

## **Cologne Dialogue on Digital Humanities**

Controversy 6: “‘Digital curation’ or ‘digital preservation’ is a topic, which has originated within the world of digital libraries; recently it has been drawn closer and closer to the Digital Humanities. What is the balance between conceptual work and technology?” Your role would be unusually clear: Basically make an explicit statement, that work on preservation should be driven by conceptual, not technological issues.

### **Placing the Horse before the Cart: Conceptual and Technical Dimensions of Digital Curation**

**Author: Helen R. Tibbo**

Good afternoon. This is a bit of change of pace as I am from information and archival science rather than the digital humanities. Manfred asked me to take the position that work on digital preservation or curation should be driven by conceptual and not technical issues. In the spirit of this event I have prepared a 30 minute talk and a bit longer paper for publication but just for a moment of sanity, I cannot conceive of how one can develop and drive research, educational programs, or practical implementations with anything but conceptual issues as a starting point. All development, including technical development starts in the conceptual realm. I understand that commercial interests often develop products and then look for audiences or have products they try to repurpose and that in these cases the existing technology may drive the application or at least the sale, but even in these cases the technological design started at a conceptual level, even if imperfectly matched with the needs of a given user community.

Perhaps a slightly better question would be: In a discipline, sub-discipline, or whatever field of endeavor you specify where there are both practitioners and theoreticians, what is the proper balance between theory being built around practice and practice flowing from theory. This is a constant tension in professional domains such as information science, law, medicine, and archival studies. How does theory inform practice and practice inform theory? This balance keeps the field relevant to practitioners, vibrant for academics, and overall useful to society, but is at times troubling to purists on either side of the equation. Even in more academic endeavors including all of the humanities, there is a tension between what can be accomplished with existing technological tools if one is clever and what tools might be developed (if computer science had enough motivation, economic or otherwise) that flow from various disciplines’ theoretical constructs and perspectives.

I would argue that this is a healthy tension that should result in collaboration and broader world views, but at the same time it can be an uneasy relationship. Now, back to my assignment to discuss digital curation.

.....

Although machine readable files can be traced to the use of punched cards in the textile industry in the Eighteen Century and the US government recognized such materials as records as early as the Records Disposition Act of 1939 [1], digital preservation, [2] and more recently, the broader domain of digital curation, are still incipient fields of effort. The phrases "digital curation" and "digital preservation" are often used interchangeably, but they have different meanings. The term "digital preservation" refers to a "series of managed activities necessary to ensure continued access to digital materials for as long as necessary" [3]. According to Hirtle, the term digital preservation was first used in 1990 in conjunction with a Cornell University Library and Xerox research project to mean "using digital technologies to reformat analog media as part of the preservation process of those media." [2] He further notes that "the concept of digital preservation originally developed in libraries, not archives, as an aid to ongoing library analog preservation efforts" and that "it initially did not concern itself with the preservation of information that was 'born digital'." [2] The term "digital curation" dates to the Lord and MacDonald e-Science Curation Report in the UK upon which the JISC (Joint Information Systems Committee) call for proposals for a Digital Curation Centre was based. [4] For the purpose of this essay, I will focus on the development of digital curation and the optimized balance of conceptual and technical components within its study and practice. While different endeavors, such an analysis of digital curation can be instructive as a model for the digital humanities.

The terms "digital curation" and "data curation" have emerged since 2003 to represent more complex and dynamic undertakings than preservation alone. "Digital curation" involves "maintaining and adding value to digital research data throughout its lifecycle." [5] The active management of research data through curation "reduces threats to their long-term research value and mitigates the risk of digital obsolescence" while making the data more sharable. [5] Curation is essential to the long-term accessibility and understandability of resources in digital formats.

Curation of digital assets, whether cultural, educational, scientific, or economic, is one of the central challenges of the early 21<sup>st</sup> century. [6-33] The last two decades have witnessed the development of several research agendas, extensive progress toward robust repository architectures, [34-49] preservation tools and strategies, [50-69] and progress toward

trustworthy and sustainable digital curation. [70-75] Key projects provide a firm foundation for ongoing research and development. [76-91] Nine years after the “It’s About Time” report (NSF/LC,2003) and six years after “Mind the Gap,” (Waller & Sharp, 2006) the need for rigorous research, the dissemination of best practices and standards information, and the training and support of a growing cadre of digital curation professionals is more pressing than ever.

