

**Web Archiving for Music History:  
The Archive of Contemporary Composers' Websites at New York University**

**by Kent Underwood & Robin Preiss**

**Archive-It Partners Meeting (Montgomery, Alabama 18 November 2014)**

**SLIDE 1: Title Page**

Good afternoon. Our presentation today is on a very twenty-first century topic: web archiving. But let us begin with some music from the eighteenth century: Bach's Cantata No. 26, "Ach wie flüchtig," composed and premiered in 1724.

**SLIDE 2: Bach**

**(BEGIN MUSIC, THEN LOWER THE VOLUME)**

The central theme of this cantata is vanity: the fragility, impermanence, and ultimate futility of human achievement. This condition is all too familiar to archivists and preservationists, whose work is a never-ending battle against the destructive forces of time.

**(STOP READING: RAISE VOLUME UNTIL MUSIC ENDS)**

We are able to enjoy this music today because Bach's manuscript has survived. That survival is the direct result of deliberate curatorial choices over the past three centuries. As we all know, Bach's music was not highly valued by most of his contemporaries when he died, and all of us here today can celebrate the curatorial wisdom that made possible the preservation of the manuscript, anticipating a time--generations later--when the worth of Bach's music would be better understood. And let us remember, too, that Bach was himself something of an archivist, who curated a unique and priceless collection of his own family's musical compositions that still survives today.

The advent of the World Wide Web has changed the very nature of historical documentation. In the field of music, many activities have moved decisively to the Internet--including promotion, publishing and distribution of scores and recordings, critical discourse, correspondence, and even performance itself. Websites have thus become an

essential component of cultural memory, and today's websites will inevitably become tomorrow's historical documents. Yet websites are by their very nature extremely unstable, fragile, and vulnerable to loss, so that anyone with an interest in the documentation of history should be concerned about the volume of significant content that disappears daily from the web and from the cultural record. For composers living today, we must think and act in new ways to make sure the primary documents of their lives and works are preserved for tomorrow, just as the documents of Bach's life and works have been preserved for us today. Conversely, in Bach's case, we know the documents are a far cry from being complete; thus we are also acutely aware of the potential consequences of curatorial neglect. Web archiving is one emerging answer to this challenge.

### **SLIDE 3: What is web archiving?**

What is web archiving? Web archiving makes static copies of a website at periodic intervals. It then saves those copies to a digital repository, thus preserving successive versions of the website that would otherwise be lost as the site evolves over time. The result is analogous to an album of photographic snapshots documenting the lifetime of a person.

### **SLIDE 4: Types of music websites**

For the documentation of music history, we can readily identify categories of websites that have potential research value as primary sources. The personal websites of individual composers are one very prominent type; there are many thousands of composers active around the world today, and our observation is that most of them have their own websites! Performing artists and ensembles, concert organizations and venues, recording and publishing companies—all of these have a very substantial and growing presence on the web, as do critics and everyday listeners who routinely comment on music through blogging and various social media platforms.

### **SLIDE 5: Lifespan of a web page**

How long does a web page last? According to the estimate published by the Internet Archive, the average lifespan is only 44-75 days. This is a shockingly short period of time compared with any previous form of written documentation, and it demonstrates that website preservation is not only important but urgent.

### **SLIDE 6: Bach**

Here is Bach again to warn us against complacency. In this passage from Movement IV of Cantata 26, fire and water metaphorically represent the destructive forces of time. Bach's word painting vividly depicts those images with surging waves of scales in the vocal and basso continuo lines, combined with falling figures in the oboes and bassoon to evoke collapse and disintegration.

### **PLAY MUSIC**

At the New York University Library we have embarked on a project to archive the personal websites of contemporary American composers. This web archive builds upon our library's long-standing commitment to contemporary music, and it serves the interests our students and faculty, who increasingly rely on Internet resources for teaching, learning, research, and scholarly communication. Furthermore, it fulfills our broader mandate as a research library to make a contribution to posterity, in accordance with libraries' traditional collective role as repositories of cultural memory and creative achievement.

### **SLIDE 7: Selection of Composers**

Our project's curatorial scope focuses on young, emerging American composers. This complements the Borrow Direct web archive (about which you will hear more in the next presentation), which concentrates mostly on the more established figures in the contemporary music world. Our first candidates were composers affiliated with New York University, and today we have expanded our circle to include over 300 individual artists. Their websites contain a wealth of historically important content, and if all goes well, our project will lay the foundation of an archival documentation strategy that will last for their entire careers.

### **SLIDE 8: Christopher Cerrone's Home Page**

Here is a screenshot of Christopher Cerrone's homepage. It really speaks for itself as to the importance of preserving composers' websites as primary historical documents, with its tabs for biography, music, calendar, media, press, and so on.

### **SLIDE 9: *Double Happiness* page from Cerrone's "Music" tab**

Here is an internal page from Christopher Cerrone's site: devoted to his composition, *Double Happiness*. It offers detailed information about the piece, along with an embedded video of a performance.

Incidentally, Christopher Cerrone is not offering his printed scores from this website, but many composers today are doing exactly that. Keeping ourselves current with composers who are self-publishing on the web is a challenge unto itself for the music library community, but that is a topic for another day.

