1

Introduction: the value, use and impact of digital collections

Lorna M. Hughes

Background and context
The aims of this volume
A key motivation for developing this volume was the need to address the ‘use’, ‘value’ and ‘impact’ of digital collections in the context of an expanding mass of digital content with tremendous potential. Specifically:

• How can we understand how digital collections are being used, and by whom?
• How do we assess their value, and add value over time, in order to make decisions about which collections to digitize or make available, and how?
• How do we assess their impact on scholarship, on knowledge transfer and on information management and access?
• How do we ‘measure’ value? What can be measured, and how?
• Is it possible to ensure their sustainability, value and impact over time?
• How might we apply indicators of use, value and impact to inform funding decisions and policy making for the future?

The explosion of digital initiatives
This volume of essays is, in many respects, a follow up and companion to Digitizing Collections: strategic issues for the information manager (Hughes, 2004). Since its publication, digitization initiatives have continued at a tremendous pace in libraries, archives and museums, as well as in higher education. Digitization of existing library, museum and archive collections is still a major priority, where funding can still be found for these initiatives. The National Library of Wales is continuing to digitize two million pages of historic Welsh newspapers and journals in a three-year project that will conclude in 2012, with funding from the Welsh Government. The British Library is partnering online publisher Brightsolid to digitize up to 40 million pages of newspapers. In the USA, the Smithsonian Institute has a remit to digitize its entire collection: a challenge considering that the Smithsonian is home to 137 million objects, 100,000 cubic feet of archival material and 1.8 million library volumes. In the UK, the Joint Information Systems Committee (JISC) has launched a new phase of digitization via its e-Content programme, for projects that will be completed in 2012. This is the era of Google, and mass digitization initiatives to put our cultural heritage online are flourishing. This has been described as a ‘data deluge’ (Hey and Trefethen, 2005), and it has a huge impact on scholarship, teaching and public engagement.
Digitization in an economic downturn

This is an auspicious time to take stock of this mass of digital content, and consider its impact, value and use. The global economic decline that began in 2007 has led to serious cuts in funding for almost all humanities and cultural heritage initiatives, including the development of, and support for, digital collections. In the arts, humanities and cultural heritage world, especially in the UK, the threats to funding have become a reality, with the closure of the Museums, Libraries and Archives Council (MLA), cuts to the Arts Council, and to Research Council funding overall. In the USA, there have been calls to withdraw federal funding from the National Endowment for the Humanities. This economic ‘austerity’ has created significant institutional and societal pressures on cultural heritage and higher education organizations.

Partly as a consequence of the reduction in funding, we have seen a sharper emphasis on the need to demonstrate the ‘impact’ of publicly funded resources and research, as a means of quantifying the value of the investment in their creation. Research councils and funding agencies, notably the Arts and Humanities Research Council (AHRC) in the UK, have placed an increasing emphasis on ‘impact’ and ‘evidence of value’ of all research that they fund (AHRC, 2006) for several years. This focus on understanding the ‘evidence of value and impact’ of digital research and collections in the arts and humanities was one of the reasons that AHRC ICT Methods Network was funded from 2005 to 2008. This was a national organization that provided a forum for the exchange and dissemination of expertise in the use of information and communication technologies (ICTs) for arts and humanities research. In its final evaluation report, the Methods Network was able to provide a considerable body of evidence that there was indeed ‘evidence of value’ of the use of digital collections for scholarship and research in the arts and humanities:

The new research that has been enabled by ICT . . . has depended upon the development of new kinds of resources, such as large corpora in literary, linguistic, musicological, and television and film studies domains, the digitization and digital-encoded representation of materials in classics, history, literature and history of art, and the creation of databases in archaeology and the performing arts. This recognition that the future generations of scholarship in the arts and humanities will depend upon the accessibility of a vast array of digital resources in digital form is becoming more widespread. (Hockey and Ross, 2008)

Since then, the Browne review (Browne, 2010) has called for evidence of the ‘value’ of the arts and humanities to society, and there is increasing pressure for scholarly research to demonstrate economic and social impact, despite the fact that the ‘economic benefit of the arts and humanities’ is a topic for which there is little hard evidence at this time. Nonetheless, the AHRC has produced two publications that set out the arguments for the ‘value’ of the arts and humanities, in the recent publication edited by Jonathan Bate, The Public Value of the Humanities (Bate, 2010) and in the AHRC report Leading the World: the economic impact of UK arts and humanities research (AHRC, 2009). A noteworthy web-based initiative, 4Humanities (http://humanistica.ualberta.ca), organized by leading digital humanities scholars, has also been set up as an advocacy organization for the value of digital collections and methods in the humanities.
The need for open and useable digital collections