Does all this activity, including but not limited to, the creation of standards; the convening of numerous conferences including DigCCurr 2007 and 2009, i-PRES, the IS&T’s annual Archiving 20xx Conferences, Personal Digital Archiving (PDI); and the development of educational programs to prepare digital curators, qualify Digital Curation as a discipline? Do similar lists of activities qualify the Digital Humanities? Rather than contemplating where the digital curation star shines in the academic firmament or whether we should consider it a proper discipline or a sub-discipline or a sub-sub-discipline of library science or information science or archival studies, it is more important to stress that it is almost always a collaborative endeavor in practice and that curation research projects often involve a wide array of individuals with various skills, knowledge, and perspectives. Perhaps with both digital curation and digital humanities the fundamental characteristics of the arenas of discourse are more important than their names on the halls of academe. That is to say, the current conversation can be useful even if the old academic guard never anoints either Digital Curation or Digital Humanities as disciplines.

So, back to our original question of what drives research and practice in digital curation – conceptual issues or technical ones? Is technology the tail wagging the dog or is action flowing from theories and principles? Perhaps the term “digital curation” itself provides us with clues. The term naturally blends both technical (digital) and conceptual (curation) facets, but which is more essential to the enterprise? It is instructive to note that “digital” is a modifier while “curation” is the head word that carries most of the weight in the term. We are talking about a particular type of curation, that involving digital objects, but nevertheless, it is curation. Some have argued, including Cliff Lynch in his wrap-up session at the 2007 DigCCurr conference in Chapel Hill, that indeed the word “digital” was unnecessary and that those concerned with preserving digital assets should just use the more inclusive term “curation.” [x] This comment was perhaps for effect and to stress that in the not so distant future most content that librarians, archivists, and museum information professionals, along with their counterparts in the governmental and commercial sectors, will manage will be digital and thus the adjective will no longer be necessary.

As in any debate or dialogue situation, the positions that Manfred asked Henry and I to take are artificial for effect. Certainly both conceptual and technical components are not only important, but are essential and fundamental to any digital curation enterprise. One cannot simply sit around and contemplate digits and preserve them in the process; theory and abstractions and examination of values and missions are requisite but they can only get us so far. These elements, of necessity, should drive the development of technical approaches, tools, and hopefully one day, solutions, as computer science is chiefly about abstractions captured in code. Nevertheless, curation of digital content requires technical knowledge and skills in the mix; there must be doing as well as conceiving, theorizing, and motivating, but the technical components of the doing in many ways are the last steps in the process and arguably the easiest to accomplish.

Not only are we drawing artificial distinctions between conceptual and technical components of digital curation, I am not entirely sanguine that we are drawing the most essential and important distinctions when we place conceptual and technical components of digital curation in opposition. If we take “conceptual” to mean “theoretical,” “intangible” or “abstract” and we only look at these elements of the endeavor along with the low-level procedural and methodological components, involving digital technology at the bits and bytes level, we miss much in the domain that involves behavioral and higher-level functional elements.

What indeed comprises digital curation? To determine the components of the digital curation endeavor in theory and practice, it is instructive to look at its conception from the beginning. In December 1994, the Commission on Preservation and Access and the Research Libraries Group created the Task Force on Digital Archiving to investigate the means of ensuring “continued access indefinitely into the future of records stored in digital electronic form.” (CPA, 1996, iii) The Taskforce, co-chaired by John Garrett and Donald Waters, was composed of individuals drawn from academe, industry, museums, archives and libraries, publishers, scholarly societies and government including Dr. Gladney. The Task Force was charged to frame the key problems “to ensure continuing access to electronic digital records indefinitely into the future.” (CPA, 1996, iii). The Task Force’s report is considered to be the first substantial consideration of digital curation and what it takes to preserve digital content for the long-term. The Task Force envisioned a national system of digital archives that would take on various functions and would be “collectively responsible for the long-term accessibility of the nation’s social, economic, cultural and intellectual heritage instantiated in digital form” as well as a system of repository audit and certification. (iii) While the report includes extensive discussion concerning building national cyberinfrastructure

for such an undertaking and involved individuals from Bell Labs and IBM, the six of the nine CPA recommendations do not focus on technology. They are:

1. Design a coordinated network of digital repositories;
2. Secure funding for a competition “to advance digital archives, particularly with respect to removing legal and economic barriers to preservation”;
3. Sponsor a white paper on the legal and institutional foundations needed for the development of effective fail-safe mechanisms to support the aggressive rescue of endangered digital information;
4. Seek creative models from professional societies from a variety of disciplines concerning creating and financing digital archives of specific bodies of information;
5. Institute work on standards development necessary for audit and certification of digital archives; and
6. Coordinate digital preservation initiatives in the United States with similar efforts abroad.

