



ROWENA LOVERANCE

**DIGITAL PRESERVATION OF USER-GENERATED
CONTENT: THE NEXT CHALLENGE
FOR MUSEUM LIBRARIES AND ARCHIVES**

ROWENA LOVERANCE

She belongs to the pioneering generation of new media professionals in UK museums. In 1991 she made the British Museum's first exhibition interactive and CDROM, *The Anglo-Saxons Interactive* and in 1996, with colleagues at Japan's National Museum of Ethnology in Osaka, she created the Global Digital Museum. She was a member of the prototype team for the British Museum's collections online multimedia public access system (COMPASS). She then set up an in-house educational multimedia team, series of online learning resources. She was a member of the UK National Museums Directors Conference subgroup and represented the national/larger museum sector on the Department of Education and Skills. She is currently working with Becta, the British Educational Communications and Technology Agency, on barriers preventing cultural institutions from creating online learning resources.

Abstract

This paper raises a new issue for museums, libraries and archives as they seek to embrace Web 2.0 technologies to enhance their online presence. Many of them are increasingly requesting their online visitors to contribute digital content. It is not clear whether institutions intend this content to have the same degree of sustainability as institutionally-generated digital content, nor whether this is part of the audience expectation in contributing it. If institutions decide to preserve this content for the future, should it be treated in the same way as institutional content, or does it present new requirements of cataloguing and resourcing. The paper surveys the current practice of several UK national institutions and outlines an approach to generating good practice guidelines for the future.

Keywords: user-generated content; digital preservation; museums, libraries and archives; audience expectation; personal memory; sustainability

1. Introduction: acquisition and disposal

Museums, libraries and archives are monuments to the preservation of content. Their whole mission is to conserve, catalogue and preserve their collections for posterity. In the UK the issue of the potential disposal of collections has only recently been raised, and that very tentatively¹.

¹ See Model Acquisition and Disposal Policy in MLA's Museum Accreditation Scheme http://www.mla.gov.uk/resources/assets/A/accreditation_model_policy_doc_5643.doc

It is still a dangerous subject, since it appears to strike at the very heart of a museum's purpose, and to threaten future donations, since many donors have declared their unwillingness to donate or bequeath to any museum which does not guarantee the long-term future of their objects. In the decision whether or not to acquire a new object, one of the factors a museum must assess is the sustainability of looking after it in the long-term, and whether it will continue to justify its place in the collection. In another part of their operations, however, museums, libraries and archives are actively seeking new content, the quality, scale and sustainability of which they are not fully able to assess. The difference is that this is digital content, which is being sought from users under the banner of new Web 2.0 technology. The hope is that by opening up their institutional websites to this new kind of content, a museum will revolutionize its relationship with its audience, both the minority who currently contribute their opinions, reminiscences or digital artworks and the majority who, though non-contributors, nevertheless enjoy accessing these sites, presumably appreciating their wider range of viewpoints and material. The potential tension between these two attitudes to content has not so far been much explored. This paper seeks to define some of the issues which it raises, and to identify a possible way forward, with the hope that some good practice guidelines can be put in place before museums go much further down the road to acquiring large quantities of user-generated content.

2. User-generated content in the cultural sector

At present no one in the cultural sector is quite sure whether the enthusiasm for user-generated content (UGC) which presently characterises the online world is the future or a passing fad; in either case, however, the general opinion is that if the sector ignores such a widespread phenomenon, it risks losing the interest of a significant proportion of its audience². The interest of the big players, such as Google and Yahoo! Inc., Microsoft and News International, in sites like YouTube, MySpace, Flickr, Technorati and del.icio.us suggests that they think we're here for the long haul. It is possible, of course, for cultural institutions to engage in this revolution without storing up problems of digital preservation. Much of the activity can take place on the social networking sites themselves, a method which has the significant advantage of reaching the audience where it is rather than expecting it to find its way to the institution: a recent dramatic example is the Museum of London's online auction on Ebay of one square metre of empty display space.

