

[Threat Level](#)

[Privacy, Crime and Security Online](#)

[Previous post](#)

[Next post](#)

Feds Charge Activist as Hacker for Downloading Millions of Academic Articles

By [Ryan Singel](#)  July 19, 2011 | 2:55 pm | Categories: [Sunshine and Secrecy](#), [The Courts](#)



Updated: The results of Swartz's arraignment were added to the story July 19, 2011; 5 p.m. EDT

Well-known coder and activist Aaron Swartz was arrested Tuesday, charged with violating federal hacking laws for downloading millions of academic articles from a subscription database service that MIT had given him access to via a guest account. If convicted, Swartz faces up to 35 years in prison and a \$1 million fine.

Swartz, the 24-year-old executive director of Demand Progress, has a history of downloading massive data sets, both to use in research and to release public domain documents from behind paywalls. Swartz, who was aware of the investigation, turned himself in Tuesday.

Disclosure: Swartz is a co-founder of Reddit¹, which like Wired.com is owned by Condé Nast. He is also a general friend of Wired.com, and has done coding work for Wired.

The grand jury indictment accuses Swartz of evading MIT's attempts to kick his laptop off the network while downloading more than four million documents from JSTOR, a not-for-profit company that provides searchable, digitized copies of academic journals. The scraping, which took place from September 2010 to January 2011 via MIT's network, was invasive enough to bring down JSTOR's servers on several occasions.

According to the U.S. attorney's office, Swartz was arraigned in U.S. District Court in Boston this morning where he pled not guilty to all counts. He is now free on a \$100,000 unsecured bond. His next court date is Sept. 9, 2011 and he's represented by Andrew Good of Good and Courmier.

The indictment alleges that Swartz, at the time a fellow at Harvard University, intended to distribute the documents on peer-to-peer networks. That did not happen, however, and all the documents have been returned to JSTOR.

JSTOR, the alleged victim in the case, did not refer the case to the feds, according to Heidi McGregor, the company's vice president of Marketing & Communications, who said the company got the documents, a mixture of both copyrighted and public domain works, back from Swartz and was content with that.

As for whether JSTOR supports the prosecution, McGregor simply said that the company was not commenting on the matter. She noted, however, that JSTOR has a program for academics who want to do big research on the corpus, but usually faculty members ask permission or contact the company after being booted off the network for too much downloading.

"This makes no sense," said Demand Progress Executive Director David Segal in a statement provided by Swartz to Wired.com before the arrest. "It's like trying to put someone in jail for allegedly checking too many books out of the library."

"It's even more strange because the alleged victim has settled any claims against Aaron, explained they've suffered no loss or damage, and asked the government not to prosecute," Segal said.

JSTOR doesn't go quite as far in [its statement on the prosecution](#) — though there are clear hints that they were not the ones who wanted a prosecution, and that they were subpoenaed to testify at the grand jury hearing by the federal government.

We stopped this downloading activity, and the individual responsible, Mr. Swartz, was identified. We secured from Mr. Swartz the content that was taken, and received confirmation that the content was not and would not be used, copied, transferred, or distributed.

The criminal investigation and today's indictment of Mr. Swartz has been directed by the United States Attorney's Office.

When asked about this, Christina Sterling, a spokeswoman for the U.S. Attorney's office said, "I can't speak specifically about this case, but fundamentally speaking, the U.S. Attorney's Office makes own independent decisions regarding prosecution based on the merits of a case."

But the feds clearly think they have a substantial hacking case on their hands, even though Swartz used guest accounts to access the network and is not accused of finding a security hole to slip through or using stolen credentials, as hacking is typically defined.

In essence, Swartz is accused of felony hacking for violating MIT and JSTOR's terms of service. That legal theory has had mixed success — a federal court judge dismissed that argument in the Lori Drew cyberbullying case, but it was later reused with more success in a case brought against ticket scalpers who used automated means to buy tickets faster from Ticketmaster's computer system.

“Stealing is stealing whether you use a computer command or a crowbar, and whether you take documents, data or dollars. It is equally harmful to the victim whether you sell what you have stolen or give it away,” said United States Attorney Carmen M. Ortiz in a [press release](#)(.pdf).

