

11-21-2013

The Guerilla Open Access Manifesto: Aaron Swartz, open access and the sharing imperative

Benjamin L. Hockenberry

St. John Fisher College, bhockenberry@sjfc.edu

Follow this and additional works at: http://fisherpub.sjfc.edu/library_pub

 Part of the [Library and Information Science Commons](#)

Publication Information

Hockenberry, Benjamin L., "The Guerilla Open Access Manifesto: Aaron Swartz, open access and the sharing imperative" (2013).
Lavery Library Faculty/Staff Publications. Paper 6.
http://fisherpub.sjfc.edu/library_pub/6

This Conference Proceeding is brought to you for free and open access by the Lavery Library at Fisher Digital Publications. It has been accepted for inclusion in Lavery Library Faculty/Staff Publications by an authorized administrator of Fisher Digital Publications.

The Guerilla Open Access Manifesto: Aaron Swartz, open access and the sharing imperative

Abstract

In my talk, I'll discuss the current state of legal Open Access as it faces challenge from publishers and regulatory establishments above and from extra-legal filesharing bubbling under the surface. Specifically, I'll discuss how the case of Aaron Swartz exposes the ethical culture of the "guerilla open-access" undercurrent that seeks to disruptively advance the OA mission. "The world's entire scientific and cultural heritage, published over centuries in books and journals, is increasingly being digitized and locked up by a handful of private corporations. Want to read the papers featuring the most famous results of the sciences? You'll need to send enormous amounts to publishers like Reed Elsevier" (Swartz 2008).

Disciplines

Library and Information Science

Comments

Paper and slides delivered at World Philosophy Day, St. John Fisher College, Nov. 21, 2013.

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 3.0 License](https://creativecommons.org/licenses/by-nc/3.0/)

The *Guerilla Open Access Manifesto*: Aaron Swartz, open access and the sharing imperative

Delivered at World Philosophy Day, St. John Fisher College, Nov. 21, 2013.

Abstract:

In my talk, I'll discuss the current state of legal Open Access as it faces challenge from publishers and regulatory establishments above and from extra-legal filesharing bubbling under the surface. Specifically, I'll discuss how the case of Aaron Swartz exposes the ethical culture of the "guerilla open-access" undercurrent that seeks to disruptively advance the OA mission.

"The world's entire scientific and cultural heritage, published over centuries in books and journals, is increasingly being digitized and locked up by a handful of private corporations. Want to read the papers featuring the most famous results of the sciences? You'll need to send enormous amounts to publishers like Reed Elsevier" (Swartz 2008).

Introduction

First, a disclaimer. I'm here today as a librarian who works on computer systems and databases for Lavery Library. I have also been involved in developing Fisher Digital Publications, SJFC's largely-Open-Access institutional repository. My biases on today's topic reflect that position, but any opinions are my own and do not reflect the College or Library.

The suicide of Aaron Swartz in January of this year, when he was only 26, brought the light of tragedy onto issues that had been coming to a boil in digital publishing for years. In this talk, I will not focus on whether Swartz was in effect "murdered" by prosecutorial overreach, as claimed by many including his father, though this is a powerful idea. I will, however, discuss the nature and possible purpose of his act, and its possible implications in the field of digital publishing. The manner of this case's end really hit me, but I've tried to steer clear of the personal aspects of this case, though like many of you likely have, I have experienced the tragedy of a close friend's suicide.

Summarizing the case

For those of you who know the story, I apologize for the rehash, and for all the details I will gloss over. Swartz had been engaged in a legal battle regarding a 2010 case in which he bulk-downloaded the archives of the database JSTOR from a guest account on MIT's network, violating terms of service. This downloading activity took place over months, largely on open wi-fi, but concluding with a span in which he entered a networking closet and plugged his laptop directly into a network jack. He faced federal charges of up to 35 years in prison, though he delivered the hard drives to JSTOR and they did not pursue a charge: JSTOR instead responded by releasing significant portions of their archives, approximately 4.5 million articles, openly. Tragically, the case ended with his suicide in January.

Legal scholar Lawrence Liang claimed Swartz as the "First Martyr of the Free Information Movement" (NewsClickIn, 2013). The nomination letter for one of the posthumous awards given to him, the American Association of Law Librarians' PAGI Award, states, "Aaron's life and work were dedicated to ensuring that technology would be leveraged to ensure the broadest possible access to information.... To judge from the accomplishments that marked his tragically short life, Aaron's guiding principle was that the public should be able to access information without needless costs or other impediments." (Feltren, 2013). Legislation has been proposed called "Aaron's law," seeking to restrict the scope of the Computer Fraud and Abuse Act which enabled the prosecution.