² Kevin von Appen, Bryan Kennedy and Jim Spadaccini, *Community Sites & Emerging Social Technologies*, in J. Trant and D. Bearman (eds.), *Museums and the Web 2006 Selected Papers from an International Conference*, Toronto: Archives & Museum Informatics, 2006, pp.197-206, also available at <http://www.archimuse.com/mw2006/papers/vonappen/vonappen.html>

Libraries are using their MySpace profile to ask their readers which books they would like them to stock, while museums are putting photographs of their objects on to Flickr to get them tagged socially rather than just curatorially, commissioning community videos on YouTube or keeping their information factsheets regularly updated on del.icio.us. A consortium of US museums are experimenting with a shared approach to folksonomy, or social tagging, at www.steve.museum⁴. Most of these institutions hope that by establishing a presence on social software sites they will lure audiences on to their own websites, where they can engage them with a wider range of online activities. Others have started by experimenting on their own sites to start with. The Powerhouse Museum in Sydney is starting to incorporate social tagging on to its collections database. The Liverpool Museum is creating a bridge from Flickr to its own site, by inviting local Flickr photographers to recreate historic photographs of the city for an online exhibition. At the UK national level, the lead in user-generated content has been taken by the V&A, who have developed a range of sites - a blog by their artist in residence, digital reminiscences for people who own V&A objects, such as clothes by celebrity designers and digital artwork sites where people can design their own Arts and Crafts tile or Vivienne Westwood tartan. The V&A also worked with Ultralab and Channel 4 on a Culture Online project, Every Object Tells a Story, which seeks to build up a community archive of objects which people particularly value.

Although the field is still very much in its infancy, other museums are following the V&A lead. The NHM invites its younger users to submit dinosaur recipes and the British Museum invites users of any age to design their own hanging mobile or interactive boardgame. The National Gallery and Tate are currently scoping out the field, prior to plunging in. Meanwhile the V&A is taking the lead in a consortium of ten UK national museums who are seeking corporately to apply this experience to their websites. The aim is to encourage a creative response to different aspects of their collections, which could then be filmed and made available online as a series of 'creative journeys'⁵.

If all this works, and sufficiently interesting topics are found which can grab the public imagination, then we could soon be looking at an explosion of UGC, from text-based content such as blogs and wikis, across the full range of media: photographs, audio, video, 3D and animations.

3. Preserving UGC: current practice

In preparing this paper, I have made an informal survey among colleagues as to whether museums, libraries and archives are addressing the question of sustainability of user-generated content. It is only a starting point, but it has produced some interesting results.

The British Museum, although it has only a sprinkling of UGC on its main site, has considerable experience going back to 2001 through the children's version of its public database, www.thebritishmuseum.ac.uk/childrensCOMPASS. Museum learning programmes for children have always had an incentive to capture examples of children's work so as to validate their own work. ChildrensCOMPASS has many more audience-related features than the main site: they include children's art work, mostly images and texts relating to museum objects, questions and expert answers, online tours curated by invited community audiences - children from a special school or from the Chinese or Sudanese community - and material generated by special exhibitions, such as identity and belief as shown on modern badges. The museum currently keeps all this material, whether published online or not, on its COMPASS CMS, tagged to museum objects where appropriate, but does not undertake any additional cataloguing of it.

The Natural History Museum relies on its CMS to archive its site regularly, so has a 'snapshot' approach to its UGC; it does not make any further attempt to catalogue or preserve access to it. The National Gallery is installing a new CMS which it hopes will improve its control over publication and de-publication, but did not find a CMS which it considered to have a satisfactory 'snapshot' archiving feature. The V&A, despite having taken the lead in generating UGC, has no special digital preservation policy. As for the V&A-led national museums' online learning project, its draft contract specifically differentiates between museum-generated content and UGC: it guarantees to keep the former for five years but makes no such undertaking re the latter.

The UK institution which seems to have thought out the issues most clearly is, appropriately, The National Archives, which has an institutional responsibility for digital preservation. Social history, with its obvious tie-in to oral history, is the other area, besides children's learning, which has always had an interest in this kind of material and The National Archives is leading on the online project, Moving Here, which seeks user-contributions from immigrant communities. For this project, they have undertaken to maintain UGC for twelve years until 2017, though this undertaking is not actually referred to on the site.⁶

⁴ Jennifer Trant, Social Classification and Folksonomy in art Museums: early data from the steve.museum tagger prototype, a paper for the ASIST-CR Social Classification Workshop, Nov. 4th, 2006, also available at <http://www.archimuse.com/papers/asist-CR-steve-0611.pdf>

⁵ http://www.vam.ac.uk/about_va/online_learning/index.html

⁶ <http://www.bbc.co.uk/guidelines/editorialguidelines/onguide/interacting/handlingusergene.shtml>

4. Preserving UGC: is it really an issue?

Some people may consider it premature to be addressing these issues, when this new field of UGC is still in its infancy. I would suggest, however, that this is precisely the time to do it, before institutions amass too much undigested content, while most of the content is in text rather than audio or video format and before the task of managing it becomes impossibly onerous.