The [indictment](#) (.pdf) accuses Swartz of repeatedly spoofing the MAC address — an identifier that is usually static — of his computer after MIT blocked his computer based on that number. Swartz also allegedly snuck an Acer laptop bought just for the downloading into a closet at MIT in order to get a persistent connection to the network.

Swartz allegedly hid his face from surveillance cameras by holding his bike helmet up to his face and looking through the ventilation holes when going in to swap out an external drive used to store the documents. Swartz also allegedly named his guest account “Gary Host,” with the nickname “Ghost.”

Why would Swartz want to download what is likely gigabytes of information? His history includes a study co-authored with Shireen Barday, which looked through thousands of law review articles looking for law professors who had been paid by industry patrons to write papers. [That study](#) was published in 2008 in the Stanford Law Review.

Swartz is [no stranger to the feds being interested in his skills at prodigious downloads](#). In 2008, the federal court system decided to try out allowing free public access to its court record search system PACER at 17 libraries across the country. Swartz went to the 7th U.S. Circuit Court of Appeals library in Chicago and installed a small PERL script he had written. The code cycled sequentially through case numbers, requesting a new document from PACER every three seconds. In this manner, Swartz got nearly 20 million pages of court documents, which his script uploaded to Amazon’s EC2 cloud computing service.

While the documents are in the public record and free to share, PACER normally charges eight cents a page.

The courts reported him to the FBI, which investigated whether the public records were “exfiltrated.” After in-depth background searches, a luckless stakeout and futile attempts to get Swartz to talk, the FBI dropped the case.

The same anti-hacking statute was used to prosecute Lori Drew, who was charged criminally for participating in a MySpace cyberbullying scheme against a 13-year-old Missouri girl who later committed suicide. The case against Drew hinged on the government’s novel argument that violating MySpace’s terms of service was the legal equivalent of computer hacking and a violation of the Computer Fraud and Abuse Act.

A federal judge who presided over the prosecution tossed the guilty verdicts in July 2009, and the government declined to appeal.

1. Swartz’s title and role re: Reddit is an object of much controversy. Swartz joined Reddit in November 2005, months after the site launched. At that point, the initial features were well-established and initial funding had been raised. Swartz’s company Infogami, like Reddit, was part of the first class of Y-Combinator. Reddit’s initial founders Steve Huffman and Alexis Ohanian absorbed Swartz and Infogami into Reddit. Swartz was given an equal chunk of ownership as Alexis and Steve, and Ohanian says he considers Swartz both an early employee and a co-owner, but not a co-founder.

Swartz maintains he was a co-founder, and Y-Combinator’s Paul Graham has referred to him as a founder. Reddit was bought by Conde Nast, Wired’s parent company, in October 2006. Swartz was fired months later, after making it clear he did not like the new environment. That [did not sit well with the rest of the Reddit crew](#). For lack of an accurate term for someone who joins a company early — but after launch — and who gets paid largely in equity, we use the term co-founder in this story.



July 19, 2011

Open-Access Advocate Is Arrested for Huge Download

By **JOHN SCHWARTZ**

A respected Harvard researcher who also is an Internet folk hero has been arrested in Boston on charges related to computer hacking, which are based on allegations that he downloaded articles that he was entitled to get free.

A federal indictment unsealed in Boston on Tuesday morning on charges that the researcher, Aaron Swartz, broke into the computer networks at the [Massachusetts Institute of Technology](#) to gain access to [JSTOR](#), a nonprofit online service for distributing scholarly articles online, and downloaded 4.8 million articles and other documents — nearly the entire library.

Mr. Swartz, 24, made his name as a member of the Internet elite as a teenager when he helped create RSS, a bit of computer code that allows people to receive automatic feeds of online notices and news. Since then, he has emerged as a civil liberties activist who crusades for open access to data.

In 2008, Mr. Swartz released a “Guerrilla Open Access Manifesto,” calling for activists to “fight back” against the sequestering of scholarly papers and information behind pay walls.

“It’s time to come into the light and, in the grand tradition of civil disobedience, declare our opposition to this private theft of public culture,” he wrote. One goal: “We need to download scientific journals and upload them to file-sharing networks.”

