Terras, M. (2010). "Should we just send a copy? Digitisation, Usefulness, and Users." Art Libraries Journal, Vol 35, no.1 "Plus ça change? Papers from ARLIS/UK & Ireland's Cambridge conference 2009: Tradition and transformation".

Should We Just Send a Copy? Digitisation, Usefulness and Users

Melissa Terras
Reader in Electronic Communication
Department of Information Studies
University College London
m.terras@ucl.ac.uk
@melissaterras

In February 1963 the Mona Lisa visited the United States of America: spending a total of seven weeks at National Gallery of Art in Washington, DC, and then the New York Metropolitan Museum of Art. In that time 1,600,000 visitors – more than 30,000 viewers per day – filed past the painting. There are reasons for this: touted as the most famous painting in existence, this was the chance for anyone across the pond to see La Gioconda's smile at a time where air fares to Europe were still prohibitive to tourist travel. In the 1960s, the quality of print colour reproduction was relatively poor, and this was a chance to see the famous painting of the wife of Francesco del Giocondo in the flesh. Beyond this, human beings are pack animals, and are designed, primarily, to look at the world around us. What better thing was there to do but to queue in the snow to see, for yourself, the most famous piece of art in the world?

Above 50 percent of the human brain is estimated to be devoted to processing visual information (Terras 2008). Images have been part of human society, psychology, and imagination since the earliest times, before the dawn of written language. The word "image" stems from the Latin *imaginem*, or the later French *imagene*, expressing ideas of imitation, copy, and likeness, but also of thought, conception, and imagination. Research into how we see, perceive, and interpret the world around us, and our reaction to images, is wide, varied, and spans an interdisciplinary reach encompassing psychology, biology, physiology, chemistry, physics, philosophy, and beyond (Gregory 1998). Our enjoyment of visual art is an extension of this complex physical and perceptual activity: we can respond to visual stimuli in various

emotional, cognitive, and physical ways. Looking at art is one way to allow our advanced brain-eye systems, which evolved to allow us to survive effectively in a hunting and hunted environment, to be exercised in a playful manner (Livingstone 2002).

In 1963, a young artist aware of the nature of images, copies and fame, who was just on the cusp of becoming an era-defining artist himself, did not bother queuing to see the Mona Lisa. Instead, Andy Warhol commented "Why don't they just have someone copy it and send the copy? No-one would know the difference" (Hughes 2006, p. 223). He went onto produce a series of serigraphs and screenprints ("Double Mona Lisa", "Two Golden Mona Lisas", "Thirty copies are as good as one") which reduced the impact of this famous painting –and therefore all famous images - through repetitive reproduction.

Of course, we've known how to produce high quality copies of images for a surprisingly long time, with realistic copies being produced since the invention of photography in the early 19th Century, and following shortly after this came the development of electronic and digital imaging technologies, tied to advancements in communications such as the telegraph and the telephone (Terras 2008). An easy way to break up an image into individual parts is to lay a grid over it, and deconstruct it into discreet dots separated into lines. If the size of the grid is known, and a numeric value is assigned to the colour in each square (or pixel, "picture element") in the grid, a numeric, or digital, representation of an image has been produced which can be transmitted, copied, and reconstructed elsewhere. Make the grid small enough – around 50 dots per inch – and the complex human eye-brain system will see the grid as a continuous tone image. Delivering an image in such an encoded format is easier than trying to transmit the continuous, analogue representation of a photograph, drawing, or document, and all digital imaging technologies fool the human eye into seeing this representation as they would the source image.

In 1963, engineers and computer scientists were also working on something special. Telecommunications links were established between the few available computers in the United States to establish data sharing between remote sites. This network

became Arpanet, then the Internet, and eventually hosted the World Wide Web. Over the next thirty years, computer scientists and technology firms developed more efficient, cheaper computers, networks and peripheral devices (Naughton 2000). By the close of the twentieth century, a number of converging technologies encouraged the uptake of digital photography and digital imaging. The availability of desktop machines to view and manipulate images, the networked environment provided by the Internet for the sharing and dissemination of images, and the availability of cheap computational peripherals, such as scanners, digital cameras, and digital mobile phones, all encouraged the production and use of digital images, by both individuals and institutions. This was followed by a corresponding rise in user expectations regarding the quality and amount of digital information that should be able to be produced by desktop machines, or be available online (Terras 2008).

