

Contemporary Composers Web Archive (CCWA): Progress in Collaboratively Collecting Composers' Websites

June 24, 2015

IAML/IMS

Anna Perricci, Columbia University

Laura Stokes, Brown University

What is CCWA?

The screenshot shows a web browser window with the address bar displaying <https://archive-it.org/collections/4019>. The browser's toolbar includes a search bar and various icons. Below the browser window, the Internet Archive website is visible. The header features the Archive-IT logo, navigation links (HOME, EXPLORE, LEARN MORE, CONTACT US), and a tagline: "The leading web archiving service for collecting and accessing cultural heritage on the web. Built at the Internet Archive".

The main content area displays the title "Contemporary Composers Web Archive (CCWA)" with a small thumbnail image. Below the title, it states "Collected by: [Columbia University Libraries Consortial Collections](#)". The "Archived since" date is "Oct, 2013". The "Description" reads: "The Contemporary Composers Web Archive is a newly launched initiative by the music librarians at Brown, Columbia, Cornell, Dartmouth, Harvard, Johns Hopkins, Princeton, and Yale universities, MIT, and the universities of Chicago and Pennsylvania (collectively known as the Borrow Direct Music Librarians Group) and operates under the auspices of Columbia University Libraries and Information Services." The "Subject" is listed as "Arts & Humanities, Society & Culture" and the "Collector" as "Borrow Direct Music Librarians Group".

Below the description, the "Narrow Your Results" section is visible. It includes a "Group" filter showing "Contemporary Composers (54)" and a "Subject" filter. There is a search bar with the placeholder text "Enter search terms here" and buttons for "Search" and "Clear".

Collaborative Project

- Borrow Direct (Ivies-Plus) Music Librarians Group
 - Brown University (Laura Stokes)
 - Columbia University (Elizabeth Davis)
 - Cornell University (Bonna Boettcher, Leonora Schneller)
 - Dartmouth College (Patricia Fisk)
 - Harvard University (Sarah Adams, Sandi-Jo Malmon)
 - Johns Hopkins University (Jennifer Ottervik)
 - Massachusetts Institute of Technology (Peter Munstedt)
 - Princeton University (Darwin Scott)
 - University of Chicago (Scott Landvatter)
 - University of Pennsylvania (Dick Griscom)
 - Yale University (Suzanne Lovejoy)
 - ...and Anna Perricci, Columbia University

Ex.: Steven Stucky's website, list of captures

The screenshot shows a web browser window with the Internet Archive Wayback Machine interface. The search bar contains the URL <http://www.stevenstucky.com/>. The search results show 5 captures between April 18, 2014, and April 28, 2015. The table below summarizes the captures.

Found 5 Captures between Apr 18, 2014 - Apr 28, 2015	
2014	2015
4 pages	1 page
Apr 18, 2014 *	Apr 28, 2015 *
Apr 28, 2014	
May 19, 2014	
Oct 28, 2014	

Home | Internet Archive

Captured page, 10/28/2014

Steven Stucky, composer

wayback.archive-it.org/4019/20141028160355/http://ww ccwa stucky

Most Visited News Popular Gmail Google Maps Dictionary / Wörter... Der digitale Grimm...

You are viewing an archived web page, collected at the request of [Columbia University Libraries Consortial Collections](#) using [Archive-It](#). This page was captured on 16:03:55 Oct 28, 2014, and is part of the [Contemporary Composers Web Archive \(CCWA\)](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. Note that this document was downloaded, and not saved because it was a duplicate of a previously captured version. HTTP headers presented here are from the original capture.

COMPOSER
Steven Stucky

Biography
Works
Performances
Recordings
Writings
Reviews
Articles/Interviews
For the Press
Contacts

Widely recognized as one of the leading composers today, Steven Stucky was awarded the 2005 Pulitzer Prize in Music for his *Second Concerto for Orchestra*. He has written commissioned works for many of the major American orchestras and ensembles.

Mr. Stucky has taught at Cornell University since 1980, where he serves as Given Foundation Professor of Composition. He has also been associated with the Los Angeles Philharmonic for more than 20 years, and is currently Consulting Composer for New Music.




Photo © Hoebermann Studio

StevenStucky.com is produced by OperaWeb.com.

stucky Highlight All Match Case 1 of 3 matches

Web Resources Archiving Collaboration

Many thanks to the Mellon Foundation

Building collaborations among

- Web archiving communities
- Other research libraries
- Users and potential users of web archives
- Website creators

Premise: we need radical collaboration and cross-institutional partnerships

Collaborative
web archiving
pilots projects
supported/furthered
by existing and
developing
cohorts of peers
focused on
cooperative collection
development
initiatives



Photo credit:
<http://www.concordia.ca/about/strategic-directions/james-neal.html>

What is a web archiving project librarian?

