for manual curation.
- Google are not crying about data deluge—they're constantly trying to get *more* data.
- Data Judo: the more data you throw, the better the filter gets.
- Don't turn off the taps, build boats.

How do we build the DcJ?

What would the scholarly communication system look like if we built it from scratch, today?

How do we build the DcJ?

What would the scholarly communication system look like if we built it from scratch, today?

Who cares, we can't do that.

Real Question: how to smoothly transition from the antiquated system we have to a modern one?

Data citation is key

- Introduces the idea of functional citation
- Blurs the line between scholarly canon and the “expanded universe.”
- Subverts the journal-based certification paths we’ve come to rely upon. Eventually, data citation (and citation of other alt-products) will *require* a decoupled journal.

The second revolution has started.

Once we have alt-citation data, it’s too useful to ignore; alternative filters and even certification paths based on this data will open.
As Peter Vinkler says, citation graph data is like Chekhov’s gun: once on stage, it has to be fired.

A wise man, that Checkov.

Questions?

Jason Priem @jasonpriem, http://jasonpriem.com