View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by IUScholarWorks

Reading, Writing, Building: the Old English Illustrated Hexateuch

Dot Porter, Royal Irish Academy

This paper was originally presented in the Culture and Technology European Seminar Series sponsored by the Humanities Advanced Technology And Information Institute (HATII) at the University of Glasgow, on January 26 2008. It was presented live in the Royal Irish Academy and was simultaneously webcast. Slide numbers refer to accompanying PDF slide show.

Please cite this paper as **Dot Porter, "Reading, Writing, Building: the Old English Illustrated Hexateuch," paper presented at the Royal Irish Academy for the Culture and Technology European Seminar Series, 26 January 2009.**

Abstract

In recent years there has been a growth amongst humanities scholars in the interest in the materiality of objects including manuscripts, printed books, and inscribed stones, as they relate to the text inscribed upon them and contained within them. This interest has shown itself in the digital humanities as well, as scholars explore how computers might be made to express the physical in the digital.

This may take many forms, including 2D images, 3D images or scans, or textual descriptions of objects. This presentation will explore how digital elements describing, expressing, or representing different aspects of a single physical object might be used to study the creation of that object. The focus will be on a manuscript commonly known as the Old English Illustrated Hexateuch (BL Cotton Claudius B.iv.), an Old English translation of the first six books of the Old Testament that includes over 400 color illustrations. In his recent book *The Illustrated Old English Hexateuch, Cotton Claudius B.iv: The Frontier of Seeing and Reading in Anglo-Saxon England* (British Library Press, 2007), Benjamin Withers describes a theory for how the relationship between the images and text prescribed both the layout of the content and the physical construction of the entire manuscript. How might Withers' theory be expressed, visualized, or tested in a digital environment? This paper is intended to be the start of a conversation, rather than the answer to a very complex and wide-ranging question.

Introduction

This paper is the first step in a major research project, so I hope to be forgiven if it is liberal in questions but deficient in answers to those questions. I want to be very clear about what this talk is, mostly for my own benefit. What I am going to attempt here is to work towards a model for thinking about digital editions of primary sources that focuses on studying, with computers, how the textual and paratextual elements contained in a text-bearing object (such as a manuscript) influence the construction of that object, and for studying the construction of the object itself.

The impetus for this paper is work that I've been doing on and off over the past few years with Dr. Benjamin Withers, Professor and Chair of the Department of

Art and Art History at the University of Kentucky. Dr. Wither's life work is dedicated to the study of the British Library Cotton Claudius B. iv, more commonly known as the Old English Illustrated Hexateuch [slide 2]. This is a photograph of the front cover and spine of the manuscript's current binding.[1] The manuscript itself is typically dated from the first half of the 11th century, but this binding is obviously much newer than that. It's common name ("Old English Illustrated Hexateuch") comes by virtue of First: its being a translation, into Old English, of the first six books of the Old Testament (the Pentateuch - Genesis, Exodus, Leviticus, Numbers, Deuteronomy - plus Joshua, the first of the "Historical" books [*see the spine*]) and Second: its incorporation of 394 color illustrations, at various levels of completeness.

Dr. Withers and I first met in 2005, when I was working at the Collaboratory for Research in Computing for Humanities at the University of Kentucky and he had just arrived as the new Chair in the Art Department. I had been working with Kevin Kiernan on the *Electronic Boethius* project; he was developing an image-based electronic edition of the three manuscript sources (two medieval, one Early Modern) for King Alfred's translation into Old English of Boethius' *Consolation of Philosophy*.[2] Dr. Kiernan's belief in the centrality of the physical object for any edition, and his work towards developing software to enable the linking of text and image of text in the editorial process, were a great influence on my own thinking and when Dr. Withers and I first sat down and started discussing his thoughts about the manuscript, I realized how very different this manuscript was from any I'd seen to that point, and what a challenge it would be to create a really *effective* edition, or project, for it.

Withers has recently published a book detailing his theories on how the scribes and artists designed the book before it was written out, and how the layout of the text and accompanying illustrations influences the manuscript's collation (how the folios are gathered into quires). The book is *The Illustrated Old English Hexateuch, Cotton Claudius B.iv: The Frontier of Seeing and Reading in Anglo-Saxon England* (British Library Press, 2007).

