

## On Archiving and Commoning the Snowden Files

by Andrew Clement

What Snowden has revealed is a complex, institutionalized system of mass surveillance that is deeply embedded within and operating through our state and corporate apparatus. Only through a major collective investigative effort drawing on multiple perspectives can we adequately come to grips with its scope, consequences and remedial possibilities. An archive such as the Digital Snowden Surveillance Archive [<https://snowdenarchive.cjfe.org>] which I developed with various colleagues would be an essential resource in this effort.

Looking at ground breaking leaks, especially with regard to how society managed (or not) to archive them, we can learn from history. For me the most relevant prior leak that had great social significance was whistleblower Daniel Ellsberg's leak of the Pentagon Papers. Making public authoritative internal documents about the Vietnam War that showed that officials were routinely lying about the motivations and state of the war played an important role in public opposition to the war and the eventual US withdrawal.

The Snowden documents have a similar potential power because they too show in detail shocking government activities and bald lying by public officials. Conditions are of course different than in 1971 when the Pentagon Papers became public. There was already a strong social movement opposing the Vietnam war **to** which the Pentagon Papers added fuel. At present, there is only a nascent, still quite weak social movement opposing state surveillance.

The potential value of the Snowden leak is to help coalescing and broadening opposition. Furthermore there appears to be have been more dissent in 1971 among the upper political strata than is the case now, making the challenge of changing direction even more formidable.

### Public education about mass surveillance

I had several motivations in initiating the Digital Snowden Surveillance Archive project, mainly having to do with helping to promote and inform the public debate around mass state surveillance. Now that we know our state security agencies are conducting fine grained surveillance of everyone's electronic activities, we as a society have very serious choices to make about the appropriate role for secretive security agencies in a democracy.

If we do nothing, then we will have accepted *de facto* that our everyday lives are open to scrutiny by unaccountable government agencies. This I believe is inimical to the foundations of democracy and we run a high risk of becoming police states. Reining in these agencies and eliminating those aspects that are not justifiable is a very difficult, but necessary task. It can only be accomplished when substantial numbers are well enough informed about the existing surveillance practices and the threats they pose, to take effective remedial action.

Given the secrecy and complexity of the practices involved, public education about mass surveillance is vital. This is something that I have been pursuing in my research for several years, especially around the IXmaps.ca project that seeks to show people the paths their data takes across the internet and where it may be intercepted by the NSA.

Firstly, I wanted a searchable archive of the Snowden documents for this research, so I could better locate and identify surveillance sites of the NSA and its Five Eyes partners that I could include in the on-going IXmaps work. It seemed like a pretty obvious idea, so was surprised I couldn't find such an archive already available. I had some research funds, and looked for someone in my Faculty's Archive and Records Management specialization who was interested in the subject matters that I could hire. I was fortunate to find George Raine, a trained archivist who had recently graduated from our masters program. George was keen to be involved in the project, had many of the necessary skills and was up for learning what else was needed.

More generally it struck me that many other opportunities were opened up by the Snowden documents that could lead to academic and journalistic research and reporting that weren't addressed by the media coverage to date. Apart from Glenn Greenwald's "No Place to Hide"

book, reporting has consisted almost entirely of sensational stories based on a relatively small handful of documents newly released with the article. The ability to see an individual document in a wider context and to pursue threads across the whole range of documents makes possible a more penetrating inquiry into the driving forces and overall nature of mass surveillance.

### **The archive's architecture**

Given my primary goal of promoting an open, informed public debate, I intended from the beginning to create a widely accessible on-line archive under free/open licences.

The Snowden archive is built using Greenstone, a suite of software for building and distributing digital library collections. It is produced by the New Zealand Digital Library Project at the University of Waikato, and developed in cooperation with UNESCO and the Human Info NGO. Being open source, it is widely used around the world for digital library initiatives, especially in developing countries. We recognize that Greenstone does not have many features of more recently developed digital archive platforms. Once we get a better sense of the needs of Archive users we may consider porting to another platform.

The Snowden Archive that is available on the Canadian Journalists for Free Expression (CFJE) website has been highly customized. Documents are described according to a custom metadata schema that is sensitive to contextual elements of the Snowden documents that are not present in most other document collections, such as security classification codes and distribution markings. The look and feel of the collection, including the format of the document descriptions have also been very heavily modified from the standard Greenstone template.