Another recent development that provides part of the background to this volume is the increasing focus on the value and use of ‘open’ scholarly resources for research. While publicly funded research outputs are intended to be freely available (even if this is not always the case, and as Robey discusses in this volume, the demise of the Arts and Humanities Date Service in the UK in 2008 has made it harder for publicly funded resources to be used and sustained), digitization that has been funded by commercial entities is frequently subject to licensing and use restrictions. Notably, in the UK the digitization and redistribution of the census materials and similar ‘name rich’ resources has been made possible by commercial entities including Ancestry.com and Find My Past.

While these efforts have digitized and made available an enormous range of primary source materials, there is some concern that resources that are not ‘open’ are less valuable for scholarship. This is of particular concern, as the availability of large-scale, distributed collections allows new approaches to scholarship in a number of disciplinary areas. The idea of taking a ‘big picture’ approach to historical and cultural issues, working with large-scale data across disciplines, is gaining ground in scholarly enquiry once more. This is the type of research highlighted in the American Council of Learned Societies report, emphasizing the opportunities ‘to reintegrate the cultural record, connecting its disparate parts and making the resulting whole available to one and all, over the network’ (ACLS, 2006). Underpinning this re-integration, of course, is the principle of freely available and open data for aggregation, use and reuse: ‘the full range of online content needs to be made available to all, quickly, easily and in a form appropriate to individuals needs’ (JISC Strategic Content Alliance, 2010).

Some organizations have made the principle of providing freely available digital content where possible (see Prescott’s discussion of the National Library of Wales in this volume), but this is not always possible, especially in the current economic climate. There are various approaches: can data be chargeable for a time to recover the costs in making it? Under what conditions can cultural heritage content be monetized? This volume is not about business models for digital content, but there is an implicit understanding that ‘price’ and ‘value’ could become interconnected – people may have to pay for valuable resources. There are also issues related to sustainability – the digital content must be sustained as long as charging models are in place – which adds to their cost. In a related development, commercial entities have raised concerns about digitization of out of copyright materials in libraries (Sabbagh, 2010). However, these discussions lack the evidence of the actual economic ‘value’ of digital resources, and work is under way to address this balance, developing a ‘deeper understanding of the social and economic impact of digitization’ (Hargreaves, 2011).

The cost of digital creativity

While the economic data is incomplete – and this volume is not a quantitative analysis of business models for digitization – one thing that is very clear is that digital collections are expensive to develop, manage and sustain over the long term. This has been
documented in two reports commissioned by the AHRC: its *Resource Enhancement Scheme* (AHRC, 2005) and *Sustainability of Digital Outputs from AHRC Resource Enhancement Projects* (Denbo, Haskins and Robey, 2008). The latter report demonstrated that ‘The average grant to projects with digital outputs was £228,155, as against £153,090 to those without: the total amount invested in those with digital outputs was therefore almost £40m.’ Furthermore, there are hidden costs to developing digital resources – institutional overheads in keeping collections online and visible, and in the allocation of resources to digital programmes, frequently at the cost of other activities. In academia, the costs are less hidden – there is a body of evidence (see, for example, Raban, 2007) demonstrating that the development of digital resources by academics can hinder their career progress, and even jeopardize promotion or tenure. It is notable the research findings presented in this volume by Ross, Terras and Motyckova show that the largest numbers of academic users of the British Museum Collections online database were either postgraduate students or professors – either those just starting out, or those secure in their promotions. Impact, of course, can be negative as well as positive, so it is timely to assess the pressures that developing, sustaining and using digital collections can place on organizations and individuals.

