Metamorphosis: Remediation in *Early English Books Online (EEBO)*

Diana Kichuk
University of Saskatchewan Library, University of Saskatchewan, Canada

Abstract

Remediation refers to the re-presentation of old media in new media. This article studies remediation in electronic products in library collections, especially the digital facsimile. *Early English Books Online (EEBO)* is a particularly interesting example, not only because of its scholarly importance, but also because of its multi-layered genesis from printed work to microfilm (*Early English Books (EEB)*) to digital (*EEBO*) facsimile, and to the text encoding initiative *EEBO-TCP*, a joint ProQuest and Text Creation Partnership (TCP) project. The article analyses the impact of filters and limits of remediation in relation to *EEBO* and its predecessor *EEB*, such as the choice to duplicate a single copy of a work as bi-tonal black and white images, and to scholarly work.

Correspondence:
Diana Kichuk,
University of Saskatchewan Library, Rm. 163, ITS, Main Library, 3 Campus Drive, Saskatoon, Saskatchewan, S7N 5A4 Canada.
E-mail:
diana.kichuk@usask.ca

New media may be 'new' because they have not existed before, but do have antecedents. They incorporate old media, transforming or remediating them (Bolter and Grusin, 2000). Even 'born digital' products emulate the form, language and syntax of older media. A blog (web log), for example, is a combination town crier performance, personal diary, and news magazine. The open access online encyclopedia *Wikipedia*[1] has the look and functionality of a print encyclopedia, albeit with the transforming power of online searching and browsing. Even the windowed and mixed media computer desktop, charged with electric speed, dynamic color and hyperactivity, recreates the research environment of the analog scholar sitting at her desk, surrounded by multiple open books, notes, articles, data, images, and a complex reference system close at hand.

Advances in new media may one day create truly 'new', original electronic products, but for now, from digital facsimiles to virtual worlds, electronic products seem oddly familiar and the hype about newness overdone. Current library electronic products fall, with few exceptions, into two categories: digital facsimiles of works originally published in 'old' media formats—print, microform, image, and audio—or products that have functionally evolved more or less closely from an analog precursor. *Early English Books Online (EEBO)* is an example of the former, and *Encyclopaedia Britannica Online*, the latter. Electronic products currently dominating library collections—abstract and indexing databases, electronic journals and books, news services, streaming video and audio, and data—are copies or sequels of analog counterparts.

Remediation is a relatively new scholarly concept rarely if ever applied in discussions on issues related to the digitization of primary texts. Subsequently any analysis and discussion related to remediation within library scholarship would be a welcome contribution to this new media studies field. This article limits its discussion to remediation and the digitized facsimile, specifically to *EEBO*. *EEBO* is particularly interesting, not only because of its considerable scholarly importance, but for its multi-layered genesis from print copy to microfilm facsimile, and from microfilm to digitized facsimile, making it a model of remediation.

Literary and Linguistic Computing, Vol. 22, No. 3, 2007.

doi:10.1093/llc/fqm018 Advance Access Published on 18 June 2007

What is the impact of remediation in digital facsimiles? Is the scholar examining the work itself—the Early English book or broadside in the case of *EEBO*—as vendors and many eminent scholars claim? Is it a faithful copy of the text and the physical object? Alternatively, does the process of digitization and remediation transform the original work into a virtual artifact with an ersatz resemblance to an original? Are there limits to the use of digital reproduction as scholarly evidence? These questions do not devalue *EEBO*. Digital facsimiles such as *EEBO* have unquestionable scholarly interest and value. However, are they what they purport to be—an exact copy of the original? Scholarship can benefit by a greater awareness of the impact of remediation and the limits of digital facsimiles. After examining the genesis of *EEBO* and the impact of remediation, this article will conclude with a brief speculation on how *EEBO* could improve its already formidable scholarly value.

1 Remediation

Jay David Bolter and Richard Grusin applied Marshall McLuhan's insight, 'the "content" of any medium is always another medium' (McLuhan, 2001), to digital media, and named the concept, *remediation*. Remediation (re-*media*-tion, with the emphasis on *media*) is the re-presentation of one medium in another. It is the appropriation or re-purposing of old media in new media (Bolter and Grusin, 2000). McLuhan referred to this aggressive takeover of one medium by another as 'hybridizing or compounding' (McLuhan, 2001). The juxtaposition of old and new media and the replacement of one medium with another through reproduction, insertion, and hypertext linking are characteristic of the persistent mosaic quality of digital media.