He also earned renown for [downloading nearly 20 million pages of court documents](#) for a project that put them free online. That brought Mr. Swartz under federal investigation. He was not indicted but later [published the resulting F.B.I. file](#) online.

He faces up to 35 years in prison and \$1 million in fines for charges related to wire fraud, computer fraud and unlawfully obtaining information from a protected computer. He surrendered to the authorities on Tuesday morning, was arraigned in Federal District Court and pleaded not guilty to all counts. He was released on \$100,000 unsecured bond.

Institutions like colleges and libraries pay for access to JSTOR, which is then available free to their

users. Supporters were quick to defend Mr. Swartz. David Segal, executive director of [Demand Progress](#), an activist group that Mr. Swartz founded, said in a statement that the arrest “makes no sense,” comparing the indictment to “trying to put someone in jail for allegedly checking too many books out of the library.” An online petition gathered 15,000 signatures in just a few hours.

In an interview, Mr. Segal said that his comments went to the principle, not to anything Mr. Swartz might have done in obtaining the documents.

“I know him as a person who cares deeply about matters of ethics and government,” Mr. Segal said. “I don’t know about the matter of what has been alleged.”

Beginning in September of last year, according to the indictment, Mr. Swartz used several methods to grab articles, even breaking into a computer-wiring closet on the M.I.T. campus and setting up a laptop with a false identity on the school network for free JSTOR access under the name Gary Host — or when shortened for the e-mail address, “ghost.” When retrieving the computer, he hid his face behind a bicycle helmet, peeking out through the ventilation holes.

The flood of downloads was so great that it crashed some JSTOR servers, the indictment stated, and JSTOR blocked access to the network from M.I.T. and its users for several days.

Ultimately Mr. Swartz returned the hard drives containing the articles to JSTOR and promised that the material would not be disseminated.

“We are not pursuing further action,” the organization’s general counsel, Nancy Kopans, said; the organization said in a statement that the criminal case “has been directed by the United States Attorney’s Office.”

As for the comments from Mr. Swartz’s supporters that he had done nothing wrong, however, Ms. Kopans said, “It’s an unfortunate situation, but I think the facts speak for themselves.”

Mr. Swartz recently completed a 10-month fellowship at the Edmond J. Safra Center for Ethics at Harvard. “Aaron has never done anything in this context for personal gain — this isn’t a hacking case, in the sense of someone trying to steal credit cards,” said Lawrence Lessig, the center’s director. “That’s something JSTOR saw, and the government obviously didn’t.”

In a statement announcing the charges, a United States attorney, Carmen M. Ortiz, said: “Stealing is stealing, whether you use a computer command or a crowbar, and whether you take documents, data or dollars. It is equally harmful to the victim whether you sell what you have stolen or give it away.”

Carl Malamud, an online activist who worked with Mr. Swartz on the court-documents project,

called Mr. Swartz “one of the Internet’s most talented programmers,” but said that “the JSTOR situation is very disturbing.”

In an e-mail exchange with a reporter, Mr. Malamud, who is [engaged in a project](#) intended to put all laws and government documents online, said: “My style, when I see a gate barring entry and that gate is sanctioned by the law, is to go up to that gate and pound on it hard and force them to open up. Others sometimes look for a back door.”

He added, “I’m not convinced that style is always effective, and it is certainly often dangerous.”

Nick Bilton contributed reporting.



MORE IN U.S. (1 OF 20 ARTICLES)

OPEN

Damage and Flooding Scar Atlantic Seaboard

[Read More »](#)

Swartz supporter dumps 18,592 JSTOR docs on the Pirate Bay

By [Timothy B. Lee](#) | Published about a month ago

A 31-year-old American who says his name is Gregory Maxwell has posted a 32GB file containing 18,592 scientific articles to BitTorrent. In a [lengthy statement](#) posted to the Pirate Bay, he says that Tuesday's [arrest](#) of onetime Reddit co-owner Aaron Swartz inspired the document release.

"All too often journals, galleries, and museums are becoming not disseminators of knowledge—as their lofty mission statements suggest—but censors of knowledge, because censoring is the one thing they do better than the Internet does," he wrote.