Most libraries, archives, museums and "memory institutions" are now involved in the process of Digitisation – the production of digital representations of analogue objects. The rhetoric of digitisation goes something like this: it will improve "access" to cultural heritage (the term "access" rarely being defined), allow the virtual reunification of dispersed collections, facilitate novel research, and encourage greater use of our institutional collections, albeit from afar (Deegan and Tanner 2002, p.32). Millions of pounds, euros and dollars have been invested in digitisation programmes since their routine rollout across the Library, Archive and Museum sector, beginning in the late 1980s (Lee 2002, Hughes 2004). We now routinely send digital copies of documents, manuscripts, art works, music, and artefacts to any user who requests them. Yet it is hard to articulate, with quantifiable evidence, the impact this had had on scholarship, and our use and understanding of cultural heritage material. Indeed, after over twenty years of digitisation, it is time to pause and ask should we just "send the copy"? What issues are emerging about our understanding of the impact, relevance, and success of our massive investment in providing digitised content?

Firstly, let us consider the nature of digitised copies. A digitised artefact can never replace an original – it is a surrogate, or a digital interpretation of an analogue object (Deegan and Tanner 2002). This, necessarily, results in some loss of information about the object itself, and usually its surrounding context. Viewing a picture of the

Terras, M. (2010). "Should we just send a copy? Digitisation, Usefulness, and Users." Art Libraries Journal, Vol 35, no.1 "Plus ça change? Papers from ARLIS/UK & Ireland's Cambridge conference 2009: Tradition and transformation".

Mona Lisa online, shrunk to a two dimensional screen, is a very different experience from visiting her in context in the Louvre. Some would say it is preferential – her Salon is usually crowded, noisy, and she is shielded behind bullet proof glass from hourdes of marauding tourists (and the occasional vandal). We do not get to experience this particular painting up close and in detail – which we can with the digital copy – but most who have seen common reproductions of the Mona Lisa are surprised by how small she is in real life: something which is difficult to convey in the digital realm. Additionally, considering the Mona Lisa as a three dimensional object is different to viewing her as a flat image. We all know what the Mona Lisa looks like – but who has seen the back of the Mona Lisa? Yet the back of this painting contains important information regarding condition which affects the appearance of the image. What do we lose in digitisation by often treating cultural and heritage artefacts as objects that can be summed up by one digital image taken from one particular angle? If we send a copy, can you tell the difference?

A further major issue, stepping away from the issue of how best to represent analogue objects in the digital realm, is how we can encourage users to find, and use the vast amounts of digitised cultural and heritage content that has been created, at great expense, by both institutions and amateur enthusiasts. The majority of the digitised content that has been created so far is offered as digital images of documents, objects and artefacts. Yet it is hard to search through banks of digital images: understanding and interpreting images is a huge field of computer and engineering science, with many problems left to solve. Despite much research, automated image processing based searching through Content Based Image Retrieval is still not a successful enough approach to allow widespread implementation. Instead, most collections of digitised content are presented to the user with a textual search box: what would you like to see? Users have to know the terms to search for, and know what they are looking for, and these terms have to be already created and stored, alongside the images in a database, for users to find.

This poses us with a fundamental disconnect between our databases and user knowledge, and between images and text. Think of the Mona Lisa. You may know the title, artist, and location, and potentially the date and medium of this particular

painting. You may also be able to think of terms that could be related to the historical and art historical context of the painting, such as Renaissance, landscape, portrait, etc. These may serve as the terms you could start with if you wanted to find this, or similar paintings, in a database. Now look at the picture presented at the top of this article. What can you tell about the title, artist, date, location, and genre? If you were to try and search for this in a database, where would you start? How can you describe something in textual terms if you do not know what terms to provide?