According to Bolter and Grusin, digital media remediate old media in several ways (Bolter and Grusin, 2000). The range of remediation types is not exclusive and a work may manifest more than one. *EEBO*, for example, displays elements of all remediation types, especially the first two. The following is an interpretative summary of the Bolter and Grusin remediation types:

(1) *The remediation is the 'real thing' or a clone with a primary focus on the old medium.* The remediation is a 'faithful' facsimile, as if the old medium could migrate to the new medium without alteration. For example, digital art image galleries or manuscript collections are often presented as real paintings on a gallery wall or manuscripts on a library study table. A publisher brochure boldly asserts that works in *EEBO* are 'as they appeared in their original printed editions.'[2]

(2) *The remediation seeks to improve the old medium.* The old medium is still the focus, but it is 'improved'. Digital remediation provides better access, adds sound, hyperlinks, search capability, and so on. The publisher disregards unintended consequences (such as scale distortion or detail loss) as if they are benign, portraying transformations to original works in digital facsimile products like *EEBO* as 'improvements'. Full bibliographic records, browse and search capability, indexed illustrations, and downloadable image files in varying sizes within an integrated access interface—add value without parallel to the original print or microfilm media.

(3) *The old medium is intentionally refashioned or changed.* The new medium refashions the older medium more or less aggressively. Unlike *EEBO*, the integrated *EEBO Text Creation Partnership (EEBO-TCP)* project represents a radical departure from the original work. *EEBO-TCP* manually transcribes selected *EEBO* works into searchable text.

(4) *The old medium is 'absorbed' into the new media without a trace.* The new medium proposes to absorb the older medium and erase it, or at least that is the intended outcome. As an example, computer games remediate cinema, with players becoming characters in an interactive film-like narrative. *EEBO* attempts to bring everything to the

> scholar's desktop—replacing the need to travel, to navigate catalogs and indexes, to access remote shelves and dusty books or unwieldy microfilm reels—and banish the appalling expenditure of time, while offering the scholar the experience of turning pages of the original as if in a carrel in a far away library.

Bolter and Grusin find that digital media display dual traits of remediation: *immediacy* and *hypermediacy* (Bolter and Grusin, 2000). On one level, digital media provide an immersive and transparent experience, an immediacy that permits the user to 'disappear' into the screen or virtual world. In an ongoing dialectic, the transparency of immediacy couples with the opacity of the hyper-mediated, windowed, fragmented and interactive reality of interfaces and multi-media. The more hyper-mediated the digital object, the more fragmented is the user's experience and the more media conscious.

2 The Genesis of *EEBO*

2.1 From print to microfilm facsimile

In the mid-1930s under the specter of world war, there were concerns about the preservation of Britain's culturally important book collections. At the time microfilm was the high-tech archive option in the western world (seventy-five years later, despite the rise of digitization projects, microfilm still has abiding relevance for archive and scholarly purposes). In 1938, University Microfilm Inc. (UMI; now a division of ProQuest Information and Learning) initiated its first microfilm facsimile collection: *Early English Books I, 1475–1640* (*EEB I*). In 1957, it launched *Early English Books II, 1641–1700* (*EEB II*). Scholars often refer to the two collections as *EEB*. *UMI* based the selection of works in *EEB* on two print catalogs often referred to as the STC or *Short-Title Catalogue of English Books*.[3] The publisher released *Thomason Tracts* and *Early English Books Tract Supplement* later as companion collections.[4] *EEB* and its companion collections contain micrographic images of at least one representative copy of potentially all printed books and tracts printed in Britain and its dependencies between 1475 and 1700.[5] While UMI completed the bulk of *EEB* by 1988, filming is still ongoing, though at a much slower rate. One of the earliest examples of facsimile production using microfilm technology, *EEB* is arguably still one of the most important microfilm facsimile collections in existence today.

From the outset, UMI used the recently perfected 35 mm microfilm camera and silver halide on acetate base 35 mm film. After the 1980s, it used the more stable polyester base film. *EEB* has three film generations: a first-generation master or preservation negative-polarity film (white on a black background), a copy negative, and multiple working copies (either negative or positive polarity) for distribution to subscribing institutions. Positive polarity images have black text or illustrations on a white background, the standard polarity of distribution copies.

