Data Publishing and Post-Publication Reviews

Tobias Kuhn

http://www.tkuhn.ch
@txkuhn

ETH Zurich

PEERE Meeting: Peer Review: Past, Present and Future
ETH Zurich
22 April 2015

Peer Review in the Digital Age

Traditionally, peer review has two purposes:

(1) Assessment of the validity, accuracy, relevance, and quality of presentation of the given scientific work ...
but this assessment is not made public (except for the fact that some papers are accepted)

(2) Distribution of a scarce resource, namely pages in printed journals, ...
but this scarce resource no longer exists (except for presentation slots at conferences)

Is the current peer review system broken?

It's at least terribly old-fashioned and inefficient.

Publishing has Gone from Print to Online

READ THE FULL ARTICLE

Subscribe to *Nature* for full access:

$199

[Subscribe]

ReadCube Access*

preview pdf

*printing and sharing restrictions apply

Purchase article full text and PDF:

$32

[Buy now]

The Increasing Importance of Digital Data Sets

(London Underground staff sorting 4M used tickets to analyse line use in 1939)

Approaches for the
Future of Scientific Publishing

Short term:

- Open post-publication reviews

- Data publishing

- ...

Longer term:

- Publish results and reviews as Linked Data

- Global reputation system that includes artificial agents ("bots")

- ...

Open Post-Publication Reviews

- Publish first — collect reviews afterwards
- Make reviews public: Open Evaluation (similar to Open Access)

Such open post-publication reviews have existed for a long time in the case of books and book reviews:

[BOOK] **Verbal behavior**
BF Skinner - 2014 - books.google.com
In 1934, at the age of 30, BF Skinner found himself at a dinner sitting next to Professor Alfred North Whitehead. Never one to lose an opportunity to promote behaviorism, Skinner expounded its main tenets to the distinguished philosopher. Whitehead acknowledged ...
Cited by 7135   Related articles   All 22 versions   Import into BibTeX   Save   More

A review of BF Skinner's **Verbal Behavior**
N Chomsky - Language, 1959 - cogprints.org
I had intended this review not specifically as a criticism of Skinner's speculations regarding language, but rather as a more general critique of behaviorist (I would now prefer to say" empiricist") speculation as to the nature of higher mental processes. My reason for ...
Cited by 4365   Related articles   All 9 versions   Import into BibTeX   Save   More

---

See e.g.: Kriegeskorte. Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. *Frontiers in Computational Neuroscience* 6, 2012.

Benefits of Open Post-Publication Reviews

- Fast publication
- Transparent reviewing
- Publicly accessible detailed assessements of scientific works (as compared to just knowing that a paper has been accepted for a given journal)
- Incentives for reviewers
- Possibility for a multitude of evaluation metrics to be used in parallel

Data Publishing

Data sets are becoming increasingly important for most scientific disciplines, but so far they are not well integrated into the publishing and reviewing process.

Current developments:

- Supplementary material for papers
- Data journals
- Data repositories: Figshare, Dryad, ...

Publishing and Reviewing Scientific Data

Scientific data that should be published and reviewed:

- Input/output data of scientific processes (raw/aggregated)
- Scientific interpretations/conclusions (e.g. causal relation between gene and disease)
- Meta-data (e.g. involved researchers, citations, used methods)
- All of the above should ideally be formatted as Linked Data
- Source code of software

Nanopublications:
Provenance-Aware Semantic Publishing

Nanopublications are small pieces of scientific data with their provenance information, represented in a machine-interpretable language (RDF).
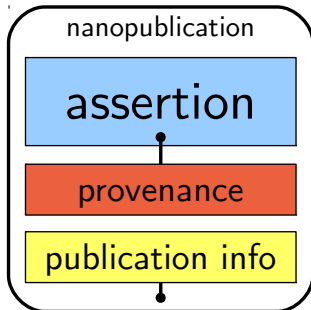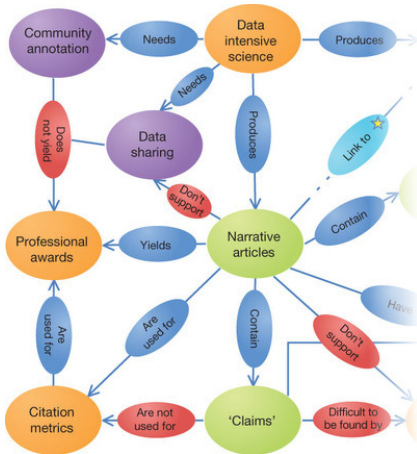