**Protecting the digital investment**

Given the current economic climate, and the pressure on funding for digital collections, addressing these issues is crucial in order to make the case for protecting existing digital heritage, and for increased digitization. New digital projects must have compelling and visible impact. It is necessary to be creative about the sort of funding available, and to look further afield, both internationally and in disciplinary focus, for funding. In this regard, the economic downturn has been a way of incentivizing collaboration, enhancing existing resources, and aggregating content, as well as consolidating expertise in a few centres. Research Council funding in the UK has also been mostly contingent on developing digital resources that are driven by research challenges, rather than resource enhancement or access, which may well create resources that have more demonstrable ‘value’ for scholarship. This has been a ‘carrot and stick’ approach – the carrot being the opportunity to apply for the limited funding still available, the stick being that this is the only way that funding can be awarded. Equally pragmatically, as it becomes more difficult to get funding for digital collections, there is a need to make better use of those we have, and to ensure that they are fully embedded in research, teaching and public engagement. There is a need to show clearly – to funders, to the public and to those responsible for their long-term sustainability – that digital resources can enable scholarship that generates new research questions and findings; that they make it easier and more efficient to carry out ‘traditional’ scholarship through better and enhanced access to resources (Lehmann and Renfro (1991) and Wiberley (2000) suggest that humanities scholars are receptive to technology as long as it demonstrates adequate savings in time or effort); and that they extend the evidence base for research. By expanding the cultural heritage collections available to the public, they also have economic and societal impacts. However, this
Evidence is not readily available, as outlined in the keynote address at the Digital Humanities Conference in 2010 (Terras, 2010), where the case was forcefully made for better, more readily available exemplars of the use and value of digital collections and research.

**Evidence of use and impact**

Some evidence exists for the use of digital resources for research, education and public engagement, thanks to initiatives like the JISC e-Content programme; the Methods Network; and the National Endowment for the Humanities Office of Digital Humanities, notably the ‘Digging into Data’ initiative, which seeks to use large digital corpora to address key research challenges in the humanities (see hwww.diggingintodata.org). The AHRC ICT Methods Network produced a body of evidence (documented in the final evaluation of the programme) that demonstrated that digital resources were having a transformative effect on scholarship, and that this was a vibrant, fast-expanding area. However, as the final evaluation stressed, the full impact of the Methods Network and, indeed, the impact of digital resources in the arts and humanities, will be over the longer term (Hockey and Ross, 2008). But, since the demise of the Network, there has been no attempt to systematically address these issues across the disciplines. A longitudinal, multi-disciplinary study on use and impact of digital resources is still needed.

**Value**

One of the key issues is being able to provide a definition of ‘value’. Digital resources are valuable to different audiences for different reasons, and some value may not be realized immediately. Digital collections come about for different reasons. Many research projects, for example, have produced digital images or digital text as a by-product of scholarship, and a need to put these digital images online for public access may not be the first priority of the project team. For example, ‘The Visual Culture of Wales’ Project (based at the University of Wales Centre for Advanced Welsh and Celtic Studies) was funded mainly through the Arts Council of Wales Lottery Unit and the University of Wales from 1996 to 2003, as a publication project. The photographic archive contains digitized photography of around 3000 works of art, dating from the fifth century up to the 1960s, in a wide variety of media. The database for the images was subsequently made available online, with minimal funding, as a by-product of the research publication project. Similarly, name-rich sources (parish records, census records, baptism records) have been digitized by the Church of Latter Day Saints for theological reasons. The digital files may be invaluable for family history, or demographic research, but this was certainly not the motivation for their creation. Other digital resources, including born-digital materials, are being preserved for future use – the Library of Congress, for example, has archived the Twitter archive, which could prove just as useful to scholars of the early 21st century as 19th-century newspapers online have proved to political, social and cultural historians of the newspaper era. In this regard, digital resources are just like any other cultural artefact – historical resources that are used for research were seldom
intended to be ‘historical resources’; they were administrative records, ephemera or cultural artefacts. They tell us more than they were ever intended to. This is a useful truth to apply to digital content. As we create more and more digital content – as part of conscious attempts to digitize analogue materials, as by-products of research or as part of our daily digital mode for work and communication – we are creating, maintaining and sustaining ‘digital content’ that may have value to future scholars.

A digital text may be valuable to a scholar because it enables the use of text mining tools to undertake historic research, or it may be valuable to family historians as it mentions an ancestor, or to scholars of material culture through its description of objects. It may gain value if linked to other digital content through ‘virtual reunification’, where collections held in disparate archives around the world can be combined in digital facsimiles. Value is subjective, changes over time and has different meanings that are contingent on external factors. Value of digital collections, and digital humanities in general, is particularly difficult to assess as ‘Digital Humanities is not a unified field but an array of convergent practices’ in multimedia configurations; using digital tools, techniques and media have altered the production and dissemination of knowledge in the arts, human and social sciences, creating ‘digital models of scholarly discourse for the newly emergent public spheres of the present era (the www, the blogosphere, digital libraries, etc.), to model excellence and innovation in these domains, and to facilitate the formation of networks of knowledge production, exchange, and dissemination that are, at once, global and local’ (Schnapp and Presner, 2009).

