# Digital Publishing Seen from the Digital Humanities

## Tobias Blanke

Tobias Blanke is a senior lecturer in the Centre for e-Research, Department of Digital Humanities, King's College London. His academic background is in philosophy and computer science. His principal research interests lie in the development and research of digital libraries and archives as well as infrastructure for research, particularly in the arts and humanities. He has led on several projects in this field, from open-source optical character recognition, open linked data, and scholarly primitives to document mining and information extraction for research. Tobias works on several international projects and committees. Most notably, he is one of the directors of the Digital Research Infrastructure for Arts and Humanities (DARIAH), a European Strategy Forum for Research Infrastructures (ESFRI) project to create an integrated research infrastructure for digital arts and humanities.

tobias.blanke@kcl.ac.uk

## Elena Pierazzo

Elena Pierazzo has a PhD in Italian Philology: her specialism is Italian Renaissance texts, digital editions of early modern and modern draft manuscripts, and text encoding. She has published and presented papers at international conferences on Renaissance literature, digital editions, text-encoding theory, and Italian linguistics. She is currently a lecturer in the Department of Digital Humanities at King's College London, where she chairs the Teaching Committee and is Director of the MA in Digital Humanities; she also teaches XML-related technologies at both undergraduate and master's level. Formerly she was a researcher and tutor at the University of Pisa. She is the Chair and CEO of the Text Encoding Initiative (TEI) and involved in the TEI user community, with a special interest in the transcription of modern and medieval manuscripts. She co-chairs the working group on digital editions of the European Network for Digital Methods in the Arts and Humanities (NeDiMAH).

elena.pierazzo@kcl.ac.uk

## Peter A. Stokes

Peter A. Stokes is a senior lecturer in the Department of Digital Humanities, King's College London, and Director and Principal Investigator of the European Research Council (ERC) funded DigiPal, 'Digital Resource and Database of Palaeography, Manuscripts and Diplomatic'. Besides palaeographical methods and the application of computing to manuscript studies, his primary interests include the vernacular English scripts of the late 10th to 12th centuries. He has also published on computing in lexicography, Anglo-Saxon charters and bounds, and early modern book collectors, and has developed software for digital humanities.

peter.stokes@kcl.ac.uk

**This article provides an overview of digital humanities activities that relate to publishing. Digital humanities is a growing scholarly domain, definitions of which vary but which generally involves the application of computers to research questions that fall within the traditional remit of the humanities. It includes many areas of research that overlap with publishing. An important aspect of digital humanities is therefore to question assumptions that digital publishing should produce faithful visual reproductions. We argue that this cannot be the only objective of digital humanities publishing, and rather that publishing needs to be understood as a range of modelling activities that aim to develop and communicate interpretations, whether consciously or not. The article introduces a selection of digital humanities publishing standards and systems that support a flexible digital representations of objects, such as the Text Encoding Initiative (TEI), which emphasizes scholarly flexibility and collaboration.**

*Keywords*: digital publishing, digital humanities, standards, systems, modelling

### Introduction

In this article, we aim to provide an overview of digital humanities activities that relate to publishing. Digital humanities is a growing scholarly domain, definitions of which vary but which generally involves the use of computers for research questions that fall within the traditional remit of the humanities. It therefore includes many areas of research that overlap with publishing, including not only ebooks but also historical and theoretical questions such as how existing books, manuscripts, and other 'publications' (in a broad sense of the term) can best be studied using digital aids and methods; how publications can best be modelled and represented; the nature of text and the relationship of text with its physical (or digital) manifestation; different forms of dissemination online; and so on. The topic cannot be treated fully in an article like this. Our goal here is simply to analyse some of the main approaches, both theoretical and practical, in the hope of promoting dialogue between publishing and digital humanities to the benefit of both.

To begin our analysis of digital publishing and digital humanities, we ask a fundamental question behind the digitization of cultural heritage, namely what it means to publish an existing object digitally in a scholarly way. Indeed, the question arises of whether it is ever possible to 'publish' an object fully, to entirely and completely translate some pre-existing physical object into a single published digital form. In general the answer has to be no. Any publication like this must involve some sort of transformation from physical to digital form; this will always involve interpretation, and this interpretation will in turn become part of the publication. Although it is very easy to think of an object—say, a book or a painting—as being 'put online', what we find online is of course not the book or painting but a representation of it, in the same way that Magritte's famous painting is not a pipe but a representation of a pipe. This is obvious in one sense, but it is also a trap in others: there is a good reason why Magritte's painting is 'The Treachery

(or Treason) of Images', not least because photorealistic images of objects create a strong illusion of the 'real thing' (Kichuk, 2007; Sutherland, 2009; Tarte, 2011). If we are to produce or even use digital resources then we must understand the difference between the two.