We'll return to Dr. Wither's book in a moment; now, let's take another look at the manuscript. [slide 3]

The Manuscript

Slide 3 is an image of Claudius B. iv folio 4 recto, containing an illustration of the creation of man and the naming of the animals. In the center of the page is the main biblical text; at the top and bottom of the page we can see Latin glosses in a later hand, dated to the 12th century. In addition to the illustrations, main text, and 12th century annotations, the manuscript also contains rubricated initials, black ink initials, *lettres d'attente*, 12 century annotations to illustrations, as well as various other later additions. It consists of 20 quires.

The Illustrations

Since the illustrations are arguably the most notable aspect of the Hexateuch,

---

[1] This and all images of the Hexateuch are from the CD provided with the Withers volume.

[2] http://beowulf.engl.uky.edu/~kiernan/eBoethius/inlad.htm

let's look at them in some detail. They appear to have been designed by a single artist, although they have been implemented by a number of artists.

The editors of the 1974 print facsimile of the manuscript,[3] C. R. Dodwell and Peter Clemoes, identified five stages of completion of the illustration, and these stages are represented by "discernible groupings" throughout the manuscript. The locations of these groupings may help give us some idea of how the various quires making up the manuscript were created. Withers describes the various stages and identifies these groupings (Withers pp. 26-27):

Stage 1: The artist sketched the outlines of his composition in dry poing (ff. 149v-159v still at this stage) [slide 4, 149v]

Stage 2: The artist painted solid blocks of colour as needed on draperies, animals, architecture, and frames (ff. 143r-149r still at this stage) [slide 5, 145v]

Stage 3: The artist outlined heads, arms, bodies, swords, frames, etc. using red ink (ff. 101-127v still at this stage) [slide 6, 110v]

Stage 4: The artist adds details to the face such as eyes, nose, and mouth (ff. 78-100v and 151r-155v still at this stage [slide 7, 86r]

Stage 5: in the fifth and final stage, the artist completes the drapery, defining folds and contours. [slide 8, 4r] The illustrations on folios 2r to 76v and in Gathering 18 have been completed to this point. But even among these sections some illustrations lack a few detials that the artist has neglected or forgotten.

Construction

I find it an interesting thought experiment to read through Withers' comments about the relationship between the artist and scribes, and the illustrations and texts they inscribed in the manuscript, and to consider how one might design a digital project – be it editing software, visualization processes, or a scholarly edition – that can visualize or clarify ideas that make a lot of sense written out on paper.

Withers, p. 27:

> "Given this close interaction of the scribes and the artist and their shared priority to reserve spaces for the illustrations, the size and placement of the pictures on the page must be understood to be the result of conscious choices made during the preliminary stages of design. As a deliberate choice, the placement reveals how the narratives were to be read or viewed. In effect, the choice and placement of illustrations in the manuscript was editorial; the person responsible for this was in a position little different from that of the translators or copilers of the Old English text who, Rebecca Barnhouse has shown, 'consciously chose to omit or edit passages in an attempt to control their readers' responses to the text.'"

"The *choice* and *placement* of illustrations in the manuscript was editorial." Both

[3] C. R. Dodwell and Peter Clemoes. *The Old Englsih Illustrated Hexateuch: British Museum Cotton Clausius B.IV.* Early English Manuscripts in Facsimile 18. Copenhagen: Rosenkilde and Bagger, 1974.

the text and the illustrations have been designed "in an attempt to control their readers' responses to the text." These are fascinating statements about what the illustrations and texts are doing. They are *active*; they aren't sitting on the page waiting to be read, they have been placed on the page, and content and position and relationships are meaningful. Can we imagine an edition of this manuscript that would allow us to test this theory? To view not only the relative positions of the illustrations and the sections of text they are illustrating across the manuscript, but also to contain within that visualization information about the content or meaning of the text and illustrations.

Withers, pp. 44-46:

> "In designing these visual narratives, the makers of Claudius B.iv demonstrate an awareness of the manuscript as a structured medium of communication. They not only give priority to the placement of the illustrations in the overall design of the manuscript, they quite frequently use the placement of the images to indicate the structure of the narrative to highlight certain episodes... This is accomplished only through a great deal of flexibility and cooperation on the part of the artist and the scribes. It is also the result of a high degree of control over the details of construction. There is clear evidence that the makers of the manuscript (perhaps under the direction of the artist) planned the initial placement of illustrations one gathering at a time, using the rectangular block of text or framed illustration as a fundamental unit in the process of design. Using this simple visual principle, the makers planned the typical gathering from its initial stages to both begin and end with either a complete block of text (an individual episode, marked by punctuation at the start and finish; the text was not allowed to run over to the next gathering) or with a framed illustration."

