UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Academia 2.0: removing the publisher middle-man while retaining impact

Raphael Poss, Sebastian Altmeyer, Mark Thompson, Rob Jelier

Angaben zur Veröffentlichung / Publication details:

Poss, Raphael, Sebastian Altmeyer, Mark Thompson, and Rob Jelier. 2014. "Academia 2.0: removing the publisher middle-man while retaining impact." In *Proceedings of the 1st ACM SIGPLAN Workshop on Reproducible Research Methodologies and New Publication Models in Computer Engineering - TRUST '14, June 2014, Edinburgh, UK*, edited by Grigori Fursin, Bruce Childers, Alex K. Jones, and Daniel Mosse, 3. New York, NY: ACM Press. https://doi.org/10.1145/2618137.2618139.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

Academia 2.0: removing the publisher middle-man while retaining impact

Raphael Poss
University of Amsterdam, NL
r.poss@uva.nl

Sebastian Altmeyer
University of Amsterdam, NL
altmeyer@uva.nl

Mark Thompson
L.U.M.C., NL
m.thompson@lumc.nl

Rob Jelier
KU Leuven, BE
rob.jelier@biw.kuleuven.be

ABSTRACT

Recent work on academic publishing has focused on transparency, to eliminate skews in favor of results channeled through already established publishers. This movement, called "open peer review", will require infrastructure. So far, proposed realizations of open peer review have relied on centralized coordinating platforms; this is unsatisfactory as this architectural choice stays vulnerable to long-term predatory commercial capture and data loss. Instead, we propose "Academia 2.0", a combination of both true peer-to-peer, distributed scientific dissemination channels, and their accompanying workflows for open peer review. It features safe decoupling of storage, indexing and search sites and supports research metrics. Our proposal relies on the existence of semantic web sites for researchers and powerful Internet search engines, an assumption which did not hold 10 years ago. We also introduce post-hoc citations, a key mechanism for quality control, impact measurement and post-hoc credit attribution for previous work. Due to the technology involved, computer engineering is likely the scientific field with the most potential to try out and evaluate our model.

Keywords

academic publishing, dissemination of science, emergent impact, semantic web, post-hoc citation

1. INTRODUCTION

What if tomorrow, all academic publishers had disappeared without a trace? This is a thought exercise. For an instant, imagine we were to wake up one morning and discover that every private organization that receives money in exchange for publishing academic knowledge had simply vanished, away with any exclusive distribution rights.

What would change? Likely, after a few days, Inter- net search engines would notice the increased load on their caches. After a few weeks, libraries would self-reorganize to handle the extra load on their lending services, mostly on- line. Researchers would start publishing their private PDF copies on their personal web pages. After a few months, printing industries would convert their contracts towards non-academic works. Mostly, science would continue unaffected for a while. Until... the turn of the fiscal year. Then, everything would change.

For one, academic administrators would start to worry about *how to assess research*: without journals, no impact factors. When pressed to prove their competitiveness, lead researchers would scramble to *find other leading work in their field* to compare against. And more significantly: libraries around the world would *release their immense subscription budgets* [1, 18, 4]; money that can be reinvested in research, and, as we would like to propose, *in new ways to curate and publish academic work*.

In this world, what could we achieve to optimize dissemination, short of recreating the publishing industry? This is what we propose to address here. Concretely, we start by identifying four recent assets of science: ubiquitous personal web sites for researchers, powerful search engines, the semantic web, and cheap computing-as-a-service. These assets are recent, developed in the last ten years. From this foundation, we propose a new, backward-compatible workflow to exchange and disseminate knowledge. In short, any author can self-publish; including reviews of other works whenever they wish. Works are securely timestamped and identified by title, author list and content hash. Identity escrow services are introduced to publish those reviews that reviewers wish to keep anonymous while retaining accountability. A new semantic object, which we call "*post-hoc citation*" can be used to assert prior work, influence or plagiarism relationships when they are discovered only after publication. A new distributed infrastructure is deployed incrementally for document indexing and lookup; it also provides public and free interfaces for search and syndication: per field of expertise, geographical area, social affinity, or relevance to a topic. We detail these interactions briefly in section 3.