In order to progress, then, we must first consider in what respects this digital object is different from the original one. There are some very obvious differences: we can no longer touch, feel, or smell it, for example. If the surrogate is published in print then we may be able to judge its size, but this becomes very difficult in digital form. Furthermore, in both print and digital forms the original colours may be disguised by problems of calibration, whether of ink, camera, or monitor. These may seem like small details but they can lead to real problems: for example, one may be seduced into studying the colours of a painting according to its online representation, but the colours on the screen may be very different from those of the object unless steps are taken to calibrate the two (Craig-McFeely, 2007–2008). Similarly, one may be tempted into trying to do things to the digital surrogate that would only make sense with the original, such as measuring its dimensions from the screen, since the digital image may well be reduced or enlarged.

The list goes on: there are many other aspects of the original which are lost in most published forms, including its historical development—its biography of alterations, additions, repairs, uses, and changes. Some of this can be recorded in a catalogue description or similar, as can the object's provenance and previous uses and owners, but this is necessarily limited. And what of the object's 'aura', which we know from Benjamin (2010) may be lost in reproduction?

That the publication is not the original also raises the question of what it *is*. It may be correct to say that it is coloured ink on a page, or coloured light on a computer screen, but this is overly reductive and no more useful than saying a pipe is an assemblage of wood and plastic. The published object is a surrogate, a representation of the original. It may include only the text, perhaps without any of the original layout, typeface, and so on. In many cases, particularly when the object itself (rather than its textual content) is the subject of interest, 'digitizing' equates to publishing digital photographs, in which case we are presented with a surro-

gate that is largely or even only a *visual* representation. This is particularly evident in large, influential projects such as Google Books but is also apparent in many others, including the Internet Archive, the UK's National Portrait Gallery website, and scholarly projects such as Early English Books Online, the British Library's Digitised Manuscripts website, the William Blake Archives, and many others.

In many of these cases there is no attempt to reproduce the whole object, but only some parts of it—the pages in the book but not its binding, for example, or the front of the painting but not the back—even though the claim may be made to reproduce 'the original' in its entirety (Kichuk, 2007; Sutherland, 2009). This model is so prevalent that it has almost become an unstated definition of 'digitization': as Ségolène Tarte has argued, the goal when publishing scholarly objects in digital form, as often stated explicitly on websites, is almost always to produce 'a faithful visual representation of the signified where the word "faithful" refers to physical appearance rather than to meaning or presence' (Tarte, 2011).

> **The risk associated with a digital surrogate is instead linked to the question of whether it truly is a near perfect substitute for physical examination, and whether its representation is, in Tarte's words, 'faithful'.**

There can be no doubt that even purely visual surrogates, when supplementing real access, have enabled a different kind of scholarship in the humanities. A digital surrogate is, for instance, useful where remote and concurrent access is required, or when tools for analysis can help discover features that a purely physical examination cannot (Unsworth, 2004; Stokes, 2012). The risk associated with a digital surrogate is instead linked to the question of whether it truly is a near perfect substitute for physical examination, and whether its repre-

sentation is, in Tarte's words, 'faithful'. However, Tarte's emphasis on—and perhaps implicit criticism of—our equation of 'faithful' with the visual opens the possibility of other forms of reproduction which may not attempt visual fidelity but may be equally true to the original in other ways.

An important aspect of digital humanities is therefore to question assumptions that the role of digitization and digital publishing is to produce faithful visual reproductions. We argue that this cannot be the only objective for digital humanities publishing, and that publishing needs to be understood as a range of modelling activities that aim to develop and communicate interpretations, whether consciously or not. We will therefore now develop this idea, presenting it along with the standards and publishing environments that people in digital humanities have developed to enable their modelling activities, as well as some current work in the area. Here, we see efforts to create an effectively infinite web-based archive and to develop new open-access models for scholarly publication.

**Not publishing but modelling**

As noted above, in very many cases, 'publishing a complex scholarly object' is interpreted as 'reproducing (some aspect of) its physical appearance', and even more as 'representing the object as a sequence of photographs'. One might, then, propose measuring the quality of the digital publication in the information loss that the digitization involves for the consuming agents. The aim here is therefore assumed to be to represent the physical object as accurately as possible, without noise. Publications looking for faithful surrogates like this are often useful, and they probably do provide the most value to the most people and so make sense for public libraries, archives, and galleries whose remit is more or less explicitly to serve as many people as possible. However, there are two main points to be made about this. First, this faithful reproduction is one very specific view of what the original object 'is', but it is not the only one. Second, the sequence of photographs is also a format that to some degree was developed out of necessity from print culture, where the possibilities were very limited compared with those available today. In print format, the only options are static images or text, and this in essence tended to lead to limited

forms of publishing objects that essentially lie on the spectrum between these two options, from primarily verbal descriptions (editions, unillustrated catalogues) through type facsimiles, illustrated catalogues, and so on, to relatively text-free photographs and facsimiles. However, as we wish to argue here, the options for digital publication are much larger. The question then becomes how best to take advantage of the new potentials that the digital brings, and how to decide which of the myriad options are best to use.

