Long-term digital preservation: a digital humanities topic?

Gladney, Henry M.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Long-Term Digital Preservation:
A Digital Humanities Topic?

Henry M. Gladney [*]

Abstract: *»Digitale Langzeitarchivierung: Ein Thema für die Digitalen Geisteswissenschaften?«*. We argue that the so-called Digital Humanities fail to meet conventional criteria to be an accredited field of study on a par with Literature, Chemistry, Computer Science, and Civil Engineering, or even a specialized professorial emphasis such as Ancient History or Nuclear Physics. The argument uses long-term digital preservation as an example to argue that Digital Humanities proponents' case for their research agenda does not merit financial support, emphasizing practical aspects over subjective theory.

Keywords: digital preservation, electronic documents.

> We are today as far into the electric age as the Elizabethans had advanced into the typographical … age. And we are experiencing the same confusions … which they had felt when living simultaneously in two contrasted forms of society and experience. [McLuhan]

> The exhaustion, the surfeit, the pressure of information have all been seen before. … This time it is different.[1] We are a half century further along and can begin to see how vast the scale and how strong the effects of connectedness. [Gleick]

> Formal academic recognition of digital work in the humanities remains problematic. Socially this has to do with the slow pace of institutional change. Intellectually it has to do with the poorly understood nature of non-verbal knowledge-bearing objects. Curatorially it raises the problem of how such knowledge-bearing objects are to be preserved for the long term. Culturally it runs afoul of the low status given to works of popular culture – multimedia, documentaries, interactive games, and [so on] – which tend to be dismissed as entertainment. The increasing number of digital humanities articles suggests … that serious attention is urgently needed for understanding and preserving digital objects.[2]

---

[*] Address all communications to: Henry M. Gladney: hgladney@gmail.com.

[1] "A considerable part of the gear and tackle of print media – now taken for granted, invisible as old wallpaper – evolved in direct response to the sense of information surfeit" (Gleick 2011, 411).

[2] Excerpted from <http://en.wikipedia.org/wiki/Digital_humanities; emphasis added>.

"Digital Humanities" (DH)[3] is the name chosen by an interest group that is promoting their activities for funding and for inclusion in university faculties. Digital document preservation is prominent among the topics proposed for investigation by this interest group.[4] For an upcoming workshop debate, Manfred Thaller asked me to present a case for denying the requested support, arguing that DH is not a worthy academic discipline by discussing research into long-term digital preservation (LDP), and requested this advance position paper.[5]

An outsider may be pardoned for murky understanding of what is meant by 'the Digital Humanities'. Even insiders are struggling with fuzzy boundaries, as might be expected of any new activity. For instance, the following excerpt[6] typifies web-accessible comments.

> Our definitions are often a little muddy. (Melissa Terras, in a keynote presentation at [the 2010] DH conference,[7] called the community to task for hemming and hawing: "It's ... kinda the intersection of...") We need to get better at this! …

> CUNY's DH Initiative has published a beginner's Resource Guide to the Digital Humanities, which includes links [to] definitions and pages [about] sample projects, basic readings, and "hot topics" in DH, …

> Patrick Svensson has a solid piece in DH Quarterly called The Landscape of Digital Humanities.

> A post by a UVa graduate student, Chris Forster, attempted to define DH … [as having] four areas of activity – (i) use of computational methods for research; (ii) new media studies; (iii) how technology reshapes the humanities classroom; and (iv) how it reshapes scholarly communication and academic roles.

A recent conference call asserts simply, "DH is the nexus of computing and the humanities".[8] And the content of Borgman (2007) suggests that much of what DHP describes is covered by Information Science faculties.

---

[3] Abbreviations used in the text might depend on the context, as follows: DH: "the Digital Humanities" or else "Digital Humanities"; a.k.a. "e-Humanities"; DHP: "DH proponents" or else "a typical DH proponent (*D*avid *H*oward *P*otter)"; DL: "digital library" or "digital libraries"; LDP: "long-term digital preservation"; SE: "science and engineering", as represented in university faculties; SWE: "software engineering" or else "a typical software engineer (*S*amuel *W*illiam *E*ast)".

[4] In fact, digital preservation is the *only* specific DH research topic I found in recent Digital Humanities Quarterly articles.

[5] This draft responds to an invitation to participate in an April 2012 debate: *The Cologne Dialogue on Digital Humanities*.

[6] <http://digitalhumanities.org/answers/topic/what-is-digital-humanities>.

[7] <http://dh2010.cch.kcl.ac.uk/>.

[8] <http://news.stanford.edu/news/2011/june/digital-humanities-conference-060911.html. See also (Anon) at <http://shapeofthings.org/resources.html>.

