Data, Information, and Knowledge
"where is the Life we have lost in living?"

Abstract:

This abstract attempts to raise the question of whether current practices in digital preservation properly address the issues of findability of digital objects. It is also intended as a starting point for discussing preservation of digital information in contrast to digital data. The abstract is exploratory and informal.

Keywords: data, information, knowledge, wisdom, preservation, appraisal, selection, findability

1. Digital preservation and archival science

Much of the practices recommended within the study of digital preservation are motivated by archival sciences, records management practices and diplomatics. As such, digital preservation tends to place emphasis on maintaining the authenticity, integrity, faithful reproduction, and reliability of information . Usability and the retrieve-ability of information is also discussed in the literature, however, these discussions are mostly associated to maintaining human-readability, unique identifiers, metadata for contextualisation, and the faithful reproduction of its content once the object has been located. The question of whether one can actually *find* the target information is largely disregarded. How many times have we searched elsewhere for information we already have on our computer because we cannot remember where we put it? In fact, it is often the case that we can locate a digital object more quickly by submitting keywords in Google than to look through our own folders (even when we know in which folder to look!). A large proportion of information loss is due to lack of *findability.* So, then should we have metadata about our metadata and so on and so forth recursively forever? And, what about the situation where we need answers to questions that require the integration, interpretation and synthesis of a range of datasets? Does the current framework support such activities? What are we preserving? Is it data, information or knowledge?

2. Information: is it filtered data?

A few of the immediately observable pitfalls of the current approach to digital preservation seems to originate from the idea that *information* is the filtered collection of recorded data which is static in itself but available for re-interpretation. This popular notion of information as a distilled version of data is reinforced by the hierarchical organisation of data, information and knowledge introduced within many disciplines (cf. Sharma [5]), e.g. in knowledge management (attributed to Ackroff [1] and Zeleny [7]), in design (Cooley [3]), and, in information science (Cleveland [2]). To support the relationship between these concepts some have even cited the poet, T. S. Eliot and the musician Frank Zappa:

"...
The endless cycle of idea and action,
Endless invention, endless experiment,
Brings knowledge of motion, but not of stillness;
Knowledge of speech, but not of silence;
Knowledge of words, and ignorance of the Word.
All our knowledge brings us nearer to our ignorance,
All our ignorance brings us nearer to death,
But nearness to death no nearer to GOD.
Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?
The cycles of Heaven in twenty centuries
Bring us farther from GOD and nearer to the Dust."

Except from the play "The Rock"[1] , *Choruses* by **T. S. Eliot (1934)**

"Information is not knowledge,
Knowledge is not wisdom,
Wisdom is not truth,
Truth is not beauty,
Beauty is not love,
Love is not music,
and Music is THE BEST."

---

[1] A pageant play by Eliot called *The Rock* was performed in 1934 for the benefit for churches in the Diocese of London.

From *Packard Goose* in album Joe's Garage: Act II & III, by **Frank Zappa (1979)**

Some have criticised the emergence of the hierarchical relationship (e.g. Weinberger [6]) as a trend towards thinking of Information as a filtered product of data, knowledge as a filtered product of information and, even further, wisdom as a filtered product of knowledge. Indeed there is something to this criticism. However, the fallacy is not so much in the hierarchy but in our assumption that this is a closed static system. I am sure T.S. Eliot did not intend to imply that knowledge is simply a filtered form of information but rather that information in itself does not lead to knowledge until it is dynamically and actively processed within context.

3. What are we preserving and why

Given a collection of data, the measurement of information contained within the collection is contextual. This is fundamental to the information theory a la Shannon (Shannon [4]) . We have strayed away from the fundamentals of information theory since information scientists (both in management and in computing science) started to propagate exponentially. In Shannon's days, the definition of information was quite explicit: my simplified take on Shannon's proposition is that the amount of information contained within a collection is measured by its capacity to reduce the entropy of a given a situation. That is, it is only when the data content is paired with a given situation that you can measure the value of data as information. So the question I would like to pose on this occasion is: in current practices in preservation of digital *information,* is it really *information* we are striving to preserve or is it *data.* Further, do current practices in appraisal, selection, preservation, and curation really support the conservation of *knowledge, culture, and heritage* and the *discovery of new knowledge and wisdom?* If not, as T.S. Eliot says, have we not lost *Life* in exchange for *living?*

To use the words from "I'm so worried" by Monty Python:

"...
And I'm so worried about the baggage retrieval
System they've got at Heathrow.
..."

Will we ever be able to "retrieve", in an informative way, what we have entrusted to a *Trusted Repository?* Perhaps we should start moving away from data preservation and be worried about the preservation and retrieval of *information* within the archival and digital preservation community.

References.

(URLS accessed 7 July, 2010)

[1] Ackoff, R. L., (1989) "From Data to Wisdom," *Journal of Applied Systems Analysis* 16:3-9.

[2] Cleveland, H., (1982) "Information as Resource," *The Futurist*, December 1982 : 34-39.

[3] Cooley, M., (1987) *Architecture or Bee?* London: The Hogarth Press.

[4] Shannon, C. E., (1948) "A Mathematical Theory of Communication",
Bell System Technical Journal, 27:379–423

[5] Sharma, N., (2008) "The Origin of the "Data Information Knowledge Wisdom" Hierarchy",
http://nsharma.people.si.umich.edu//dikw_origin.htm
version 4 Feb, 2008

[6] Weinberger, D., (2010) "The Problem with the Data-Information-Knowledge-Wisdom Hierarchy",
blog post, 2 Feb, 2010, Harvard business Review,
http://blogs.hbr.org/cs/2010/02/data_is_to_info_as_info_is_not.html

[7] Milan Zeleny, (1987) "Management Support Systems: Towards Integrated Knowledge Management,"
Human Systems Management 7, 1: 59-70.