A useful approach to both questions is to think about digital publication, particularly but not exclusively, in a scholarly context, not just in terms of surrogates but also in terms of models (McCarty, 2004). Human beings are constantly constructing models of the real world. We normally do this unconsciously, but it is a critical part of our cognitive processes. This is true with computers, too: to use a computer, we must make explicit the cognitive models we carry in our head, and we must be able to communicate these models to the machine in order for it to do what we want. But—crucially—the point of a model is *not* to reproduce the original in full. The only way to access the original is to visit the library, gallery, or whatever, and we should not pretend that the digital can achieve the same thing. The model's role should be to simplify, to make a complex case tractable so that we can analyse, manipulate, and communicate it more effectively and not get bogged down in irrelevant details.

In this sense, any publication of an object is also a model of it: it is a simplification, but one designed for a specific audience and purpose. When we 'publish a complex scholarly object', we are making decisions about what information to include and what to omit. Rather than reproducing the original, our aim in this context should be to do things we could not do with the original. That may be simply to allow people to engage with (a representation of) the object they could not access directly, for reasons of fragility for example, or to disseminate that representation to a wider and geographically distant audience. But this discussion also raises the question of moving beyond mimesis. How else could we model these objects?

One example approach to this question from digital humanities is the DigiPal project at King's College London (http://www.digipal.eu/). It involves 'publish-

ing manuscripts online' in a sense, but it makes no claim to reproduce them mimetically in full. The aim is almost precisely the opposite, to deconstruct historical manuscripts into smaller and smaller parts in an attempt to assist knowledge creation through enabling active analysis, comparison, and manipulation. Another interactive alternative is 'reflectance transformation imaging', which rejects the simple image as 'faithful reproduction' (Mudge et al. 2005) and allows the viewer to change the lighting and colour through the way the image is filtered, in order to highlight the three-dimensional shape of the object, including its texture, among other things. Three-dimensional imaging is another obvious possibility, as is 'virtual reality' modelling, where the emphasis is on immersive engagement rather than necessarily producing the most accurate visual image (Beraldin et al. 2000).

In cases like these, the model of the object is very different from what we have seen above. In the earlier example, a book is represented as a set of distinct colour images (which we think of as pages), in a fixed order, which perhaps can be zoomed and panned and which are associated with basic metadata such as a title or shelfmark. In the case of reflectance transformation imaging, however, a 'book' is also a set of colour images, again with zooming and panning, but in this case the images are not of different pages but are instead of the same page photographed under different lighting conditions and with different possibilities of manipulating the different channels. The resulting images typically have very different colours and appearance from the original and so might not be considered a 'faithful reproduction': although it is in fact a highly faithful reproduction of one particular view, this view is not one that most people consider appropriate for satisfactory viewing of the object. However, by removing or changing the colour—in other words, by deliberately altering precisely that which most digitization projects consider essential—we can see the original object in a very different way and gain an entirely new understanding of it.

Once it is understood that humanities publishing can be seen as its own form of modelling, the next question is what kind of formats, standards, and applications do we need to make the digital surrogates more than just faithful representations and to allow

scholarly reflections to emerge. To explore this further, we must move our focus from objects themselves to the texts they contain.

### Modelling standards

From its early days, digital humanities has involved a developing understanding of the importance of standardizing procedures and methodologies in digital publishing, even if framing it as modelling. One of the most important of these standardization efforts in digital humanities is the Text Encoding Initiative (TEI; http://www.tei-c.org/index.xml), the experience of which demonstrates the potential tension between creative scholarly modelling and the need for standardization. In this section, we will discuss the TEI as a short case study.

In the 1980s, the advent of markup languages and their intrinsic flexibility were seen as a threat to the possibility of sharing large data sets, as this flexibility resulted in a proliferation of vocabularies and data models. A group of scholars therefore organized themselves into the Text Encoding Initiative in an attempt to address this difficulty in data-sharing. They obtained support in the form of a research grant from the US Mellon Foundation, the US National Endowment for the Humanities, and the European Union. Since this foundation in 1987, the TEI has continued as 'an international consortium which is dedicated to maintaining the TEI Guidelines as a recommended standard for textual markup' (Cummings, 2008). It was set up with the explicit goal of providing a standardized format for text encoding and is now the de facto standard for text encoding not only in digital humanities but also in many related fields.