http://nanopub.org / @nanopub_org

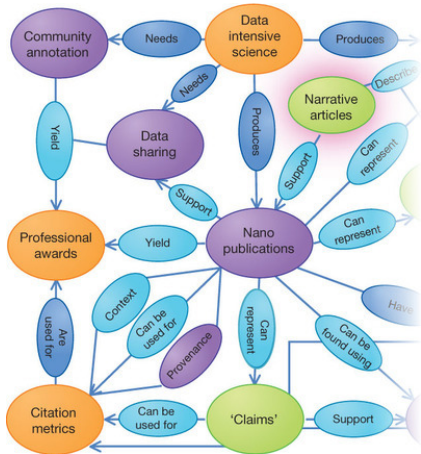# Vision: Changing Scholarly Communication

Now
Narrative articles at the center
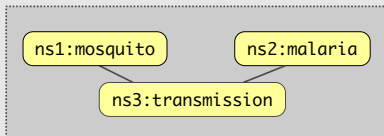
Future
Nanopublications at the center

Images from Mons et al. The value of data. *Nature genetics*, 43(4):281–283, 2011

Nanopublications:
Provenance-Aware Semantic Publishing

Nanopub0001

Assertion:

ns1:mosquito   ns2:malaria

ns3:transmission

Provenance:

`prov:wasDerivedFrom d:DataSetX`

Publication Information:

`dc:created "2013-01-01"`
`pav:createdBy p:Isabelle_Dubois`

Nanopub0042

Assertion:

p:Giuseppe   pub:123456

r:givesNegativeReviewFor
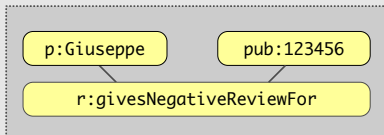
Provenance

`prov:wasInfluencedBy s:CommentX`

Publication Information:

`dc:created "2013-05-01"`
`pav:createdBy p:Giuseppe`

`http://nanopub.org` / `@nanopub_org`

How to Publish Data?

Published data should be:

- Verifiable (Is this really the data I am looking for?)
- Immutable (Can I be sure that it hasn't been modified?)
- Permanent (Will it be available in 1, 5, 20 years from now?)
- Trustworthy (Can I trust the source?)
- Reliable (Can it be efficiently retrieved whenever needed?)
- Granular (Can I refer to individual data entries?)

These points are important (much more so than for papers) because data can and will be consumed and produced automatically by algorithms ("bots").

Trusty URIs: Cryptographic Hash Values for Verifiable and Immutable Web Identifiers

Example:

`http://example.org/r1.` `RA` `5AbXdpz5DcaYXCh9l3eI9ruBosiL5XDU3rxBbBaUO70` `.trig`
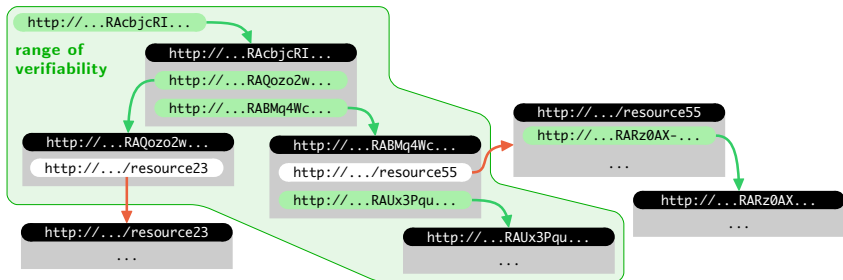
Kuhn, Dumontier. Making Digital Artifacts on the Web Verifiable and Reliable. IEEE Transactions on Knowledge and Data Engineering. To appear. / Kuhn, Dumontier. Trusty URIs: Verifiable, Immutable, and Permanent Digital Artifacts for Linked Data. ESWC 2014.