Besides the removal of distribution middle men and the resulting lower cost of *access* to science, our proposal offers the following benefits: it makes richer formats possible; it lifts some current incentives to hold research results captive; it makes the dissemination network robust to the disappearance of some distribution centers; and as we suggest in section 4 it even enables new research metrics—a goal arguably distasteful, but one which we recognize as a necessary evil in a transition phase away from the current system. After outlining in section 5 how we intend to execute our proposal, we discuss in section 6 a few relevant scenarios and related work in section 7. We conclude in section 8.

© ACM 2014. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in TRUST '14, June 12, 2014, Edinburgh, U.K.

http://dx.doi.org/10.1145/2618137.2618139

2. MOTIVATIONS AND GOALS

Our goal is to *optimize dissemination*, with three subgoals: 1) make dissemination faster 2) promote open peer review 3) make dissemination as cheap as the propagation of free speech in society and 4) guarantee persistence and availability of content over time.

Multiple proposals already exist from previous work towards goals #1 and #2. We review some in section 7 and in a separate report [15]. We believe however that our proposal is the first that aims for these 4 goals simultaneously.

Profit-based publishing works against #3 by charging unacceptable [14, 4] prices for publication (OA) and/or for access (traditional). The publishing costs imposed to authors (by OA) or libraries and non-academic readers (by traditional publishers) does not reflect the low price of distributed online publishing and search. We propose to integrate the cost of scientific dissemination in the price society is already paying for access to the Internet.

Centralized publishing also works against goal #4, as well as any system that grants responsibility for distribution to a particular organization. History has taught us that organizations live and die over time, and the end of an organization that "owns" the *sources* of scientific knowledge implies partial loss of said knowledge[1]. We propose a system where persistence and availability are guaranteed by a fully distributed, world-wide peer-to-peer data store.

3. PUBLICATION MODEL

We identify the following actors: *submitters* who publish new works; *data stores* in charge of content archival and retrieval, and metadata extraction; *time authorities* that provide timestamping; *users* who search for works, read/reuse works, or wish to subscribe to notifications about new works in their field; and *query engines* that perform lookups and synthetic aggregations. Time authorities, data stores and query engines are the new actors in our proposal. We envision that university libraries bear the bulk of the cost of running data stores and query engines, although we also propose that the corresponding algorithms and infrastructure be free to use by any independent group or individual.

3.1 Document submission

Our document model is simple and unsurprising: any digital object potentially constitutes a document. This includes academic articles, but also nanopublications [5, 6], source code, data sets, interactive web pages, movies, etc. Document handles have three fields: a searchable encoding of the *title and author list*; a cryptographically secure *content hash*; and a *signed timestamp* that binds the content hash to an earliest attested date of publication.

To publish a document, anyone can compute the first two fields from the contents, then request a signed timestamp from a time authority. Provided that the title and author list are also embedded inside the document, any external observer can subsequently independently verify that the document handle is consistent with the document. Handle generation is thus fully distributed, secure, independent from centrally managed (and often expensive) registration services like DOI. Handles can be generated either by contributors themselves, or by third parties *a posteriori*; this makes it possible to insert works published prior to this proposal.

Splitting handles into three fields also makes it later possible to 1) perform separate metadata lookups by content hash, title or author; and 2) index and search for multiple versions (or revisions) of a document with the same title and author list side-by-side.

Time authorities must also be distributed and independent. We propose using distributed algorithms for proof-of-work authentication of document existence[2], similar to transaction authentication in the Bitcoin network. Failing this, we propose to use of a distributed network of timestamping authorities, coupled to the global NTP[3] network, and with signature keys validated via web-of-trust. We explicitly discourage the use of centralized services like Crossref's Crossmark[4] which would again introduce a single point of failure (Crossref's web front-end).