The TEI Consortium maintains and develops the TEI Guidelines, an open-ended repository of suggested methods for encoding textual features, primarily for scholarly analysis and publication. In their own words, 'the TEI Guidelines are not only a guide to best practice, but are also an evolving historical record of the concerns of the field of Humanities Computing' (Cummings, 2008). The Guidelines are expressed in XML and allow personalized vocabularies of entities to be encoded. They represent the efforts of a community of scholars who are interested in communicating their modelling activities. It is impossible to name all the

people who have contributed to the TEI Guidelines over its lifespan of more than twenty years, not simply because the people who are officially in charge of editing them have changed repeatedly, but more because the editorial process is community based. Users request, debate, and analyse the need for new elements, attributes, etc. through the TEI mailing list. The TEI Council, an elected organ, then implements those requests that find an appropriate level of community support.

The TEI Guidelines are therefore the product of a large community of scholars coming from all the disciplines that articulate the humanities. Thus, the TEI provides support for the encoding of literary texts, editorial practices, transcriptions of primary sources, description of manuscripts, linguistic corpora and dictionaries, among others. The TEI community of users is organized in special interest groups (SIG), which represent sub-communities within the TEI community; the list of SIGs includes Manuscripts, Libraries, Education, Music, Scholarly Publishing, Correspondence, and several others. For the TEI, diversity is a strength.

The TEI has evolved over the past 25 years from a community solely concerned with standards into a scholarly community that has embraced digital scholarship in order to edit and publish scholarly editions as models and interpretations. This is no ordinary achievement: the field of textual scholarship is characterized by divergent and conflicting theories and methodologies, and yet practitioners of varied convictions have adopted the TEI as their modelling choice. This result has been made possible thanks to what can be called the TEI's 'ecumenical approach', namely the conscious principle that the TEI does not establish what one must do, but, rather, if one wants to do a particular thing then the TEI recommends how it should be done following a scholarly approach to modelling historical texts. For example, if one wishes to label all the personal names in a text, then the TEI states that the XML tag <persName> should be used, thereby ensuring that all personal names are marked in a consistent way, but it is entirely up to the individual whether or not personal names should be marked at all. On the one hand, this allows different scholars to record different pieces of information about the same text—to construct different models of the text, in other words—but it also allows those scholars who do share a common interest

or model to share their material much more easily.

This key choice has made it possible to use the TEI for many different purposes and scholarly approaches, which is precisely what digital publishing understood as modelling requires. Paradoxically, in spite of such an agnostic approach, the Guidelines have succeeded in shaping the way that scholars in digital humanities say and think about editing, despite being in different countries and holding different ideological positions. Allen Renear has summarized it thus:

> The principal goal of the TEI, developing an interchange language that would allow scholars to exchange information, was ambitious enough. But the TEI succeeded not only in this, but at a far more difficult project, the development of a new data description language that substantially improves our ability to describe textual features, not just our ability to exchange descriptions based on current practice. (Renear, 2004)

Renear concludes that the 'TEI is now itself a research community' (Mylonas and Renear, 1999).

The evolution of the TEI from a standard to a scholarly modelling community has enabled the creation of highly sophisticated, interactive scholarly digital editions, which not only respond to scholarly research questions, but also shape them. The Samuel Beckett Digital Manuscript project (http://www.beckettarchive.org/) is a good example. It enables the inspection and investigation of complicated draft manuscripts, presents transcriptions next to or on top of an image, and allows the comparison of different versions of the text at page or sentence level. A recent publication of a small prototype of Marcel Proust manuscripts hints at future developments, which will include the provision of interactive facsimiles and the involvement of users in deciding sequences of authoring and reading (http://research.cch.kcl.ac.uk/proust_prototype/).

The TEI is not the only standard used for the production of scholarly digital editions, but in our experience it is the most productive modelling tool for scholarly work. DocBook is another XML-based standard; although designed for the production of technical documentation (it is used for the documentation of Linux, for instance), it has found other applications, though

even these seem confined to the field of online self-publishing. Perhaps the most commonly used standard for scholarly editions destined for print publication is LaTeX, in its many manifestations. Based on the procedural markup language TeX, LaTeX has been adopted by many scientific journals, which usually maintain (and expect their authors to use) slightly customized versions of LaTex. The use of LaTeX in the production of papers of various kinds is also widely integrated within teaching in scientific disciplines, where many students learn to use LaTeX as part of their core skills. Although widely used in the scientific community, however, LaTeX has found much less use in the humanities and in the production of scholarly editions. Here, the TEI has been perceived to be more apt for the modelling and publication (especially online) of complex scholarly objects, as a result of its emphasis on the semantics rather than presentation of the text.