### **SLIDE 10: Wayback Machine vs. Archive-It**

For our archiving project we have partnered with the Internet Archive, which is well known as one of the most ambitious web archiving enterprises in the world. The Internet Archive has two web archiving tools. The first, called the Wayback Machine, began operation in 2001, and since then it has archived over 435 billion web pages. The Wayback Machine is essentially a passive curation tool. Despite the massive size of the Wayback Machine's archive, the idea that it is automatically copying and backing up the entire World Wide Web is a misconception. In fact, the Wayback Machine's algorithm evaluates and prioritizes web pages automatically, based on the amount of traffic those pages encounter on the live Internet. Consequently, the Wayback Machine does not contain archival copies of every website in existence. Furthermore, for the sites it does include, the Wayback Machine's chronological distribution of historical copies is inconsistent, and the degree of completeness with which it captures internal sub-pages is also variable.

The Internet Archive's own solution to the limitations of the Wayback Machine's passive approach is a product called Archive-It. Archive-It is now in use at over 275 partner institutions in the United States and throughout the world. Archive-It is a tool for active curation. Instead of putting critical scoping decisions into the hands of a robot, it enables

archivists to perform their accustomed, traditional roles in defining, organizing, shaping, describing, and managing collections. It gives us the choice of which sites to capture; control over the frequency and duration of the web crawls; and control over the inclusion or exclusion of sub-domains, of embedded files (such as audio, video, and PDF), and of material linked from external websites. These flexible curatorial features made Archive-It the obvious choice of software for our project.

In the initial phase of solicitation we sent email surveys to 311 composers to request their participation in the project and to ask them a series of questions about the content and management of their websites. Here are some highlights from the questionnaire results:

**SLIDE 11.** Do you wish your site to be included?

- Although the response rate as a whole was just under half (a generic problem with online surveys!), 97% of those who did respond said yes to being included in the web archive.

**SLIDE 12.** Are you taking measures yourself to preserve your site?

- Only 31% said yes

**SLIDE 13.** Are you administering and updating your site yourself?

- 85% said yes

**SLIDE 14.** How often do you add material?

- The majority (68%) add material monthly or quarterly

**SLIDE 15.** How often do you delete material?

- The majority ranges between quarterly and annually
- 29% never deletes!

The architectural capstone of our project is a discovery portal that will interface with the NYU Libraries' local system to provide researchers with guided access to the archive's contents. This is still a work in progress and it not ready for public view yet.

Web archiving is still in its infancy as a technology, and we will touch now on some of its limitations and unsolved problems.

We have used the word "snapshot" to describe each constituent object in a web archive, but the reality is by no means as straightforward as that term "snapshot" implies. The most obvious complication is that a snapshot, to the extent that it is a frozen object, is incompatible with the very nature of the World Wide Web as a dynamic, interconnected system. Indeed, preserving the contextual relationships between your archived site and other sites to which it is linked is by far the most formidable inherent challenge to the enterprise of web archiving. Successful web archiving is about preserving both the content and the medium of the web. Achieving that goal will ultimately require us to maintain both the integrity of the digital files and the relationships among the elements within a web sphere.

**SLIDE 16: Web spheres**

The term "web sphere" was coined by scholars in the early 2000's to represent the innate quality of interconnectedness that defines the web. Steven Schneider and Kirsten Foot's definition of a web sphere is "an interactive community of content producers and consumers, manifesting in a collection of sites that are united by a common interest: a common theme, central activity or event." As curators, we have recognized the relevance of web sphere analysis to the building of our archive. Most of our sites contain hyperlinks to the pages of other artists and composers, revealing collaborative and synergistic

relationships. Pages such as Florent Ghys's cohort of "friends" (which we see in the slide here) have enabled us both to shape our selection process and to make at least some headway in preserving the identities and activities of web communities to which individual composers belong.

Our time allotment today permits only a glancing mention of the audio, video, and animated graphical elements that are so prevalent in music-related websites. Suffice it to say that these types of features often exceed the technical capabilities of web archiving software as it currently stands, and that we and the Internet Archive are actively investigating and developing solutions to those limitations.

### **SLIDE 17: Bach 3**

To conclude our presentation, let us return to Bach and the final chorale of Cantata No. 26:

#### **PLAY MUSIC**

Bach ends on a note of theological expectancy, yet we hope that he would permit us also to suggest a message of temporal optimism that is especially relevant to us in the archival and library professions. Time's destructive power is indeed relentless, but thanks forward-thinking curators who have made the right decisions since 1724, Bach's music lives on.

As it was in 1724, so it is today. For music archivists and for the composers of our own time, the Internet is part of our world. Today's websites will become tomorrow's historical documents, and archival websites must certainly be an integral part of tomorrow's libraries. But websites are fragile and impermanent, and they cannot endure as historical documents without active curatorial attention and intervention. We must act quickly to curate and preserve the memory of the Internet now, while we have the chance, so that researchers of tomorrow will have the opportunity to discover their own past. The decisions and actions that we take today in web archiving will be crucial in determining what our descendants know and understand about their musical history and culture.

**THE END**