

April 2009

From Dark Archive to Open Access: CLOCKSS Trigger Event Lessons

Victoria Reich

Stanford University, vreich@stanford.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



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

Recommended Citation

Reich, Victoria (2009) "From Dark Archive to Open Access: CLOCKSS Trigger Event Lessons," *Against the Grain*: Vol. 21: Iss. 2, Article 9.

DOI: <https://doi.org/10.7771/2380-176X.2552>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

From Dark Archive to Open Access: CLOCKSS Trigger Event Lessons



by **Victoria Reich** (Director, LOCKSS Program, Stanford University Libraries, 1450 Page Mill Road, Palo Alto, CA 94304; Phone: 650-725-1134) <vreich@stanford.edu> <http://www.lockss.org> <http://www.clockss.org>

The **CLOCKSS¹ Archive** is singular among digital Archives. Its governing board is comprised equally of librarian directors and publisher directors. Every institution supporting the archive has a seat on either the board or the Advisory Council. The archive is global, with archive nodes — libraries that are holding, and actively preserving, the physical digital bits on a server called a “**CLOCKSS box**.” These nodes are situated across four continents: in the United States and Canada; Scotland; Hong Kong and Japan; and Australia.

In 2008, two **SAGE Publication** titles, *Graft* and *Auto/Biography* were triggered from the archive and made available to everyone on the Web, at no charge.

What is a Trigger Event?

The **CLOCKSS Archive** formally defines a trigger event as follows:

A Trigger Event occurs when either the owner of all rights to the content gives unconditional consent to the release of such content to the general public, or the content is determined in good faith by the board to be unavailable from any publisher for at least six consecutive months and there are no successor interests or reversions or transfers of rights known to the board at the time of the determination.

Trigger Events include, but are not limited to, situations of non-availability of archived content in which:

1. **Publisher No Longer in Business.** Publisher is no longer in business or is no longer in the business of publishing Content or providing access to previously published Content and there are no successor interests or reversions or transfers of rights.
2. **Title No Longer Offered.** Publisher has stopped publishing and is no longer providing access to the Content and there are no interests or reversion or transfer of rights.
3. **Back Issues No Longer Available.** Publisher has stopped offering or providing access to some or all of the back issues of the Content and there are no successor interests or reversion or transfer of rights; or
4. **Catastrophic Failure.** While still publishing Content, Publisher is not able to provide access to the Content electronically due to technical or similar catastrophic and permanent failure.

The First Trigger Event

Three volumes of *Graft* are preserved in the **CLOCKSS Archive** (from 2001 to 2003).

When **SAGE** announced this title would no longer be available from the **HighWire** hosting platform, the board voted to approve the trigger event and agreed to make the content freely available. (See <http://www.clockss.org/clockss/Graft>).

The Second Trigger Event

SAGE ceased to publish *Auto/Biography* in 2006; however, **IngentaConnect** continued to host the title until 2008. When **SAGE** announced **Ingenta** would be taking *Auto/Biography* off-line, **CLOCKSS** experienced its second trigger event. Again, with board approval, the content was made available to anyone with a Web browser for free. (See <http://www.clockss.org/clockss/Auto/Biography>).

Carol Richman, **SAGE's** Director of Licensing, said, “As these titles did not have a viable subscription base, **SAGE** thought it a good opportunity to offer the community real trigger event experiences.” Indeed, the trigger events were an excellent learning experience. They validated the **CLOCKSS** policy of making triggered content available open access; and are providing data about how readers use triggered content.

Why Open Access?

Content is most likely to be triggered when:

- it is not garnering enough subscription revenue or ad revenue to continue to earn money for the publisher, or
- a catastrophic disaster has occurred.

Content without a viable subscription base for the publisher is unlikely to have a viable subscription base for an archive, therefore the **CLOCKSS** board agreed to make content triggered from **CLOCKSS** freely available.

If a catastrophic disaster befalls one or more publishers, it's likely that this disaster will have other wide reaching consequences. The **CLOCKSS** board agreed in times of catastrophic disaster, generosity was an appropriate response and again agreed the content would be available for free.