Withers then introduces a table showing the design of gatherings in Claudius B.iv, showing what each gathering begins with (an illustration or some particular kind of text), which books are represented in the gathering, and what the gathering ends with.

The **Illustrations** section of this paper describes the stages of completion of the illustrations, and what sections of the manuscript have illustrations still unfinished. Then in the first quotation we learn that the physical layout of the text and illustrations may serve to guide how a reader will understand the meanings of both. In the second quotation we are reminded that the layout has certain restrictions, based on the physical limitations of the manuscript and the restrictions those who planned and created the manuscript placed on themselves. There is only so much space: each quire has a limited number of folios, and each folio has a limited amount of space for text and illustrations. Larger illustrations means less room for text. Text may not overlap quires, and quires must start and end with specific elements – however books may start and end in the middle of quires.

These physical limitations and relationships between elements will need to be accounted for in our theoretical model. Making these concepts digital will require a combination of text encoding and digital images.

Elements: Text and Images

Text Encoding

What is "text encoding"? The concept should be familiar to anyone who has worked with handwritten materials. [slide 9] This is a photograph of the first folio of Beowulf (British Library MS Cotton Vitellius A.xv, 129r), an Old English poem known from a single manuscript, which was written in the early decades of the 11th century (although the poem may or may not be much older than the record). A quick glance can tell us several things, even if we don't know much (or anything) about Old English, Old English poetry, or medieval manuscripts.

First, the manuscript has been damaged around the edges, and this damage impacts some of those words on the right side of the folio. This manuscript (along with all others in the Cotton collection, including the Old English Hexateuch) were in a fire in 1731. The Hexateuch escaped relatively unscathed; The Beowulf manuscript suffered minor damage around the edges; other manuscripts were completely destroyed. Conservators at the British Museum in the 19th century disbound many of these damaged mauscripts and taped their folios onto paper frames - if we could see the other side of the folio, we would notice that the frame actually covers up some of the text around the edge. So (1) - the text is not entirely legible.

A second thing we might notice (even if we are not familiar with Old English poetic conventions) is that this "doesn't look like poetry". It's written from margin to margin. There is no poetic structure to the way the text is written on the page, although we can understand that since this is a poem, there must be structure inherent in the text. In this instance, then, we have two parallel or perhaps overlapping sets of encoding – one that describes the layout of the text, or its external structure, the other that describes its internal structure (the *meaning* of the text would be yet another type of encoding).

A third thing we might notice is the shape of the letters as they are written. Some of them will look familiar (see "a", "T" in the top line, "h" in the third line); some of them will look different but will perhaps be recognable (see "g" and "r" in *ingear dagum* on line 2; "f" and "r" in *frunon* on line 3, "g" and "s" in *egsode* in penultimate line), and some letters you may not recognize because we do not use them in the modern English alphabet (see "Ƿ" (wynn) on the first line, " þ" (thorn) and "ð" (eth) in line three, "Æ" (ash) in the first line and in *æthelingas* line 3). There are also scribal elements you might notice - special notation used to indicate abbreviated words (such as "ū" standing in for "um" in *monegū* line 5) and letters that share strokes, called "ligatures" ("ea" in *ingear* line 2; "eg" in *egsode* and "eo" in *eorl*, penultimate line).

All of these "things" - the condition of the substrate, the arrangement of the text on the page, and the appearance of the text itself - are *encoded* into this text, by its very nature. They are things inherent to the text that we notice and can describe in some way that makes sense to us.

Electronic text encoding serves to take these things that we can see, understand, and describe, and make them explicit to a computer. Ideally, this enables us to take what *we* can see, understand, and describe, and search, process, and

visualize it in ways that are either incredibly time-consuming or impossible for a human being. As I hope to show later in this paper, however, the ability to really use and process text encoding lags behind the level of detail we can provide in the text encoding itself. I am not sure that it's a lack of method or technology; there are robust languages for processing text encoding; but I don't think that these methods and technologies are currently being used to their greatest advantage with regards to the needs of the humanities scholar.

Digital Images

The first generation of scholarly electronic archives and editions (1992-2000) often brought together encoded texts and images for parallel viewing and study.[4] In 2002, *Computers and the Humanities* dedicated a special issue to "Image-based Humanities Computing" at a time when "a majority of first generation image-based humanities computing projects [had] reached at least an initial plateau of completion."[5] Articles in that issue described both image-based projects and software for developing such projects. This issue marked a watershed moment, illustrating a recognition within the digital humanities community that images are - or should be - an integral part of scholarly digital editing. Since that time interest in incorporating primary source images into digital editions has blossomed, as has the number of image-based projects completed or under development.