Despite these apparent advantages, the extensiveness and flexibility of the TEI have attracted a range of criticisms, in particular from the more commercially oriented publishing world, which often needs a more rigid approach. In response to this, the TEI also provides scaled-down versions for specific purposes, such as TEI Lite (http://www.tei-c.org/Guidelines/Customization/Lite/) and TEI Tite (http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html). Despite these customizations, however, it seems that the world of commercial publishing is still reluctant to adopt the TEI. Even though XML is the technology at the centre of many commercial publishing workflows, the TEI has been largely ignored by professional publishers. This owes perhaps to its focus on the scholarly aspect of publishing and the ability to model different interpretations and representations, which is not relevant or even necessarily desirable for commercial publishers. Whatever the reason, this has led in turn to an ideological, if not technological, separation between scholarly and professional editing. This difference has led to scholarly communities developing their own solutions for digital publishing, designed to allow more flexible workflows and to publish the kinds of scholarly models that the community needs. The next section describes some of these efforts in more detail.

**Modelling systems**

Returning to the principles of the first section of this paper, if we do away with the simple understanding of digital surrogates as lossless representations of the real object then the task for systems and applications in digital humanities is to allow the kind of dynamic flexibility that publishing understood as modelling requires. However, this is not the only requirement. The system must also 'understand' the deep annotation work that is enabled by the TEI and that is involved in the creation of digital surrogates in the humanities. In other words, these texts and elements need to be linked in a machine-readable way if we are to create a web-based version of the 'infinite archive' that humanities scholars have dreamt about since the early days of the Enlightenment (on which see Goldstein, 2004). We shall return to this ideal in the final section, but first we concentrate on the experiments and applications in digital humanities which have sought to realize this dream.

> **... these texts and elements need to be linked in a machine-readable way if we are to create a web-based version of the 'infinite archive' that humanities scholars have dreamt about since the early days of the Enlightenment ...**

Organizing this infinite archive is the task of the many digital library and digital archive projects that we are witnessing today. The first generation of digital libraries tried to reproduce traditional libraries, focusing on digitization technologies for 'faithful' digital surrogates and corresponding metadata schemes. The second generation included dynamically enhanced content using Web 2.0 techniques; the general focus, however, was still very much on 'faithful' digital surrogates. Now, we see the third generation emerging, which aims to make digital libraries part of the larger whole of the web of information.

Considering the historically close relationship between the humanities and libraries, it may come as no surprise that humanities researchers have been

engaged with the development of digital libraries and archives from quite early on. Libraries have been the traditional research infrastructure for humanities researchers, who have in turn been active supporters of library developments in their own places of work and study and beyond. Digital humanities has added to the research on digital libraries by concentrating on dynamic, evolving digital archives that enable different interpretations and models of the same digital publications to emerge. The community has fully embraced the second generation of digital libraries and is now actively contributing to research of the third generation.

Although no digital libraries projects for the humanities have yet had the long-lasting impact of some of the more famous science ones such as arxiv.org or PubMed Central, the projects have gained worldwide recognition and are often based on long-term community contributions. The Perseus digital library (http://www.perseus.tufts.edu/hopper/), for instance, is one of the most famous examples of functioning community libraries to date and has gained recognition far beyond its narrower field of Classics. Another example, the Oxford Text Archive (http://ota.ahds.ac.uk/), was founded in 1976 and can therefore be considered one of the earliest attempts to establish an academic digital library. Both projects see themselves as examples of digital libraries dedicated to publishing scholarly models and interpretations, to which scholars are actively encouraged to contribute. From the beginning, they have been their own versions of second-generation digital libraries, as they concentrated on faithful representations but added community input.

The Oxford Text Archive was also part of the now defunct UK Arts and Humanities Data Services (AHDS; http://www.ahds.ac.uk/), one of the first national services dedicated to publishing research materials. With the AHDS, the humanities communities were at the forefront of the use of digital archives not only for publishing texts, articles, and the like, but also for preserving and disseminating research data. The need to publish models and interpretations as well as texts led the AHDS, shortly before its demise, to adopt the digital repository system Fedora (Flexible Extensible Digital Object Repository Architecture; http://www.fedora-commons.org/). Originally developed at the Cornell Digital Library, Fedora has seen widespread adoption

in digital humanities communities for online publication. It was also the main architecture used to support the original US National Science Digital Library (http://www.nsdl.org/) in order to provide online access not only to sources in the sciences but also the tools to work with them.

Fedora's main advantages include its strong 'disseminators', which transform stored data sets on the fly into various formats and outputs (Payette and Lagoze 1998). These are standard Fedora services and have proven to be a core component of the architecture for digital publishing as digital modelling. For instance, an XML document can be linked to services containing XSLT stylesheets in order to transform it into various standard publication formats such as HTML or PDF. More significantly, however, they can also be used to link entirely different models of a digital surrogate. In a previous study (Blanke et al., 2012), we demonstrated how to use Fedora's disseminator infrastructure to store its optical character recognition (OCR) workflow, along with a digitization image. This way we can cluster digitization surrogates in a repository infrastructure according to their OCR representation, where the OCR workflow is realized as another interpretation of the original object.

