Pandora's Box: The Future of Cultural Heritage on the World Wide Web

Leif Isaksen
Moderator, Antiquist Community
leif.isaksen@cantab.net

Abstract

This paper argues that discussions about managing control of heritage resources have been rendered redundant by the runaway success of the World Wide Web and its increasing impact on global culture. If heritage professionals and memory institutions are to maintain their relevance it will be vital for them to use new media to engage with both special interest groups and society at large. A willingness to embrace participatory technologies will play a fundamental role in this process. It is also argued that promising experimentation with Semantic Web technologies is likewise being impeded by a failure to properly integrate information with the Web through the use of stable URIs.

The Open Box

Opinions about cultural heritage and the Web generally fall into two categories. One camp is excited by the opportunities afforded for dissemination and sharing in the Brave New World of the internet. The other camp is also excited by the opportunities afforded for dissemination and sharing in the Brave New World of the internet but is also concerned that, if it is not done 'right', we may open *Pandora's Box*. This particular Pandora's Box contains all the evils of the cultural world – plagiarism, lack of transparency, misleading or inaccurate information, even outright fraudulence. Yet neither side seems willing to admit to the Elephant in the Room:

Pandora's Box is already open.

It was open long before Tim Berners-Lee thought up the World Wide Web or the American military invented the internet. It was open before Gutenberg created his printing press. It was open when Herodotus penned his *Histories* and we can be sure it was open even in the days of Homer himself. Because it is impossible to control cultural discourse. As long as humans are interested in their past there will those who want to relate it to them and they will use whatever means they have available. If you are not willing to provide them with the 'authorized' version of events they'll go out and get the 'unauthorized' version instead. Culture does not lie in the material record or paper archives but only in the stories people tell about them, regardless of whether they include references, citations or apparatus. Alternative histories have always existed but they were often invisible to those in authority. Now that the internet has

begun to both render and remix them, we can begin to see just how far we, the curators and academics, are falling behind.

Let us take the example of one of the West's most famous artefacts, housed in one of its most well-known, well-funded and well-respected repositories, the Mona Lisa in the Louvre. The first six results of a Google Search for 'Mona Lisa' are given in their return order below, along with their Google PageRank.

URL	Google PageRank
http://en.wikipedia.org/wiki/Mona_Lisa	6
http://www.youtube.com/watch?v=uk2sPl_Z7ZU	6
http://www.monalisamania.com/	5
http://www.ibiblio.org/wm/paint/auth/vinci/joconde/	5
http://www.bbc.co.uk/science/leonardo/gallery/monalisa.shtml	5
http://www.louvre.fr/llv/dossiers/detail_oal.jsp?CONTENT%3C%3Ecnt_id=10134198673229908&CURRENT_LLV_OAL%3C%3Ecnt_id=10134198673229908&bmLocale=en	5

The pages which outrank the (highly informative) Louvre website are: 1) the ubiquitous Wikipedia article, 2) a YouTube video showing the Mona Lisa being drawn with Microsoft Paint, 3) a personal fansite based in the US, 4) the WebMuseum, and 5) a BBC factsheet. Not only does the Louvre come sixth but the first two have a higher PageRank, currently the most commonly accepted rule-of-thumb for measuring web influence. It uses an algorithm based on incoming links and the results are on a log scale which informal estimates put between 6 and 8. There are therefore at least two websites of entirely public-generated content which are considerably more influential than that of the host institution. Searches for 'La Joconde' and 'La Gioconda' (its French and Italian appellations) give similar results. Let us repeat, then, that it is no longer worth arguing whether we should try to *control* cultural information. Clinging on to artefacts and archives, restricting or obscuring data, claiming academic authority – these are of little or no use to us if we are struggling even to make ourselves heard.