Now, in 2009, most typically digital projects will include 2D images of resources taken under "natural light". [slide 10] This image represents folio 12 recto in the tenth-century Marcianus Graecus Z. 454 (= 822), commonly known as Venetus A. This is the earliest complete copy of the *Iliad* (this folio shows the opening of the first Book of the *Iliad*) and is the version of the text on which all modern editions are based. In addition to the main *Iliad* text (the larger text in the center of the folio) the manuscript contains several layers of various types of scholarly commentary – annotations and variant readings – called *scholia*. The *scholia* represent the vast history of scholarship on the *Iliad* dating back to the library of Ptolemaic Alexandria in the second and first centuries BCE. This image was taken in May 2007 as part of the Homer Multitext Project (centered at Harvard's Center for Hellenic Studies), and the complete images are available freely online.[6] The *scholia* are of major interest to scholars, and these images are of high enough resolution that even the smallest texts (which are tiny) are visible. Work is currently underway to encode all the texts (main text and all *scholia*).

---

[4] Valley of the Shadow: Two Communities in the American Civil War, Virginia Center for Digital History, University of Virginia (http://valley.vcdh.virginia.edu/); The Complete Writings and Pictures of Dante Gabriel Rossetti, A Hypermedia Archive, edited by Jerome J. McGann, University of Virginia (http://www.rossettiarchive.org/); Dickinson Electronic Archives, edited by Martha Nell Smith, Online. Institute for Advanced Technology in the Humanities (IATH), University of Virginia (http://www.emilydickinson.org/); The William Blake Archive. Ed. Morris Eaves, Robert N. Essick, and Joseph Viscomi. (http://www.blakearchive.org/); *Wife of Bath's Prologue on CDROM* (Chaucer 1996); *Electronic Beowulf* (Kiernan 1999); *Piers Plowman Electronic Archive*, Vol. 1, (Langland 2000).

[5] Matthew G. Kirschenbaum, "Editor's Introduction: Image-Based Humanities Computing" *Computing and the Humanities* 36:1 (February 2002), 3-6.

[6] Homer Multitext Project site: http://chs.harvard.edu/chs/homer_multitext

The vast majority of online manuscript collections will include images similar to these (and to the Hexateuch images we've already seen). Although it takes time and equipment to create these kinds of digital collections, the cost of that technology has dropped over the past several years, as has the cost of large amounts of digital storage. Once an institution has invested in the equipment and has developed work-able workflows it becomes easier to digitize more materials. If the primary sources are legible and in good condition, 2D imaging is fine.

Some sources, however, require more than 2D imaging can provide. Earlier I mentioned manuscripts damaged in a fire in 1731; for some of those, multi-spectral imaging (taking photographs under different lighting conditions, from ultra-violet to infra-red, in order to bring out lost text, and then combining the various images to create the best view for all text on a page) can be a very effective tool for helping with legibility. An extreme example of a manuscript that has been helped by various sorts of special imaging is the Archimedes Palimpsest, which has recently been released through GoogleBooks (all of the raw data, including encoded transcriptions, are available from the Archimedes Palimpsest website).[7] [slide 11] The Archimedes manuscript was likely copied during the 9th century. In the 13th century, along with four other manuscripts, it was disbound, scraped clean, the pages turned 90 degrees, and a Greek prayer book was written upon it. Scholars and scientists have spent the last ten years imaging and editing the palimpsest; the view in Google Books shows only the palimpsistic text. This is just one example of a text that would be completely illegible were it not for digital imaging.

Returning to the Venetus A [slide 12]: We can see even in this image that the folio is not flat. Over time conditions have caused the folios throughout the manuscript to become wavy – the technical term for this is "cockling." A team at the Center for Visualization at the University of Kentucky, under the direction of Brent Seales, has been developing technologies to virtually flatten and unroll materials in order to help make them legible for scholars. [slide 13] In this proof-of-concept shot, we have a piece of paper that has been crumpled into a ball and then flattened, and [slide 14] the same picture after the virtual flattening technique.

