

PERSONAL ARCHIVING: STRIKING A BALANCE TO REACH THE PUBLIC

EXTENDED ABSTRACT

William G. LeFurgy

Library of Congress
101 Independence Ave, SE
Washington, DC 20007

ABSTRACT

Never have so many people documented so much about their lives. Digital technology has empowered individuals to build large, rich collections of photographs, videos, e-mail, documents and other information. But the ability to create digital content is far outstripping personal capacity to manage and keep it over time. Looking ahead over the next decade, it is possible to foresee two consequences for libraries and archives. The most obvious and certain is that digital accessions of personal materials will supplement, and eventually surpass, traditional analog materials. Another outcome is more subtle and speculative: people seeking trusted guidance about how best to manage their important digital items. Memory organizations are reasonable places for people to go in search of such guidance, and this presents an opportunity to provide a valuable—and highly visible—public service. The Library of Congress National Digital Information Infrastructure and Preservation Program is undertaking a project to provide guidance aimed at the general public in connection with personal digital archiving. The project focuses on interacting with people through several different channels, including web-based written instructions, video productions, and social media. The Library is also exploring use of public events such as “Personal Archiving Day” to engage directly with people. In developing a strategy for this program, the Library has to balance professional practice with the need to clearly communicate with non-specialists in a Web 2.0 environment.

RISING TIDE OF PERSONAL DIGITAL INFORMATION

Residents of the developed world are generating an astonishing amount of personal digital information. Reliable figures are hard to come by, but the Twitter archive is estimated to consist of five terabytes¹; Flickr has and estimated 4.3 billion pictures and Facebook may have anywhere from 15-60 billion pictures.² Millions of digital cameras (and phones with cameras) are in circulation. Just one of many computer manufacturers expects to sell 25 million personal computers itself over the next year.³ An information technology market analysis claims that 70 percent—or about 880 petabytes-

-of the annual “digital universe” is generated by personal users.⁴ However the phenomenon is looked at, it is clear that 1) a typical consumer has accumulated a staggering quantity of data, and 2) the trend line is headed up, probably dramatically so.

At the same time, it appears that categories of newly generated analog personal information are in steep decline. Kodak, an iconographic analog stalwart, has struggled in recent years, and one business analyst recently forecast more trouble for the firm due to its “exposure to the secular decline in analog film.” The company has, since 2004, pinned its hopes on “digital photography services and printers and away from photographic film.”⁵ The U.S. Postal Service also faces economic challenges that stem in part from “digital alternatives such as electronic bill payment [and] e-mail document delivery.”⁶ Home movies on 8mm film and other analog formats have given way to digital video recorders.

It seems that many people are putting all their personal information eggs in a virtual digital basket. This is, of course, a risky strategy. A shoebox filled with photographic prints, letters, home movies and the like can easily last for years with minimal care and easily pass from one generation to the next. Personal digital content presents a whole new challenge. It is often scattered across a variety of websites, devices and storage media. Content is frequently disorganized and subject to spotty strategies for selection, replication and metadata. Commercial services are frequently used to store personal content despite the fact that such services are under no long-term obligation to keep or provide access to data. And even if a user keeps a copy of their data, digital media at this point are fundamentally non-archival.

CONSEQUENCES FOR MEMORY ORGANIZATIONS: EXTENDING CURATORIAL PRACTICE

Looking ahead over the next decade, it is possible to foresee two consequences in connection with personal digital information for libraries, archives and other

memory organizations. The most obvious and certain is that digital accessions of personal materials will supplement, and eventually surpass, personal analog materials. Salman Rushdie's donation of his personal papers recently received attention because they included "four Apple computers (one ruined by a spilled Coke) [and] 18 gigabytes of data."⁷ But there will come a day when such donations are routine, as the generation that first adapted to digital technology start offering the fruits of their labors to collecting institutions. This assumes, of course, that the fruits are preserved in the first place. There is plenty of room for concern that e-mail correspondence, for example, will disappear into the ether.⁸ But it would seem that in the future, where there are materials to donate, they will be increasingly digital.

Given this, it makes sense for memory organizations to develop a strategy for dealing with the situation. Actually, there might parallel strategies. One could focus on applying traditional curatorial approaches to prospective digital collections. While more work is needed, some attention has been devoted to this area in connection with ideas about Trusted Digital Repositories and other approaches. It does remain to be seen just how ready most institutions—even the biggest—are for adapting their practices to bringing in personal digital collections.

Another strategy needs to focus on first understanding how personal collections of scholarly interest are built and managed, and then exploring how to provide guidance, tools and services to help creators build viable personal digital archives. This constitutes a more proactive role prior to making institutional stewardship arrangements, and would seek to extend curatorial action or influence in a pre-custodial (or is it post-custodial?) manner. The Digital Lives Research Project at the British Library is a premier example of such an effort.⁹

CONSEQUENCES FOR MEMORY ORGANIZATIONS: PROVIDING BROAD-BASED GUIDANCE

The other big consequence looming as a result of the growth of personal digital holdings is that millions of people are going to need advice about how to save important parts of their collections and pass them on to family members or other interested parties. Memory organizations are reasonable places for people to go in search of such guidance, and this presents an opportunity to provide a valuable—and highly visible—public service.

