What does Academia_edu's success mean for Open Access? The data-driven world of search engines and social networking

blogs.lse.ac.uk/impactofsocialsciences/2015/10/22/does-academia-edu-mean-open-access-is-becoming-irrelevant/

10/22/2015

With over 36 million visitors each month, the massive popularity of Academia.edu is uncontested. But posting on Academia.edu is far from being ethically and politically equivalent to using an institutional open access repository, argues **Gary Hall**. *Academia.edu's financial rationale rests on exploiting the data flows generated by the academics who use the platform. The open access movement is in danger of being outflanked, if not rendered irrelevant by centralised entities like Academia.edu who can capture, analyse and exploit extremely large amounts of data.*

At the Radical Open Access conference at Coventry University in June, I spoke briefly about Academia.edu as part of a session with Stuart Lawson and David Harvie on Radical Accountability. A number of people asked afterwards if I would be publishing a written-up version of those comments. Then, at The Sociality of Sharing event at the Centre for Interdisciplinary Studies, University of Warwick, in September, some of the participants found they had each been approached separately by Academia.edu to join their 'Editor Program' (i.e. act as an unpaid editor for Academia.edu, recommending publications appearing on the platform to others in their areas of research expertise), and were keen to know more about its philosophy and business model. So here are my brief thoughts on the subject. I will post them on Academia.edu using the title, 'Should This Be the Last Thing You Read on Academia.edu?'

A brief discussion took place this month on the Association of Internet Researchers air-l listserve concerning a new book from the publishers Edward Elgar: *Handbook of Digital Politics*. Edited by Stephen Coleman and Deen Freelon, this 512 page volume features contributions from Peter Dahlgren, Nick Couldry, Christian Fuchs, Fadi Hirzalla and Liesbet van Zoonen, among numerous others. The discussion was provoked, however, not by something one of its many contributors had actually written about digital politics, but by the book's cost: $240 on Amazon in the US. (In the UK the hardback is £150.00 on Amazon. *Handbook of Digital Politics* is also available online direct from the publishers for £135.00, with the ebook available for £40.) As one of those on the list commented, 'I'd love to buy it, but not at that price' – to which another participant in the discussion responded: 'I encourage everyone to use the preprint option to post their piece on ssrn.com and academia.edu, perhaps others have other open access suggestions (e.g. Institutional Repositories of individual universities)'. Now, to be fair, the idea that is implied by this suggestion – that the Academia.edu platform for sharing research represents just another form of open access – is a common one. Yet posting on Academia.edu is far from being ethically and politically equivalent to using an institutional open access repository.

This week is International Open Access Week 2015, an annual event designed to promote the importance of making academic research available online to scholars and the general public free of charge. But when it comes achieving this goal, is the open access movement in danger of being somewhat outflanked by Academia.edu? Has the latter not better understood the importance of both scale and centralisation to a media environment that is rapidly changing from being content-driven to being more and more data-driven?

Image credit: Yuri Samoilov CC BY

Launched in 2008, Academia.edu is a San Francisco-based technology company whose platform displays many of the same features as professional social networking sites such as LinkedIn. Users have an individual 'real-name' profile page, complete with their picture, CV, details of their professional affiliations, biography and employment history. The main difference in Academia.edu's case is that these features are accompanied by the user's academic research interests and a list of publications – generally the associated metadata but also quite regularly now the actual full texts themselves (often in the form of the author's pre- or post-print manuscript, if not the final published pdf) – that others in the network can bookmark or download from the platform.

Academia.edu also enables users to send messages to one another on the site, post drafts of papers they would like feedback on, and receive updates when new texts are uploaded – either by those on the platform they are following or in specific areas of research in which they have expressed an interest. In addition, a set of metrics is provided detailing the number of followers a user has, together with an Analytics Dashboard that allows academics to monitor the total number and profile of the views their work has received: page view counts, download counts, and so on. The platform even breaks these 'deep-analytics' down by country.

Yet for all Academia.edu describes itself as a 'social networking service' for academics that 'enables its users, including graduate students … to connect with other users… around the world with the same research interests', it operates increasingly as 'a platform for academics to share research'. 26,281,552 academics have signed up to Academia.edu as of October 18, 2015, the site claims, having collectively added 6,972,536 papers and 1,730,462 research interests. In fact, academics are using it to share their research – both journal articles and books – to such an extent that shortly after it purchased the rival social network for researchers Mendeley in 2013, Elsevier sent 2,800 Digital Millennium Copyright Act (DMCA) takedown notices to Academia.edu regarding papers published on the site that the academic publishing giant claimed infringed its copyright.