Islandora takes Fedora to the next level of an advanced scholarly modelling environment by combining it with the popular Drupal Content Management System to develop a platform to deploy virtual research environments (VRE) and complete out-of-the box solutions for digital asset management. Virtual research environments are being actively researched at the moment with dedicated funding from many national and international funding bodies. A full discussion of VREs is beyond the scope of this paper, but a humanities use case is given by Blanke et al. (2010a). Another example is TextGrid (http://textgrid.de/), a VRE dedicated to supporting digital publishing activities in the humanities and funded by the German Bundesministerium für Bildung und Forschung (BMBF).

The trend towards applications that enable reflective humanities publications continues with the digital humanities project Omeka (http://omeka.org/), produced by the Roy Rosenzweig Center for History and New Media at George Mason University. Although the architecture is far less complex than, for instance,

Fedora or Islandora, Omeka has proven very successful in supporting publications from the many small to medium-sized archives that are so typical in the humanities, since it combines simplicity with a strong focus on the needs of humanities researchers. Omeka archives are designed to model the stories of particular historical events, in effect exhibiting them on the web. A good example is the Memorial Stadium project (http://brickhouse.lib.umn.edu/), which narrates the history of the Minnesota Memorial Stadium in the archives of the University of Minnesota. Another example is 'Lincoln at 200' (http://publications.newberry.org/lincoln/), which explores the legacy of Abraham Lincoln 200 years after his birth. In these kinds of projects, the technology must be flexible but also easy to use, since many humanities researchers lack the institutional technical support to run projects with long-term large-scale development.

Humanities collections are often dispersed around the world, and so typical to digital humanities are many small-scale, often distributed digital publications from libraries and archives. Zeega (http://zeega.com/), a tool co-developed by metaLAB from Harvard, is one application that targets this domain, promising to revolutionize distributed storytelling on the web. In contrast to Omeka, it is designed to integrate into the stories documentation and sources from sites across the web. One example of this (http://zeega.com/93075) covers the interactive publication of Underground Explorations by Steve Duncan. Users can browse through Duncan's pictures and receive information on individual parts of them; they can follow different pathways and comment on particular parts of pictures.

Zeega is truly a digital library application of the third generation, insofar as it organizes information from the whole web and is closely embedded in the architecture to understand web content. It is an example of how related content can be brought together and of how digital library services have begun to concentrate on the whole web of information, as discussed above. This idea of the 'web of information' has led to the organization and development of very large databases for the publication of cultural heritage. In the US, digital humanists and the HathiTrust have developed the HathiTrust Research Center in order to provide computational access to public domain publications for edu-

cational and research use (http://www.hathitrust.org/htrc), and the newly founded Public Library of America has in Dan Cohen a prominent digital humanist as its first executive director. European examples include the Europeana project, which collaborates with digital humanities on various levels, currently mainly through Digital Manuscripts to Europeana (http://dm2e.eu/) and the Europeana Cloud (http://pro.europeana.eu/web/europeana-cloud). Alongside these are many more collaborations on local and national levels, such as those based around the German Digitale Bibliothek (https://www.deutsche-digitale-bibliothek.de/), many of which have run as stable services for years.

### Related current discussions

In the final section of this paper, we shall now consider two larger current discussions in digital humanities and digital publishing. As discussed, publishing in digital humanities is typically characterized by many small online publications, each focused on representing a particular model or interpretation or delivering a particular story or view on the web. As we have also seen, many of these models draw on publications in various distributed archives in order for their story to be told completely. The integration of material from dispersed archives is therefore one of the main current research interests of digital humanities. Together, these form an effectively infinite archive, but, in order to be integrated seamlessly, these archives and publications need to be open and easily accessible. The second major theme we want to discuss is work in the domain of open access in the humanities, since this allows completely new forms of digital publications if research data and articles are combined in enhanced publications.

To create an 'infinite archive' of distributed humanities models on the web, these models need to be connected. Researchers in digital humanities have therefore been heavily involved in developing and deploying many distributed digital library and publication solutions. One notable example is the Blacklight tool (http://projectblacklight.org/), developed in Virginia, which allows the integration of heterogeneous data from (potentially) dispersed collections. It provides a new type of enhanced union catalogue and has been successfully used and deployed in numerous archives and libraries, such as the US Holocaust

Memorial Museum. Humanities researchers have also been involved in setting standards for the online distribution and publication of library and archive resources. One of the originators of the widely used archival encoding standard EAD (Pitti, 1999), Daniel Pitti, is also based at the Institute for Advanced Technology in the Humanities at the University of Virginia. The main aim here was to integrate the many smaller digital archives and link them together.