Before proceeding further, we should compare the following definition and description of Information Science (IS),[9] a collection of topics that has been recognized as an academic discipline for approximately thirty years, i.e., much earlier than any mention of DH!

Information science (or *information studies*) is an interdisciplinary field primarily concerned with the analysis, collection, classification, manipulation, storage, retrieval and dissemination of information. Practitioners within the field study the application and usage of knowledge in organizations, along with the interaction between people, organizations and any existing information systems, with the aim of creating, replacing, improving or understanding information systems. Information science is often (mistakenly) considered a branch of computer science. However, it is actually a broad, interdisciplinary field, incorporating not only aspects of computer science, but often diverse fields such as archival science, cognitive science, commerce, communications, law, library science, museology, management, mathematics, philosophy, public policy, and the social sciences.

Information science focuses on understanding problems from the perspective of the stakeholders involved and then applying information and other technologies as needed. In other words, it tackles systemic problems first rather than individual pieces of technology within that system. In this respect, information science can be seen as a response to …, the belief that technology "develops by its own laws, that it realizes its own potential, limited only by the material resources available, and must therefore be regarded as an autonomous system controlling and ultimately permeating all other subsystems of society." Within information science, attention has been given in recent years to human-computer interaction, groupware, the semantic web, value sensitive design, iterative design processes and to the ways people generate, use and find information. Today this field is called the Field of Information, and there are a growing number of Schools and Colleges of Information.

Comparison of the definitions of DH and IS suggests that DH is an unneeded invention!

Any scholarly group may reasonably name its shared topics however it pleases, provided only that the chosen name does not mislead. So we have little reason to challenge the naming. The substantial issue instead is whether or not DH deserves to be ranked together with long-established university faculties such as History or sub-faculties such as Analytical Chemistry. Or perhaps, instead of judging what is deserved, we should consider whether it will attract respect from the established faculties, and also funding that it seeks from government institutions, such as the U.S. National Endowment for the Humanities.[10]

---

[9] Adapted from <http://en.wikipedia.org/wiki/Information_science>.
[10] Funding issues are made more important than they might otherwise be by current cutbacks that threaten established university faculties (Economist 2011), (Underwood 2001). This circumstance makes it appropriate to ask each DH funding applicant questions along the following lines.

> NEH supports … training programs for scholars … to extend their knowledge of digital humanities. … NEH seeks to increase the number of humanities scholars using digital technology in their research and to disseminate knowledge about [relevant] advanced technology … and methodologies.

> Today, complex data – its form, manipulation, and interpretation – are as important to humanities study as more traditional research materials. … digitized historical records … [and] multimedia collections … are increasing in number due to the … affordability of mass data storage devices, … extensive networking capabilities, and sophisticated [software] … improving interactive access to and analysis of these data … The Advanced Topics in the Digital Humanities program seeks to enable humanities scholars … to incorporate [such] advances into their scholarship and teaching.[11]

To judge the merits, we should consider several DH activities: instruction, proposed research, tools development, and analysis of social behavior. The current article examines only technical aspects, leaving other aspects to other commentators. It emphasizes objective over subjective aspects because, whenever doing so is sensible, these tend more rapidly towards debate closure.

> When Coleridge tried to define beauty, he returned always to one deep thought: beauty, he said, is 'unity in variety.' Science is nothing else than the search to discover unity in the wild variety of nature – or more exactly, in the variety of our experience. Poetry, painting, the arts are the same search, in Coleridge's phrase, for unity in variety.[12] (Bronowski 1965, 16)

What is it that computer scientists and software engineers do? Their projects begin (logically) with abstraction.[13] However, "Some people ... think that the current abstractions of Computer Science ... [and] algorithms handling [them] need to be adapted to fit the requirements of the Humanities."[14] To react to such an assertion, we need specific descriptions of the adaptations they call for – descriptions seemingly not yet available. With these in hand, we would surely ask, "What skills are needed to provide what's called for? Should we find an e-Humanist for such work, or should we find a software engineer?"

Imagine a debate between a prototypical digital humanist, DHP, and a software engineer, SWE – a debate in which SWE responds to some vague DHP assertion by asking for specific, relatively objective examples. DHP might respond in some way that does not satisfy SWE, leading him to request more specificity/objectivity. If this process continues for several rounds, DHP might respond angrily along the lines of, "Dr. SWE, your background seems insufficient for you to understand!"

---

[11] Extract from <http://www.neh.gov/grants/guidelines/IATDH.html>.

[12] Coleridge traced [this definition] back to Pythagorus: "The safest definition … of Beauty, as well as the oldest, is that of Pythagorus: The reduction of many to one" (Bronowski 1965, 22).

[13] Bronowski (1965) p.11 ff. provides an eloquent characterization of abstraction and its social role.

[14] From Thaller's notes with his workshop announcement.