There is no reason for anyone to know the title, date, or provenance of the above oil painting. It was picked up for pennies in a flea market in Latvia (so if anyone has any ideas, please get in touch) but serves to demonstrate an important issue: "an image… is beautiful, but not very useful, before it is connected to a description" (Thaller 1992, p. 2). By providing structured terms about an information object in defined fields, the condensed representation of the object can be used to find, search, and manage digital objects (Lazinger 2001). Over the past thirty years, metadata ("information about information") has become a busy, broad field of interest to information professionals and data managers, and a great deal of work has gone into resolving issues of metadata and cross record searching, and comprehensive access to information for users. Many different metadata schemes and structured vocabularies have been created to provide efficient and effective description of objects to enable resource discovery (Baca 2002). However, providing descriptions of digitised content in the creation of structured metadata is often as time consuming as creating the digital image surrogate of an object itself.

Additionally, studies have confirmed that this provided metadata does not often match with the terms users enter when searching for material. In 2005 the Metropolitan Museum of Art carried out an experiment where volunteers supplied keywords for thirty images of popular artworks. When compared with the museums curatorial documentation, it became apart that 80 percent of the terms provided by the general public were not represented in the structured metadata. For example, Joachim Friess's sculpture "Diana and the Stag," was tagged by the users with the expected "antler," "archery" and "huntress," but was also tagged "precious" and "luxury." (O'Connell, 2007). As a result

> our collections databases isolate objects; … collections documentation is recorded in the specialist discourse of curators and registrars, a language unfamiliar to members of the general public, whose searches against that information often fail to yield results users expect or understand…. Simply put, the access offered by the Web hasn't translated into accessibility" (Chun 2006).

It turns out we are not actually sending our copies very far. Memory institutions have been very reticent to publish statistics regarding the use of digitised online content, but research indicates that a large proportion – perhaps as much as 50 percent – of digitised online cultural and heritage content is never, ever accessed (Warwick, et al 2008). We expect users to brave our clunky interfaces, and guess what terms to provide, to guess what objects there may or may not be within our collections. It is rare for collections to provide other routes into their digitised collections, such as daily blogs, or updates, cherry picking items for users to engage with. Institutions are just beginning to realise the power of many web 2.0 technologies in aiding them to disseminate information about their collections.

Additionally, in a world where the general public is used to shopping in supermarkets for their provisions, and going to centralised search engines such as Google for all their information requirements, our digitised content often sits behind institutional interfaces, disconnected and hidden. Our investment has created a virtual high street of the butchers, the bakers, and the candlestick makers, where the shop content is hidden behind a virtual shop keeper and high counter. We expect users of cultural and heritage material to make a monumental and predictive effort to find any information that could possibly be of interest to them.

And what do we know of our users who do brave our clunky, individual interfaces? Very little. It is exceptionally difficult to gather information about the users of online resources. Online surveys suffer from a low return rate, and the fact that only the keenest users self report. Other qualitative methods – focus groups, institutional visitor surveys, can be difficult to organise, with no guarantee that the views of the small constituency questioned can be extrapolated to cover the opinions of the potentially wide internet audience (Gorman and Clayton 2004). Quantitative

methods, such as log analysis of web server records, or provided log services such as Google Analytics, that record every visit to a website and the IP address of the computer making the request, can provide statistics regarding use. These can be useful in counting visits, determining the path users take through a website, seeing how long visitors stay in particular areas of a resource, and ascertaining which areas of a resource are not found, used, or useful. However, identifying individuals, rather than the address of the computer they log in from, knowing the purpose of the visit, and establishing any resulting offline use of the online information are all impossible using log analysis methods. Additionally, the fact that many institutions are reluctant to provide their web servers logs for further scrutiny, means that user statistics generated from log analytics are not generally available (Warwick et al 2008). There is also no methodological, quantitative or qualitative evidence to suggest what either the general public or academic researchers do with cultural and heritage information they find online. The anecdotal evidence suggests that while some projects become popular and well used, others become relatively quickly forgotten.