The EAD standard and the Blacklight tool are examples of research in digital humanities the results of which have already been widely adopted. However, connecting archives often requires consideration of the computational models that underlie the web-based publications to be so connected. This is especially relevant and necessary to enable machines to access and process the underlying information in these sources. How to do this effectively has been the subject of much research. For instance, Blanke et al. (2010b) address how digital surrogates can be linked for machines to consume them, following a strict resource-oriented approach. Furthermore, machines are often best served by representations and models that are not appropriate for or tractable to humans. A digital surrogate is attractive because it allows virtually infinite reproduction of the object and eliminates the necessity for physical delivery. In this respect, it should be presented online according to models that are appropriate for humanities researchers to interact with and understand. Digital surrogates are also useful for connecting machines in order to publish scholarly objects on the web for the infinite archive, but the form of the surrogate in this case, the model underlying it and the way in which it is structured, is likely to be very different.

Blanke et al. (2010b) offer three alternative types of digital surrogate. The first used deep annotations in humanities content in order to be found by web search engines. The second presented resources in such a way that they can be processed by machine agents, and the third used semantic web standards to make humanities research visible in the emerging web of scholarly communications. The approaches proposed in this study were found to be useful particularly in cases where existing rights and ethical issues prevented the publications from being fully available in open access, since the digital surrogates were developed in such a way as to prevent the full reconstruction of the original, thereby avoiding those parts which could not be published.

The question of constraints in online publishing leads us to the second major theme for contemporary scholarly publications in the humanities, namely the question of open-access mandates and the future of existing types of journal publishing. Whatever one's personal opinion on the future reach of open-access publications and their impact on the humanities in particular, it seems clear that journal publications will not be the only forms of scholarly communications in the future; most likely not even the most important ones. Scholars and their publishers have already lost the monopoly on research content. Researchers in digital humanities have therefore concentrated on new tools and services to support answers to the perceived crisis in publishing. We have described the transformative impact the digital has had on a whole range of traditional humanities activities. However, the biggest immediate impact on the widest range of humanities research will perhaps come from the transformation of traditional forms of scholarly publication. This implies that digital humanities have a key role to play in redefining the landscape of scholarly communication. Prescott (2012) even sees this as the main 'killer application' that digital humanities will be able to deliver in the next few years. It is no wonder, therefore, that researchers in digital humanities are involved in many projects that map out this new landscape. The French platform revues.org, for instance, explicitly sees itself as an answer to the increasing commercialization of journal and book publishing and offers humanities researchers their own platform to deliver not only high-quality open-access journals and books, but also more informal publications such as research blogs. It has been highly successful and at the time of writing publishes over ten book series and over 370 journals, all of which are open access.

These open-access publication services already serve a large community of scholars. There are also more experimental projects. In the Netherlands, the Dutch Data Networking and Archive Service is working on integrating research data from archaeology and other humanities disciplines into formal online publications of articles. These 'enhanced publications'

(Jankowski et al., 2012) allow a historian to directly embed a historical record from an archive in their journal publication and make full use of the digital's potential to link and integrate. Another good example for experiments in publishing is the Max Planck Digital Library Living Reviews Platform (http://www.livingreviews.org/), which supports open-access review journals in the humanities and sciences. Review journals provide surveys of recent developments in a field, but the distinct feature of Living Reviews is that authors can 'regularly update their articles to incorporate the latest developments in the field. Living Reviews are available online only, enhanced by web features like movies, downloadable source code, or cross-linking to other resources' (http://www.livingreviews.org/).

**Conclusion**

This paper has argued that, for digital humanities at least, publishing a research object has to mean more than relying on 'faithful reproductions'. Publishing objects implies thinking about the audience and purpose; thinking about what we are trying to achieve and how best to do it. McCarty has argued that, 'since a model is pragmatic, the worth of a model must be judged by its fruitfulness' (McCarty, 2003) and not by its verisimilitude. The value of a model, of any model including a publication, is not to reproduce the original but to make it tractable, to communicate something about it: to reduce it to the key aspects that are appropriate for the audience and purpose in question. The digital format allows us to do this in many different and innovative ways. Freeing ourselves from equating the object with accurate photography or skeuomorphic representations will help a great deal in exploiting these new possibilities.