3.2 Document storage and indexing

Our proposal does not assume any particular location for documents, as long as they are reachable from data stores and query engines. With current technology, they could be exposed eg. as indexable URLs on any researcher's personal web pages, as files in the public document repositories of their host organization, and as yearly archives of files that can be seeded through peer-to-peer networks, eg. Bittorrent. Since document handles are location-independent, document migration and multi-homing is never an issue, although it is the responsibility of data stores to maintain (public) directories that map document handles to the actual content locations, eg. using decentralized, distributed hash tables (DHTs[5]). This scheme is intended to be integrated in the MAGNET-URI scheme[6].

The other and main responsibility of data stores is metadata extraction. Next to fully automated extraction of the usual metadata relevant for research (abstract, keywords, classifiers, supporting organizations and bibliographical references), we require an additional effort to automatically extract *geolocation* for the authors and supporting institutions, and *citation contexts*: the surrounding text where citations appear in the body of documents, with a minimum of two natural sentences and the nearest section heading if available. Even though this extraction would be best performed over structured, semantic formats [13], the state of the art in OCR technology makes it also possible to capture digitized printed materials or PDF documents.

Finally, to ensure openness and transparency, the software for indexing and metadata extraction should be available publicly so that any group or individual can run their own data stores. This, combined with the fact that any data store can be fully reconstructed from document contents and handles found elsewhere, also ensures that a particular data store owner cannot exert exclusive control over the publication workflow or the contents it has archived so far.

3.3 Post-hoc citations

We propose a new type of semantic object that can be published as standalone document, or part of another doc-

[1]Even if print copies are still available, original data sets, typesetting sources, etc. would be lost by shutting down a publishing organization.

[2]`http://www.proofofexistence.com/about`

[3]Network Time Protocol

[4]`http://www.crossref.org/crossmark/`

[5]`http://ast-deim.urv.cat/cpairot/dhts.html`

[6]`http://magnet-uri.sourceforge.net/`

http://dx.doi.org/10.1145/2618137.2618139

ument: the *post-hoc citation*. Its purpose is to *assert an influence or similarity relationship between two or more other works* after both have been published. An object of this type should be structured so as to clearly expose the handles of the works referenced, together with a descriptive assessment of the relationship by the author of the post-hoc citation.

Further exploration will determine the best formats to represent post-hoc citations; any eventual solution must ensure that data stores can unambiguously extract the following metadata: for influence relationships, the *direction* ("these works seem to have influenced those other works"), and for similarity relationships the *clustering* ("these works seem to explore similar issues in similar ways"). Natural language analysis over the accompanying descriptive assessment can also optionally extract degrees of certainty and expertise from the assessor.

We also suggest extending data store indexing algorithms to automatically recognize survey articles and register them as post-hoc citations automatically whenever appropriate.

3.4 User queries

Data stores should expose their document and metadata directory using public-facing APIs, and optionally register their presence in a federally organized network or using peer-to-peer gossiping. The key concept is that query engines should be able to perform *automatic discovery* of the data stores near them, or well-known data stores worldwide. A number of existing standard Internet protocols can be leveraged for this purpose, including DNS, IRC and XMPP. In this environment, the responsibility of a query engine is to *translate user queries* into API requests throughout the federated network, supported by an adequate network of caching proxies. Given the metadata described previously, the reader can readily satisfy themselves that the usual queries can be directly served: "match by document handle, keyword, title, author, date range", "related work using direct citations (cited by / citing) and post-hoc citations (explicitly related)" or "all works on a given topic in a given geographical area, or from all areas a contributor has published from."