This is, of course, a common issue across the arts and humanities: funded research often does not show its ‘value’ until it is broadly disseminated, shared, cited and becomes common across the disciplines. This can take time. It is the same with digital resources, and funders often fail to appreciate that the ‘value’ of digital collections and the scholarship they enable may take time to emerge. The chapters by Prescott, Hudson and Oliver are interesting in this regard as they take a longer view of the impact of digital collections and programmes on the institutions that host them.

Some of the factors that make digital collections ‘valuable’ to libraries, archives, museums and higher education were addressed in Digitizing Collections: strategic issues for the information manager (Hughes, 2004). These included:

- access, both broader access to a global audience via the internet, and enhanced access, by making aspects of collections searchable, findable and linked to related materials
- supporting preservation, by providing digital surrogates of rare and fragile materials
- collections development, by enabling organizations to develop cataloguing and records management around digital objects, and by enabling the ‘virtual reunification’ of collections that are physically separated
- institutional and strategic benefits, such as professional development of staff; the prestige and PR value to the institution; and enabling the institution to fulfil its goals of access and outreach
• supporting research and education across the disciplines.

The book also investigated the impact of digital collections on institutions, including:

• the need for new business models to support and develop digital collections, and how this was forcing institutions to consider the institutional costs and benefits of digital collections, grappling with the contradiction that there can be indirect cost savings from digital delivery of services, but that these are offset by the increased costs of digital access and preservation
• the resources and intellectual implications of changes to the way that information is used and managed
• the implications of supporting entirely new approaches to scholarship and access.

The book concluded that there was no definitive evidence base that could provide concrete numbers about the economic ‘value’ of digital collections, but there were some interesting debates around the topics at the time, for example, ‘The Economics of Digitizing Library and Other Cultural Materials’, which identified the ‘costs’ of digitization: institutional, technological and legal (Waters, 2004).

The structure and content of this volume
This volume has three parts. The first, ‘Digital transformations in libraries, museums and archives’, describes the use and impact of digital collections in libraries, museums and archives. Andrew Prescott, Claire Hudson and Gillian Oliver provide valuable examples that capture practice and experience in a variety of organizations, discussing ways that the digital revolution has transformed the mission and organization of services and collections. The second part, ‘Understanding and measuring the use, impact and value of digital collections’, presents different approaches to measuring and understanding value. Ben Showers describes some approaches developed through the JISC e-Content programme to measure impact and to embed digital collections more thoughtfully within user communities. Milena Dobreva, Andy O’Dwyer and Leo Konstantelos present a series of approaches to user testing that can inform the development of more valuable resources, and discuss how this can impact the development of business models. Claire Ross, Melissa Terras and Vera Motyckova discuss a survey of users of the British Museum Collections Online Database, and develop an analysis of the use of digital content by academic researchers. Simon Tanner concludes this section with an overview of the opportunities and impacts that digitized resources have made for learning, teaching, research and society.

The third part, ‘Enhancing the future impact and value of digital collections’, presents some approaches that can add value to digital collections: Lorna M. Hughes describes ways in which they can effect transformative research via the use of ICT tools and methods; Ann Borda and Lyle Winton discuss the ways in which e-infrastructures can enable collaborative scholarship and shared resources, considering how preservation and
research infrastructures and strategies developed for the sciences can be applicable to other disciplines, and citing the exemplary work that has been funded in Australia in order to support this; and David Robey discusses sustainability and the long-term value of digital resources.