At the same time the 2D photographs of the Venetus A manuscript were taken, full 3D laser scans were also taken. These could then be mapped together and put through the virtual flattening technique. [slide 15] Here is a screenshot showing a cross-view of the 3D image of the folio. A video of the virtual flattening process available on the same website as this paper (example3d.mov) shows how the wavy, cockled folios may be virtually flattened in order to straighten out text that may not be legible in regular 2D imaging. This isn't the most extreme example – the cockling of the Venetus A is not so severe as to render text completely illegible – but the virtual flattening technique could be useful for more seriously damaged materials. A related technique is virtual unrolling, in which papyrus scrolls are non-invasively scanned (using such technologies as CT or X-Ray scanning), the ink is located (by differentiating the chemicals in the ink from the composition of the substrate), and the computer virtually unrolls the

[7] Archimedes Palimpsest website: http://www.archimedespalimpsest.org/

scroll, uncovering missing text. There are two screenshots showing the process, the first a long view [slide 16] and the second a close-up illustrating the differentiation of ink and substrate [slide 17]. Both screenshots are from a video of the virtual unrolling process available on the same website as this paper (rotate-1.mov).

So there are many potentially useful types of digital images that may be incorporated into a project: 2D images in natural light, which are fine if the materials are in good condition; 2D image in special lighting, which should be a requirement if the text on the surface is otherwise illegible or difficult to read; 3D images and methods for manipulating them – another method for clarifying text; or a combination.

But images will only get you so far. Images – and the physical objects represented by them – need to by studied by scholars, who from them will develop theories, descriptions, and notes. These notes need to be written, and in a digital context they need to be encoded in some way for them to be presented and perhaps processed by computers.

That is the purpose of the Text Encoding Initiative.

The TEI: Standards for text encoding

The Text Encoding Initiative Consortium is a consortium of individuals and institutions that promulgates a set of guidelines for humanities text encoding.[8] These guidelines are commonly referred to as "TEI". TEI is the de facto standard for humanities text encoding, widely used in Europe, Canada, and the United States.

I'd like to take a few minutes to discuss some chapters from the TEI Guidelines that are relevant to this discussion.

1. Manuscript Description (Chapter 10)[9]

The chapter on Manuscript Description provides recommendations for describing manuscripts in the TEI. Descriptions may be given in narrative paragraph form, or using a more structured design, depending on how the descriptions will be used (narrative description to be read on a website, structure for computation). TEI Manuscript Description provides sections for identifying the manuscript (library, collection, shelfmark, alternative numbers or names, etc.), its textual contents, and its history and provenance. Most importantly for our purposes, the chapter includes extensive methods for describing the manuscript physically: its binding, decorations, scribal hands, musical notation, etc., as well as descriptors for its “object-ness” – its layout, or how the elements are laid out on the folios, and its support, that is information about the material comprising the manuscript, its condition, its foliation or how

[8] *P5: Guidelines for Electronic Text Encoding and Interchange*, edited by Lou Burnard and Syd Bauman. Text Encoding Initiative, 2007.

[9] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html

the pages are numbered, and its collation or how the leaves or quires are physically arranged.

2. Representation of Primary Sources (Chapter 11)[10]

The chapter on Representation of Primary Sources consists of two main sections. The first one I'll get to in a minute. The Second one provides recommendations for encoding detailed transcriptions of primary sources. Quoting from the Guidelines, they are designed to provide:

- first, methods of recording editorial or other alterations to the text, such as expansion of abbreviations, corrections, conjectures, etc. (section 11.3 Altered, Corrected, and Erroneous Texts)
- then, methods of describing important extra-linguistic phenomena in the source: unusual spaces, lines, page and line breaks, change of manuscript hand, etc. (section 11.4 Hands and Responsibility)
- finally, a method of recording material such as running heads, catch-words, and the like (section 11.7 Headers, Footers, and Similar Matter)

Notably for our concerns, the chapter also includes recommendations on describing how damage to the substrate and ink impacts on the legibility of the text (and how to indicate text that has been supplied by an editor).

3. Digital Facsimiles (subsection 11.1 of Chapter 11)[11]

The first section of the chapter on the Respresentation of Primary Sources is called "Digital Facsimiles" and describes methods for incorporating references to image files into TEI files. These include methods for listing all the image files associated with a manuscript, and for defining where elements (including illustrations and text) appear on those images.

4. Linking Structures (mainly described in Chapter 16, "Linking, Segmentation and Alignment")[12]

TEI linking structures: link everything together

So the TEI gives us methods for structuring lots of information about manuscripts.

Returning to the Old English Hexateuch as our central example:

---

[10] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html

[11] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHFAX

[12] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html

Using the TEI, we can create a manuscript description that contains both narrative description and more structured information about how the text and illustrations appear on the page, in relation to one another; how the quires are organized and fit together (the collation), and how these elements relate to one another.