The advantage of our proposal comes from the exploitation of citation contexts (section 3.2). Using natural language analysis or explicit semantic qualifications in a structured format, it becomes possible for query engines to *qualify* citations: whether a citation is perfunctory, serves as witness for an assumption, for related work, previous work, as criticism, as appreciation, etc. Thanks to the computational power available at low cost in contemporary datacenters, such analyses can be performed systematically without much overhead. Using this additional knowledge, we can not only extend usual queries for related work, but also search for impact (cf. also section 4 below):

- works that use a given work's outcome as starting assumption (for dependency analysis);
- works that are well-liked;
- works that support/confirm a given work, or works that are not yet confirmed;
- works that contradict a given work, or works that are not yet contradicted;
- controversial works that are supported by some derived works but criticized by others.

As an optional extension we suggest that query engines also register with social network providers, and extend their query language to account for the social relationships between researchers. This would suitably extend the scope of queries for related work, as we can assume that researchers who advertise a close work relationship on a social network are likely to influence each other in their research.

The combination of citation contexts with geolocation data and the social graph would enable also higher-level queries, such as "identifying topics and keywords for clusters of related works that mostly gather consensus in one geographical area but are contradicted or criticized in another" or "identify clusters of related works that also happen to be produced by contributors in social clusters" which would be relevant to scientific directors or policy makers when determining long-term research plans. We have only barely started to explore the wealth of data mining opportunities that would arise from such an infrastructure.

Nevertheless, we highlight again that the software and infrastructure for query engines should be public and open, so as to enable the creation of new query sites by any group or individual over time. Although suitable protocols must ensure that the infrastructure remains robust to malicious behavior (eg. denial of service attacks), we emphasize that the resulting ecosystem must stay fully distributed so as to resist concentration of influence by a few organizations.

3.5 Distribution channels

Search results can be presented directly to querying users, but a decade of experience with search engines has taught us two other ways to use *saved queries*: notifications and syndication. Notification (of changes in search results) is a trivial application; we are especially interested in syndication here, as *syndication can be advantageously used to complement or eventually replace journals.*

We should recognize here a strong and practical need of researchers, one that is only partially and poorly served by current publication channels: the subtle combination of filtering (selecting the "good" works in an ocean of mediocre scientific outputs) and widening: selecting works of unknown quality but "potentially interesting" by virtue of being unconventional, controversial, etc.

Here, syndication would shine. Syndication is the automatic merging of the stream of results produced over time from multiple different queries in a search engine; a number of web standards for syndication (eg. RSS, Atom) already exist. When coupled with an existing syndication reader tool, syndicated results can be packaged weekly or monthly like a journal or magazine: featuring well-liked and well-referenced works in the forefront, followed by potentially interesting works, and so forth. Again, university libraries could take the responsibility to maintain access portals with pre-defined queries and syndication channels, ready for use by their guest researchers and the general public.

4. METRICS

Contemporary impact metrics for journals and individual researchers are commonly based on some properties of their citation graph (h-index, g-index, c-index, etc.).

Perhaps surprisingly, these metrics translate trivially. Instead of per-journal impact factors, we can determine impact per topic cluster (identified by the most often used queries), by measuring citations to the works that most often appear in search results or syndication channels. Additionally, impact computations can be further refined by citation context analysis, ie. whether the citations are impactful or perfunc-

http://dx.doi.org/10.1145/2618137.2618139

tory, and by post-hoc citations. All existing citation graph metrics can be computed directly from data stores using their usual definitions. Similarly, most of the metrics defined by the Altmetrics project [16, 3] would also adapt to our proposal transparently. Since altmetrics are based on the "external use" of documents (ie. how they are embedded in non-academic works), they would automatically extend to our proposed environment without changes.

An interestic feature that emerges from the use of syndication is *self-clustering*: the metrics are only defined over the queries that users actually use, or even define themselves. Impact emerges from the clustering of multiple users interested in the same topics (and thus using similar queries). Instead of journals whose editorial board decides by fiat what topics should be interesting, our proposed ecosystem will group works using the actual scientific interest in the field.