I would not however wish to argue that preservation is the most immediate issue with regard to UGC. Most institutions, both cultural institutions and the news organizations who also have a huge stake in UGC, have rightly seen the primary issues as IPR - making sure that generators of UGC have permission to publish the content they are submitting - and moderation - ensuring that the content is legal and appropriate. For this reason, a clause ensuring that publishers have the right to remove UGC from their sites at any time is standard in the Terms of Reference of media companies such as Guardian Unlimited. One could argue that having such a clause also legally covers the issue of digital preservation, but it does not address the need for institutions to formulate their own policy, nor to communicate that policy to their content-generating users.

It could well be argued that the question of preserving UGC can be addressed as part of the wider archiving question, whether of website published material or of unpublished material held on a CMS. As indicated above, this is currently the policy pursued consciously by the NHM, and unconsciously by most other UK national institutions. At first sight there may not appear to be any economic drivers to address the issue separately - storage capacity is cheap, and UGC constitutes only a small proportion of a museum's requirements. Given the short life of digital delivery platforms, there will be a need - and probably a charge - for migrating content into a different format, but once again UGC is no different in this respect to the rest of an institution's digital content. However one possible complication is the difficulty of anticipating an institution's future digital archiving policy. Presently, most UK national museums undertake their own archiving, but it is possible they could decide to switch to the use of third-party archival services, which would presumably be chargeable. The BBC addresses this point, without entirely resolving it, in their editorial guidelines for online content, which state that contributors should be warned at the time of contribution that material collected online may be handed over in due course to a third party as a permanent archive, but it is not clear whether they may withdraw their material at a later date, for instance if it is likely to be commercially exploited by the archive.

5. Preserving UGC: user expectations

One way of viewing UGC is as a particular kind of third-party content. What is so far unknown, however, is anything about users' motivations in contributing content. Work is starting to be

done on user incentives for social tagging, which has identified at least six motivations: future retrieval, contribution and sharing, attracting attention, play and competition, self-presentation and opinion expression⁷. It is likely that the same motivating factors are in play with people contributing UGC to cultural sites. Each of these motivating factors is likely to have a different implication for user expectations about the longevity of their contributions. In the case of future retrieval, it would be useful to know what scale of future people have in mind. Does self-presentation motivate people purely with regard to their peers, or are they thinking of future audiences? Would people be more likely to contribute digital material if its online longevity was guaranteed by a reputable institution? Would a guarantee of longevity result in a consequent improvement in the quality of material submitted?

Another area which would benefit from more research is the relationship, if any, between UGC on cultural sites and people's own systems for preserving digital content. At a practical level, individuals are unlikely to have the degree of expertise in digital storage which is being acquired by cultural institutions, so may prefer to have institutions look after the digital storage of particularly valuable items for them. At a more subtle level, museums' familiarity with issues of identity and selection, disposal and forgetting, may be useful to people as they face the increasing necessity to formulate their own strategies for digital preservation, now that a lifetime's worth of memories can be stored in digital form - photographs, emails, financial accounts, blogs and video diaries⁸. At the root of these issues is a question of trust: is the trust which people have in museums as cultural repositories transferable into the digital era?⁹ And what standards of behaviour would museums have to adopt to ensure that it is? It is a comparable issue, after all, with the lengths to which museums go to ensure that donor information appears on object labels, even if other apparently more useful details then have to be squeezed off.

Cultural institutions are starting to publish terms and conditions which users accept in offering their content for publication, though most do not yet address sustainability. Museums are familiar with the process of producing a Freedom of Information publication policy; a UGC digital preservation policy could perhaps be viewed in the same light, and even incorporated into it.

⁷ C. Marlow, M. Naaman, M. Davis, D. Boyd, 'HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read', *Proceedings of Hypertext 2006*, New York: ACM Press; also available at <http://alumni.media.mit.edu/~cameron/cv/pubs/2006-ht06-tagging-paper>

⁸ Kieron O'Hara, Richard Morris, Nigel Shadbolt, Graham J. Hitch, Wendy Hall and Neil Beagrie, 'Memories for life: a review of the science and technology', *Interface, Journal of the Royal Society* (2006), 351-365

⁹ Kieron O'Hara, 'The ethics of cyber trust', in ed. R Mansell and B.S. Collins, *Trust and Crime in Information Societies*, Cheltenham, UK: Edward Elgar, pp. 442-449.

