

Critical Competencies for Digital Curation:
Perspectives from 30 Years in the Trenches and on the Mountain Top

DigCCur2007 Conference
University of North Carolina
April 20,2007

Kenneth Thibodeau

Today I will talk not about what we are doing at NARA, but about what other people need to do in order for us to be successful. But first let me establish my bona fides for asserting the broad spectrum of perspectives stated in the title. I have spent considerable time in the trenches, starting with writing FORTRAN programs in the early 1970's; managing annual acquisitions of equipment such as memory typewriters and dedicated word processing machines in the late 70's, but piecemeal acquisitions typically for one or a few pieces of equipment; providing support for users of first generations of PCs in the early 80's; responding to a demanding judge and plaintiffs in the Armstrong v. EOP case in the early 1990's. And I have been on the heights. I had the privilege and pleasure of organizing NARA's Center for Electronic Records, starting in 1989, and of leading the development of the DoD standard for Records Management Applications in 1996. Currently, I'm in charge of building the biggest archives system in the world. Without going into details on the system, if you're not familiar with what we're doing, you can get an idea of the scope of the project from the fact that to date we have had \$130,000,000 in funding and the President has asked Congress to give us an additional \$58,000,000 in 2008. In addition to an archival system, as defined in the OAIS standard, the Electronic Records Archives is the basis for business process modernization for NARA's core functions for managing the lifecycle of all types of records in the federal government.

For us to successfully implement the system, we will need staff who are competent to use it and realize its value. Thinking of what it will take to educate practitioners who are capable of realizing the potential benefits technology is offering, I think that there is a fundamental dilemma we face in educating digital curators: how do you teach something you don't know? Worse, how do you teach something no one knows? I don't believe we are anywhere near complete or definitive answers as to how to preserve and provide sustained access to digital information assets over time. I've spent over 30 years managing, preserving and providing access to digital records. I have happily observed a large increase in last 10 years in the level of attention and in both the quantity and the caliber of information and insights it has produced and is producing. However, I have seen nothing which would lead me to change an opinion I expressed in the first public talk I gave as the Director of NARA's Center for Electronic Records in 1989. Giving a plenary talk at a conference on Advanced Computation for the Social Sciences, I asserted the need to recognize that we are facing a future characterized by continuing, open-ended change, change which to a significant degree is unpredictable.

The validity of that opinion, at least in the time frame around 1989, has been confirmed by experience. In 1988, I had been down in the trenches, planning local area networks at the National Institutes of Health. We were very concerned with internetworking, but not in connectivity to the Internet. Our biggest concern with connectivity was not with the Internet, but with how we would get the LANs we were planning for the 12 buildings occupied by units of the Office of the Director of NIH to talk to the LANs supporting the 22 bureaus, institutes, and divisions filling the rest of the forty some buildings on

the NIH campus. At that time the Internet simply wasn't significant. Consider just a few historical facts. The first dial-up connection to the Internet would not be made until 1991. Tim Berners-Lee wouldn't release the World Wide Web until 1993. Gopher hadn't even been invented yet. At NIH we ignored the Internet. Other people paid a lot of attention to it, devoting a lot of attention to planning for internetworking. This led to the development of the ISO Open System Interconnection (OSI) standard. It promised to define the path to expanded and improved digital communication. In 1988, the Department of Defense adopted the OSI standard for digital communications, qualifying TCP/IP as an interim solution. The whole federal government followed suit, establishing GOSIP, the Government OSI Protocol as its standard. The "interim" in which TCP/IP has served as the foundation has now lasted two decades and there's no evidence it will end.

I believe my 1989 opinion is still valid today. There is no reason to assume that change will stop or even slow down in the foreseeable future. We have to accept what we know about IT today may be outdated and irrelevant in a few years. We have to accept that what may be simple changes in IT, such as changes not in the characteristics or capabilities of IT, but merely in its quantities can have profound implications. In 1989, the world surpassed 100,000 Internet hosts, rapidly growing from 10,000 two years earlier. This year, 2007, there are over 430,000,000 million. A world with over 400,000,000 Internet hosts is very different than one with a hundred thousand. A world with the World Wide Web is very different than one which relies primarily on ftp to move digital information. A world with ready access to Google, MySpace, and YouTube is very different from one in which the miniscule fraction of people who accessed the Internet relied on Gopher and WAIS.