Culture 2.0

Tim O'Reilly, the man officially credited with introducing the phrase 'Web 2.0' is also famous for noting that winners and losers [on the web] will be designated simply by "who figures out how to use the network" (The Economist 2007). Web 1.0 was a case of pitching up on the internet and hoping that your brand was strong enough for people to find you and your content was interesting enough for them to return. The revelation of Web 2.0 is that 'the network' extends beyond the internet and into the

populace. Services that encourage *involvement* get richer over time and keep people coming back. It is perhaps precisely because academic and heritage discourse often tends to keep the layman at arm's length (a policy of 'look but don't touch') that it has proved so poor at utilising a technology which has seen extraordinary take-up elsewhere. And there's the rub - cultural custodianship is an inherently elitist concept, whereas the internet is initiating the first great age of Culture Democracy. Plato would be horrified.

As self-appointed guardians of 'heritage' it is easy to perceive ourselves as protecting the torch of truth and civilization from barbarous hordes intent on 'dumbing down'. This view betrays an awkward conflict within the Humanities. We are happy to talk about past cultures in an ostensibly objective fashion, regardless of distinctions in race, gender or class. Yet when it comes to our own we try to distinguish between an innate, sophisticated 'high culture' and a banal and transient 'media'. This is a dangerous mistake even if we were to talk only amongst ourselves. (How many of us have seen *Gladiator*? And how many have read the *Historia Augusta*? Thought so.) Stories are valuable only so long as they are told and retold.

So how can we engage with this information Leviathan that threatens to smother us in its embrace? There is no Silver Bullet, for reasons which will shortly become clear, but there are a number of guiding rules to the new information paradigm which can help illuminate our path. We can begin with the eight core patterns of Web 2.0 listed by O'Reilly et al. (Musser 2006). Although they were identified with specifically enterprise objectives in mind it will do us well to look at the relevance they have for the cultural sphere.

- Harnessing Collective Intelligence – We have to accept that not only do we not know everything, but that cultural perspectives are infinite in principle. Nonetheless, there are plenty of tools available that allow us to monitor and moderate input from 'the crowd' to our own sites.
- Data is the Next "Intel Inside" – Think provision, not possession. Syndication has made information an infinitely sharable resource resulting in two kinds of benefactor: those who use it to aggregate get brand recognition. Those who are aggregated get their message across.
- Innovation in Assembly – The tendency to build from the ground up and ignore the myriad of freely available services and APIs is strong in our discipline. Maps, timelines, lightboxes and other widgets are frequently just a few lines of Javascript away.
- Rich User Experiences – You can never second guess the needs of your audience because your audience is never the same. Give them options.
- Platform Agnosticism – "It doesn't work on Firefox? On Linux? On my iPhone? It doesn't matter, Wikipedia does."
- The Long Tail – There may be very few experts in a given domain, but the large number of people who come into contact with it can provide vital contributions. Forums, mailing lists and online communities can be powerful ways to bring them together.
- Lightweight Software and Scalability – large, complex and monolithic systems are unveiled with pride at heritage sector conferences every year. Few are heard of a second time. 'Agile' approaches are far preferable because the Web is now in…

- Perpetual Beta – there is no Silver Bullet because Web 2.0 requires continuous adaptation to change. YouTube, Facebook, Google Maps and Wikipedia were unimagined by most of us less than 10 years ago. They may well be replaced by Vimeo, OpenSocial, Geonames and Freebase in the next five years and after that…who knows? Successful utilisers of the contemporary Web will be those who plan for the unknown.

So much for the principles, what about the practicalities? Dion Hinchcliffe, another notable Web 2.0 pundit, gives a useful breakdown of the different Web 2.0 media, listing websites, syndication (RSS/ATOM), Web 2.0 applications, Open APIs/webservices, web widgets (embeddable apps), social networking applications and the Semantic Web. Most importantly he notes that on the present day Web, *"having a website is usually the least interesting thing"*. (Hinchcliffe 2008) It is much more powerful to have your content discover users than the other way around. Syndication, as mentioned above, is one way to do this. So is tagging your content with relevant terms, or utilising networks of friends and colleagues. There are several non-commercial websites that use social networking to rate, review, compare and suggest books (Nations 2008). Why is there no such site for academic texts?