Titles triggered from the **CLOCKSS Archive**, are assigned a **Creative Commons** license. For *Graft* and *Auto/Biography*, the content is copyright **SAGE** and licensed under a **Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License**. The **Creative Commons** license clarifies how people can use this content. This particular **Creative Commons** license permits users to share (i.e., to copy, distribute and transmit the work) under the following conditions:

- Attribution. You must attribute the work in the manner specified by the author or

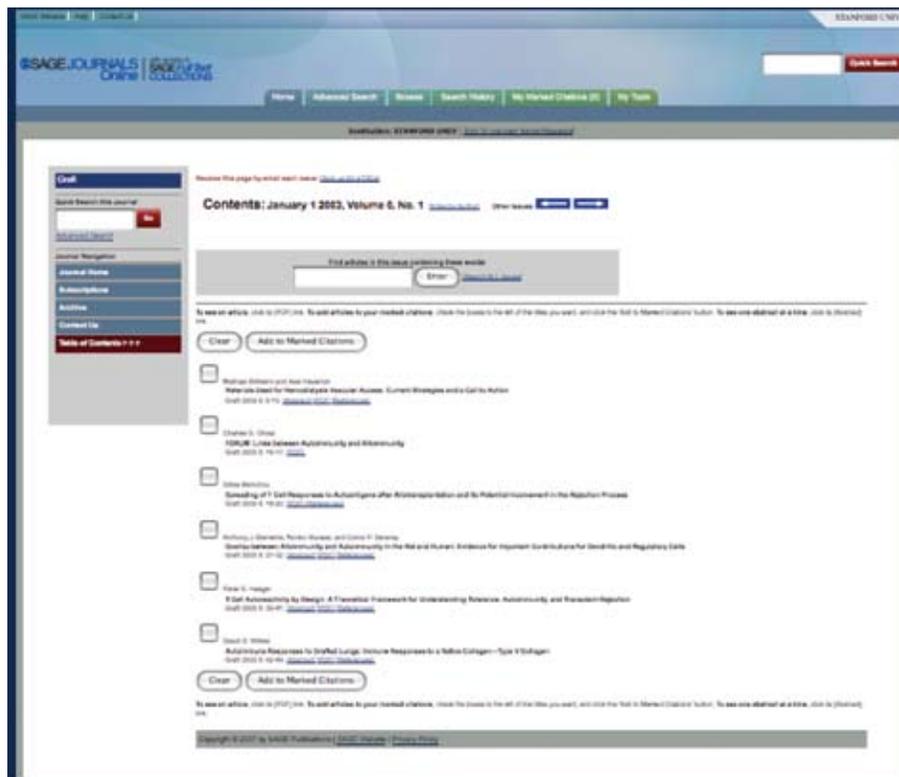


Figure 1: Screenshot of **CLOCKSS** triggered content, *Graft*.

continued on page 26

licensor (but not in any way that suggests that they endorse you or your use of the work).

- Non-commercial. You may not use this work for commercial purposes.
- No Derivative Works. You may not alter, transform, or build upon this work.

The **Creative Commons** license is important for archive interoperability, and the continued preservation of this content. It permits those archives that use a preservation method that preserves the original content — those, which do not alter the content upon ingest — to take this content into that archive with no further negotiation or contact with the copyright holder.

How is it Accessed?

The triggered content was copied from the archive and posted by **CLOCKSS** host institutions, the **University of Edinburgh EDINA** data centre and **Stanford University**. The content is not heavily used. For the month of November 2008, for example, the **Stanford** server for *Graft* delivered about eight URLs per hour. Excluding obvious search engine crawlers, readers downloaded 64 PDFs to 36 distinct IP addresses, of which 31 had domain names that could be found by reverse DNS lookup. Nine were identifiably academic. 28 of the IP addresses found the content via **Google**. Seven of them found the content via the **CLOCKSS** Website. In March 2008, a few months after the initial release of the content, about 7% of the access came via OpenURL resolvers. In November no accesses were recorded.

The *Auto/Biography* content is even less used. In the same month, only two PDFs were downloaded, both by a crawler.