The Task Force did include 3 technology recommendations:

1. Commission follow-on case studies of digital archiving to identify current best practices and to benchmark costs in the following areas: (a) design of systems that facilitate archiving at the creation stage; (b) storage of massive quantities of culturally valuable digital information; (c) requirements and standards for describing and managing digital information; and (d) migration paths for digital preservation of culturally valuable digital information.
2. Foster practical experiments or demonstration projects in the archival application of technologies and services, such as hardware and software emulation algorithms, transaction systems for property rights and authentication mechanisms, which promise to facilitate the preservation of the cultural record in digital form.
3. Engage in national policy efforts to design and develop the national information infrastructure to ensure that longevity of information is an explicit goal.

Looking back sixteen years we can see how the Task Force set the stage for much of what has developed in digital curation including the publishing of ISO Standard #16363 “Audit and Certification of Trustworthy Digital Repositories” in early 2012. We can also see a blend of emphasis on technical, legal, social, commercial, and behavioral issues. This is a

truly remarkable document and one which has provided a foundation for much subsequent research and development that balance technical and conceptual components. Here the development of technology is always in service to the larger goal of preserving digital content for future users. Technology is the means, not the goal in and of itself. It is not the driver but the tool.

To help us further determine what that balance point should be and whether technical or conceptual elements drive digital curation it is useful to examine organizations and efforts that provide instruction and guidance on long-term preservation information professionals. The Digital Curation Centre in the UK provides an apt starting point as it has largely been responsible for defining the field since its inception in 2004 and providing extensive guidance and education. Their Curation Lifecycle Model is perhaps the best known of many data/content lifecycle models. It encompasses an extensive array of activities and considerations from the conception of a digital object all the way through its active life and transfer into a repository to its potential reuse and regeneration into new digital products. This model involves many functions and activities, including conceptualizing digital objects and their lives; creating and receiving digital objects; appraising and selecting them; ingesting them into the repository performing some sort of preservation action on them (migrating or emulating perhaps); storing the objects; disseminating them; and possibly transforming them. It also specifies the creation and maintenance of representation information or metadata as well as preservation planning and community watch activities. While many of these functions have technical aspects, none of them can be accomplished by technology alone. Many require extensive human intervention and judgment. Given the role of the DCC in shaping the international digital curation endeavor for the past 8 years and given the centrality of the DCC Lifecycle Model one must conclude that the conceptual, in terms of preservation planning, content creator education, and workflow development, drive digital curation activities.

The IMLS-funded DigCCurr (Digital Curation Curriculum) projects at the SILS at UNC-Chapel Hill are building curriculum tools for master's, doctoral, and professional level education in digital curation. A core product from these efforts is the DigCCurr Matrix which builds from The DCC Lifecycle Model as well as the Australian Continuum Model (digital objects fluidly flow from one life stage to the next and back throughout their lifespan without rigid demarcation across stages). The DigCCurr Matrix has seven dimensions, most of which have little to do with technology. This Matrix is the underlying structure of the digital curation graduate programs at UNC SILS and the DigCCurr Professional Institute. The first dimension is "mandates, Values, and Principles." These are the "core reasons why the digital curation functions

and skills should be carried out and should serve as the basis for criteria to evaluate whether the digital curation activities have been carried out responsibly and appropriately.” The second dimension is “Functions and Skills.” This is "Know how," as opposed to the conceptual, attitudinal or declarative knowledge that dominates several of the other matrix dimensions” but even here technology is not the primary driver. The third dimension is “Professional, Disciplinary, Institutional, Organizational, or Cultural Context” Here we focus on students understanding “challenges, opportunities and characteristics of particular disciplines or institutions (e.g. social science data archive in a university, commercial collection of scanned page images, state archives, serving a population with specific cultural norms).” While technological components may impinge on this dimension, most of the drivers here are social. Dimension 4 is “Type of Resource” which can involve technological considerations as well as may indeed involve technical knowledge but that is only one type of knowledge students need before embarking on study of digital curation. Finally, Dimension 6 is “Transition Point in Information Continuum.” This is the “points of transition that span from pre-creation design and planning all the way to secondary use environments.” These points may have technical implications but that is not their primary thrust.

