

Nancy Sims

# Library licensing and criminal law

## The Aaron Swartz case

According to the allegations of a federal indictment, in the fall of 2010 a guest user logged on to the MIT campus network and began systematically downloading articles from the JSTOR archive. This activity continued over the next few months. Both MIT and JSTOR noticed the unusual traffic (which allegedly caused some server overloading) and took steps to disable the automated access, but the user was able to restore the connection. Both Massachusetts state police and the FBI became involved in investigating the situation, and in January 2011, activist and programmer Aaron Swartz was detained by police on the MIT campus.

At that time, Swartz was a research fellow at Harvard University's Safra Center for Ethics, and, although only 24 years old, he has a long history of activism and advocacy on issues related to openness and sharing. As early as age 14, he was a member of an international working group that developed early specifications for the Web-content syndication protocol, RSS. He was an early employee of the social-news site Reddit. He worked on projects with Wikipedia and the Internet Archive, and founded the progressive activist group DemandProgress.

Swartz has been involved in similar incidents of large-scale downloading in the past. In 2008, during a trial run of a program providing free public access terminals for the government's Public Access to Court Electronic Records (PACER) system, he used an automated script to download millions of pages (approximately 20 percent of all the records in that system), and shared them with the public. PACER records are in the public domain, but Swartz's activities

resulted in the trial public-access program being put on hold, and attracted the attention of federal investigators. No charges ever resulted from that incident.

In the present case, however, serious charges have resulted—charges carrying a potential prison term of up to 35 years and a fine of up to \$1 million. While the public discussion around this case has focused on the articles Swartz allegedly copied, the charges have little to do with the downloading itself. They primarily relate to the actions Swartz took in order to have access to JSTOR through the MIT network—actions including online evasions of network management, and physical-world unauthorized access to areas of MIT buildings. The charges brought against Swartz for these actions are primarily wire fraud (18 U.S.C. § 1343) and computer fraud (18 U.S.C. § 1030). In several instances, the criminal charges are based specifically on the fact that Swartz violated MIT and JSTOR user policies—and in those instances, these charges raise some significant issues that the academic library community should be concerned about.

This is a relatively new model of criminal liability, growing out of a piece of legislation from 1986 (the Computer Fraud and Abuse Act) that prohibits unauthorized access to computer networks. Starting in

---

Nancy Sims is copyright program librarian at the University of Minnesota Libraries, e-mail: [nasims@umn.edu](mailto:nasims@umn.edu)

Contact Mike Furlough—series editor, assistant dean for scholarly communications, and codirector of the Office of Digital Scholarly Publishing at Penn State University—with article ideas, e-mail: [mfurlough@psu.edu](mailto:mfurlough@psu.edu)

© 2011 Nancy Sims

the early 2000s, prosecutors in some cases began arguing that any access to a computer network in violation of its terms of use is “unauthorized” access, and can amount to federal computer fraud. This strategy has been hotly contested by legal scholars and many in the technology and Internet sectors, because under this theory, criminal liability can arise regardless of the actual terms in the terms of use, and without regard for whether they match up with any other commonly understood elements of criminal behavior.

Under this theory, creating a Facebook or Google+ profile under a pseudonym (which is explicitly against the terms of service on each site) could serve as the basis for federal criminal charges. That may sound farfetched, except that one of the first cases using this theory of liability was about exactly that: the creation of a fake MySpace profile that was used to harass a young teen who eventually committed suicide. The harassment, though deplorable, was not sufficient to meet the definitions of any existing crime, so prosecutors argued for federal computer fraud liability for violation of the MySpace terms of service.

The judge in that case refused to apply the new theory: he deplored the harassment, but said that interpreting any violation of terms of service as grounds for criminal liability would be “overwhelmingly overbroad” and constitutionally problematic. Nevertheless, the theory has been increasingly used by prosecutors when few other laws would directly enable prosecution, and a number of courts have accepted it.

The constitutional and other troubling issues raised by this novel and evolving theory of criminal liability are seriously problematic, but probably best handled by lawyers, law enforcement, legislators, and policymakers. However, there are specific issues about this theory’s application in Swartz’s case that should be of deep concern to the academic community: the terms of use that MIT created for its own community, and the terms of use that JSTOR created and MIT accepted on behalf of its users are being used as the

basis of criminal prosecutions. Moreover, the prosecution is proceeding when JSTOR and MIT seem to have chosen not to pursue civil litigation against Swartz. According to a public statement released by JSTOR shortly after news of the indictment became public, JSTOR had resolved their concerns with Swartz before the federal prosecution came into play. And although MIT has been less forthcoming about the case, it too seems to have chosen not to pursue civil litigation against Swartz. Despite this, the United States Attorney for the state of Massachusetts chose to prosecute Swartz.

The U.S. Attorney’s office is well within its rights in choosing to prosecute despite a lack of action from potential civil litigants, which is known as “prosecutorial discretion.” In theory, we distinguish crimes from civil offenses because the former cause harms to society as a whole. Therefore, it is up to prosecutors in any criminal case to decide whether to bring charges against someone, regardless of the choices of the direct victims of the crime. Leaving aside the question of whether there was a societal harm here, and whether JSTOR and MIT were the “victims,” we should be very aware that the terms we impose on our own networks’ use, and the terms we agree to in contracts with vendors, may be able to serve as the basis of criminal charges against our users. This may happen even when we, and our vendors, might prefer to resolve any conflicts privately. It is, of course, unlikely that most library user activities will ever attract the attention of federal law enforcement, but when and if they do, the enforcement of our policies, and the licensing terms to which we have agreed, may be out of our hands.