How might SWE respond? It seems likely that he might say (or think, even if he is too polite to say), "Well, since you seem unable to explain it for students, you are not qualified for a DH professorship!"

A likely outcome of such debate is that, while the e-Humanist community considers such questions, perhaps even writing articles about them, software engineers will provide responsive tools – ones that even address human factors not even identified. And these engineers are likely to finish and deploy their work earlier than the e-Humanists reach consensus about their opinions!

This is likely because any objective specification of what's wanted is surprisingly close to specification of satisfying software. And turning specifications into implementations is what software engineers do!

An Example: Long-Term Digital Preservation

A 2012-Jan-9 invitation included a conference description[15] asserting: "Preserving digital artefacts is a global challenge, which has not been solved conclusively as yet." Burgess and Hamming (2011) elaborate as follows:

> Institutional interest in exploring the possibilities for digital scholarship, after an initial flurry of activity followed by something of a hiatus, seems to be gaining impetus again. We have recently seen the establishment of new granting initiatives … as well as a general "buzz" about digital scholarship epitomized by articles in the Chronicle of Higher Education and elsewhere, culminating in standing room only panels on digital humanities at the MLA conferences … Innovative work … is gaining ground among a growing cohort of digital scholars.
>
> …
>
> Scholars in the digital humanities are now starting to explore … technical and rhetorical problems of … preserving "born digital" creative works … But what about "born digital" scholarship … that never had a print analog? Very few theorists have attended to this category … The work of new media researchers in the humanities tends to get lumped into a single category rather than … distinct categories of scholarship rendered in new media and scholarship about new media. Institutionally, this distinction is crucial for upcoming scholars, since much of the contention centers around originality of content: if the multimedia format of the work is essential to … the argument it presents, where should it count – as a work of scholarship … or as a reworking of an existing argument? Thus it is important to distinguish … between 'scholarly multimedia' and other terms frequently used …. By scholarly multimedia we specifically mean critical scholarly works – interpretive and argumentative, as opposed to creative or archival – that are produced, and [perhaps] performed, in multimedia form. These works represent a new rhetorical genre of scholarship … that differs from multimedia art or hypertext fiction …

---

[15] See <http://computerspielemuseum.de/documents_public/Veranstaltungen/KEEP_Emulation _Expert_workshop_Berlin.pdf>.

Such excerpts suggest questions that, as far as I know, have not been adequately answered in any professional publication.[16]

1) What criteria must be satisfied for a digital preservation method to be judged a solution in principle?[17]

2) Over and above an answer to (1), what criteria must be satisfied for a digital preservation method to be judged a practical solution?

3) Over and above answers to (1) and (2), what criteria must be satisfied for world-wide digital preservation practice to be judged socially satisfactory?

Epistemological Bases

An article that I no longer can identify referred to a "technical hard core to preservation, rather than just librarianship". Such phraseology suggests the importance of named topics being clearly identified and overlapping minimally. Terms such as 'digital library' and 'archiving' had been used for two decades before anybody mentioned 'digital preservation'. The current article, therefore, limits 'digital preservation' to extensions beyond digital document management suggested by Gladney (1993).

Figure 1: Human and Machine Roles in Sharing Documents
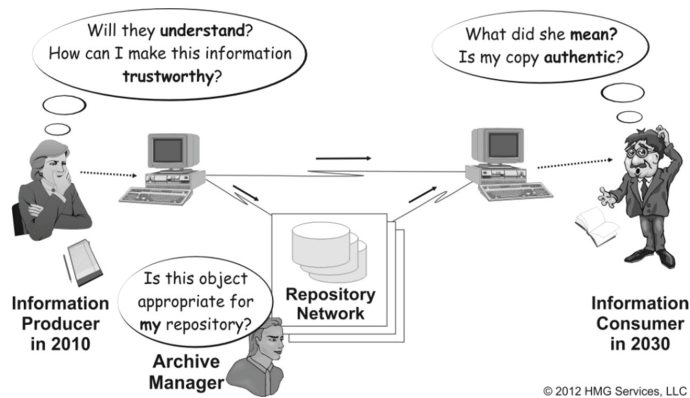
simplification of Gladney (2009) (Figure 1); see also (OAIS).

---

[16] Anybody who disagrees with "not adequately answered" is invited to cite contradictory articles.

[17] A 'solution in principle' is a methodological prescription that, were it to be implemented by software engineers and repository managers, would be adequate. A 'practical solution' is an implementation that pilot installations have demonstrated to be satisfactory. 'Socially satisfactory' calls for managed infrastructure (perhaps within a digital repository network) that satisfies anybody who wants some particular information to endure for some specified period. (S)he would be satisfied if (s)he deemed reliable institutional promises for the service alluded to, and if fees for such service were reasonable.