6. Preserving UGC: institutional expectations

From the institutional point of view, what would be the requirements of such a policy? Would institutions wish, for instance, to be able to make a quality distinction in deciding whether or not to keep particular items of UGC, in addition to the quality judgement of whether or not to publish them in the first place? Whereas in acquiring physical collections, museums would expect to decide on grounds of quality whether or not to acquire material in the first place, it may be that value judgements by curators are by definition inappropriate in the case of UGC. However, if preservation is once decided upon, digital archivists disagree as to whether further selection is then permissible¹⁰, nor is it likely that principles designed for appraising institutionally-generated content can be applied unaltered to UGC. If value judgements were to be made, would institutions have to make public their criteria? And how can museums ensure that UGC offered to them and preserved in good faith is unique, and is not simultaneously being preserved elsewhere¹¹. Currently most Terms and Conditions for UGC, in making clear that copyright remains with the creator, specifically allow it to be republished elsewhere. If this remained standard, it is not clear how it would accord with museums' requirements not to replicate one another's collections.

The main issue from the institution's point of view in undertaking to preserve UGC is likely to be in ensuring its future accessibility, without which, apart from its possible use as a timecapsule, there is little point in starting down the road of preservation. One difficulty here is in knowing, in such a new field, what the future retrieval requirements are likely to be. Another is the problem that this content does not generally lend itself to a museum's traditional metadata categories. Establishing digital categories for traditional material has all but exhausted many museum's resources, and as already discussed, the task of socially tagging established collections has only just begun. One can barely begin to guess at the kind of categories, no doubt highly subjective, which would be required to enable future audiences to find their way around this material¹², or at the commitment of resources involved in setting them up. Social

tagging might once again be the answer, but at the moment would be a distraction from the more urgent job of establishing folksonomies of museum, library and archive collections which is barely underway.

7. Preserving UGC: a way forward?

This paper has argued that the cultural sector needs to begin to establish a digital preservation strategy for UGC; which also needs to incorporate a publication statement for the guidance of contributors. Given the collaborative nature of the material, it is clearly essential to involve contributors and users as much as possible in creating this strategy¹³. Such a dialogue could usefully focus on questions of user motivation in creating the material, what use subsequent users would be likely to make of it and how they might wish to access it, whether value judgements in selection for preservation would be appropriate, and if so, by whom, and which, if any, categories of UGC are likely to have enduring value.

Clearly cultural institutions will wish to reserve the final judgement in determining a UGC preservation policy, so as to ensure it aligns with the core strategic objectives of the institution. However, once determined, it is hoped that it carry the same degree of transparency and trust in the Web 2.0 world as is conveyed in the old world by the 'Adopt a Book' scheme at the British Library, where contributors receive the undertaking, presumably intended to inspire maximum confidence: 'We will insert a permanent bookplate into your chosen book, recording your generous support for as long as the Library exists.'¹⁴

¹⁰ For the view that digital objects should be preserved only for as long as they are judged to have continuing value, see the National Library of Australia's portal Preserving Access to Digital Information, *Selection*, <http://www.nla.gov.au/padi/topics/9.html>; for the contrary view that once material is selected for preservation it should be kept for ever, see Monica Blake, David Haynes, Tanya Jowett, David Streatfield, *Responsibility for Digital Archiving and Long Term Access to Digital Data*, British Library Research and Innovation Report 67, Joint Information Systems Committee of the Higher Education Funding Councils, 1997, also available at www.ukoln.ac.uk/services/elib/papers/supporting/pdf/digital-preservation

¹¹ For BBC Terms and Conditions, see <http://www.bbc.co.uk/guidelines/editorialguidelines/advice/videoaudioandstills/paymentformater.shtml>

¹² S. Greenfield, *Tomorrow's People: how 21st century technology is changing the way we think and feel*, London: Allen Lane, 2003

¹³ For advice in engaging in such a dialogue with information creators, see Laurie Hunter, 'Investment in an Intangible Asset', DCC Digital Creation Manual, S. Ross, M. Day (eds), 2006, from <http://www.dcc.ac.uk/resource/curation-manual/chapters/intangible-asset>

¹⁴ <http://www.bl.uk/about/cooperation/adoptstep2.html>