A concept with considerable impact is the ‘holistic’ approach to digitization, which sees the whole digital life cycle as significant and interdependent, rather than as a series of individual phases or stages. For illustrations of the concept, see, for example, the work of the Digital Curation Coalition in the UK (www.dcc.ac.uk), and the recent report by the Strategic Content Alliance, *Developing Digipedia: a guide to the digital content lifecycle* (JISC Strategic Content Alliance, 2010). Taking this approach at the outset of a digitization project should make the resource more ‘valuable’, as it will enable use and reuse to be embedded into the resource at each phase of development (Hughes, 2008). However, many digitization projects do not develop this approach. Therefore, at a practical level, this volume covers a number of useful and valuable things that the managers of digital collections can do at any stage of the digital life cycle in order to develop a better understanding of, and to increase, the use, value and impact of their resource. In terms of documenting resources, making their impact more visible and providing badly needed use cases (and guidelines for use, which are equally important) to demonstrate the transformative effect of digital collections, the five modes of value suggested by Tanner are an extremely useful taxonomy that could be useful for categorizing and documenting projects and their outputs. This model could well be adopted as a means of documenting ‘impact’ while projects are in progress, or applying to published digital collections retrospectively. Similarly, the ‘impact framework’ that Showers describes could well be adopted by project managers as an internal documentation exercise, or for dissemination. Showers also describes the implementation of the useful Toolkit for the Impact of Digitised Scholarly Resources (TIDSR) toolkit by projects funded by the JISC e-Content Impact and Embedding Programme. This is an easily adoptable methodology for using both qualitative and quantitative methods for assessing the use of a digital resource, which can produce useful findings. Similarly, the information manager might adopt the approach described in detail by Ross, Terras and Motyckova to understand user information-seeking behaviours, in order to get a deeper sense of usage patterns. Ross, Terras and Motyckova, and Showers describe approaches that can be applied at any time after a resource has been published, while Dobрева, O’Dwyer and Konstantelos describe in detail the value of applying usability and user engagement methods into the development of a digital resource. The importance of this cannot be underestimated, as usability is inconsistently factored into the project development cycle.

However, it is also possible to make resources more valuable and useful after they have been publicly available for some time. The JISC projects described by Showers are examples of ways in which projects can be enhanced by becoming more embedded in research, teaching and public engagement. Hughes describes some of the ways in which digital resources can underpin scholarly use through the application of ICT tools and methods. From a strategic perspective, Borda describes how data management and use
can be supported through research infrastructures and in virtual research environments, to foster collaborative and individual use and reuse of existing collections. At the heart of all this work is of course the expectation that resources are open to allow use and reuse for unforeseen purposes. It also supposes that the resources will be available to all. The FEDORA (Flexible Extensible Digital Object and Repository Architecture) commons community (http://fedora-commons.org) now favours the term ‘durability’ to describe robust digital resources that are sustained and preserved over the long term for use by many communities and purposes – and this approach would indeed add ‘value’ to digital resources. Sustaining durable digital resources over the long term will add institutional overhead costs to digital collections. However, the visibility of digital collections means that they must be addressed, unlike the forgotten ‘unseen’ costs to libraries of preserving and managing the monograph.

Conducting user assessments and studies, developing use cases and developing ways of embedding digital collections more effectively in teaching and research are, of course, activities that require the time and effort of staff, and therefore additional investment after a resource has been published. This is often a daunting prospect – digital projects are usually developed through short-term funding, with staff on short-term contracts. Therefore, after they are launched, they are seldom given any more than the most cursory technical attention. The projects described by Showers have attempted to redress this, by providing small amounts of JISC funding to ‘revisit’ projects to enhance usability or embedding, and this approach is an extremely useful one. Institutions may look to ways that they can themselves underwrite this sort of activity – possibly by using volunteers, interns or students to work on the enhancement of digital collections. It is also the sort of activity that users themselves can contribute to – every project has a core community of users, and identifying who they are and working with them to enhance resources could be a valuable opportunity to create a ‘virtuous circle’ of collaboration, engagement and outreach, integrating people and collections.

At the heart of the volume is the theme of digitization as a process with many stages – a digital life cycle – during which issues related to use, value and impact can be factored in. Digital culture is far more complex and encompasses more realities than textual cultures. At an Expert Seminar at King’s College London where preliminary ideas for the volume were discussed, Jean-Claude Guédon introduced this idea with the useful analogy that working with digital documents brings us back to the ways of working that were more familiar in a pre-print time, especially that of individual processing of documents, rather than mass printing. The production of medieval manuscripts is also a useful metaphor for the process of developing digital materials. In the world of manuscripts, communities took responsibility for maintaining texts carefully as a collective endeavour, as texts and the effort that had gone into their creation were easily lost. Texts were maintained due to the efforts of the community to sustain them. Churches had a role as supporters of these communities to preserve texts, and community involvement was essential for preservation: an early prototype of ‘collective intelligence’ before Wikipedia. ‘Use’ was key to sustainability of the resources – texts from antiquity were