Just how important and useful this tactic is can be seen by considering difficulties in Burgess and Hamming (2011). Many of these simply disappear if one partitions communication processes into steps and intermediate message representations describing how an information bundle moves from the mind and space of its author to those of its eventual recipient(s).

Part of what makes for clear analytical description is explicit attention to distinctions taught by 20th-century epistemology (Coffa 1993). Compare the style of Bootz, Szoniecky and Bargaoui (2009) to analyses of communication steps hidden behind what (Figure 1) suggests. Bootz et al. make no use of helpful basic distinctions:

- Between objects and values: in most information preservation, what is to be preserved is some pattern (a value) inherent in one or more representations, each embodied in an object that can be transmitted (Nimmer 1998). Multiple representations can reduce (without eliminating) ambiguity between which information is essential and which is accidental.
- Between accidental and essential information, an obviously subjective distinction. For instance, a poet might or might not intend page layout to be important. Although common conventions emphasize artists' intentions, sometimes observers' intentions dominate a discussion, such as when an observer is trying to achieve something practical, as might occur in deciding whether a painting is indeed from the purported artist.
- Between analog and digital information representations and, for the former, questions of precision and noise. Digital information can be transmitted without any error whatsoever. In contrast, moving information between human beings and human beings usually has steps with analog signals and therefore cannot avoid distortions and subjective decisions about what is good enough.

What should a digital preservation solution accomplish? As a minimum, it should:
- Ensure that a copy of every preserved document survives as long as it might interest somebody;
- Ensure that authorized consumers can find and use any preserved document as its producers intended, avoiding errors introduced by third parties that include archivists and editors;
- Ensure that any consumer can reliably decide whether information received is sufficiently trustworthy for his intended application;
- Hide technical complexity from end users; and
- Replace human effort by automatic procedures whenever feasible.

Conceptual Difficulties

> Digital data … is analogous to infrastructure in the physical world … And like physical infrastructure, we want our data infrastructure to be stable, predictable, cost-effective, and sustainable. Creating systems with these and other critical characteristics … involves tackling a spectrum of technical, policy, economic, research, education, and social issues. The management, organization, access, and preservation of digital data is arguably a "grand challenge" of the information age. (Berman)

> Published difficulties of long-term digital preservation prove to be largely confusions with language. Similar difficulties were addressed in early twentieth-century philosophy. We describe prominent confusions, show how to clarify the issues, and summarize a method that solves all the technical challenges described in the literature. Other reports provide detailed design and analysis of the [proposed] TDO method. A purpose of the current article is to invite searching public criticism before anyone invests significant resources in creating preservation data objects. (Gladney 2006)

Before addressing technology, we need to understand what people mean by 'document preservation', or at least achieve clarity about different concepts used by different communities. Such concepts can be independent of the document media, i.e., the same for documents on paper, audio and video recordings on magnetic media and vinyl platters,[18] and for digital objects that are shared.

Early digital archive literature is full of misunderstandings of basic concepts. For instance, articles about 'Trusted Digital Repositories' betray problems that call their direction into question. Confusion between 'trusted' and 'trustworthy' misled investigators into focusing on repositories rather than on content objects.[19] For instance, Beagrie et al (2002) call for certification that an institution has correctly executed sound preservation practices.

Repository-centric proposals have unavoidable weaknesses:

- They depend on an unexpressed premise – that exposing an archive's procedures can persuade its clients that its content deliveries will be authentic. Such procedures have not yet been described, much less justified as achieving what their proponents seem to assume.
- Audits of an archive – no matter how frequent these are – cannot demonstrate that its contents have not been improperly altered years before a sensitive document is accessed.

---

[18] What is intended here are analog recordings such as those of the first half of the 20th century.

[19] We know how to make information trustworthy for specified applications, but do not know how to ensure that information deliveries are trusted by eventual recipients.

In a century or so, nobody will care about the capabilities and weaknesses of today's repositories. Instead, what people will want to know whether digital content they can fetch is credibly authentic.

In casual conversation, we often say that the copy of a recording is authentic if it closely resembles the original. But consider, for example, an orchestral performance, with sound reflected from walls entering imperfect microphones, signal changes in electronic circuits, and so on, until we finally hear the soundtrack of a television rendering. Which of many different signal versions is 'the original'?

Difficulties with 'original' and 'authentic' are conceptual. Nobody creates an artifact in an indivisible act. What is an acceptable original is somebody's subjective choice. When such an original has been chosen, we can describe it objectively with provenance metadata expressing everything important about the creation event. We can then judge authenticity relative to that version, and be understood.

Conventional definitions, such as "authentic: of undisputed origin; genuine." (Concise Oxford English Dictionary), do not help much. For signals, for material artifacts, and even for natural entities, the following definition captures what people mean when they say 'authentic'.