Remarkably, the distributed nature of the proposed infrastructure makes it unavoidable to acknowledge that each sample of a metric is fundamentally dependent on the location it was sampled from. Indeed, a work can be highly influencial in a local area, reflected by metrics in the local query engines, and relatively irrelevant in another. We consider this feature to be a new opportunity that will enrich the depth and quality of interactions between research staff and administrators. Meanwhile, we also highlight that the wealth of metadata made available by our proposed infrastructure is also likely to enable new metrics. We intend to explore this opportunity with peers organically over time.

5. IMPLEMENTATION & ADOPTION PATH

At the time of this writing, the desired course of action is a suitable gathering of bright minds to scrutinize this proposal and discuss use scenarios. Assuming that our proposal stands the test of criticism and emerges from it stronger, we foresee the possibility of an incremental adoption based on the following strategy:

1. iron out the secure timestamping protocol, since public trust in the time ordering of publications is entirely dependent on it, by appropriate academic interactions with the computer security community;

2. start implementing metadata extraction algorithms for data stores; and simultaneously, research extraction from existing document stores and advertise the use of semantically structured document types, eg. nanopublications [13];

3. exploit the metadata extraction to proactively generate searchable and queriable metadata stores over existing repositories of documents, eg. the private archive of the libraries of a few universities;

4. implement a first generation of query engines and integrate existing syndication technology; over the archives prepared in the previous steps, implement metrics and start advertising how they correlate and can supersede the metrics available from current publishers;

5. from this point, use networking effects to strengthen the platform and gather contributors to its technology.

We highlight here that this approach does not require the disappearance of academic publishers. It can be freely applied in complement to the existing ecosystem, and is especially suitable for "grassroots", organic growth directly from the researchers interested to take direct action.

It is also paramount that the implementation work be supported by public funds, so as to ensure no commercial interests re-introduce the problematic situations that we are trying hard to get rid of. Hopefully, we foresee that most if not all of the "core" technology components are already available as libraries, either from the cryptographic or the text mining communities.

6. DISCUSSION

In the process of refining our proposal, we have considered a number of scenarios involved in academic publishing, especially those where complex incentive systems (see [8] and its bibliographic references) require a careful balance of features in the publication workflow. Reviewers to an earlier version of this article have also raised a number of relevant questions. Due to space limits, we cannot reproduce all our answers here; instead we detail them in a separate technical report [15] which we intend to grow over time in response to further refinement of "Academia 2.0" by the community.

Meanwhile, we start here by focusing on two major topics dear to all audiences: how to properly deal with *fear of credit loss* and *reviewer independence*. Our complementary report also already includes a discussion of double blind reviews, the file drawer effect, and dealing with bad reviews.

6.1 Fear of credit loss

Some researchers have strong feelings against the mandatory publication of software, tools and dataset next to research results, arguing that this openness and transparency will enable competitor researchers to exploit said tools and then publish results earlier than the original author. When this happens, the first researcher has to bear the main cost of the research (time, effort) but cannot reap the profits (a lot of emphasis in reward systems is put on which researcher is first to publish), which is quite unfair indeed. This incentive to avoid publication hinges on both the reality of unscrupulous researchers, and the reality of the long publication time for journal articles (commonly up to one or two years between submission and final acceptance).

Obviously, our proposal eradicates the second factor as publication then becomes essentially instantaneous. As for unscrupulous researchers, our proposal has a number of built-in features which are relevant.

First, secure timestamping certificates can be obtained by a researcher on his or her preliminary results, *before they are actually available publicly*: our proposed time authorities can issue a timestamp based only on the hash of a document's content, so the actual content can remain private until a later date. This enables a researcher to operate using the following workflow:

1. prepare the tools;
2. prepare some preliminary results using them;
3. privately obtain a timestamp certificate, which attests the work's existence although it is not published yet;
4. publish the tools;
5. later, when more results are obtained, publish the high-level outcomes of the research.