Who was Swartz?

Several of the technologies and standards that have shaped the way we interact with the web, including RSS and the Creative Commons, might not be what they are today without his involvement. He was a cofounder of Infogami, which later became Reddit. His activist side centered in later years around fighting the Stop Online Piracy Act (SOPA) through his activist organization, DemandProgress.

Aaron took several approaches to change-making: the speaking, activism, and software development for which he was praised; and the outside-the-system actions that led to his arrest. His activity in bulk-downloading from JSTOR was a high-profile case, but he had been involved in other cases of bulk download. He conducted a bulk download in 2009 via a free trial to the US Government's Public Access to Court Electronic Records (or PACER) system (Schwartz, 2009). In 2006, he downloaded the entire catalog of the Library of Congress to serve as a backbone for a project called OpenLibrary. These cases differed from the JSTOR affair in that they contained material produced by the US Government, not covered by copyright.

Copyright

So here is the topic at hand: copyright. We live in an age of unprecedented ease of copying and transferring information, with high-speed connections working over wifi and 4G. At the same time, today's copyright is defined by the the Sonny Bono Copyright Term Extension Act: works of single authorship are protected for 70 years after the death of the author, and works under corporate authorship are protected for between 95 and 120 years. The divide between the extreme length of copyright and the ease of its violation is a recipe for rampant lawlessness.

Just as in the Napster situation of the turn of the millennium, relatively simple technologies have the capacity to completely upend an industry. I used the phrase "bulk download" to describe Swartz's activities -- not hacking. Per Alex Stamos, an expert witness for the defense, Swartz 'did not "hack" the JSTOR website for all reasonable definitions of "hack": he wrote a basic script that found the internet addresses of journal articles (which are in numerical order), then used a command that's the equivalent of "right-click-save-as" to request them (Stamos, 2013).

Libraries Part 1: JSTOR and MIT

The conditions that made the bulk download possible were largely in JSTOR and MIT's hands. In most academic libraries, users can access databases from within an institution's IP range. According to terms of service, the downloaded copies are able to be shared in limited ways, including Fair Uses under copyright and interlibrary loan.

In July of 2013, JSTOR released their version of events, in which network traffic was identified and steps were taken to restrict access for MIT users. They later had to take a seldom-precedented step (for JSTOR): requiring individual user identification on MIT's network. Doing this prohibited guest users, including alumni, from using library resources. JSTOR saw this exception as very serious, and did not make the decision lightly. Part of the service philosophy in many academic libraries, including Lavery Library (and most libraries in our area), has been to ensure that just as a guest or alumna could visit a library and physically browse journal pages, she could read a journal at a library PC. The license terms libraries sign with publishers permit this activity.

Needless to say, libraries must not be complacent to the implications of this JSTOR case. In the era of electronic resources, databases are not free in any sense of the word: cost of access is an issue, but so are terms of use. Violation of terms of use (such as a method of download...or use for unintended or commercial purposes) can be construed as wire fraud. At the same time, 5 years ago, McDonald and Cranor (2008) estimated that more than 200 hours would be required for the average user of online services to read the privacy policies to which they agree.

After Swartz's death, librarian Bohyun Kim wrote on the fact that electronic resources librarians have watered down our professional ethics. She writes:

“Electronic resource vendors’ and publishers’ pricing has become more and more closely tied to the number of patrons who can access their products. Academic libraries [have] been dealing with the escalating costs for electronic resources by filtering out library patrons and limiting the access to those in a specific disciplines” (2013).

Do only pharmacy students need access to this resource? Limit access only to users with their major codes: suddenly, the number of “FTE” used to calculate the price of the resource drops significantly. Boon! But what is the trade-off? Kim writes, “if those resources were in the print format, no librarian would think that it is OK to permanently limit its use to those in medical or health science disciplines only.” We grapple with this here. Lavery Library makes conscious decisions to select and pay for resources on the basis of the most college-spanning access policies, but as the college's FTE grows and publishers charge more, how can we sustain this?