Maxwell says he is a technologist living in the Washginton, DC area. He tells Ars that "I'm not an academic, but I sometimes play one on the Internet. I am a hobbyist scientist and many of the things I work on leave me often consuming scientific papers."

He rarely has trouble gaining access to the documents he needs because "my social circle is stuffed full of academics." But he worries that others who aren't as well connected will have trouble getting access to academic research.

It's safe to say that Swartz would approve of Maxwell's actions. In a [2008 "Guerrilla Open Access Manifesto"](#) first reported by the *New York Times*, Swartz wrote that "we need to download scientific journals and upload them to file-sharing networks."

Maxwell says the documents he released were all published before 1923, which means that they should be in the public domain under US copyright law. He says he acquired the documents through "rather boring and lawful means," and wanted to publish them on a public site such as [Wikisource](#). However, he worried that incumbent publishers would "claim that their slavish reproduction—scanning the documents—created a new copyright interest."

This is not an unreasonable fear. Publishers of public domain works have long argued that [token modifications to public domain works](#) could transform them into copyrighted works.

So for "a long time," Maxwell opted not to publish the documents. But now, in the wake of Swartz's arrest, Maxwell says he feels that was the wrong decision.

"If I can remove even one dollar of ill-gained income from a poisonous industry which acts to suppress scientific and historic understanding, then whatever personal cost I suffer will be justified—it will be one less dollar spent in the war against knowledge."

Swartz and Maxwell's actions are part of a broader campaign to open up public access to academic work. In February, the prominent computer security researcher Matt Blaze [blasted](#) ACM and IEEE, the two most important professional societies in his profession, for restrictive copyright policies.

"I will no longer serve as a program chair, program committee member, editorial board member, referee or reviewer for any conference or journal that does not make its papers freely available on the Web or at least allow authors to do so themselves," Blaze wrote. He called on his fellow academics to join him. Ironically, the faculty of MIT—where Swartz allegedly did his downloading—[made a similar pledge](#) in 2009.

[Photograph by Anna Creech](#)



February 13, 2009

An Effort to Upgrade a Court Archive System to Free and Easy

By [JOHN SCHWARTZ](#)

Americans have grown accustomed to finding just about anything they want online fast, and free. But for those searching for federal court decisions, briefs and other legal papers, there is no [Google](#).

Instead, there is Pacer, the government-run [Public Access to Court Electronic Records](#) system designed in the bygone days of screechy telephone modems. Cumbersome, arcane and not free, it is everything that Google is not.

Recently, however, a small group of dedicated open-government activists teamed up to push the court records system into the 21st century — by simply grabbing enormous chunks of the database and giving the documents away, to the great annoyance of the government.

“Pacer is just so awful,” said Carl Malamud, the leader of the effort and founder of a nonprofit group, [Public.Resource.org](#). “The system is 15 to 20 years out of date.”

Worse, Mr. Malamud said, Pacer takes information that he believes should be free — government-produced documents are not covered by copyright — and charges 8 cents a page. Most of the private services that make searching easier, like Westlaw and Lexis-Nexis, charge far more, while relative newcomers like [AltLaw.org](#), [Fastcase.com](#) and [Justia.com](#), offer some records cheaply or even free. But even the seemingly cheap cost of Pacer adds up, when court records can run to thousands of pages. Fees get plowed back to the courts to finance technology, but the system runs a budget surplus of some \$150 million, according to recent court reports.

To Mr. Malamud, putting the nation’s legal system behind a wall of cash and kludge separates the people from what he calls the “operating system for democracy.” So, using \$600,000 in contributions in 2008, he bought a 50-year archive of papers from the federal appellate courts and placed them online. By this year, he was ready to take on the larger database of district courts.

Those courts, with the help of the Government Printing Office, had opened a free trial of Pacer at 17 libraries around the country. Mr. Malamud urged fellow activists to go to those libraries, download as many court documents as they could, and send them to him for republication on the Web, where Google could get to them.

Aaron Swartz, a 22-year-old Stanford dropout and entrepreneur who read Mr. Malamud’s appeal, managed to download an estimated 20 percent of the entire database: 19,856,160 pages of text.

Then on Sept. 29, all of the free servers stopped serving. The government, it turns out, was not pleased.

