

Editorial: Reflections on What We Mean

by “Forever”

Marc Truitt

What do we mean when we tell people that we want or intend to preserve content or an object “forever”?

A couple of weeks ago, I attended the Fall Meeting of the Preservation and Archiving Special Interest Group (PASIG) in San Francisco. The group, generously sponsored by Sun Microsystems, is the brainchild of Art Pasquinelli of Sun and Michael Keller of Stanford.

First, a confession on my part. Since the University of Alberta (UA) was one of the founding members of PASIG, I had occasion to attend the first several PASIG meetings. In the beginning, there were just a handful of—perhaps fewer than ten—institutions represented. It seemed at the first couple of meetings, when the group was still finding its direction, that the content was slim, repetitious, and overly focused on Sun’s own solutions in the digital preservation and archiving (DPA) arena. Since we had other attendees ably representing UA, I stayed away from the following several meetings.

Well, PASIG has grown up. The attendee list for this meeting boasted nearly two hundred persons representing more than thirty institutions. Among the attendees were many of the leading lights in DPA and the profession generally. Institutions represented included several North American and European national libraries, as well as ARLs, memory institutions, and a host of companies and consultants offering a range of DPA solutions. Yes, PASIG has arrived, and we have Art, Mike, and Sun to thank for this.

If I have one real remaining complaint about PASIG, it’s that the group is still overly focused on Sun’s solutions. True, other vendors such as ExLibris and VTLS attended, but their solutions don’t compete; rather, they build on Sun’s offerings. And while Microsoft also was in attendance for the first time, its presentation focused not so much on DPA solutions—it has none—as on a raft of interesting and useful plug-ins whose purpose is to facilitate preservation of content created in Microsoft products such as Word, Excel, PowerPoint, etc. Other large vendors of DPA solutions—think IBM, for one—remain conspicuously absent.

It’s time for Sun to do the “right thing” and “open source” PASIG. If Sun wishes to continue to sponsor PASIG by lending administrative and organizational expertise, that would be great. Indeed, a leading but not controlling role in PASIG would be entirely consistent with the company’s new focus on support of open-source efforts such as MySQL, OpenOffice, and OpenSolaris.

So, what about the title of this editorial? When we talk of digital preservation, just how *long* are we thinking of preserving an object? Ask any twenty specialists in DPA, and chances are that you’ll get at least ten different answers. For some, the timeframe can be as short as five to twenty years. For others, it’s fifty or perhaps one hundred years. At PASIG, at least one presenter described an

organizational business model that envisions preserving content for *five hundred years*. And there are even some in our profession who glibly use what one might call “the DPA F-word,” although fortunately none of them seemed to be in attendance at this fall’s PASIG.

What does this mean in a very practical, nuts-and-bolts IT sense? Chris Wood of Sun gave a presentation at the 2008 PASIG Spring Meeting in which he estimated that the cost to supply *power and cooling alone* to maintain a petabyte (1,000 TB) of disk-based digital content for a mere ten years would easily exceed \$1 million.¹ Refining his figures downward somewhat, Wood noted a few months later at the following PASIG meeting that for a 1 TB drive, the five-year estimated power and cooling for 2008–12 could be estimated at approximately \$320, or \$640,000 per petabyte over ten years, still a considerable sum.²

Add to this the costs of migration—consider that a modern spinning disk is generally thought to have a useful lifespan of about five years, and tape may have two or three decades—and the need regular integrity-checking of digital content for “bit-rot,” and you have the stuff of a sustainability nightmare. These challenges don’t even include the messy question of preserving an object so that it is *usable* in a century or five. While we probably will be able to read Word and Excel files for the foreseeable future, there are already countless files created with now-defunct PC applications of the 1980s and 1990s; many are stored on all kinds of obsolete media and today are skating on the edge of inaccessibility.

Already we are seeing concern expressed at institutions with significant digital library and digitization commitments that curating, migrating, and ensuring the integrity and usability of growing petabytes of content over centuries may be unsustainable in both dollars and staff.³ Can we even imagine the possible maintenance burden for our descendants, say, 250 or 500 years from now?

In 2006, Alexander Stille observed that “one of the great ironies of the information age is that, while the late twentieth century will undoubtedly have recorded more data than any other period in history, it will also almost certainly have lost more information than any previous era.”⁴ How are we to deal with this? Can we meaningfully plan for the preservation of digital content over centuries given our poor track record over just the past few decades?