Swartz's Motivations

Speaking of money, we come back around to the possible motivations for Swartz's actions. Per Harvard's Sam Klein (2011), Swartz was intensely concerned with money and institutional influence. In 2008, he founded a site called watchdog.net to analyze and visualize where political contributions come from. Swartz also worked with Dr. Shireen Barday at Stanford Law

to identify when law review articles were funded by corporations. He did this by (get this) downloading and analyzing a corpus of articles from academic databases, then conducting textual analysis to determine funding sources (Klein, 2011). It is not much of a stretch to imagine the corpus of JSTOR articles being used in a similar manner.

An alternate hypothesis for his actions is presented in Swartz's 2008 "Guerilla Open Access Manifesto." I will read a few paragraphs:

"Those with access to these resources — students, librarians, scientists — you have been given a privilege. You get to feed at this banquet of knowledge while the rest of the world is locked out. But you need not — indeed, morally, you cannot — keep this privilege for yourselves. You have a duty to share it with the world. And you have: trading passwords with colleagues, filling download requests for friends. Meanwhile, those who have been locked out are not standing idly by. You have been sneaking through holes and climbing over fences, liberating the information locked up by the publishers and sharing them with your friends.

But all of this action goes on in the dark, hidden underground. It's called stealing or piracy, as if sharing a wealth of knowledge were the moral equivalent of plundering a ship and murdering its crew. But sharing isn't immoral — it's a moral imperative. Only those blinded by greed would refuse to let a friend make a copy."

He closes with a call for action that became damning evidence:

"We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that's out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access."

He did not post the downloaded content from JSTOR to filesharing networks, though in the wake of his arrest, others did so with their personally-downloaded archives of journal articles out of protest. The culture of sharing is so strong on the web that one can post a request with hashtag #ICanHazPDF to Twitter and receive the requested article from a complete stranger. Piratebay, a bittorrent site, is brimming with content downloaded, scanned or ripped from legal sources. Estimates of online piracy, particularly in the developing world, are astronomical (Social Science Research Council, 2011).

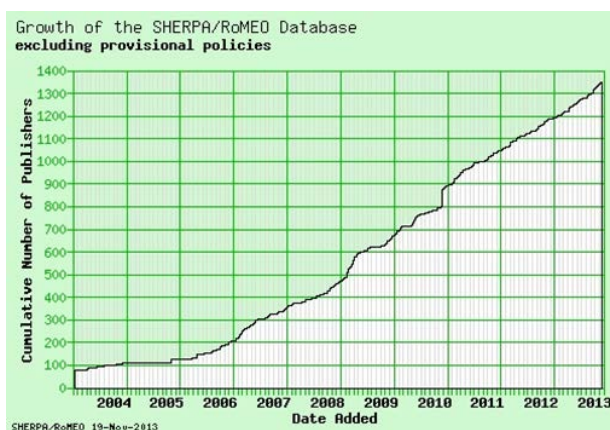
We will never know whether Swartz truly intended to post the JSTOR content to filesharing sites himself, to create an indexed archive of the content, or to apply analytical methods to the data. In either case, this high-profile case brought a spotlight onto the "paywalls" behind which the fruits of scholarly research are so often hidden.

Libraries Part 2: Open Access

Libraries, through our licenses with the content industry, are the largest customers of these databases and are, by agreeing to terms and conditions, at some level accountable the growth of these paywalls. But why do libraries exist? To share: we share content and share costs.

The Open Access movement, supported by scholars in law and other academic disciplines, computer people and (yes) librarians, has been seeking to change the paywall model for over a decade, focusing on legal ways to open up doors to scholarship. Authors and institutions supporting their research have wrestled with publishers to get legitimate rights to control dissemination of their work, and a growing number of journals. Librarians have become involved out of a combination of supporting faculty scholarship at our institutions, concerns about inflating prices for electronic resources, and a professional inclination toward sharing the world's knowledge.

Many institutions (like the NIH, MIT and Harvard) have implemented Open Access mandates, requiring research the institution helps fund to be made openly available through open access journals or in an institutional repository like our Fisher Digital Publications.



(Source: <http://www.sherpa.ac.uk/romeo/statistics.php>)

Since 2004, the SHERPA/RoMEO database of publisher's policies for self-archiving has grown from a small batch of fewer than 100 publishers to a trusted collection of over 1400 publishers across the globe. Over 70% of the publishers documented therein have a formal policy allowing for archival and dissemination of an author's work from an institutional or personal site (<http://www.sherpa.ac.uk/romeo/statistics.php>). This growth of publisher acceptance of alternate models of content distribution would not be possible without the OA movement.