Nanopublication Indexes

Sets of nanopublications can be described as indexes that are
nanopublications themselves. This allows us to define arbitrary sets,
while keeping the individual addressability of the data entries:
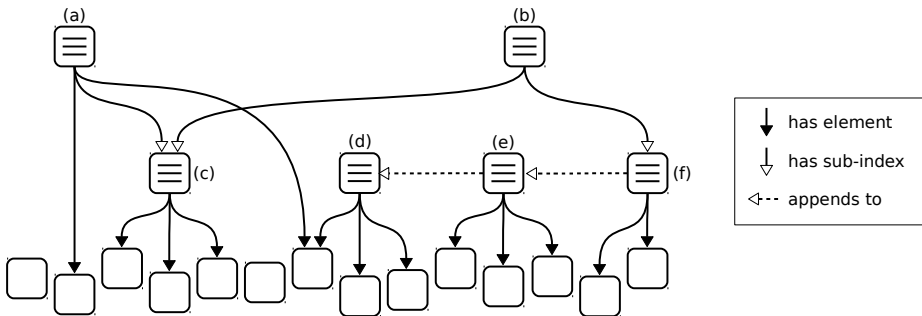
Kuhn et al. Publishing without Publishers: a Decentralized Approach to Dissemination, Retrieval, and Archiving of
Data. arXiv:1411.2749.

Nanopublication Server Network
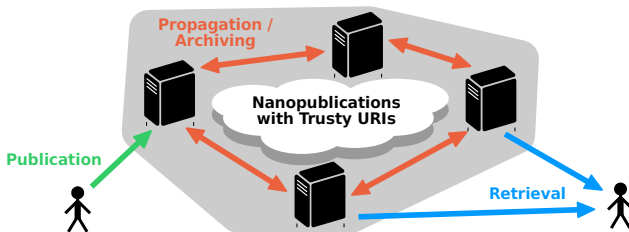
http://npmonitor.inn.ac

Kuhn et al. Publishing without Publishers: a Decentralized Approach to Dissemination, Retrieval, and Archiving of Data. arXiv:1411.2749.

How to Review Data?

- Large data sets only allow for samples of data entries to be manually reviewed
- Data is often produced and consumed by algorithms: How can these data be linked to the algorithms. How should the data and the algorithms be reviewed?
- Certain kinds of technical reviewing can be automated!
- Can all this be achieved in an fashion that is decentralized, open, and real-time?

Bots and Reputation Mechanisms

Robust automatic calculation of reputation metrics in a decentralized and open system:
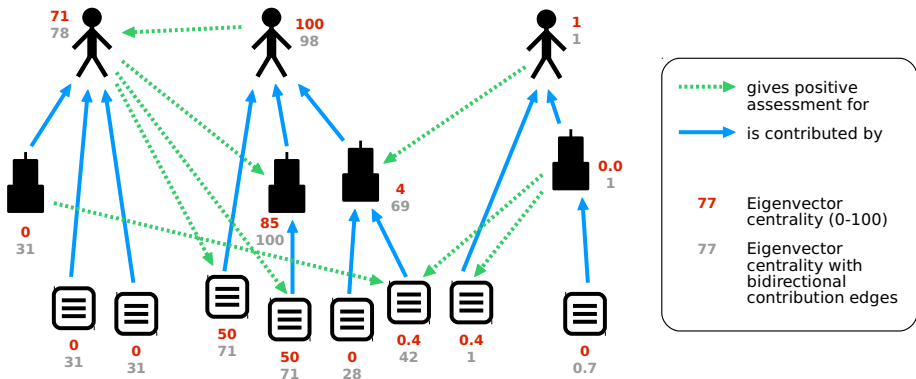
Kuhn. Science Bots: A Model for the Future of Scientific Computation? SAVE-SD 2015.

Thank you for your attention!

Questions?