It's not unreasonable to assume that technological innovations across the next twenty years will be at least as novel and as significant as any we've seen in the last twenty. Some changes we can foresee. We can be sure the Internet will continue to grow and its character will continue to change. One basis for that prediction: today 40 % of people use the Internet as their primary source of news, but only 10% of the money invested in advertising goes to the internet. Advertising dollars will follow the customer, intensifying the kind of impacts we've already seen from advertising. The pace of technological change, even if it does not accelerate, has a compounding effect. Stratton Scavos, the Chairman and CEO of VeriSign reports that they need not to update or refresh hardware or software, but to re-architect their system at a frequency of 5-7 years.

I'm sure many of you have read the white paper, *The Expanding Digital Universe*, published by IDC last month. IDC reported that last year "the amount of digital information created, captured, and replicated was 161 exabytes or 161 billion gigabytes. This is about 3 million times the information in all the books ever written." IDC goes on to predict that "Between 2006 and 2010, the information added annually to the digital universe will increase more than six fold from 161 exabytes to 988 exabytes." The amount of new technical information in particular is estimated to double every 2 years. That means for a student starting a four-year degree half of what they learn in their first year of study will be outdated by their third year of study. Even if it were only 20% that became outmoded, it would have major implications for education.¹

According to former Secretary of Education, Richard Riley, the top 10 jobs that will be most in-demand in 2010 didn't exist in 2004. Thus, schools are currently preparing students for jobs that don't yet exist using technologies that haven't been invented in order to solve problems we can't even define.

How can you educate individuals to curate digital materials in this context? I would like to suggest that there are five critical competencies digital curators need to operate in this context. They are not all the competencies required. Nor are they necessarily more valuable than other competencies. But they are essential in the current and foreseeable environment.

1. Abstraction

Digital curators need to integrate the expectation of change into their basic approach. The volume of digital information is already too big and it's growing too fast; the problems of preserving and providing sustained access to digital assets in the global village are too complex; change is too fast and too widespread. What we know how to do will be rapidly outdated. We need to accept that there is not and never will be a permanent solution to the challenge of digital preservation; that trying to find any solution is a futile effort without the recognition that digital preservation is inseparable from sustained and ideally unfettered access, so that preservation solutions must be such that they enable future generations to take advantage of improvements in technologies for information discovery, delivery, and quality.

To do that, digital curators must be able to step back from any situation they face, whether it relates to the formats of the information assets they need to preserve, the relationships of preservers to producers, the tools used for curation, or the access demands they currently face, to step back and analyze problems and requirements in the abstract. The complexities we face necessitate an ability to articulate problems and solutions at multiple levels of abstraction. The goal is to be able to articulate approaches and strategies, which not only address known problems but to articulate approaches that are viable beyond immediate circumstances, across different or changing technologies, and in different organizations and communities. Such are capable of adapting and evolving in a future fraught with both the certainty of change and the uncertainty of its exact nature.

2. Application

Obviously, approaches and solutions articulated in the abstract need to be realized in action. While knowledge of a variety of fields such as archival science, library science, information science, computer science and so on is critical, in the end the work of archives, libraries, data centers and other organizations which curate information is a practicum. Curators must be able to apply their knowledge and skills in concrete situations. Thus, educators must lead students not to know digital curation, but to be digital curators.

3. Agility

To survive, hopefully to prosper in an ever changing world, digital curators must be agile. They must be able to address not only the problems of obsolescence, but also the two sided challenges of new technology. In appraisal, ingest and preservation, they have to be able to deal with formats and also with hardware and software they have never seen before, but they also have to be able both to recognize and capitalize on opportunities offered by new technologies. Most importantly, they will need be able to acquire and often to produce new knowledge and new skills. Professional digital curators should be able

to apply their knowledge and skill in a variety of different contexts, in interactions with various communities.