Using this process, if a competitor exploits the published materials from step #4 and claims original work similar to #3, the first researcher can assert his or a her prior work *a posteriori* using the timestamp certificate.

Moreover, supposing the original researcher did not realize that further results have been published by competitors without attribution or failed to request timestamp certificates early, the post-hoc citation mechanism can be used as well: the research or even a peer can publish a statement

http://dx.doi.org/10.1145/2618137.2618139

of influence, from the "real" original work to the competitor work, after both have been published and *regardless of publication order*. In practice:

1. some work is discovered to be likely "heavily inspired" on some other work;
2. one or more researchers publish post-hoc citations declaring the same;
3. the social network of the interested parties scrutinizes the relationship and, when deemed relevant, strengthen the weight of the post-hoc citations in the search network by adding *positive reviews* for the post-hoc citations themselves.

Using these steps, both original and derived work become linked by the post-hoc citations. Agreement by peers strengthen the post-hoc citations. To fully exploit this opportunity, we envision that query engines list both sides of post-hoc citations when a search query would otherwise only return one side, together with the relevant context (including comments) around post-hoc citations.

6.2 Independence of reviews

Since all knowledge queries and evaluations in our proposed ecosystem must be based on information publicly available, it is only natural to require reviews to be published as well. Indeed, we envision reviews to be simply structured documents subject to identification, storage and indexing in the same way as any other document.

While the idea of open reviews is not new [7], we are making a step forward from previous work that assumes that the repository of review objects is centrally stored (cf. section 7): we suggest that any researcher should be able to self-publish reviews and let those reviews organically impact ranking, search results and evaluation over the distributed network of query engines. In each query engine, existing ranking algorithms from open peer evaluation can be reused (cf. section 7 for links); however we highlight the opportunity to use natural language analysis on the review text itself using the citation contexts mandated by our proposal, to account for subjectivity when weighing numerical scores.

Moreover, the distributed nature of our proposal helps overcome two known obstacles to open peer review. For one, showing the public identity of reviewers may impact the integrity of researchers: one may not be able to objectively and publicly criticize poor work performed by colleagues, or a potential future colleague in a hiring position. To compensate for this, we suggest that public organizations (eg. libraries) propose *review anonymization* as an optional service. Using this service, the reviews are public but published under a pseudonym, and the library is responsible for keeping track of real identities. This escrow service for researcher identities would also protect accountability: consistently poor reviews could then be tracked to their real authors after suitably authorized investigations [15].

The question also exists of how to organize blind reviews, ie. independent reviews on a new work. Even with open peer evaluation, the process involves a third party authority to both *call* the reviewers, and *manage* the review process. With our proposal, the third party is only needed to call, not to manage. This is because reviewers can self-publish their reviews. By virtue of reviews being public, interactions between reviewers can be more readily scrutinized by third parties afterwards. The *opportunity* for this scrutiny would bear sufficient pressure on reviewers to behave ethically. This is a general advantage of open peer review. The particular benefit of our proposal is that secure timestamping would help reviewers to coordinate directly: they could agree to review within a time period, register the timestamp when they complete a review, and actually release the reviews publicly only at the previously agreed later date.

Beyond these points, we have explored [15] other review-related scenarios, together with the reward systems currently in place for revies. We are confident that we are not introducing significant new issues in this area that do not already exist in the current ecosystem, although we obviously invite external experts to chime in on this topic.

7. RELATED WORK

Our proposal shares goals with previous initiatives, projects and commercial products, although we believe no previous work share all the goals simultaneously. For example, the idea of document handles independent from the physical location of a document, necessary for persistence, is featured by Internet URLs and DOIs; yet URLs are still dependent on the DNS system and DOIs on a central mapping database, both incurring registration costs, which in turn run against cheap dissemination. So far we know, our proposal is the first to use content-based addressing for scientific works.