A notice went out from the Government Printing Office that the free Pacer pilot program was suspended, “pending an evaluation.” A couple of weeks later, a Government Printing Office official, Richard G. Davis, told librarians that “the security of the Pacer service was compromised. The [F.B.I.](#) is conducting an investigation.”

Lawyers for Mr. Malamud and Mr. Swartz told them that they appeared to have broken no laws, noting nonetheless that it was impossible to say what angry government officials might do.

At the administrative office of the courts, a spokeswoman, Karen Redmond, said she could not comment on the fate of the free trial of Pacer, or whether there had been a criminal investigation into the mass download.

The free program “is not terminated,” Ms. Redmond said. “We’ll just have to see what happens after the evaluation.” As for the system’s cost, she said: “We’re about as cheap as we can get it. We’re talking pennies a page.”

Meanwhile, the 50 years of appellate decisions remain online and Google-friendly, and the 20 million pages of lower court decisions are available in bulk form, but are not yet easily searchable. “I want the whole database in 2009,” Mr. Malamud said.

Mr. Malamud, 49, has a long record of trying to balance openness with privacy, and has also pushed the Securities and Exchange Commission and the Patent and Trademark Office to put their records online free. But the issue is a thorny one with court documents, which often contain personal information.

Daniel J. Solove, a professor at the [George Washington University](#) Law School, noted that marketers skim court records for personal data, and making records easier to troll will put even more data at risk. “It’s taking away this middle ground that offered a lot of protection, practically, and throwing it into this radically wide open box,” he said.

But this argument for what is known as “practical obscurity” does not convince Peter A. Winn, a privacy expert who is an assistant United States attorney in Washington State. Noting that he was speaking only for himself, he argued that the courts developed rules over the last 400 years to protect privacy.

“It worked in the bricks-and-mortar age — it should work in the electronic age,” Mr. Winn said. The administrative office of the courts, he said, should take on the role of policing privacy on its databases. “This is going to take focus and a lot of hard work,” he said.

Mr. Malamud agrees that the court system needs to do a better job of protecting privacy. He found thousands of documents in which the lawyers and courts had not properly redacted personal information like [Social Security](#) numbers, a violation of the courts’ own rules. There was data on children in Washington, names of Secret Service agents, members of pension funds and more.

“They’re pretty spectacular blunders,” he said. He sent letters to the clerks of individual courts around the

country. After some initial inaction, and repeated and increasingly spirited notices from Mr. Malamud, most of the offending documents were pulled from the databases to be redacted.

Ms. Redmond, of the administrative office of the courts, said the courts comb through the documents “on a regular basis” and tell lawyers to redact confidential information. The number of violations, she noted, was relatively small.

Mr. Malamud scoffed at that. “This is a large number of transgressions, and this is illegal,” he said. “The law doesn’t say that you should only publish a small number of Social Security numbers!”

Mr. Malamud said his years of activism had led him to set a long-shot goal: serving in the [Obama administration](#), perhaps even as head of the Government Printing Office. The thought might seem far-fetched — Mr. Malamud is, by admission, more of an at-the-barricades guy than a behind-the-desk guy. But he noted that he published more pages online last year than the printing office did.

Mr. Malamud represents a perspective of openness and transparency that is much in tune with the new administration’s, said [Lawrence Lessig](#), a law professor at Harvard who is a leading advocate for free culture. “The principles are those that Carl has been at the center of defining,” he said.

The idea also seems to have a measure of appeal for [John D. Podesta](#), a longtime fan of Mr. Malamud and head of the Obama transition team, who stopped short, however, of anything resembling an endorsement. “He would certainly shake things up,” Mr. Podesta said, laughing.

Mr. Malamud says he is not counting on the new administration’s being quite that bold. Besides, he said, he keeps himself awfully busy doing what he believes the government ought to be doing anyway.

“If called, I will certainly serve,” he said. “But if not called, I will probably serve anyway.”

Copyright 2009 [The New York Times Company](#)

[Privacy Policy](#) | [Search](#) | [Corrections](#) | [RSS](#) | [First Look](#) | [Help](#) | [Contact Us](#) | [Work for Us](#) | [Site Map](#)