Roles and goals:

- **Web** (education/outreach about web collecting, coordinating & supporting selection of web resources to archive)
- **Archiving** (helping to prioritize, obtain and save content for immediate and long-term access)
- **Project** (project manager for many moving pieces and partners)
- **Librarian** (technical services librarian and liaison roles)

Project elements

- Incentives grants to advance web archiving tools
- Collaborative collection building through Ivy Plus / Borrow Direct
 - CCWA; Collaborative Architecture, Urbanism and Sustainability Web Archive (CAUSEWAY); and climate change curation experiment
- Outreach to site creators
- Best practices for site creators
- Citation analysis
 - Humans testing 2061 URLs from citations in scholarship on Human Rights published in 9 major journals in this field in 2010
 - Around 50% don't work or have major content drift/lead to cited content—now what?
 - Leveraging APIs to determine if cited content in the Human Right Web Archive and/or the Internet Archive; part of this process involved assistants looking for missing content on live web
- Interviews with scholars to enrich use case development

A brief introduction to web archiving

- Web archiving entails a multifaceted approach to preserve web-based materials (e.g. websites) and ensure ongoing access to collected content
- The main elements of web archiving are
 - Selection
 - Collection
 - Metadata assignment/cataloging
 - Quality assurance
 - Access
 - Long-term stewardship
- Columbia University Libraries Web Resources Collection Program
 - All of the above steps plus a policy to ask permission before collecting

Selection

From collection development policy to subject expertise

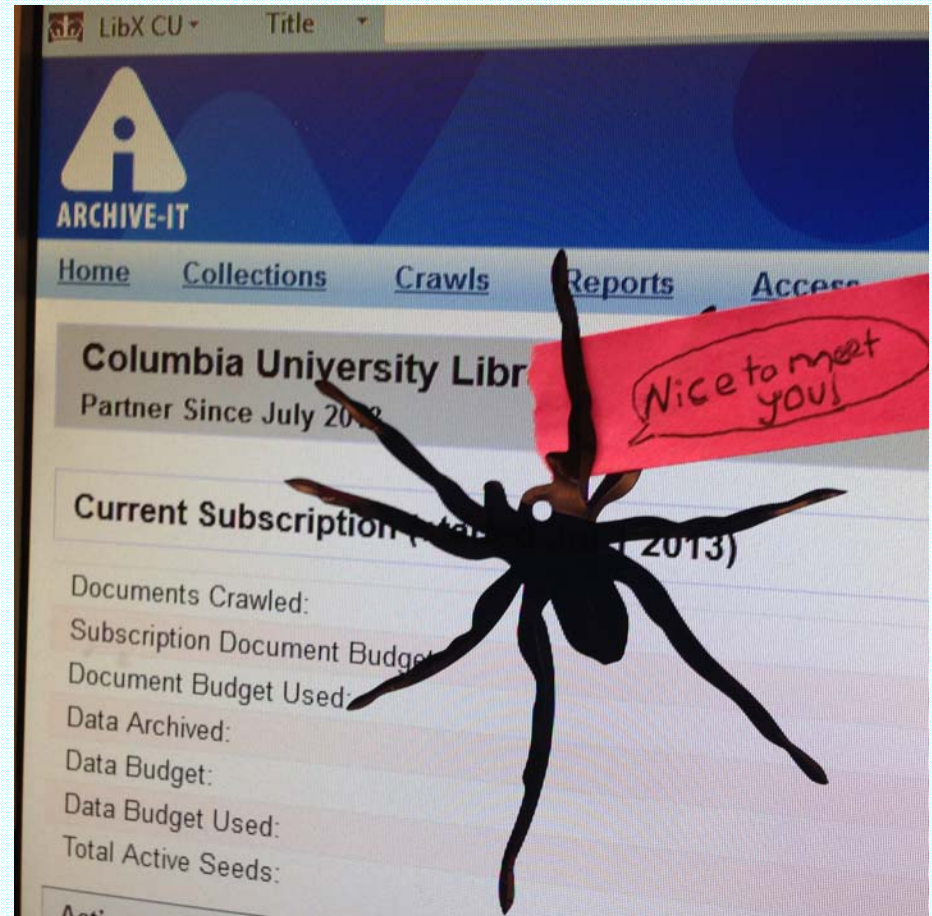
- First questions: what to collect, for whom and why?
- Collection development policy
- Defining themes and goals
 - For CCWA this was an extension of cooperative collecting of contemporary composers' scores
- Engaged selectors



Photo credit: Anna Perricci

Collection / harvest using software

- How do we get the content for our collections of archived websites?
 - We harvest sites using software, often known as a crawler, spider or robot
 - For collecting and access we use Archive-It, a subscription-based service of the Internet Archive