Fortunately, we are beginning to see experimentation in this arena. In the UK, the Wessex Archaeology contract unit have been a surprising fieldleader, using a wide range of widgets and services including tagging, syndication and photo sharing. The result was a big upsurge in both web traffic and public profile. Site blogs are becoming increasingly commonplace as well, enabling those with passing interest to be alerted to daily developments. The Portable Antiquities Scheme at the British Museum now has a Web API which enables anyone to get (degraded) spatial data from their extensive find database. Finally, open source web-based recording systems like ARK and OpenContext are already providing archaeologists with a range of opportunities for making excavation data available online. Whilst these trends are new and relatively untested within our sector, they suggest that there is both the will and capacity to innovate. Nevertheless, despite these pioneering efforts, we have yet to see an emergence of 'the network effect'.

Putting the 'Web' back into 'Semantic Web'

One of the greatest strengths of the web has been its fault tolerance. The ability to create unidirectional links to resources meant that persistence was unnecessary. As these links have moved from providing context to critical content or even services, stability has become a much greater issue. Nowhere is this more apparent than in what is set to become the *non plus ultra* of information integration: the Semantic Web.

First, a quick digression. The Semantic Web is not, as is frequently supposed, directly related to Web 2.0. The latter is based on the observance of internet trends since circa 2003. Although the phenomena mentioned above have been identified *a posteriori* as key factors, its ultimate definition is simply 'what people are doing with the Web today'. In contrast, the Semantic Web (capitalised) is a very specific project, undertaken by the W3C in order to improve the machine-readability of information on the internet. It uses specific technologies such as URIs and RDF, and well-defined

vocabularies including RDFS, OWL and SKOS (of which, more later). It is sometimes referred to as the Giant Global Graph (GGG) because it is a web of concepts (classes, relationships and things) rather than documents or services. It is important to bear this in mind because both paradigms have drawbacks as well as advantages and conflating them does both a disservice. However, as both Hinchcliffe and Kansa (this volume) note, it is also the direction in which many Web trendsetters are slowly gravitating.

The Semantic Web is mentioned in several papers at this conference and I do not wish to unnecessarily repeat what is likely to have been said already. Nevertheless, I wish to raise a point that, until now, seems to have been sorely missing from many so-called Semantic Web applications in our field. We can begin by looking at Berners-Lee's famous 'Layer Cake' diagram (Berners-Lee 2002). Its purpose is to show the technology stack which makes the Semantic Web possible. It is surprising then, that Uniform Resource Identifiers (URIs), one of the two foundational elements of the entire project, get virtually no attention in almost any of the cultural heritage Semantic Web projects I have seen to date.

We are all familiar with the concept of URIs as the web addresses in our internet browsers. In the Semantic Web they also form the atomic concepts of the {Subject, Predicate, Object} statements which combine to create an RDF graph. It is their uniqueness which makes them perfect for establishing semantic continuity across datasets. Basic merges between graphs can be done by simply identifying common URIs (hence 'Giant Global Graph'). Unfortunately, many applications fail to appreciate that for the Semantic *Web* to function URIs must also be both public and stable (and ideally resolvable). The current tendency to build semantic 'stovepipe' applications which neither provide external URIs, nor draw upon them, illustrates the degree to which we have yet to fully comprehend the Semantic Web's potential. Likewise, ontologies and Knowledge Organisation Systems are merely information design patterns until they are a) implemented in the appropriate vocabularies (The Web Ontology Language (OWL) and Simple Knowledge Organisation System (SKOS)) and b) given fixed URIs for their concepts.

Once again, we can identify a small numbers of positive developments – The CIDOC CRM has been implemented in OWL although it still appears to have no fixed URIs for its concepts. The MultimediaN E-Culture project has converted the Getty thesauri into SKOS, but this has in turn highlighted another danger. Linked Data is not possible (or at least, highly retarded) when concepts are proprietary. The University of Glamorgan STAR Project has had related success with the English Heritage thesauri (but these are still to be made public). Lastly, the Pleiades Project has taken the forward-thinking step of establishing URIs for its entire dataset of ancient locations even before it is able to provide locations for them.