This is a tricky business for memory organizations, however. Curators are by definition experts with deep and arcane knowledge. They focus on material that is of scholarly interest, which typically is a mere fraction of the larger information universe. To the extent that advances have been made in digital preservation and

curatorial, they are tightly bound with specialized curatorial and technical concerns relating to complex issues relating to authenticity, metadata, validation and verification and fixity. While strides have been made in generalizing and simplifying some digital preservation methods, much current practice remains institution specific and opaque to the average person.

There is a further complication to providing archiving guidance to non-specialists. Web 2.0 has brought about a hunger for information that is quickly found, read and understood. Many people have limited patience for carefully nuanced, specialized information sources. When writing for the web, authors are given advice to write text that is easily scanned, as opposed to read word by word. This means bullets and many fewer words than conventional writing.

If memory organizations want to meet the challenge of providing broad-based digital archiving guidance they need to identify the bare bones of good practice, seek out effective channels to present information and work to engage users in different ways. And while much of this work will take place through computer mediated methods, they is very much a place for meeting with people directly to exchange information.

THE LIBRARY OF CONGRESS AND PERSONAL DIGITAL ARCHIVING

The Library of Congress National Digital Information Infrastructure and Preservation Program is undertaking a project to provide guidance aimed at the general public in connection with personal digital archiving. The project focuses on interacting with people through several different channels, including web-based written instructions, video productions, and social media. The Library is also exploring use of public events such as "Personal Archiving Day" to engage directly with people. In developing a strategy for this program, the Library has to balance professional practice with the need to clearly communicate with non-specialists in a Web 2.0 environment.

In May 2010, NDIIPP posted new and expanded guidance for personal digital archiving on the program website.¹⁰ The information focus on six categories: digital photographs, digital video, digital audio, e-mail, personal digital documents and websites and social media. Each category is build on a basic structure of advice that is distilled from professional practice. Specific issues relating to a particular category of personal digital information are highlighted. The advice is described as "basic and is meant to be a place to get started"—it makes no claim to be one-stop shopping for everything that an individual needs to know for digital preservation.

At this stage, the NDIIPP personal digital archiving guidance is regarded as a “beta”: it should work as intended but future enhancements are expected, if not required. The program will conduct user testing and will seek comments about how to improve the guidance while adhering to its intentional Web 2.0 presentation. The current guidance framework is built on the following components:

- Identify the full scope of your collection
- Decide which parts of it you want to save
- Organize and describe what you selected
- Make copies and store them in different places

A fifth component, export selected items from individual programs and services, is used when discussing e-mail and other content with specific dependencies.

NDIIPP has also developed a Digital Preservation Video Series to convey information in a YouTube-friendly format.¹¹ The videos are, in fact, posted on YouTube as well.¹² They deal with an number of digital preservation issues and are meant to be engaging and informative. Individual videos are planned for each of the six content categories covered in the personal archiving guidance.

Also in May 2010, NDIIPP launched a Facebook page to engage with the public about digital preservation.¹³ The intent is to use the page as a separate channel for distributing information and most especially as a way to interact and with interested people and field specific questions and concerns. Despite the newness of the page, the Library is impressed with how rapidly people are “liking” it; expectations are that the page will play a key role in bringing NDIIPP to many more individuals than ever before.

Despite the pull of Web 2.0, the Library learned that place is still important—particularly a place for people to come and deal directly with curators and other experts about preservation practices. On May 10, 2010, the Library held its first “Personal Archiving Day” for the public.¹⁴ About 200 people came to the James Madison Building on Capitol Hill in Washington, DC, to attend the event. Library staff gave brief talks on steps people can take to save their digital—and non-digital—information, and staff also were available at content-specific tables to answer questions and talk about preservation issues. The event was held in conjunction with the American Library Association’s inaugural “Preservation Week.”¹⁵ Interest shown in the work of ALA and the Library in connection with digital preservation is a solid indication that the public is eager for trustworthy advice—especially advice that flows through the right channels.

¹ Library Journal

<http://www.libraryjournal.com/article/CA6726233.html>

²

<http://www.personalarchiving.com/2010/02/conference-notes/>

³<http://www.google.com/hostednews/ap/article/ALeqM5iEIzDh7j2jM9lXEuW7TFxLWKmrIQD9FKL1H80>

⁴<http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm>

⁵<http://www.reuters.com/article/idUSN2923844820100429>

⁶http://articles.sfgate.com/2010-04-19/news/20855490_1_postmaster-general-john-potter-postal-service-mail

⁷<http://www.nytimes.com/2010/03/16/books/16archive.html>

⁸<http://www.nytimes.com/2005/09/04/books/review/04DONADIO.html?pagewanted=all>

⁹ <http://www.bl.uk/digital-lives/>

¹⁰ <http://www.digitalpreservation.gov/you/>

¹¹ <http://www.digitalpreservation.gov/videos/>

¹² <http://www.youtube.com/user/LibraryOfCongress>

¹³ <http://www.facebook.com/#!/digitalpreservation>

¹⁴ <http://www.digitalpreservation.gov/news/events/presweek2010/index.html>

¹⁵ <http://www.ala.org/ala/mgrps/divs/alcts/confevents/pre/swk/index.cfm>