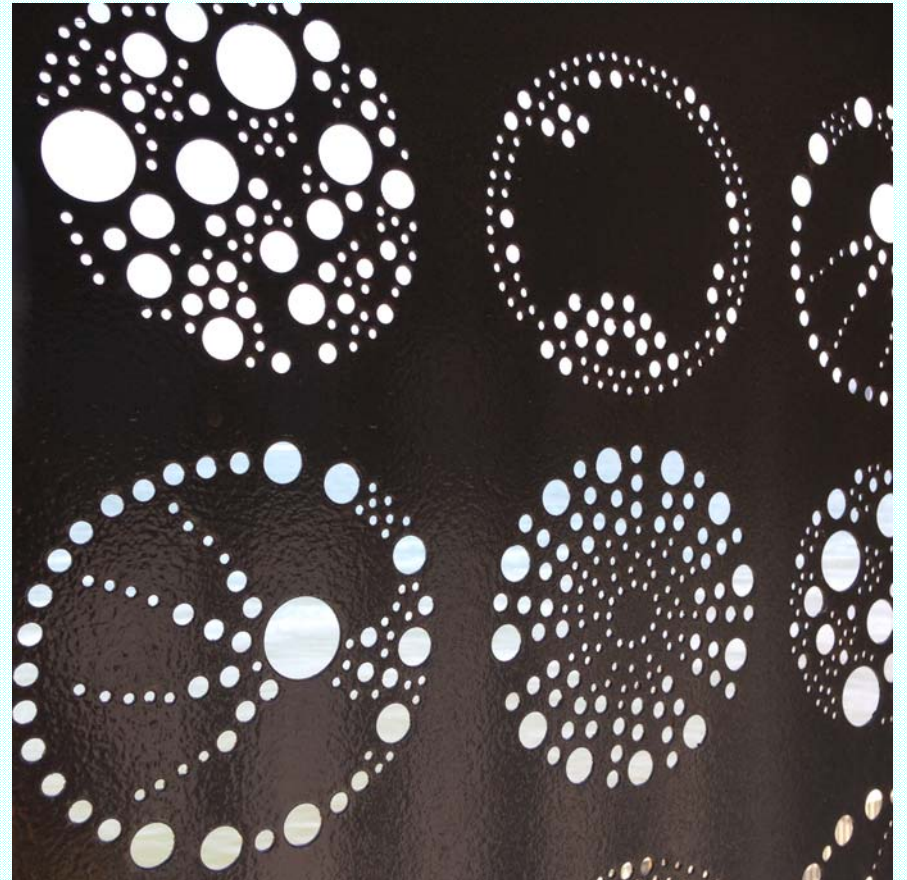
Technical elements of curation

- Fostering and managing collaboration is a multi-faceted pursuit
- Tools for collecting, track progress and manage communications
 - Archive-It
 - Google Drive
 - Basecamp
 - Email aliases & folder rules



Cataloging & Quality Assurance

- Cataloging / Metadata assignment essential to discoverability
 - Recognition of Russell Merritt, a skilled music cataloger at Columbia, who has cataloged most of the sites in CCWA and lead the effort to create a process for cataloging the archived websites of contemporary composers
- Quality assurance testing
 - Many thanks to music librarians who did QA testing last year!
 - Sometimes errors can be corrected through patch crawls/troubleshooting



Access

- Who will use what we have collected and how will they access it?
 - We need more use cases
 - We need to make web archives more accessible to get use cases
- Archive-It is an access system with limitations



This is a server at the Internet Archive!

When? Now! Plus long-term stewardship...

- We need to collect websites before they disappear but we also must ensure their long-term survival and maintain access to them over time
- So far we are saving websites in the WARC file format (the preservation standard) and temporarily relying on the Internet Archive to store the files until a repository framework can be established/chosen



Web archives in the media

- Jill Lepore, “The Cobweb: Can the Internet Be Archived?” *The New Yorker* (January 26, 2015)
 - Lepore got a lot of things right but there was a severely incorrect assertion that most everything on the internet is already in the Internet Archive (it’s not!)
 - There was no recognition of curated web archives and their tendency to be of higher quality/fidelity to the original resource
 - Archive-It provides a set of tools to focus and improve crawls



Advocating for curated collections

- Curated collections are
 - focused rather than haphazard
 - guided by a collection development policy
 - informed by skilled selectors
- Re-crawling sites at regular intervals can show patterns and maintain a consistent flow of information
- Because we ask permission, we ignore a file that blocks crawlers (so we get things that the Internet Archive otherwise would pass over as a matter of policy)




Discoverability & facilitating access

Most Visited News Popular Gmail Google Maps Dictionary / Wörter... Der digitale Grimm... Kurrent Alphabete ...