Many users of subscription resources do make uses of the systems that are questionable in relation to the official terms of use. Often such questionable uses are ignored or tolerated by campus and vendor system administrators. One such use is computational analysis of the content in such databases, most subscription resources prohibit mass-downloading for such purposes, but

some allow it when permission is sought separately.

In the Swartz case, the prosecution alleges that Swartz intended to distribute the downloaded documents publicly. And although the prosecution offers no evidence in support of this, some of Swartz's own public statements suggest he could have had such an intent. But Swartz has also engaged recently in scholarly research projects doing computational analysis on article texts, and certainly could have intended only such a use for the materials he downloaded. JSTOR's own Data for Research program is still in beta, and will likely evolve in response to the needs of researchers who take advantage of the program. Is it possible that if MIT's license with JSTOR had provided for research access to the corpus in a way that satisfied his specific research needs, Swartz would never even have begun his systematic downloading?

Much of the discussion around this case highlights public misperceptions of libraries and of licensed content. Both Swartz's supporters and the prosecution have focused on Swartz's copying of articles, despite the fact that the charges focus on fraud, not theft or copyright violations. His supporters ridiculed the prosecution by suggesting his alleged downloading was like checking "too many" books out of the library. The prosecution repeatedly made reference to Swartz's actions as "stealing," despite the fact that the charges they brought against Swartz focus on fraud. Most academic library employees could readily explain to Swartz's supporters that subscription license terms are often more restrictive than the "First Sale" doctrine that enables library lending, and would probably also point out that most libraries do impose some limits on book borrowing. Similarly, library employees could explain to the prosecution that although Swartz may have acquired copies of millions of JSTOR documents, there was never any erasure or removal of content from JSTOR's servers, so invoking the rhetoric of "theft" is a bit problematic.

The fact that both the prosecution and Swartz's supporters can talk about the case

with such superficial, and deeply incorrect, messages about libraries and licensed content, highlights the significant disparity between what our users understand our services to be, and what we agree to when we sign contracts for licensed resources. Most of our users honestly don't understand that our contracts, rather than providing free and unlimited access to licensed resources for all purposes, often provide access only up to usage limits implemented by vendors, and often provide access only for personal use. Perhaps this means that we need to negotiate license terms that are more in line with user expectations for our services; perhaps it means that we need to better educate our users about what access our contracts really provide. The former approach is an extremely tall order; the latter is unlikely to be very palatable to our users. It is, however, very clear that licensing terms, which govern an increasingly large proportion of our collections, are a fundamental issue in the present and future usability of library resources by our campus populations.

Another strong thread in the public discussion of this case has been a denunciation of restricted access to scholarship (and tangentially, to public domain documents). Swartz's 2008 "Guerilla Open Access Manifesto" speaks out against "[t]he world's entire scientific and cultural heritage, published over centuries in books and journals, [being] increasingly being digitized and locked up by a handful of private corporations." Researchers and scholars commenting on the case speculated on whether many of the authors of the articles Swartz downloaded would have objected to his activities, and others deplored the "gatekeeper" function that even nonprofit aggregators like JSTOR serve.

One individual, Greg Maxwell, uploaded a collection of more than 18,000 public domain documents legally acquired from JSTOR to the file-sharing site The Pirate Bay, as a direct response to the indictment against Swartz. In a public statement attached to the document set, Maxwell said, "I've been afraid that if I published them I would be subject to unjust legal harassment by those who profit from

controlling access to these works. I now feel that I've been making the wrong decision.”

There is a cost in maintaining access to these resources, and some commentators have expressed support for JSTOR's non-profit mission, but even among these supporters, many suggested that public domain documents should be more freely available than in-copyright works. A number of commentators criticized Swartz's alleged actions while expressing support for the underlying idea that scholarship should be more open and free.

Open society activist Carl Malamud, at whose suggestion Swartz conceived his earlier project scraping content from the public-domain PACER database, spoke in the *New York Times* of his own efforts to “force” open gates around content, but characterized Swartz's alleged efforts as searching for a “back door” and questioned the effectiveness of such pursuits.<sup>1</sup> Legal scholar and activist Larry Lessig, who has worked with Swartz for several years, made a statement in the comments at the “Media Freedom” blog that questioned whether the alleged activities truly constituted a crime, but affirmed that, if the allegations proved true, Swartz's actions crossed an ethical line.<sup>2</sup>

Swartz has pled not guilty to all charges, and was due back in court on September 9, 2011. We may not know the status of the criminal prosecution for some time to come. Nonetheless his activities, and the public reactions they have generated, highlight some of the most troubled, and troubling, legal and ethical issues in academic licensing, open access, and scholarly communication. Whatever the outcome of the criminal prosecution, we must not ignore the importance of these issues for the academic and research library community.

## Notes

1. [www.nytimes.com/2011/07/20/us/20compute.html](http://www.nytimes.com/2011/07/20/us/20compute.html).
2. <http://mediafreedom.org/2011/07/larry-lessig-responds-says-swartzs-alleged-actions-crossed-ethical-line/>. ↗

**SerialsSolutions®**

A ProQuest® Company

“Easy to find stuff.”

“Lets me focus on the information.”

“Very Simple.”

**Discover Summon™**  
for your library.



**Summon™**

[www.serialssolutions.com/summon](http://www.serialssolutions.com/summon)