We can create a transcription of the text in the manuscript that differentiates between the Old English biblical text, the 12th century glosses, and any later notes that might appear - with information about letter forms, abbreviations, and notes about the different scribal hands. It could also include information about the condition of the substrate and how that impacts the text - are there holes that text is written around? Are there newer holes that obliterate text? Does the text go into the gutter, or like the Beowulf example is the text cut off at the edge?

We can use facsimile not only to list all the manuscript images, but to note the coordinates of areas on the page that may be of special interest - for example, the locations of all the separate textual areas and illustrations.

And finally, we can take advantage of the linking structures to virtually bind all of this information together: [slide 18]

> [slide 19] Linking descriptions of illustrations in the manuscript description to the locations of those illustrations on the pages as defined in the facsimiles.
>
> [slide 20] Linking transcriptions of the text to the locations of that text on the pages as defined in the facsimiles.
>
> [slide 21] Linking descriptions of different "types" of scribal elements - abbreviations, ligatures, etc. - from the manuscript description, to notable (or all) instances of those elements in the transcription.
>
> [slide 22] Linking transcriptions of illegible or editorial-supplied text to the description of the condition of the substrate in the manuscript description, and to locations of that text on the pages as defined in the facsimiles.
>
> [slide 23] In effect: Linking each folio image to all of the information contained on that folio (transcriptions, illustrations, marginal notes, as well as descriptive information from the manuscript description).
>
> [slide 24] Linking the collation formula to the folio images - and from there we have access to all the information contained within the TEI encoding - information about the texts, the illustrations, and their locations.

Do you see where we are going with this? With this information, we have almost everything we need to construct a full model of the Old English Illustrated Hexateuch.

Almost. Because - as I expect most anyone who has ever used the TEI will attest - getting your information encoded is only the first step. In order to really use or

take advantage of this information, you have to be able to process it effectively. And I am unaware of any existing implementations that can process the kind of TEI structures that I am talking about in such a way that would be useful for scholars studying the structural relationship between a text and the surface - indeed the entire object - on which is it inscribed; its physical context. Let's look at how manuscripts are usually presented online currently.

Processing

There are of course many methods for processing texts and images. There are relatively simple languages for converting TEI to be viewed through a web browser.[13] Other languages can be used for searching TEI-encoded texts for particular descriptors (allowing such searches as "show me all words that have been damaged on the edge of a page" or "show me all folios that contain illustrations that only have sketching and blocking"). [14]

The next set of slides show several examples of displaying text, images, and descriptions of single manuscripts together. The Pembroke 25 project links manuscript description to images [slide 25], as well as linking transcription to images [slide 26].[15] The Thomas MacGreevy Archive presents text and images facing, as well as nice navigation through the images [slide 27].[16] The e-codices Virtual Manuscript Library of Switzerland, as of 1 February 2009, presents digital images of 363 manuscripts from 16 different libraries across Switzerland. [slide 28] Through the e-codices site, you can search all libraries and collections, or within individual collections, as well as searching by manuscript. Through the site you can view manuscript descriptions [slide 29], folio images [slide 30], and images of binding [slide 31], but not full text and there is no connection amongst the manuscript descriptions and folio images.

Another example of manuscript visualization is represented by the Turning the Pages™ software, which was originally developed by the British Library, has received support from Microsoft, and is currently being marketed to libraries, museums and educational institutions specifically for presentation to the public (mainly through the sale of CDs, also in museum kiosks).[17] Turning the Pages™, is now available in three versions: Turning the Pages™ 2.0, Turning the Pages™ 3D, and Turning the Pages™ VirtualBook. This screenshot is a page from St. Chad's Gospel as presented through Turning the Pages™ VirtualBook (image from the Turning the Pages™ website) [slide 32].

Here I will quote extensively from the description of the Turning the Pages™ VirtualBook on the website (all italics are added):[18]

---

[13] Cascading Stylesheet Language (CSS) (http://www.w3.org/Style/CSS/); Extensible Stylesheet Language (XSL) (http://www.w3.org/Style/XSL/); XML Path Language (XPath) (http://www.w3.org/TR/xpath)

[14] XML Query Language (XQuery) (http://www.w3.org/TR/xquery/)

[15] Pembroke 25 Project (http://www.rch.uky.edu/Pembroke25/)

[16] The Thomas MacGreevy Archive (http://www.macgreevy.org/index.jsp)

[17] Turning the Pages (http://www.turningthepages.com/)