There are multitudes of concern with Open Access:

- Maintenance of access/archival systems
- Confusion over "archivable" versions and version of record
- Rank & Tenure policies

- Future publishing opportunities (“the book deal”)
- Potential unauthorized or uncredited reuse of the material for commercial or other purposes
- Author-pays model may lead to sham journals/publishers, as in Bohannon’s (2013) sting in *Science*, Beall’s Predatory List (<http://scholarlyoa.com/publishers/>), etc.
- ...and many more

But these are not the concerns of an idealist manifesto. Open Access advocates have been working within a complex system that has, for a long time, been rigged toward publishers, not users or (often) even content creators. This puts stumbling blocks at every turn. By calling for mass piracy, this manifesto seeks to disrupt that model.

Here is where, as a librarian, I grow concerned. Libraries exist to share content, but we want to do so in legal bounds, with a vested interest in keeping our physical and virtual doors open to serve our users. We want to be able to guide these users to the best content, and help them do things with it. We have an interest in protecting content for future generations, which we’ve been able to do for centuries in the physical world and have made significant steps in the digital world. But protection, in our vernacular, sometimes can turn into “gatekeeping:” keeping unwanted users out. This is anathema to library philosophy: Ranganathan’s seminal statement, “every person needs his or her book,” means not to turn a user away. This links into Stewart Brand’s oft-quoted statement and credo of hacker culture: “Information wants to be free... information also wants to be expensive.” If all content were actually free, would we need libraries? If the libraries (digital or virtual) disappeared, how long would that freedom last?

I’ll close with some more questions, for which I don’t have answers.

- Could the legitimate Open Access movement have gotten JSTOR to provide access to more than 4.5 million articles, without this case?
- JSTOR is a non-profit. What fallout from other, more commercially-oriented, publishers and vendors will follow?
- Do OA advocates have the same goals as copyright infringers (pirates)?
- Can libraries truly strike a balance between gatekeeping and sharing?

References

- Barday, S.A. (2008). Punitive damages, remunerated research, and the legal profession. *Stanford Law Review*, 61(3), 711.
- Bohannon, J. (2013). Who's afraid of peer review? *Science*, 342(6154), 60–65.
doi:10.1126/science.342.6154.60
- Brand, S. (Ed.) (1985, May). "Keep designing:" How the information economy is being created and shaped by the hacker ethic. *Whole Earth Review* 46, p. 44-55. Retrieved from <http://www.wholeearth.com/issue-electronic-edition.php?iss=2046>
- Feltren, E. (2013). Congratulations to the 2013 recipients of the PAGI and Oakley Advocacy Awards. *AALL Spectrum*, 17, 8.
- Kim, B. (2013, February 11). Aaron Swartz and too-comfortable research libraries. *ACRL TechConnect Blog*. Retrieved from <http://acrl.ala.org/techconnect/?p=2823>
- Klein, S. (2011, July 24). Aaron Swartz vs. United States. *The Longest Now*. Blog. Retrieved from <https://blogs.law.harvard.edu/sj/2011/07/24/aaron-swartz-v-united-states/>
- JSTOR. (July 30, 2013). *JSTOR Evidence in United States vs. Aaron Swartz*. Retrieved from <http://docs.jstor.org/>
- McDonald, A. and Cranor, L. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*. Author's version available at <http://www.aleecia.com/authors-drafts/readingPolicyCost-AV.pdf>
- NewsClickIn. (2013, January 19). Aaron Swartz: *The First Martyr of the Free Information Movement*. Retrieved from <http://www.youtube.com/watch?v=Bq87SR0TRw4>
- Ranganathan, S.R. (1931). *The Five Laws of Library Science*. Madras: Madras Library Association. Retrieved from <http://catalog.hathitrust.org/Record/001661182>
- Schwartz, J. (2009, February 13). An effort to upgrade a court archive system to free and easy. *The New York Times*, p. A16. New York.
- Social Science Research Council (2011, March). *Media Piracy in Emerging Economies*. Retrieved from <http://piracy.americanassembly.org/the-report/>.
- Stamos, A. (2013, January 12). The truth about Aaron Swartz's "crime." *Unhandled Exception*. Blog. Retrieved from <http://unhandled.com/2013/01/12/the-truth-about-aaron-swartzs-crime/>
- Swartz, A. (2008). *Guerilla Open Access Manifesto*. Retrieved from <http://archive.org/details/GuerillaOpenAccessManifesto>

See also:

Aaron Swartz collection on the Internet Archive: <http://archive.org/details/aaronsw>