Given a derivation statement R,   "V is a copy of Y $(V=C(Y))$",
     a provenance statement S,   "X said or created Y as part of event Z", and
     a copy function,   "$C(y) = T_n (…(T_2(T_1(y)))),$"
we say that V is a *derivative* of Y if V is related to Y according to R.

We say that "by X as part of Z" is a *true provenance* of V if R and S are true.
We say that V is *sufficiently faithful* to Y if C conforms to social conventions for the genre and for the circumstances at hand.

We say that V is an *authentic copy* of Y if it is a *sufficiently faithful derivative* with *true provenance.*

Here 'copy' means either "later instance" or "conforming to a specific conceptual object". Each tk represents a transformation that is part of a figure 1 transmission step and that potentially alters the information carried. To preserve authenticity, the metadata accompanying the input in each transmission step should be extended by a tk description. These metadata should identify who made each tk choice and all other aspects important to consumers' judgments of authenticity.

> … reflecting on the challenge … for ensuring the reliability and authenticity of records that lack a stable form and content. The ease with which [dynamic documents] can be manipulated has given … a new reason for keeping them: 'repurposing'. … We have to consider the possibility of substituting the characteristics of completeness, stability and fixity with the capacity of the [repositories] to trace and preserve each change the record has undergone. And perhaps we may look at the record as existing in one of two modes, as an entity in becoming … and as a fixed entity at any given time the record is used. …

strategies must be developed … for both the creators and preservers …
(Duranti 2004)

We disagree! Neither our careful definition of 'authenticity' nor any other work suggests that 'dynamic documents' (representations of artistic and other performances) present a new or difficult preservation problem. What is different for different object kinds is merely the ease and frequency of change and of copying.

A repeat of an earlier performance would be called authentic if it were a faithful copy except for a constant time-shift. This can describe any kind of performance. Its meaning is simpler for digital documents than for analog recordings or live performances because digital files already reflect the sampling errors of recording performances that are continuous in time.

The authors expressing difficulty with dynamic digital objects do not express similar uncertainty about analog recordings of music or television performances. Perhaps their confusion is misunderstanding of language, as suggested by Wittgenstein (1921, 4.003).

> The "digital curation" concept is still evolving. [Lee] defines it as follows: Digital curation involves selection and appraisal by creators and archivists; evolving provision of intellectual access; redundant storage; data transformations; and, for some materials, a commitment to long-term preservation. Digital curation is stewardship that provides for the reproducibility and re-use of authentic digital data and other digital assets. Development of trustworthy and durable digital repositories; principles of sound metadata creation and capture; use of open standards for file formats and data encoding; and the promotion of information management literacy are all essential to the longevity of digital resources and the success of curation efforts.

> Digital preservation is typically regarded as a key subset of digital curation. (Bailey 2008)

The social challenge and the essence of its solution are conceptually simple. Without careful management, recorded information gradually would become inaccessible (Rosenthal et al 2005). Impediments include changing language. For works on paper, it might take centuries before readers are no longer comfortable with the language used.[20] For digital documents, this period is today much shorter, partly because rendering technology is still changing rapidly and partly because usability expectations are higher than for information on paper.[21]

---

[20] This need not be because a book is written in Latin. It can also be because key expressions, idioms, and metaphors are no longer commonly understood as their authors intended.

[21] The most sensitive examples are computer programs, for which a single changed bit might impede use.

Both the social and technical structure of any LDP solution should parallel that for documents on paper. The only exceptions should address aspects for which we can identify reasons for deviation.[22]

The traditional roles of repositories include acquiring, saving (including redundant copies), and sharing "interesting" content objects. They include editing that content and associated metadata only if available sources cannot make satisfactory copies available. This occurs for no more than a tiny fraction of the worthwhile literature. Instead, editing and describing documents and records are traditional responsibilities of outside communities, such as those of authors, editors, and publishers.

Surprisingly, librarians, archivists, and their faculty colleagues do not seem to see it this way.[23] Many of their published articles propose work on or methods for preserving digital content by extending the role of repository institutions, and prominent members of the DH community call LDP a "grand challenge" (Lee and Tibbo 2007; Berman 2008).

We disagree. An LDP solution will not be a prescription for repository management, but instead a method for making digital objects durably useful, readily sharable, and durably trustworthy – a scheme for representing content. The next section sketches one such scheme.

An unsolved challenge is caused by immense increase in the number of books, papers, periodicals, memorabilia, technical data (Berriman and Groom 2011), and other digital objects published.[24] The fraction of this flood meeting any dispassionate quality criterion has probably decreased, so that what one needs to read to be well-informed has not grown nearly as quickly.[25] And information technologists have provided, and are refining, tools that make finding the answer to any well-formulated question – if that question has in fact been addressed – much easier than it was either a decade ago or a century ago.