Perhaps we’re thinking too big when we speak of “forever.” Maybe we need to begin by conceptualizing and implementing on a more manageable scale. Or, to adopt a phrase that seemed to become the informal mantra of

Marc Truitt (marc.truitt@ualberta.ca) is Associate University Librarian, Bibliographic and Information Technology Services, University of Alberta Libraries, Edmonton, Alberta, Canada, and Editor of *ITAL*.

both this year's PASIG and the immediately preceding iPres meeting, "To get to forever you have to get to five years first."⁵

About this issue of *ITAL*

A few months ago, while she was still working at the University of Nevada Las Vegas, *ITAL*'s longtime managing editor, Judith Carter, shared with me the program for Discovery Mini-Conference that had just been held at UNLV. The presentations, originally cast as poster sessions, suggested a diverse and fascinating collection of insights deserving of wider attention. I suggested to Judith that she and her colleagues had the makings of a great *ITAL* theme issue, and I'm pleased that they accepted my invitation to rework the presentations into a form suitable for publication here. I hope that you will find the results of their work interesting—I certainly do. They've done a superb job!

Bravo to Judith and the presenters at the UNLV Discovery Mini-Conference!

Corrigenda

In our September issue, in an article by Kathleen Carlson, we inadvertently characterized Camtasia Studio as an open-source product. It is not. Camtasia Studio is published by TechSmith Corporation. You can find out more at the product website (<http://www.techsmith.com/camtasia.asp>).

Also, in the same article, we provided a URL to a Flash tutorial titled "How to Order an Article that ASU Does Not Own." Ms. Carlson has recently advised us that the tutorial in question is no longer available.

References and Notes

1. Chris Wood, "The Billion File Problem and Other Archive issues" (presentation, Spring Meeting of the Sun Preservation and Archiving Special Interest Group [PASIG], San Francisco, California, May 28, 2008), http://events-at-sun.com/pasig_spring/presentations/ChrisWood_MassiveArchive.pdf (accessed Oct. 22, 2009).

2. Chris Wood, "Archive and Preservation: Emerging Storage: Technologies & Trends" (presentation, Fall Meeting of

PASIG, Baltimore, Maryland, Nov. 19, 2008), http://events-at-sun.com/pasig_fall08/presentations/PASIG_Wood.pdf (accessed Oct. 22, 2009).

3. Consider, for example, the following extract from a recent posting to the Syslib-L electronic discussion list by the head of library systems at the University of North Carolina at Chapel Hill:

I'm exaggerating a little in my subject line, but it's been less than 4 years since we purchased our first large (5TB) storage array. We now have a raw 65TB online, and 84TB on order—although a considerable chunk of that 84 is going to replace storage that's going out of warranty/maintenance and is more cost effective to replace (Apple XRAIDs, for instance). In the end, though we'll net out with 100TB or thereabouts by the end of next year.

A great deal of this space is going to digitization projects—no surprise there. We have over 20TB now in our "digital archive," storage I consider dim, if not dark. We need a heck of a lot of space for staging backups, given [sic] how much we write to tape in a 24-hour period. Individual staff aren't abusing our lack of quotas—it's really almost all legitimate, project-driven work that's eating us up. What's scarier is that we're now talking seriously about moving from project-driven work to programmatic work: the latest large photographic archive we acquired is being scanned as part of the acquisition/processing workflow. We're looking at ways to prioritize the scanning of our manuscript collections. Donors increasingly expect to see their gifts online. And we're not even yet supporting an "institutional repository."

Will Owen, "0 to 60 in Three Years: Mass Storage Management," online posting, Dec. 8, 2008, Syslib-L@listserv.indiana.edu, <https://listserv.indiana.edu/cgi-bin/wa-iub.exe?A0=SYSLIB-L> (account required; accessed Oct. 22, 2009).

4. Alexander Stille, "Are we losing our memory? or, The Museum of Obsolete Technology," *Lost Magazine*, no. 3 (Feb. 2006), <http://www.lostmag.com/issue3/memory.php> (accessed Oct. 22, 2009). While Stille was referring in this quotation to both digital and nondigital materials, his comments are but part of a larger debate positing that the latter half of the twentieth century could well come to be known in the future as a "digital dark age" because of the vast quantity of at-risk digital content, recently estimated by one expert at some 369 exabytes (369 billion GB) worth of data. Physorg.com, "'Digital Dark Age' May Doom Some Data," <http://www.physorg.com/news144343006.html> (accessed Oct. 22, 2009).

5. Ed Summers, "IPRES, IIPC, PASIC Roundup/Brain-dump," online posting, Oct. 14, 2009, inkdroid, <http://inkdroid.org/journal/2009/10/14/ipres-iipc-pasig-roundupbrain-dump/> (accessed Oct. 22, 2009).