The vast majority of documents released by the media are PDF files. In their original form, there were a lot of powerpoint files and other proprietary formats. The newspapers did work for us by releasing them in PDF and PDF/A, which are both very widely used, open-source formats. We determined that there was little likelihood that PDF files would become obsolete in the foreseeable future (it is an extremely widely used, open-source standard). If they do, it is easy to retrieve the documents from the collection and re-upload them in a different, more widely used format.

### **Linking to offline archives**

There is an initiative to develop an offline "Snowden Archive in a Box" developed by Evan Light at Concordia University's Mobile Media Lab where he works on privacy, surveillance and telecom issues.

The Portable Snowden Surveillance Archive is an autonomous version of the fully text-searchable Internet-based archive Snowden Digital Surveillance Archive created by Canadian Journalists for Free Expression and researchers at the University of Toronto. It is a stand-alone wifi network and web server that permits you to research all files leaked by Edward Snowden and subsequently published by the press. The purpose of the portable archive is to provide end-users with a secure off-line method for individuals to use this database without the threat of mass surveillance.

The Portable Snowden Surveillance Archive began as part of an evolving and touring European project called Performigrations [<http://www.performigrations.eu>] which focuses on migration/immigration and was launched in Montreal at the Blue Metropolis literary festival in April 2015. An evolving project in its own right, a current version of the Portable Archive also includes a surveillance demonstration apparatus that monitors wifi traffic around it and plays it back to the public. In June, it will be showcased at the Biografilm festival in Bologna, Italy – in partnership with Performigrations – and at the Citizenship and Surveillance Conference in Cardiff, Wales. The Portable Archive may appear in future Performigrations iterations in Europe and Canada.

### **The role of public libraries**

I would like to see the Snowden Archive become more than a passive resource, but also a site for collaborative research and deliberation. Libraries certainly have an important role to play, especially public libraries as they go beyond their more conventional role of making materials

accessible to devote more attention to facilitating discussion and deliberation within the communities they serve based on these materials.

My own university library contacted me about archiving materials related to the Snowden Archive (specifically the media articles that published the documents). We're now working to have the library host a mirror of the entire Archive. Establishing mirroring sites is desirable in several ways. Besides improving accessibility and technical stability through redundancy, it also provides local users (say students) access to the collection without exposing their search traffic to internet interception and expresses solidarity with the ideals of open access to controversial materials.

We've approached other universities as potential mirror sites, but so far this has been bogged down by the fact that the documents represent 'stolen goods' and so possessing them would be a criminal violation (at least in Canada). While the chance of prosecution is very small, legal departments in a couple of universities are balking. Going directly through the libraries themselves looks to be a better prospect as they both have the necessary technical capabilities and appear more oriented than university administrations to preserving academic freedoms around contentious holdings.

While our current focus is on ensuring that the Archive is accessible to all, reliable, easy to use, accurate and updated as new documents are published, to fulfill its potential as a 'knowledge commons' around the issue of state surveillance, it also needs a community of engaged users who will conduct research based on the Archive and give wider public meaning to its contents. Ideally this would include people who can provide insightful annotations, contribute additional relevant documents, host mirrors, stimulate conversations, initiate collective research ventures, ... While extending the software to support such distributed collaboration and animating the wider conversation is beyond our abilities at the moment, hopefully there are others who are willing and able to take this on.

*\* This protocol is based on an interview conducted by the editorial team of the Berliner Gazette in the context of the annual focal point UN|COMMONS ( <http://berlinergazette.de/archivierung-der-snowden-dokumente> ) in the context of which the Berliner Gazette is hosting an international conference October 22-24, 2015 in Berlin. The German translation of the text has been published here: <http://berlinergazette.de/archivierung-der-snowden-dokumente/>*

*\*\* George Raine, the archivist who designed and built the "Digital Snowden Surveillance Archive", provided additional info on the archive technology. Furthermore some information by Evan Light are included in the text. The projects mentioned in the text are presented at the "Digital Citizenship and Surveillance Conference" in Cardiff, UK, June 18-19.*

*\*\*\* Andrew Clement is a Professor in the Faculty of Information at the University of Toronto. He is a co-founder of the Identity, Privacy and Security Institute. His research and teaching interests are in the social implications of information/communications technology and human-centred systems development. Recent work focusses on public information policy for guiding the development of Canada's information infrastructure, digitally mediated surveillance, privacy; digital identity constructions, public participation in information/communication infrastructures development, and community networking. He has also written papers and co-edited books in such areas as: internet use in everyday life, computer supported cooperative work; participatory design; workplace surveillance; women, work and computerization; end user computing; and the 'information society' more generally. See: <http://iprp.ischool.utoronto.ca/>*