Not only does the DigCCurr Matrix underlie the digital curation graduate coursework, the new graduate Certificate in Digital Curation, and continuing education offerings at SILS, it is influencing digital curation education outside of Chapel Hill. [x] DigCurV or Digital Curator Vocational Education Europe is a project funded by the European Commission’s Leonardo da Vinci program to establish a curriculum framework for vocational training in digital curation is using the Matrix in part to design its Evaluation Framework. This framework is being created to help individuals and organizations design, provide or assess digital curation curricula.[x] The Society of American Archivist’s Digital Archives Specialist Certificate Program is also in part based on the DigCCurr Matrix. [x] Thus, a growing body of digital curation continuing education offerings and the framework that will influence future curricular development in Europe are placing their primary emphases on the conceptual, social, legal, and behavioral components of the field. This is not to say that there is no recognition of the importance of technological knowledge, but technology is seen as being in service to the users of information and not a driver in the equation.

One of the most successful and long-standing continuing education offerings in digital curation is the Digital Preservation Management (DPM) workshop. [x] Anne Kenney and Nancy McGovern initiated this workshop at the Cornell University Library in 2003. McGovern further developed this program and brought it to the Interuniversity Consortium for Political and Social Research (ICPSR) of the University of Michigan, in 2008, and this year to the

Massachusetts Institute of Technology (MIT) Libraries. The Digital Preservation Management Workshops “incorporate community standards and exemplars of good practice to provide practical guidance for developing effective digital preservation programs” [20]. The National Endowment for the Humanities (NEH) has partially funded the workshops. Kenney and McGovern have taken t-The DPM workshops around the world and University of London’s Computer Centre’s Digital Preservation Training Programme (DPTP) builds from the DPM workshop series.

One of the hallmarks of the DPM workshops is the Three-Legged Stool Model. This model holds that “A fully implemented and viable preservation program addresses organizational issues, technological concerns, and funding questions, balancing them like a three-legged stool.” [x] While technology is one of the legs in this model, it is only one out of three and every organization needs to have all three legs of the stool well developed in order for the stool to stand. In this model, Organizational Infrastructure includes “the policies, procedures, practices, people—the elements that any programmatic area needs in order to thrive, but specialized to address digital preservation requirements.” Technological Infrastructure consists of the requisite equipment, software, hardware, a secure environment, and skills to establish and maintain the digital preservation program. The Resources Framework component of the DPM model addresses the requisite startup, ongoing, and contingency funding to enable and sustain the digital preservation program. Thus, this highly influential professional education program provides a vision of digital preservation work that is driven by organizational capacities and institutional and user requirements.

In January 2007 representatives from the Digital Curation Center (U.K), DigitalPreservationEurope, NESTOR (Germany), and the Center for Research Libraries(CRL) (North America) convened at the Center for Research Libraries in Chicago to seek consensus on core criteria for digital preservation repositories. The attendees of this meeting identified what they believed were ten basic characteristics of digital preservation repositories. Note that technology is only mentioned in the last point. In this model the repository:

1. Commits to continuing maintenance of digital objects for identified community/communities.
2. Demonstrates organizational fitness (including financial, staffing, and processes) to fulfill its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.

7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
8. Fulfills requisite dissemination requirements.
9. Has a strategic program for preservation planning and action.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

Here we see technology relegated to even a smaller role in the digital curation enterprise.

We could continue with many more examples and analysis of several research agendas for digital preservation and curation but my time is limited so I will stop here. Both digital curation practitioners and researchers recognize that technology is but one component of successful curation. Practitioners and researchers both desire technological tools and the integration of such tools into efficient and effective workflows. But technology itself does not make for successful curation solutions. Technology is only one component in very complex social and behavioral systems that comprise long-term curation of digital assets.