We promote open peer review, an old concept [7] not yet commonplace: processes and technology where reviews are published and the review process is fully transparent. The current consensus is that open reviews are beneficial [9, 8, 17, 2] although anonymity may be a desirable feature [10]. Our proposal strongly promotes open reviews while enabling accountable reviewer anonymity (cf. section 6.2, [15]). There also exist different forms of open peer review; FCN has published in 2011-2012 a series [12] on this topic. For example, the proposal by N. Kriegeskorte [11] largely overlaps with our proposal regarding reviews; it also discusses extensively the related benefits, pitfalls and incentives for reviewers, all of which applies just as well within our proposal.

However, we found that most previous work on open peer review has focused on process and requirements, and was silent on the topic of platforms. The questions "*where* does the process take place?" and "where to find the works?" if at all answered, suggest either OA journal portals, or web platforms with a unique entry point. We believe these authors assume the documents and their open reviews must be identified in listings that must in turn be managed by centralized authorities. We have not been able to identify previous work that embraces the concept of peer-to-peer distribution, content-based addressing and distributed hash tables, and thus couples the process of open peer review to a truly distributed storage and query system as we propose.

8. CONCLUSIONS AND FUTURE WORK

We have presented Academia 2.0, a model for an ecosystem for the dissemination of scientific results that supports arbitrary document formats and does not require academic publishers. It is based on a fully distributed workflow over the Internet, where primary storage, indexing and search can be implemented in different locations by arbitrary parties, so as to reduce incentives to concentrate control over distribution into few hands. Documents are identified independently from their physical location and timestamped securely. Metadata extraction also registers citation contexts and geolocation for authors and institutions. A new

http://dx.doi.org/10.1145/2618137.2618139

semantic object, post-hoc citations, reveals previous work or similarity relationships even after both the related works have been published. Distributed query engines and especially syndication are leveraged to customize dissemination channels to the research interests of each possible audience, either researchers or from the general public.

Careful analysis suggests that our proposal does not suffer from incentives to avoid early publication exhibited by the current ecosystem, including Open Access. The specific contributing features are post-hoc citations, indexable reviews and secure timestamping. Although our proposal may stay sensitive to some drawbacks already found in the current ecosystem, we strongly believe that it avoids multiple well-known and well-understood issues in academic publishing.

With this presentation, we hope to attract the attention of the computer science community: we consider this is the community most likely to be able to tolerate installing early versions of the necessary distributed support software while playing both the role of authors and knowledge users.

Also, despite the inflammatory premise, Academia 2.0 can be introduced gradually, side-by-side with the existing ecosystem. The proposed data stores and query infrastructure can be already developed and prototyped, capturing the research documents already available on the personal web sites of individual researchers. In other words, we could start working on it today.

Acknowledgements

The authors would like to thank the anonymous reviewers for their early feedback, as well as A. Oprescu, M. Verstraaten and R. Piscitelli for their thoughtful suggestions.

9. REFERENCES

[1] T. Bergstrom. Free labor for costly journals. *Journal of Economic Perspectives*, 15(3):183–198, 2001. URL: http://www.escholarship.org/uc/item/5w0865j3.

[2] E. Dzeng. How academia and publishing are destroying scientific innovation: A conversation with Sydney Brenner. *King's Review*, February 2014. URL: http://kingsreview.co.uk/magazine/blog/2014/02/24/how-academia-and-publishing-are-destroying-scientific-innovation-a-conversation-with-sydney-brenner/.

[3] F. Galligan and S. Dyas-Correia. Altmetrics: Rethinking the way we measure. *Serials Review*, 39(1): 56–61, 2013. ISSN 0098-7913. doi:10.1016/j.serrev.2013.01.003.

[4] T. Gowers. Elsevier journals — some facts [online]. April 2014. URL: https://gowers.wordpress.com/2014/04/24/elsevier-journals-some-facts/.