[18] Turning the Pages™ VirtualBook

> "Turning the Pages™ VirtualBook is the ultimate in *creating the illusion* of actually using a fully 3-dimensional book.
>
> "Using the *exact dimensions* of the chosen book we create a *precise 3-D computer model of it*. At this point in the process we normally film a curator turning the pages to provide *reference material for the 3D modelling*. This means the shape the pages make as they turn can be *exactly the same as the original*: 8th century vellum turns very differently to 17th century paper for example. Pages of the book are then photographed, and scanned at high resolution. Following in-depth consultation, we retouch and colour-correct each one to the curators satisfaction. This part is crucial to the realisation of a high-quality end product and we process the images with the most faithful attention to the original.
>
> "We then integrate the processed graphics and the computer model to create the animation of the turning pages. The *magic of the illusion* really takes shape when we add lighting and shadows and even adjust the reflection values and texturing to bring out certain details such as gold leaf and embossed leather.
>
> Text commentary, audio voiceover, interactive hotspots and any additional audio assets are combined with the 3D animation in a custom-written interactive package resulting in a media-rich and immersive interactive environment. The programme is then loaded onto a touchscreen computer (this version does not work on the web). [*ed: although the other versions do*]
>
> "The result is astonishing - a *true digital facsimile*, allowing visitors to experience *the book as it really is*, but with all the added Turning the Pages™ features to make their experience educational and enjoyable."

I'd like to take a minute to deconstruct the language used here to describe what exactly it is that Turning the Pages™ VirtualBook is doing. In the first paragraph, they state that they are *creating the illusion* of a real book. In the second paragraph, they state that in order to create this illusion, they use the *exact dimensions* of the physical book to create a *precise 3D model*. So this is not only an "illusion of a book"; this is to some extent a replica – to the extent that they duplicate the way the substrate "moves" in real life. High-resolution digital images provide the look of the layout.

The next step is actually not mentioned at all in this process, but it is the one I find the most interesting. The next paragraph mentions a "computer model" which is integrated with the processed graphics to create the pages animation. This computer model is a "generic book model" that is part of the software, and upon which any imported information – digital images and metrics for the weight of the parchment or paper – are placed.[19] What this means is that this "model of this specific book" – following the *exact dimensions of the original* – is not, I would say in any way at all, a model of this specific book. It may *look* like that book, and its pages may turn in a way that is similar to the pages on that book, but so many things that make the real, physical book what it is – the shape of the pages around the edges (think of the 3D image of the folio from the

(http://www.armadillosystems.com/ttp_commercial/products_virtual.html)

[19] Personal correspondence with Michael Stocking, Managing Director, Armadillo Systems, 26 January 2009.

Venetus A manuscript), the particular way that this book opens and moves, how the book is constructed – its quires, adding pages, missing pages – all this is lost. What we have instead is a digital "thing" that is trying very hard to be "this book", but just isn't. When I see this, it makes me feel as though I have entered the "uncanny valley" of digital texts.

The term "uncanny valley" was first described in 1970 by Japanese roboticist Masahiro Mori in a short article published in the journal *Energy* and translated into English in 2005.[20] In this article, Mori posits that as robots become more lifelike – specifically, more *human* – we (real live humans) see them as familiar and become more comfortable with them. There is, however, a point at which robots become so humanoid that when we notice they are *not* in fact humans, we find it strange, or *uncanny*. One example he give, and which I'll read it total here, is of prosthetics:

> Some readers may have felt sympathy for handicapped people they have seen who attach a prosthetic arm or leg to replace a missing limb. But recently prosthetic hands have improved greatly, and we cannot distinguish them from real hands at a glance. Some prosthetic hands attempt to simulate veins, muscles, tendons, finger nails, and finger prints, and their color resembles human pigmentation. So maybe the prosthetic arm has achieved a degree of human verisimilitude on par with false teeth. But this kind of prosthetic hand is too real and when we notice it is prosthetic, we have a sense of strangeness. So if we shake the hand, we are surprised by the lack of soft tissue and cold temperature. In this case, there is no longer a sense of familiarity. It is uncanny. In mathematical terms, strangeness can be represented by negative familiarity, so the prosthetic hand is at the bottom of the valley. So in this case, the appearance is quite human like, but the familiarity is negative. This is the uncanny valley.

This figure [slide 33] (published in the online translation of the article – a simplified version of the original figure) charts Mori's conception of the lay of the "uncanny valley", with industrial robots (manufacturing, for example) on one end, and healthy persons (you and me) on the other end. His prosthetic hand example lies towards the bottom of the valley, *bunraku* puppets (used in traditional Japanese theatre) are heading up to familiar territory.