The remaining problems are social: making the tools easy to use, teaching the public how to do so, and choosing criteria for repository accession. The last does not seem to call for research, because it will be a matter of subjective choice by each repository community. The solution for these challenges cannot be hurried, but instead will be worked out socially over a few decades.

---

[22] An abstract reason for this assertion is Occam's Razor compliance. Practical reasons include that doing so will take advantage of library management practice developed over more than a century and that the resulting procedures can be designed to seem familiar to repository personnel and their clients.
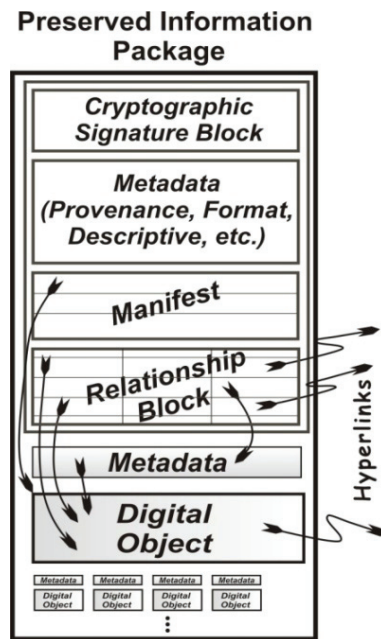
[23] At least, their publications suggest this.

[24] The hyperbolic phrase "exponential growth" has lost much of its original force. However, published information has, in fact, experienced exponential growth.

[25] Objective judgment of this fraction would be difficult, even if one could achieve consensus about subjectively chosen criteria. The assertion might, however, be agreed by thoughtful critics who have experienced a growing flood of scholarly articles that teach us little that we did not already know and also wanted to know.

If there is a big problem, it seems to be that the DH community has perhaps not noticed, perhaps not comprehended, and certainly not acknowledged manifest continuing technical progress.[26] C.P. Snow's gap between "two cultures" is still evident!

A Technical Solution

Figure 2: Schema for a Preserved Information Package

Preserved Information Package

Cryptographic Signature Block

Metadata (Provenance, Format, Descriptive, etc.)

Manifest

Relationship Block

Metadata

Digital Object

Metadata · Digital Object · Metadata · Digital Object · Metadata · Digital Object · Metadata · Digital Object

Hyperlinks

© 2011 HMG Services, LLC

An in-principle solution for the LDP requirements summarized above was published as early as 2005. Later, Gladney (2009) disagreed with most other preservation authors by asserting that the technical core could not be procedures for managing digital repositories, but instead had to be a scheme in which a single file could package a "complete" information corpus.[27]

The scheme for such a "Trustworthy Digital Object" (TDO), which represents some document together with subjectively chosen critical context, is suggested by figure 2. Its most important properties follow.

---

[26] Supportive evidence can easily be gathered by inspecting citations made by DH authors.
[27] Gladney (2009) provides a more thorough description and analysis than that in the following synopsis.

- Representing bit-strings are packaged with registered schema.
- The package includes or links reliably to all metadata needed for interpretation and as authenticity evidence.
- These bit-strings and metadata are encoded to be platform-independent and durably intelligible.[28]
- Every critical link to another TDO is secured by a cryptographic message authentication code.
- All this is sealed using cryptographic certificates based on public-key message authentication, with each cryptographic certificate authenticated by a recursive certificate chain grounded in a public reliable source.

Several articles describing this work requested public criticism wanted before implementing pilots to test and demonstrate the ideas' correctness and practicality.

I paused, waiting for reactions. Over eight years later, almost nobody has commented, nothing distinctly different and workable has been published,[29] and repetitive preservation conferences seem remarkably similar to their counterparts of a decade earlier (Gladney 2011). How could this happen? Surely part of the problem is DH community inattention to software engineering literature.

Summary and Conclusion

An aspect mostly missing from LDP literature is a sense of history-in-the-making. A few commentators, following Marshal McLuhan's The Gutenberg Galaxy,[30] suggest that the "digital revolution" has a precedent sometimes called "the Gutenberg Revolution". They point out that the social changes stimulated by the invention of movable type required about a century to play out.

Only 30 years have passed since e-mail became available, and only 15 years since the first digital libraries were deployed. If we are indeed experiencing a digital revolution, it is only its early days. If so, it might be silly for scholars to debate how it should work out. Society will, over time, decide. A tiny group of scholars can sometimes influence society. But is the current issue such a case?

We further wonder, "Why might scientists and engineers intuitively feel that DH does not merit high respect?" This might be because some DH publications

---

[28] For some data classes, representations approaching obsolescence might have to be superseded, perhaps as often as every decade. A fail-safe way of doing this is known. Implementations can be executed as batch processes that use "waste" computer cycles.