At the time of the trigger event the **CrossRef** DOI resolver could map a DOI to only a single URL. **Portico** claimed the DOI, so it

against the grain
people profile

Director, LOCKSS Program, Stanford University Libraries
1450 Page Mill Road, Palo Alto, CA 94304
Phone: (650) 725-1134 • <vreich@stanford.edu>
www.lockss.org • www.clockss.org
http://www.lockss.org/lockss/Vicky_Reich

Victoria Reich

PROFESSIONAL CAREER AND ACTIVITIES: I've been privileged to work in great libraries with outstanding professionals. I learned how to do reference at the **University of Michigan** and further honed these skills at the **Library of Congress' Science Division**. I was Head of Serials and Acquisitions, first at the **National Agricultural Library** and then at **Stanford University**. My introduction to digital materials, and digital preservation came in 1982. I worked in the Office of the **Librarian of Congress** and one of my many assigned projects was to participate in the "**LC Optical Disk Pilot Project**." It would take twenty years for digital preservation to become a general concern. In the 90s I helped to start **Stanford University Libraries' HighWire Press**, the digital imprint of some of the best STM journals in the world. In 1998, with the help of a brilliant engineer, a small grant from the **NSF**, and the support of **Stanford University Librarian, Michael Keller**, we launched the **LOCKSS Program**. Today libraries are using the technology to preserve a broader spectrum of digital materials than we could ever have imagined. In October 2008, a terrific team of publishers and librarians successfully launched the **CLOCKSS Archive** as a non-profit organization. **CLOCKSS** is the only archive governed and owned by publishers and librarians.

BOOK I AM READING NOW: *Here Comes Everybody: The Power of Organizing Without Organizations* by **Clay Shirky**.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: **Paul Saffo** often remarks, "Never mistake a clear view for a short distance." I can't predict where the profession will be in five years but I can suggest steps we can take to ensure libraries remain vibrant, relevant institutions. **ATG** readers may be interested in an opinion piece I co-authored with **Michael Keller** and **Dr. David Rosenthal**, "'Just in Time' in Difficult Times: Lessons to be Learned." You can download it from the **CNI** Website at <http://www.cni.org/firms/2009a.spring/abstracts/PB-just-reich.html> — click on "Handout." 🌱

pointed (and still points) users to the **Portico** copy of the PDF, which is available only to **Portico** subscribers.

The **CLOCKSS** experience led to the discovery that the DOIs for triggered content

should be owned and managed by a community organization, not by a single Archive.

"The availability of *Graft* content in **CLOCKSS** prompted **CrossRef** to create an implementation of **CrossRef Multiple Resolution** since the content was available in more than one archive. The end result is that different archive URLs can be registered with the *Graft* DOIs so that users can easily find all the options available for the content. **CLOCKSS**, **Portico**, the **KB** and **CrossRef** have worked together closely to put a solution in place" — **Ed Pentz**, Executive Director, **CrossRef**.

Who Uses it?

The statistics above, reflecting one month's usage of one of the two servers concerned, show that triggered content gets little use, and that the majority of the use (75% in this case) is not identifiably academic. This is not surprising; the reason the content was triggered was that it was not generating enough use from academic subscribers to justify the costs of making it accessible.

Experience thus validates the decision by the **CLOCKSS** board to make triggered content open access, and the use of the **Creative Commons** license to do so. Charging for users



Figure 2: Screenshot of CLOCKSS triggered content, *Graft*.

continued on page 28

to access the content would likely reduce usage considerably. It would probably eliminate most of the accesses via Google, which are consistently the vast majority.

What Does it Look Like?

Graft was hosted on HighWire Press, and the content was ingested into the archive directly from HighWire Press. SAGE deposited into the CLOCKSS Archive exactly what was published. Hence, the preserved copy is what the readers saw in 2008, the look and feel, the publisher branding is preserved. (See figures 1 thru 3, on pages 24, 26, 28.)

The *Auto/Biography* files ingested into the CLOCKSS Archive were the “pre-publication” files (sometimes called “source files”). The content was not available to the CLOCKSS Archive for direct collection from the publisher’s Website. These pre-publication files are preserved in the CLOCKSS Archive. To prepare the volumes for the hosting platforms, the content had to be published. The look and feel for this title is not preserved.

What has CLOCKSS Learned?