The popularity with academics of the Academia.edu social network – its founder and CEO of Academia.edu Richard Price goes so far as to maintain it is the 'largest social-publishing network for scientists', and 'larger than all its competitors put together' – clearly raises a number of questions for the open access movement. After all, compared to the general sluggishness (and at times overt resistance) with which the call to make research available on an open access basis has been met, Academia.edu's success in getting scholars to *share* suggests that, for many, the priority may not be so much making their work openly available free of charge so it can be disseminated as widely

and as quickly as possible, as building their careers and reputations in an individualistic, self-promoting, self-quantifying, self-marketing fashion. Nor is this state of affairs particularly surprising, given the precarious situation in which much of the academic profession finds itself today.

But does it mean that any open access venture hoping to meet with similar success would be well advised to adopt many of the same subjectivising features that are used by Academia.edu and other social networks to help users connect and develop their individual profiles as 'personal brands': real-name policies, personal pictures, CVs and biographies, 'credibility metrics', analytics dashboards, quantifying deep analytics and so on. (Some open access projects have already done so, of course, including PLoS, whose journals provide Article-Level Metrics, Rich Citations, and other indicators relating to usage data.) Perhaps even more dauntingly, would such an open access venture also need to be capable of spending a similar amount of money designing and maintaining an easy-to-use social networking interface as Academia.edu, the latter having raised $17.7 million dollars from investors at the time of this writing?

The key aspect of Academia.edu to be aware of in this respect *is* its business model. Unlike that of *some* for-profit publishers, this is not based on academic authors, their institutions, or their funders *paying* a fee for their research to be made available on a free and open basis: what's known as author-pays or an article processing charge (APC). Its financial rationale rests instead on the ability of the angel-investor and venture-capital-funded *professional* entrepreneurs who run Academia.edu to exploit the data flows generated by the academics who use the platform as an intermediary for sharing and discovering research. In the words of CEO Richard Price:

> *The goal is to provide trending research data to R&D institutions that can improve the quality of their decisions by 10-20%. The kind of algorithm that R&D companies are looking for is a 'trending papers' algorithm, analogous to Twitter's trending topics algorithm. A trending papers algorithm would tell an R&D company which are the most impactful papers in a given research area in the last 24 hours, 7 days, 30 days, or any time period. Historically it's been very difficult to get this kind of data. Scientists have printed papers out, and read them in their labs in un-trackable ways. As scientific activity is moving online, it's becoming easier to track which papers are getting more attention from the top scientists.*
>
> *There is also an opportunity to make a large economic impact. Around $1 trillion a year is spent on R&D globally: about $200 billion in the academic sector, and about $800 billion in the private sector (pharmaceutical companies, and other R&D companies).*

Of course, the majority of academics who are part of Academia.edu's social network are the product of the state-regulated, public higher education system, as is their research (a system, it should be said, from which public funding is steadily being withdrawn). But just as Airbnb and Uber are parasitic on the public 'infrastructure and the investment' that was 'made by cities a generation ago' (roads, buildings, street lighting, etc.), so Academia.edu has a parasitical relationship to the public education system, in that these academics are labouring for it for free to help build its privately-owned for-profit platform by providing the aggregated input, data and attention value. We can thus see that posting on Academia.edu is not ethically and politically equivalent to making research available using an institutional open access repository at all.