Digital humanities is thus interested in specific publishing activities and modelling. The article has introduced a selection of its publishing standards and systems that support modelling and flexible digital realizations of objects, such as the TEI, which emphasizes scholarly flexibility and collaboration. As part of their work on publishing systems and repositories, researchers in digital humanities have created environments that allow their models to be published online. All the examples we have discussed have demonstrated the intense ongoing experimentation around how to make use of the new possibilities in digital publishing that digital humanities is spearheading. We have, finally, presented how new forms of scholarly publications will develop alongside traditional articles. These will be open access and embedded in a range of humanities models that are published digitally and that include research data alongside or even embedded in articles. As such, many areas of research in digital humanities seem to overlap with those in publishing. It is our hope that this paper contributes to a long and fruitful dialogue.                    ∎

**References**

Benjamin, W., 2010. 'The Work of Art in the Age of Its Technological Reproducibility', *Grey Room*, 39, pp. 11–37

Beraldin, J., Blais, F., Boulanger, P., Cournoyer, L., Domey, J., El-Hakim, S., Godin, G., Rioux, M. and Taylor, J., 2000. 'Real World Modelling through High Resolution Digital 3D Imaging of Objects and Structures', *ISPRS Journal of Photogrammetry and Remote Sensing*, 55 (4), pp. 230–250

Blanke, T., Bryant, M. and Hedges, M., 2012. 'Open Source Optical Character Recognition for Historical Research',

*Journal of Documentation*, 68 (5), pp. 659–683

Blanke, T., Candela, L., Hedges, M., Priddy, M. and Simeoni, F., 2010a. 'Deploying General-Purpose Virtual Research Environments for Humanities Research', *Philosophical Transactions of the Royal Society A: Mathematical, physical and engineering sciences*, 368 (1925), pp. 3813–3828

Blanke, T., Hedges, M. and Palmer, R., 2010b. 'Digital Surrogates for Humanities Research. Digital ecosystems and technologies (DEST)', paper presented at 4th IEEE

International Conference, Dubai, April 2010

Craig-McFeely, J., 2007–2008. 'Digital Image Archive of Medieval Music: The evolution of a digital resource', *Digital Medievalist*, 3

Cummings, J., 2008. 'The Text Encoding Initiative and the Study of Literature', in *A Companion to Digital Literary Studies*, ed. R. Siemens and S. Schreibman, pp. 451–476 (Oxford: Blackwell)

Goldstein, H., 2004. 'The Infinite Archive', *IEEE Spectrum*, 41 (1)

Jankowski, N., Scharnhorst, A., Tatum, C. and Tatum, Z., 2012. 'Enhancing Scholarly Publications: Developing hybrid monographs in the humanities and social sciences'

Kichuk, D., 2007. 'Metamorphosis: Remediation in Early English Books Online (EEBO)', *Literary and Linguistic Computing*, 22 (3), pp. 291–303

McCarty, W., 2003. '"Knowing True Things by What Their Mockeries Be": Modelling in the humanities', *CH Working Papers*, 1 (1)

McCarty, W., 2004. 'Modeling: A study in words and meanings', in A *Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth, pp. 254–270 (Oxford: Blackwwell)

Mudge, M., Voutaz, J.-P., Schroer, C. and Lum, M., 2005. 'Reflection Transformation Imaging and Virtual Representations of Coins from the Hospice of the Grand St. Bernard', in *Proceedings of the 6th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (*VAST'05*), ed. M. Mudge, N. Ryan, and R. Scopigno, pp. 29–39 (Aire-la-Ville, Switzerland: Eurographics Association)

Mylonas, E. and Renear, A., 1999. 'The Text Encoding Initiative at 10: Not just an interchange format anymore—but a new research community', *Computers and the Humanities*, 33 (1–2), pp. 1–9

Payette, S. and Lagoze, C., 1998. 'Flexible and Extensible Digital Object and Repository Architecture (FEDORA)', paper presented at European Conference on Research and Advanced Technology for Digital Libraries, Heraklion, 21–23 September

Pitti, D. V., 1999. 'Encoded Archival Description: An introduction and overview', *New Review of Information Networking*, 5 (1), pp. 61–69

Prescott, A., 2012. 'Consumers, Creators or Commentators? Problems of audience and mission in the digital humanities', *Arts and Humanities in Higher Education*, 11 (1–2), pp. 61–75

Renear, A. H., 2004. 'Text Encoding' , in *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth, 218–239 (Oxford: Blackwwell)

Stokes, P. A., 2012. 'Palaeography and the "Virtual Library"', in *Digitizing Medieval and Early Modern Material Culture*, ed. B. Nelson and M. Terras, pp. 137–169 (Toronto: Iter)

Sutherland, K., 2009. 'Being Critical: Paper-based editing and the digital environment', in *Text Editing, Print and the Digital World*, ed. M. Deegan and K. Sutherland (Farnham: Ashgate), pp. 13–25

Tarte, S., 2011. 'Digital Visual Representations in Papyrology: Implications on the nature of digital artefacts'

Unsworth, J., 2004. 'The Value of Digitization for Libraries and Humanities Scholarship', paper presented at Innodata Isogen Symposium, Newberry Library, Chicago, 17 May