The *Graft* and *Auto/Biography* trigger events validated the CLOCKSS board deci-

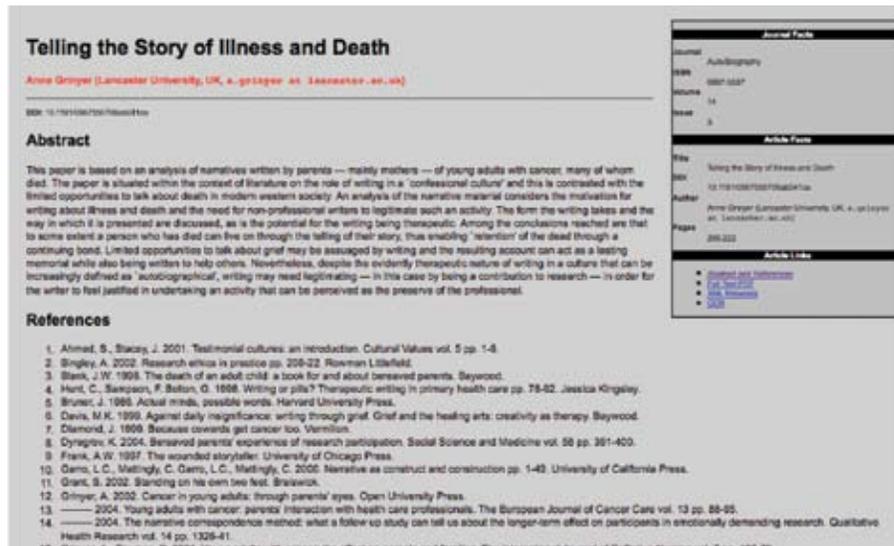


Figure 3. Screenshot of CLOCKSS triggered content, *Auto/Biography*.

sion make triggered content Open Access, accompanied by a Creative Commons license. The Creative Commons license clearly states how this content may and may not be used. As expected, use of this content is relatively low.

Endnotes

1. CLOCKSS stands for **Controlled LOCKSS (Lots of Copies Keep Stuff Safe)**.

Federal Depository Library Program: Services and Collections

by **James A. Jacobs** (Data Services Librarian, Emeritus, University of California San Diego, 7050 Condon Drive, San Diego, CA 92122; Phone: 858-452-9704)
<jajacobs@ucsd.edu> <http://www.freegovinfo.info>

In the age of digital information, libraries and librarians are struggling to define their proper roles. In a time of financial uncertainty and economic crisis, many libraries are facing decisions that will have long-term implications and consequences. More than ever, it is particularly important that we have a clear vision of a sustainable role for libraries.

The issues libraries face can be seen very clearly in a proposal by the **Depository Library Council**, which advises on matters related to the **Federal Depository Library Program (FDLP)**. It has recommended that the **Government Printing Office (GPO)** should “prepare depository libraries for a digital **Federal Depository Library** system that is not centered on collections.” The **Council** is suggesting that government depository libraries should focus on services *instead of* collections.

With this recommendation, the **Council** has reached its own implicit conclusions about the roles of librarians and libraries in society. The **Council** is saying that the role of *librarians* is to provide information services and the role of *libraries* (collections) should be in the hands of **GPO**, the **National Archives and Records**

Administration (NARA), and individual government agencies.

There are at least two reasons that this decision is a troubling one in these tumultuous times. First, it seems counter-intuitive to claim that the best future we can imagine for libraries in the digital age is “libraries without collections.” Second, it is not clear that government agencies have or should have the role that the **Council** wants for them.

The Role of Librarians

An emphasis on service at the expense of collections comes mostly from a view that users are overwhelmed by an information glut and need information professionals to help them navigate a bewildering array of choices. Although this view is a bit paternalistic, implying that librarians know better than users what they need, it is at least based on an understanding of the complex and difficult job of finding the right information on the Web today. In this view, librarianship would be about helping people navigate a complex, networked maze of shifting, changing information. There is nothing wrong with the view that libraries should provide information services and there is in

fact much to recommend it, but this service-only model misses a key role for libraries. It is a view of librarians without libraries.

This view assumes an unorganized, undifferentiated Web of information controlled by information providers (e.g., government agencies, commercial vendors, information aggregators, publishers), visible only through the information silos and portals created by those providers. It accepts that libraries will not build digital collections to fit the needs of their users but will simply provide services for information over which librarians have no control.

Librarians, in this view, are valuable precisely because they have no control over information.

This view also accepts that information will be tightly controlled by producers and distributors. What is available, who can use it, under what conditions it may be used, and when it becomes unavailable will be controlled