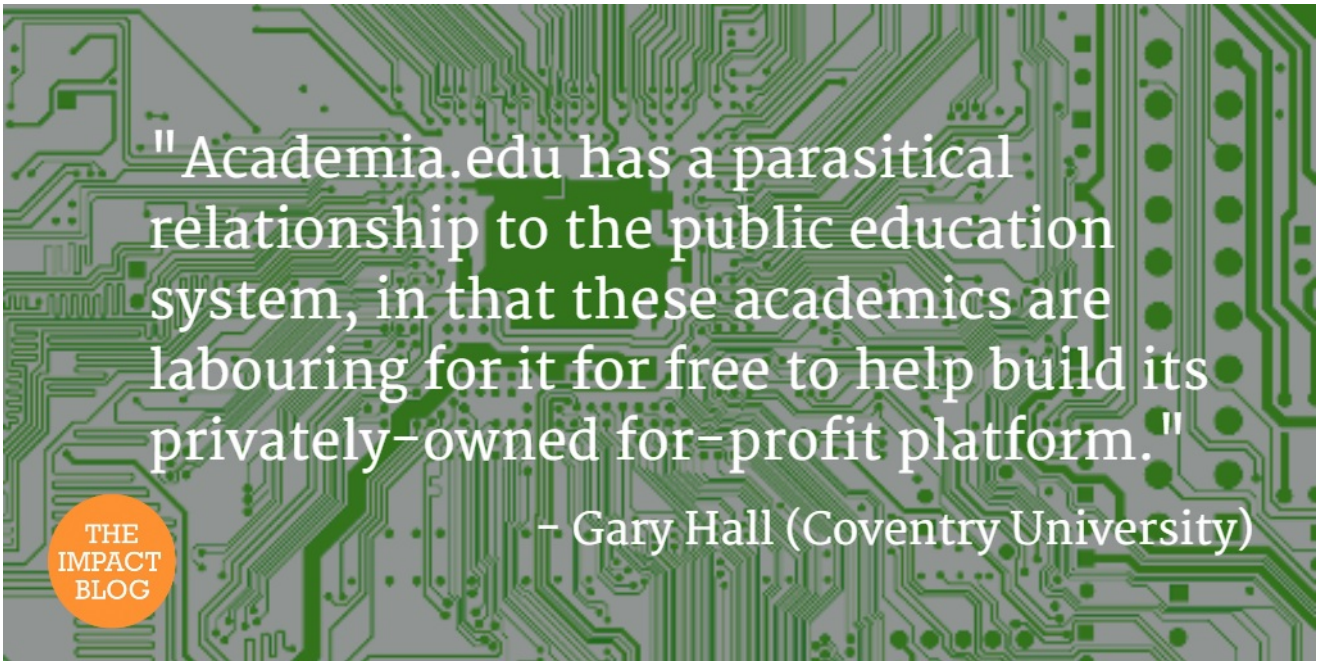

Image credit: Motherboard OpenClipartVectors (Public Domain, Pixabay)

Indeed, the reason it's so crucial to understand Academia.edu's business model is because it highlights just how much the situation regarding the publication and dissemination of academic research has changed since the open access movement first began to take shape in the 1990s and early 2000s. Without doubt the argument of this movement, that publicly-funded research should be made openly available online free of charge, is extremely pertinent to the content-driven world of profit-maximising academic publishers such as Reed Elsevier, Springer, Wiley-Blackwell, and Taylor & Francis/Informa, with their high journal subscription charges and book cover prices, 'Big Deal' library contract bundling strategies, and protection of copyright and licensing restrictions. But this argument isn't anywhere near as relevant to the data-driven world of search engines, social media and social networking. That's because for the likes of Google, Twitter and Academia.edu free content is what for-profit technology empires are built on. In this world *who gate-keeps access to* (and so can extract maximum value from) *content* is less important, because that access is already free, than *who gate-keeps* (and so can extract maximum value from) *the data generated around the use of that content*, which is used more *because* access to it is free.

Accordingly, the relevant arguments here are more those over the ownership and control of the platforms, together with the 'black-boxed' computer programmes, software, algorithms and the associated IP that are making access to the free content possible. How are these data and information management intermediaries structured? What data do they capture? How are they able to manipulate it? Who does what with this data and the resulting metrics and analytics? (Is it sold it to advertisers and other commercial companies? Shared with the NSA and GCHQ for surveillance purposes?) And as environments that encourage users to be self-disciplining, self-managing and self-monitoring, what forms of subjectivisation and subjectivity do they produce?

This is why I raised the question of whether the open access movement is in danger of being outflanked, if not rendered irrelevant, as a result of our media environment changing from being content-driven to being increasingly data-driven. For the data-driven world is one in which the *data centre* dominates. This in turn brings us to the issue of scale, as there is an obvious reason for this domination of the *data centre*. Quite simply, the larger your data sample, the more relevant data you can capture, store, process, mine and manipulate, the more accurate your data analytics. (It's not because Google has better algorithms that it has a 90-95% share of the European market for search, according to Peter Norvig, its Director of Research: it's because it has more data. This is also why such companies strive to become monopolies: because it's harder for them to scale to the massive extent that's needed to produce the best data analyses if they have rivals who are capturing a significant portion of the relevant data.)

Now the kind of decentralised infrastructure that is represented by the open access movement's wide variety of different journals, megajournals, repositories, book publishers, open source software tools, websites, portals and directories may be entirely appropriate to achieving its goal of making large amounts of different kinds of research content available for free, online, by providing green, gold and even platinum open access alternatives to a closed access publishing industry that is itself relatively decentered. The increasing importance of being able to create massive data sets, however, means that such decentralised infrastructure is in the process of gradually being replaced by what Rachel O'Dwyer, in a recent article on blockchains, describes as a 'recentralisation of infrastructure'. Lots of content may be freely accessible, but this access is now being mediated by *centralised* entities. The result is that those rich and powerful international companies who *are* able to capture, analyse and exploit extremely large amounts of data are coming to act as the gatekeepers of our media and communications networks; and this includes our scholarly communications networks, as the 36 million visitors who are apparently attracted to the Academia.edu research sharing platform each month bear witness.

For more on this topic, also see Kathleen Fitzpatrick's piece Academia, Not Edu and Martin Eve's Academia.edu's peer-review experiments.

This piece originally appeared on Gary Hall's open notebook Media Gifts. Content on Media Gifts can be distributed, reproduced, transmitted, translated, modified, remixed, built upon, used and 'pirated' in any medium, even without indication of 'origin'. (This does not affect any rights others may have in it or in how it is used.)

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our Comments Policy if you have any concerns on posting a comment below.

About the Author

Gary Hall is Research Professor of Media and Performing Arts at Coventry University.