[5] A. J. G. Gray, C. Chichester, K. Burger, S. Kotoulas, A. Loizou, V. Tkachenko, A. Waagmeester, S. Askjaer, S. Pettifer, L. Harland, C. Haupt, C. Batchelor, M. Vazquez, J. M. Fernández, J. Saito, A. Gibson, L. Wich, T. Kuhn, and J. van Dam. Nanopublication guidelines. Technical Report Working Draft, Concept Web Alliance, December 2013. URL: http://www.nanopub.org/2013/WD-guidelines-20131215/.

[6] P. Groth, A. Gibson, and J. Velterop. The anatomy of a nanopublication. *Information Services and Use*, 30 (1-2):51–56, January 2010. ISSN 0167-5265. doi:10.3233/ISU-2010-0613.

[7] T. Gura. Scientific publishing: Peer review, unmasked. *Nature*, 416(6878):258–260, March 2002. ISSN 0028-0836. doi:10.1038/416258a.

[8] S. Harnad. Implementing peer review on the net: Scientific quality control in scholarly electronic journals. In R. Peek and G. Newby, editors, *Scholarly Publishing: The Electronic Frontier*, pages 103–118. MIT Press, 1996. Address: Cambridge Ma. URL: http://eprints.soton.ac.uk/252900/.

[9] D. F. Horrobin. The philosophical basis of peer review and the suppression of innovation. *JAMA*, 263(10): 1438–1441, 1990. doi:10.1001/jama.1990.03440100162024.

[10] T. Jefferson, P. Alderson, E. Wager, and F. Davidoff. Effects of editorial peer review: A systematic review. *JAMA*, 287(21):2784–2786, 2002. doi:10.1001/jama.287.21.2784.

[11] N. Kriegeskorte. Open evaluation (OE): A vision for entirely transparent post-publication peer review and rating for science. *Frontiers in Computational Neuroscience*, 6(79), 2012. ISSN 1662-5188. doi:10.3389/fncom.2012.00079.

[12] N. Kriegeskorte, A. Walther, and D. Deca. An emerging consensus for open evaluation: 18 visions for the future of scientific publishing. *Frontiers in Computational Neuroscience*, 6(94), 2012. ISSN 1662-5188. doi:10.3389/fncom.2012.00094.

[13] B. Mons, H. van Haagen, C. Chichester, P.-B. 't Hoen, J. T. den Dunnen, G. van Ommen, E. van Mulligen, B. Singh, R. Hooft, M. Roos, J. Hammond, B. Kiesel, B. Giardine, J. Velterop, P. Groth, and E. Schultes. The value of data. *Nature Genetetics*, 43(4):281–283, April 2011. ISSN 1061-4036. doi:10.1038/ng0411-281.

[14] J. Naughton. Academic publishing doesn't add up. *The Guardian / The Observer*, April 2012. URL: http://www.theguardian.com/technology/2012/apr/22/academic-publishing-monopoly-challenged.

[15] R. Poss, S. Altmeyer, M. Thompson, and R. Jelier. Aca 2.0: Questions and answers. Technical report, University of Amsterdam, May 2014. arXiv:1404.7753.

[16] J. Priem, D. Tarborelli, P. Groth, and C. Neylon. altmetrics: a manifesto, 2011. URL: http://altmetrics.org/manifesto/.

[17] S. van Rooyen, F. Godlee, S. Evans, N. Black, and R. Smith. Effect of open peer review on quality of reviews and on reviewers recommendations: a randomised trial. *BMJ*, 318(7175):23–27, January 1999. doi:10.1136/bmj.318.7175.23.

[18] L. Westberg. Purdue re-signs contract for online scholastic access. *The Exponent Online*, December 2011. URL: http://www.purdueexponent.org/campus/article_779ca6a6-2839-11e1-bf49-001a4bcf6878.html.