Home Search

Create lists, bibliographies and reviews: [Sign in](#) or [create a free account](#)


 **WorldCat®**

[Advanced Search](#) [Find a Library](#)

[<< Return to Search Results](#)

[Cite/Export](#) [Print](#) [E-mail](#) [Share](#) [Permalink](#)

[Add to list](#) [Add tags](#) [Write a review](#) Rate this item: ☆☆☆☆☆






Steven Stucky, composer.

Author: [Steven Stucky](#)

Publisher: (United States) : [publisher not identified] (2004-1

Series: [Contemporary Composers Web Archive \(CCWA\)](#)

Edition/Format:  Website : Document : Updating website  Continually Updated Resource  Computer File : English


Database: WorldCat

Summary: Composer website; includes biography; works; performances; recordings; writings; reviews; articles/interviews; information for the press; contacts (including publisher, personal representation, and public relations).

Rating: ☆☆☆☆☆ (not yet rated) [0 with reviews - Be the first.](#)

Subjects: [Stucky, Steven.](#)
[Stucky, Steven -- Bibliography.](#)
[Stucky, Steven -- Discography.](#)

More like this [Similar Items](#)

 **Nearby libraries**
to 10961

[freshpages MHMK](#)
[Bibliothek Berlin](#)
Berlin 10961, Germany
< 1 m / km

[Zentral- und Landesbibliothek Berlin,](#)
[Haus Amerika-](#)
[Gedenkbibliothek und](#)
[Haus Berliner](#)
[Stadtbibliothek sowie:](#)
[Senatsbibliothek Berlin](#)
Berlin 10961, Germany
< 1 m / km

Librarian? [Claim your library](#)

[Borrow / obtain a copy](#)

[Find a copy online](#)

CCWA in a library catalog

The screenshot shows a web browser window with the address bar displaying `clio.columbia.edu/catalog/11264180?counter=1`. The browser's address bar also shows the page title "Steven Stucky, composer - ...". The browser's toolbar includes a search bar, a home button, a star icon, a shield icon, a magnifying glass icon, a printer icon, a speech bubble icon, and a menu icon. The browser's tabs bar shows several tabs: "Most Visited", "Getting Started", "Latest Headlines", "Brown EARS", "OCRA - Online Co...", "Collection devel...", and "Scholarly Links". The browser's status bar shows the text "COLUMBIA UNIVERSITY LIBRARIES / INFORMATION SERVICES".

The main content area of the browser window displays the library catalog entry for "Steven Stucky, composer". The entry is titled "Steven Stucky, composer" and includes the following metadata:

- Title: Steven Stucky, composer.
- Published: [United States] : [publisher not identified], [2004-]
- Description: 1 online resource.
- Series: Contemporary Composers Web Archive (CCWA)
- Subjects: Stucky, Steven. Stucky, Steven > Bibliography. Stucky, Steven > Discography.
- Also Listed Under: Stucky, Steven.
- Summary: Composer website; includes biography; works; performances; recordings; writings; reviews; articles/interviews; information for the press; contacts (including publisher, personal representation, and public relations).
- Notes: Preserved in: Contemporary Composers Web Archive (CCWA).
- Language: English
- Format: Online
- Acquired On: February 24, 2015
- Bookmark As: <http://clio.columbia.edu/catalog/11264180>

The entry also includes a "Virtual Shelf Browse" section with buttons for "Show", "Hide", and "Full Screen".

On the right side of the entry, there is a section titled "Available from:" which lists the following options:

- Online
- ✓ Current site
- ✓ Archived site
- ✓ Full CCWA Archive-It collection (related resource)

Collaborative work

- Selection
- Automation
- Curation / Quality control
 - This brings up content questions still to be resolved!
- Challenges faced
 - Ownership and costs
 - Leadership and consensus, esp. as the project grows

Research implications

- Loss and destruction of information in the digital world
- Curation and canon-formation
- How to retain original order when materials are removed from a linked context?



Effect on research

Do digital resources and tools lead to fundamentally different questions, or do they provide the basis for new answers to existing questions that could not be answered without newly emerging digital tools?

Or:

Is web archiving philosophically different from other kinds of archiving?



Concluding remarks

- This pilot project has:
 - Effectively created a robust technical framework for curated web archiving
 - Demonstrated some of the advantages (and a few pitfalls) of a collaborative approach
- Current and future media coverage and scholarly research needs demonstrate why this work is critically necessary

Thank you!

Anna Perricci

Web Archiving Project Librarian,
Columbia University

anna.perricci@columbia.edu

Laura Stokes

Performing Arts Librarian, Brown
University

laura_stokes@brown.edu