We know a little of what users want from a web resource. A study undertaken over ten years ago revealed that users wish for fast retrieval, image and text versions (to conduct full text searches of documents), completeness and legibility of images without scrolling down the page, multiple sizes of images to support differing research needs, consistency across image databases, tools (such as zoom and pan) to deal with large images, simple user interfaces, hypertextual links between structural and intellectual content, and the ability to manipulate digital images (such as zoom, being able to examine images side by side, save searches, results and annotations). Users also need a variety of search functions, which provide a variety of user options to enable cross database searching and, more importantly, a standardised interface and query language across different systems, as different and bespoke implementations of systems can yield very different search results with the same terms, depending on the cataloguing and metadata captured (Sandore 2000). However, a decade on from this study, can we say that our online systems fulfil these basic requirements? Are we doing all we can to help users find our digital copies?

Why does all this matter? Digitisation is a costly business, requiring both capital and temporal investment, for memory institutions who could probably spend the money many times over on other structural, staffing, or collections issues. As a current rule of thumb, the total, full economic cost of digitising large scale collections, of one thousand objects or more, comes in between £70 and £100 per digital image created. Digitisation projects are often only funded for the creation of the digital resource, including metadata and search interface, and the cost of sustainability of the digitised collection falls upon the institution once short term funding is exhausted. Given the costs involved, it is imperative that we do all we can to ensure we are creating resources that are both useful, useable, and used. Without understanding more of the nature and purpose of the use of online material, we are at risk of wasting scarce resources, particularly in the current economic climate.

The funding councils recognise this. Evidence of use is a thorny issue, and one which the research councils and funders are becoming increasingly twitchy about. In the UK, the Arts and Humanities Research Council has recently cited "Evidence of Value of ICT" (Information and Communication Technologies) as one of two main review areas of its research programme (the other being the problematic issue of the sustainability of digital resources after project funding ends), declaring that

> Collecting and highlighting all the evidence relating to this is important not just to persuade funders to support ICT use in the future, but also to increase uptake in the academic community. It should be the main basis for the discussion of future needs and opportunities (AHRC 2007a).

Lack of evidence of use, or uptake of services, means that funding can and will be withdrawn from funding programmes and support services. In May 2007, the AHRC announced it was withdrawing funding from the Arts and Humanities Data Service from March 2008 (AHRC 2007b), although it was a world-leading service and provided much needed advice to those undertaking digitisation, and other ICT related projects, within the UK. It may be that, as further evidence of use fails to materialise, we see a further scaling back of research programmes and services tied to digitisation, after this period of strong growth, and proliferation of digitised content. The question of who will pay to maintain digital collections once they are created, and who will ensure that they are available for use in the future (particularly given the present

funding crisis) is one which is also unresolved. If users cannot find objects in digitised collections, how can they be used? If collections are not used, should they been maintained? If they are not maintained, how can they be used? Is saying that users like to see cultural and heritage material really enough? How can we defend our approach of increasing "access" when we do not know how effective our efforts actually are? Should we just send a copy?

Of course there are other issues that are important when we digitise historical and cultural objects, such as issues of veracity, authenticity, and technical issues such as colour management, longevity of resources, etc (Terras 2008). However, issues of use and usefulness are by far the most pressing. We are at a point where we need to engage with the focus, purpose, and success of our efforts to "send a copy" of cultural and heritage objects to the user community via digital technologies. We need to articulate how and why creating digital surrogates of cultural and heritage material is important, useful, and an essential activity for memory institutions. We need to make better efforts to understand how users can find and use online digitised resources online, and design systems accordingly to ensure their relevance. Otherwise we risk creating and maintaining costly online resources which are not useful, and never used. Furthermore, we are at risk of losing the fund which would allow us to create and maintain our digitised cultural heritage.

There is evidence that there is a voracious appetite for high quality online resources of heritage material when they are made available, but that users do not know what to search for or how to access the riches we can offer them. On the 20th November 2008 the Europeana website (http://www.europeana.eu/) was launched, providing a single point of access to the digital collection of many of Europe's largest and most important memory institutions. By 2010, the site hopes to provide access to 10 million different digitised objects and their related records. However, within a few hours of its launch, the website ground to a crash as more than 10 million users attempted to access it (BBC 2008). Quite clearly, people like looking at shared culture and heritage. People want to be able to find such resources easily online, and a centralised service is progress towards joining up our individual institutional efforts.