[29] While writing these notes, I discovered philosophical support for my position in Bootz, Szoniecky and Bargaoui (2009).

[30] McLuhan did not write about digital media, but rather about electronic communication of any kind.

display appalling inattention to prior work,[31] such as seminal epistemology Coffa (1993), McDonough (1986), Pincock (2009) that has long provided fundamentals of their topics.

It also might be because DH does not have the richness and complexity of topics such as nuclear physics and nuclear engineering.[32] For instance, I seem able to make meaningful comments on most DH papers and expect that I could, with a few weeks of self-education, even publish in a DH periodical. I could not manage an equivalent feat in any scientific or engineering field, not even in those of my formal education!

Who might be harmed by considering DH to be a discipline in itself? Perhaps the community that will suffer the greatest practical disadvantages will be the strongest proponents of an independent DH! Many of these might overlook the immense IS literature and its solutions to what they see as DH research challenges, possibly "solving" already-solved questions again. And their articles, labeled as DH literature, are likely to be overlooked by most other scholars, mostly because these never notice the existence of DH and, after their attention is directed to periodicals such as DHQ, deciding that the just-mentioned weakness of the field merits ignoring DH literature.[33]

The current article illustrates the weakness of proposed Digital Humanities research agendas by showing that Long-time Digital Preservation – the most prominently featured specific topic in recent DH articles, is a solved challenge for which all that still needs attention is software creation and deployment. Unless the DH community can identify other research topics of significant depth and scope,[34] we must conclude that *there exists no persuasive DH re-*

---

[31] For LDP, this has been illustrated by Burgess and Hamming (2011). A more egregious example occurs in Gochenour's discussion of mathematical graphs (2011). This fails to cite Carnap's 1928 *The Logical Structure of the World* (Pincock 2009) – a seminal epistemology text to which Gochenour adds nothing new.

[32] Established academic practices demand varied high skills, ranging from deep conceptual thinking to relatively routine mechanical tasks. Consider, for instance, chemical physics. It calls for laboratory skills – use of glassware, balances, spectroscopes, and more sophisticated instruments that are essential to most chemistry practice. Learning these skills typically occupies 50% of an undergraduate's scheduled hours, and some chemists spend much time extending or refining such tools. And yet almost nobody confuses these aspects with "being a chemist" or contributing to human knowledge.

Skill with digital tools is surely necessary for Humanities practice, and might require significant time to acquire (either as an undergraduate or, for today's children, in elementary school). However, this is insufficient reason to conflate such mechanistic aspects with what is needed to be a Professor of Humanities.

[33] The current author was unaware of DH until Manfred Thaller proposed the DH debate, illustrating the first problem. The LDP example described above illustrates the second problem.

[34] I have not discovered such topics in my DH readings; if they exist, the DH community needs to communicate them as part of seeking funding support and respect.

search agenda – and therefore insufficient reason for establishing DH facul-ties.[35]

References

[Anonymous]. 2009. Online Humanities Scholarship: The Shape of Things to Come, a tabulation of DH-DL relationship resources. *Digital Humanities Quarterly* 3 (2). <http://shapeofthings.org/resources.html>.

Bailey, Charles W. 2008. Scholarly Electronic Publishing Bibliography, §6.4, Library Issues: Information Integrity and Preservation. <http://www.digital-scholarship.com/sepb/>, <http://www.digital-scholarship.com/sepb/lbinteg.htm>.

Beagrie, Neil, Meg Bellinger, Robin Dale, Marianne Doerr, Margaret Hedstrom, Maggie Jones, Anne Kenney, Catherine Lupovici, Kelly Russell, Colin Webb, and Deborah Woodyard. 2002. Trusted Digital Repositories: Attributes and Responsibilities. <http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf>.

Bentley, Paul. 2010. Mastering digital lives: cultural heritage institutions tackle the Tower of Babel. Online Currents. <http://www.twf.org.au/research/Masteringdigitallives.html>, (Accessed February 19, 2012).

Berman, Francine. 2008. Got Data? A Guide to Data Preservation in the Information Age. *Communications of the ACM* 51(12): 50-6.

Berriman, G. Bruce and Steven L. Groom. 2011. How Will Astronomy Archives Survive the Data Tsunami. *Communications of the ACM* 54(12): 52-6.

Bootz, Philippe, Samuel Szoniecky, and Abderrahim Bargaoui. 2009. Entity/ Identity: A tool designed to index documents about digital poetry. Paper presented at the Symposium e-poetry, Barcelona, May 24-27. <http://archivesic.ccsd.cnrs.fr/sic_00603588/en/>.

Borgman, Christine L. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.

Bronowski, Jacob. 1965. *Science and Human Values*. New York: Harper & Row.