- Different Contexts

Personal contexts: According to the U.S. Department of Labor 1 out of 4 workers today is working for a company where they have been employed for less than one year. More than 1 out of 2 have worked for less than five years in their current organization. We have to educate digital curators so that they can apply their knowledge and skills across a mobile career path in different jobs and working for different employers.

Technical contexts: They need to be able to apply their knowledge and skills working with different technologies both within any time period and over the long term. They must be able to address short, medium and long term needs. They should be able to organize and carry out projects that span the system life cycle, but also to address needs at particular phases in it.

- Different Communities

Digital curators must be able to apply their knowledge and skills in different sectors: academic, government, for-profit and others. They must be able to work in different types of organizations including not only archives, libraries, and museums, but also information centers, data centers, and line operations.

They will need to participate both in networks of practice and in communities of practice both within the workplace and in virtual organizations that cross institutional, national and cultural boundaries. We should assume that many communities of practice will consist of specialists from a variety of disciplines. A basic requirement for digital curators is to be able to converse with, learn from and inform other types of experts.

The community of practice must be a learning community. Critical role of formal education is to prepare practitioners to be life-long learners, not simply accretion, expansion or deepening of knowledge or skills learned in schools, but even and especially to learn new skills to be able to adapt and replace previous knowledge.

4. Professional Expertise

Practitioners must not merely be able to put their knowledge to use. They must perform at the level of professional experts. Digital curation does not exist, and should not be practiced, for its own sake. While it may – and hopefully will – mature, we have not yet reached a state of knowledge and expertise about the nature, management, and preservation of digital information that we can separate digital curation as a distinct profession or independent area of practice. It should work towards objectives which are externally defined. It should pursue those objectives richly informed by the concepts, understanding and methods for curation of records, library materials, data sets, or museum collections imparted by the established disciplines which focus on those types of assets. Nor, at this stage could digital curation be crafted as a multidisciplinary amalgam of knowledge and skills drawing

elements from several established fields. To the contrary, the challenges posed by changing information technology and changes in how society uses the technology argue for substantial enhancement within the established disciplines. We need more explicit and precise knowledge of both the commonalities and the differences among digital data, records, publications, artifacts, and other digital materials, and more and more refined techniques and better understanding of their applicability, efficacy and limits when applied to these different classes of objects before we can educate students to practice digital curation without professional competency in archival science, library or information science, data management, museum studies, or some other relevant specialization.

For example, archival science is a substantial and profound body of knowledge, matured over centuries. If you try to practice in this field without a solid grounding in its foundation knowledge, you will face the consequences of ignorance. Consequences such as those characterized by engineers as “10% solutions;” that is, the kind of ‘solutions’ which are devised by people who only understand 10% of the problem. Beyond that, archivists, both practitioners and theorists, recognize that archival science needs to advance considerably in order to confront the challenges posed by electronic records successfully. Professional curation of electronic records demands professional competency to absorb and integrate new knowledge generated in archival science and to assess and correctly implement new methods, techniques, and practices.

Similar arguments can be advanced about the other disciplines that are relevant to the curation of information assets.

5. Cost and Finance

In an environment of open-ended change, working in a variety of positions, in different types of institutions, I believe digital curators must have solid understanding of costs and effective skills in both obtaining and managing finances. While costs and finances are not inherent or specific to digital curation, one aspect of the immaturity of digital curation is that professionals in this area cannot assume that the organizations for which they work or will work have established and effective methods for estimating and controlling costs and realizing benefits. As Shelby Sannett points out in an unpublished work, digital curators must be able to apply “methods to determine costs, select appropriate cost elements, use cost-benefit versus risk-benefit analysis, develop and analyze cost frameworks and models, and integrate this information with institutional decision-making strategies.” I would add the obvious. Rather than bemoaning lack of resources they should be aggressive and effective in seeking and managing finances.

ⁱ IDC. The Expanding Digital Universe. A Forecast of Worldwide Information Growth through 2010. March 2007. http://www.emc.com/about/destination/digital_universe/index.jsp?hpid=1