Whilst of pre-eminent importance for the Semantic Web, the issue of stability is no less important for Web 2.0 technologies. Simple, meaningful URI strings (known as 'Cool URIs') are an important aspect of this. Look again at the Wikipedia and the Louvre website locations. Which one looks likely to be around in 5 years time? Which one should you link to? Technologies such as Persistent URLs (PURLs) make simple, stable URIs available even to small institutions so there are no longer

technical reasons for avoiding it. Microformats and RDFa are other valuable ways to integrate URI concepts into otherwise standard webpages.

Conclusion

There is a great deal of work yet to be done in building an environment in which next-generation web content can flourish in the cultural heritage sector. I have only mentioned two – First, that we need to accept that an isolationist policy amongst heritage professionals is not only foolish but ultimately meaningless because culture will carry on without us. We must play by the rules of our times and that means reaching out to society at large. Secondly, if the future of information lies with the Semantic Web, we need to understand that it is still fundamentally a web of URIs. Sharing concepts means putting them out there on the internet.

Both of these are areas in which we can, to a large degree, work independently. There are areas however, where we must work together. In the final analysis, Many of these boil down to ensuring that digital approaches are given due credit. We need to find a workable system for digital peer review to break the restrictive practices of journal publishers in the Humanities. It is a disgrace that tax-funded scholarly work is unavailable to those who have paid for it. We must also work hard at improving computational literacy and interfaces. The digital divide in academia is increasing and it is no coincidence that many of the achievements listed above have been due to the work of junior academics and practitioners. Recalling that the network includes its users, it is vital to make it more accessible to everyone. We must lobby hard in both political and academic circles in order to have digital resources accepted as mainstream output. The insanity of framing complex data-producing projects as print publications is of benefit to nobody. Finally, perhaps it is time to debate what our purpose is in an age where free, twenty-four hour access to non-commercial information is increasingly considered a human right. Whose agenda do we serve?

I began this talk by proclaiming that Pandora's Box was already open. The myth goes on to say that once the world's evils had escaped, only one trait remained – hope. Perhaps I am an optimist but I believe there is much to be hopeful about. We have been telling each other about our communal past ever since we learned to talk. In doing so we frequently invent, embellish, mislead, forget, and plagiarise. Neither storytelling, literature, theatre, the printing press, newspapers, the telephone, radio, television, email or the World Wide Web have made the slightest difference to these fundamentally human processes. We in the Humanities have our own traditions, based on methods which were first developed here in Athens two and half thousand years ago. It is vital they are heard because they introduce great richness, sophistication and, just possibly, a measure of veracity into the global conversation but it is a fallacy to believe that we own, control or protect *culture*. Technology will continue to plough its own unpredictable furrow and we can only scatter the seeds of our thought upon it and see what grows.

References

Berners-Lee, T., 2002. Enabling Standards & Technologies - Layer Cake". *The Semantic Web*. Available at: http://www.w3.org/2002/Talks/04-sweb/slide12-0.html [Accessed August 28, 2008].

Hinchcliffe, D., 2008. Building Modern Web Apps? Better Have A Deep Competency in Web 2.0, Open APIs, Widgets, Social Apps, and Much More. *Dion Hinchcliffe's Web 2.0 Blog*. Available at: http://web2.socialcomputingmagazine.com/building_modern_web_apps_better_a_have_deep_competency_in_w.htm [Accessed August 28, 2008].

Musser, J., 2006. *Web 2.0 Principles and Best Practices*, O'Reilly. Available at: http://radar.oreilly.com/research/web2-report.html?CMP=ILC-GG7423313304&ATT=web2report.

Nations, D., 2008. Book Social Networks - A List of Social Networks About Books and Reading. *About.com*. Available at: http://webtrends.about.com/od/socialnetworks/tp/book-social-network-list.htm [Accessed August 28, 2008].

The Economist, 2007. *Serious Business: Web 2.0 Goes Corporate*, The Economist Intelligence Unit. Available at: http://socialmediagroup.ca/wp-content/uploads/2007/06/smg_eiu_web20.pdf [Accessed August 29, 2008].