Burgess, Helen J. and Jeanne Hamming. 2011. New Media in the Academy; Labor and the Production of Knowledge in Scholarly Multimedia. *Digital Humanities Quarterly* 6 (3). <http://www.digitalhumanities.org/dhq/vol/5/3/000102/000102.html>.

Coffa, J. Alberto. 1993. *The Semantic Tradition from Kant to Carnap to the Vienna Station*. Cambridge: Cambridge University Press.

Digital Humanities Quarterly at <http://digitalhumanities.org/dhq/vol/1/1/index.html>. [The Alliance of DH Organizations promotes and supports digital research and teaching across all arts and humanities disciplines, acting as a com-

---

[35] We must differentiate creation of a DH university (sub)department from appointment by existing departments of individual faculty whose incumbents choose to focus on DH topics. The former would be part of some administrative agenda. Proscribing the latter would be an invasion of faculty independence which could reasonably be interpreted as a violation of ethical policy.

munity-based advisory force, and supporting excellence in research, publication, collaboration and training].

Duranti, Luciana. 2004. The Long-term Preservation of the Dynamic and Interactive Records of the Arts, Sciences and E-Government: InterPARES 2. *Documents Numérique* 8 (1): 1-14.

Economist. 2011. Social Media in the 16th Century: How Luther went Viral. <http://www.economist.com/blogs/babbage/2011/12/social-media-16th-century>, (Accessed December, 2011).

Economist. 2011. University Challenge: Slim down, focus and embrace technology: American universities need to be more businesslike. <http://www.economist.com/node/21541398/print>, (Accessed November 10, 2011).

Gladney, Henry M. 2000. Are Intellectual Property Rights a Digital Dilemma? Controversial Topics and International Aspects. *iMP Magazine* (February 22).

Gladney, Henry M. 1993. A Storage Subsystem for Image and Records Management. *IBM Systems Journal* 32: 512-40.

Gladney, Henry M. 2011. Long-Term Digital Preservation: Why is Progress Lagging? <http://www-e.uni-magdeburg.de/predoiu/sda2011/Gladney.pdf>. Paper submitted to the Nestor Workshop on Semantic Digital Archives, Berlin, September 9.

Gladney, Henry M. 2009. Long-term Preservation of Digital Records: Trustworthy Digital Objects. *American Archivist* 72 (2): 401-35.

Gladney, Henry M. 2006. Principles for Digital Preservation. *Communications of the ACM 49* (2): 111-16.

Gleick, James. 2011. *The Information*. New York: Pantheon Books.

Gochenour, Phillip H. 2011. Nodalism. *Digital Humanities Quarterly* 5 (3). <http://www.digitalhumanities.org/dhq/vol/5/3/000105/000105.html>.

Lee, Christopher A., and Helen R. Tibbo. 2007. Digital Curation and Trusted Repositories: Steps Toward Success. *Journal of Digital Information* 8 (2). <http://journals.tdl.org/jodi/article/view/229/183>, (Accessed February 19, 2012).

McDonough, Richard M. 1986. *The Argument of the Tractatus: Its Relevance to Contemporary Theories of Logic, Language, Mind, and Philosophical Truth*. Albany: State University of New York Press.

McLuhan, Marshall. 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. New York, NY: New American Library.

Nimmer, David. 1998. Adams and Bits: of Jewish Kings and Copyrights. *Southern California Law Review* 71: 219-45.

OAIS Reference Model - ISO 14721. (2009). <http://digitalcurationexchange.org/?q=node/1079>.

Pincock, Christopher. 2009. Carnap's Logical Structure of the World. <http://philsci-archive.pitt.edu/4569/1/pincock_aufbau_draft.pdf>.

Rosenthal, David S. H., Thomas S. Robertson, Tom Lipkis, Vicky Reich, and Seth Morabito. 2005. Requirements for Digital Preservation Systems: A Bottom-Up Approach. *D-Lib Magazine* 11 (11).

Tibbo, Helen R., Carolyn Hank, Christopher A. Lee, and Rachael Clemens, eds. 2009. Digital Curation: Practice, Promise & Prospects. Proceedings of DigCCurr 2009. Chapel Hill, NC, April 1-3, 2009. <http://www.ils.unc.edu/digccurr2009>.

Underwood, Sarah. 2001. British Computer Scientists Reboot. *Communications of the ACM* 54 (4): 23.

von Baeyer, Hans Christian. 2003. *Information: the New Language of Science*. London: Weidenfeld & Nicolson.

"Digital humanities". Wikipedia. 2011. <http://en.wikipedia.org/wiki/Digital_humanities>, (Accessed December 2011).

Wittgenstein, Ludwig. 1921. *Tractatus LogicoPhilosophicus*. London: Routledge.