Terras, M. (2010). “Should we just send a copy? Digitisation, Usefulness, and Users.” Art Libraries Journal, Vol 35, no.1 “Plus ça change? Papers from ARLIS/UK & Ireland’s Cambridge conference 2009: Tradition and transformation”.

But what were thousands of users searching for simultaneously, contributing to the crash (ibid)? The Mona Lisa.

References

Arts and Humanities Research Council (2007a). “Review Activities”. ICT in Arts and Humanities Research. http://www.ahrcict.rdg.ac.uk/activities/review/index.htm.

Arts and Humanities Research Council (2007b). “Important information for applicants to AHRC June deadline”. AHRC News & Press Releases, 14th May 2007. http://www.ahrc.ac.uk/news/news_pr/2007/information_for_applicants_to_AHRC_june_deadline.asp.

Baca, M. (ed) (2002b). “Introduction to Art Image Access Tools, Standards, and Strategies” Getty Research Institute, Los Angeles. Available at http://www.getty.edu/research/conducting_research/standards/intro_aia/

BBC (2008), “European Online Library Crashes”. Friday 21st November 2008. http://news.bbc.co.uk/1/hi/world/europe/7742390.stm.

Chun, S., Cherry, R., Hiwiller, R., Trant, J., Wyman,B. “Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums”, in J. Trant and D. Bearman (eds.). Museums and the Web 2006: Proceedings, Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2006/papers/wyman/wyman.html.

Deegan, M. and Tanner, S. (2002) “Digital Futures: Strategies for the Information Age”. Digital Futures Series. Library Association Publishing, London.

Gorman, G. E. and Clayton, P. (2004). “Qualitative Research for the Information Professional: A Practical Handbook”. Facet, London.

Gregory, R. L. (1998). “Eye and Brain, the Psychology of Seeing”. Princeton University Press, New Jersey.

Terras, M. (2010). “Should we just send a copy? Digitisation, Usefulness, and Users.” Art Libraries Journal, Vol 35, no.1 “Plus ça change? Papers from ARLIS/UK & Ireland’s Cambridge conference 2009: Tradition and transformation”.

Hughes, L. (2004). Digitizing collections: strategic issues for the information manager. London: Facet Publishing.

Hughes, R. (2006). “Things I Didn’t Know”. Knopf. London.

Lazinger, S. S. (2001). Digital Preservation and Metadata. History, Theory and Practice. Libraries Unlimited, Greenwood Publishing Group, Inc. Englewood, Colorado.

Lee, S. (2002). Digital Imaging, a Practical Handbook. Facet Publishing, London.

Livingstone, M. (2002). Vision and Art, The Biology of Seeing. Harry N. Abrams, New York.

Naughton, J. (2000). A Brief History of the Future: Origins of the Internet. Phoenix Press, London.

O’Connell, P. L. 2007. “One Picture, 1,000 Tags”. New York Times, March 28th, 2007. http://www.nytimes.com/2007/03/28/arts/artsspecial/28social.html.

Sandore, B. (2000) “What Users Want from Digital Image Collections”. In Kenney, A. R. and O. Y. Reiger (2000). “Moving Theory into Practice: digital imaging for libraries and archives”. Mountain View CA: Research Libraries Group, p. 4-5.

Thaller, M. (1992b) “The Processing of Manuscripts”. In Thaller, M. (ed). (1992). “Images and Manuscripts in Historical Computing”. Proceedings of a workshop at International University Institute, Firenze, November 15th 1991. Göttingen, Max Planck Institute. P. 41- 72.

Terras, M. (2008). “Digital Images for the Information Professional”. Ashgate, London.

Warwick, C. Terras, M., Huntington, P., and Pappa, N. (2008). "If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data". Literary and Linguistic Computing.23(1), 85-102.

3897 words

3502 excluding references

Terras, M. (2010). “Should we just send a copy? Digitisation, Usefulness, and Users.” Art Libraries Journal, Vol 35, no.1 “Plus ça change? Papers from ARLIS/UK & Ireland’s Cambridge conference 2009: Tradition and transformation”.