In my mind, the models created by Turning the Pages™ fall at the nadir of the "uncanny valley of digital texts" – which has perhaps a plain text transcription at one end and the original manuscript at the other end, with print facsimiles and editions, and the various digital displays and visualizations presented earlier in this paper falling somewhere between the plain text and the lip above the chasm. Despite the entertainment value of this technology, and how useful it might be for libraries and museums in drumming up interest and bringing in money, it is most definitely not scholarly and should not under any circumstances be mistaken for such. A Turning the Pages™ model of a manuscript is not, in fact, a *true digital facsimile*, as claimed in the last paragraph of the blurb on the website. It is a collection of pretty pictures that move pleasantly, and that is all.

It is entirely possible that there are scholarly "true digital facsimiles" to be created, correct in all aspects, including the actual shape of the pages (as we saw

---

[20] Masahiro Mori, "The Uncanny Valley" Energy, 7(4), 1970, pp. 33-35. Translated by Karl F. MacDorman and Takashi Minato, 2005 (http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html)

in the example from the Homer manuscript) and taking into account the collation of the manuscript and the manner of its construction. However, I'm not convinced that such a thing by itself would be very much more useful than Turning the Pages™. Through the use of the TEI, vast amounts of data that can be interpreted from a text or a physical object by a human scholar, then encoded in ways that can be both read by humans and processed by computers. A true scholarly digital facsimile, linked somehow to this information, and able to be disbound, rearranged, contemplated whole or in parts, with consideration of every aspect of the physical object: the substrate, text and illustrations inscribed upon the substrate, the condition of the substrate and likewise the condition of the text and illustration, even (after all this discussion of the physical elements) the meaning of the text – and, perhaps, methods for interpreting or studying how the meaning of the text in this manuscript might be influenced by the physical appearance of that manuscript: This, I think, is the future of image-based digital editions and the concept around which a model needs to be developed. At the same time, however, we need to be very mindful of the difference between a physical object and the digital versions of that object. We want to avoid the "uncanny valley" of the digital representation of our objects. Creating or attempting to create exact representations of manuscripts and other original resources is exactly what we should not be trying to build.

Conclusions, Propositions and Future Work

To finish up, I'd like to present some conclusions, propositions, and plans for future work.

1. A Proposition: the development of a model for digital projects that consciously recognizes and accounts for the different elements of primary sources [slide 34]. These elements, as I have discussed through the course of this paper, I see falling into six types: writing/inscription (that is, the shape of the the writing as it appears on the source); illustration/decoration; substrate/condition (the material the source is made of, its condition – there will be relationships between this type and the two previous types as the condition of the substrate impacts the appearance of text and decoration inscribed upon it); construction of the object (as I hope to have made clear in this paper, in many instances the first three elements will impact on the construction of the object); and the meaning of the text. I'm not sure how the "meaning of text" relates to the last type here, "contemporary readers". In this paper I haven't said anything about how the history of reading (how primary sources were read in their own time) might influence the development of digital sources, but that is something I plan to investigate in the future.

Within this model , these different types would in some way need to be able to be "separated out" and worked/viewed on its own (in both the editorial process and as a final finished project).

A question that I have is where does scholarly interpretation fit in? Is it up beside the six areas, or is it better conceived as part of them? Different scholars focus their work on these different areas; they tend to be specialized: art history, paleography, manuscript curators/scholars, textual scholars, and scholars who study the history of reading. Our model would need to have methods for these

scholars to work both independently, together (within their limited fields) and together (multi-fields). There may also be different "levels" or "types" for some of these.

2. "Digital facsimiles" will be central to this model, but they cannot *be* the model. The model will clarify rather than mimic, providing scholars methods for describing their knowledge and scholars and students methods for investigating that knowldge, comparing encoded information with images etc.

3. The model will enable clear visualizations of information that is not normally presented visually. Digital images are obvious in the sense that we can look at them and immediately recognize them – they are also technically easy to present online. Scholars trained in specialized areas will understand aspects of them more than the rest of us. To share their knowledge with us, these scholars can write down their thoughts, and we can read them. In a digital environment, there are tools available for attaching notes to images – this is also technically relatively easy. Some other information is also fairly easily visualized (manuscript collations can be described in formula, line drawings of stone fragments being fit together). It would even be possible to create a visualization of a collation from a collation formula. But how do we visualize the points at which these elements intersect? How can we visualize - as Withers theorizes - where the text lies on one page, a related illustration on a later page, and finally how those elements together influence the collation of the manuscript as a whole? This is the kind of visualization that I envision for this proposed model