Operators filmed the pages of bound works sequentially in an 'open book' layout—two facing pages ('two-up') with a gutter shadow down the center. Until standards and technology improved, early images in particular display varying degrees of distortion—for example, poor registration of the work within the frame or under or overexposed images. Early image quality varied dramatically from one image or reel to another, depending on equipment, applied standards, and individual technical skill or discretion of human operators in libraries doing the filming. Filmed in black and white at low resolution, the content images are essentially bi-tonal. Gradations of original color or grayscale, for example print variations in hand-press incunabula, are remediated within a narrow range of black to 'less black' on a white ground. Rubricated letters and marks show up as mottled black shapes with only softened edges and reduced contrast to distinguish them from adjacent text. Operators excluded images that might not convert well to black and white, for example, the frontispiece paintings in works from the incunabula period. Operators systematically excluded front matter and end papers, including handwritten notes added by a reader. Operators used image cropping liberally to fit the 'open book' image into the available film area or in a misguided attempt to correct the registration of a seriously misaligned

print text or illustration. As a result, there are marginalia information losses including margin width and page size.

Although its major scholarly contribution may be to preserve primary sources and distribute them to a broad swath of scholars who could never hope to travel and examine the original print, *EEB* took on a value and life of its own. It developed its own mystique and loyalists. *EEB* became a 'must have' acquisition for academic collections of a certain size and pretension. Scholars became adept at consulting works on backlit screens of varying light quality, and scrolling microfilm reels on readers of various vintages and technical advancement. Doing research with microfilm has its own unique set of preparations, etiquette, and inimitable ambience. The row on row of labeled boxes, reels with restraining strings and rubber bands, the fussy set-up. The whirring sound at the start and fast forward and back and clack-clack-clack of the disengaging film at the end of a session; the careful handling by the fragile film edges and smell of dust and off-gassing. The opacity of gray acetate and transparency of polyester film, image scratches and blotches from damaged or disintegrating film, specks of dust, blurred text and images, cropped and skewed pages, and shifting registers and scales. The unforgettable feeling of motion sickness as the film moves slowly by and sense of infantile helplessness as a film unravels or breaks.

Historian Barbara Tuchman expressed the ambivalence scholars feel about microfilm in Terry Sanders 2001 film, *Slow Fires: On the Preservation of the Human Record*—'I don't really like microfilm except for the fact that it's often indispensable' (Review, 2001). *EEB* became a necessary evil that no Early English scholar could do without. Scholarship that was unlikely or impossible while print copies were rare, access forbidden or restricted, and travel costs prohibitive, became not only possible but also acceptable based on a microfilm surrogate without the scholar ever consulting the original.

2.2 From microfilm to digital facsimile

EEBO is a major digital facsimile collection of Early English works based on the digitization of *EEB* and the *Thomason Tracts*. UMI began the digitization project in the summer of 1998. Chadwyck-Healey launched its customized online access interface in 2003. By early 2005, *EEBO* included over 100,000 of the over 125,000 targeted titles and had extended coverage back to 1473. ProQuest will add *Early English Books Tract Supplement* content in 2006. While *EEBO* has the stated goal of including 'virtually every work printed' in England and its dependencies from 1473 to 1700, the completed content will represent only '80 percent of the surviving print record in English between 1475 and 1700' (Schmitt, 2003). As in *EEB* and its companion collections, the works cover a broad range of subjects, including English literature, history, philosophy, linguistics, theology, music, fine arts, education, mathematics, science, and women's studies. Scholars can browse the collection by author name, Thomason tract in the order collected, and periodical year, issue, and title. The basic or advanced search is run against the text of *EEBO* bibliographic record fields. Institutional subscribers to the companion *EEBO-TCP* database (described below) can also search and download ASCII text transcriptions of a targeted 25,000 *EEBO* titles.

A key specification of *EEBO* required that 'the images must be delivered quickly to end-users' (McLean, 2001). Lower resolution ensured acceptable download transmission rates but sacrificed image detail. *EEBO*'s access files are bi-tonal 1-bit[6] depth images with a resolution of 400 PPI (pixels per inch). In an unusual move, ProQuest made both the access and master files the same resolution. At 400 PPI, the master copy is neither a preservation nor replacement quality facsimile. ProQuest produced *EEBO* from a second-generation negative polarity *EEB* master in order to avoid subjecting the original master to wear and tear (Schmitt, 2003). The project team thought, 'The difference between the originals and the second generation is undetectable at 400 PPI' (Pack, 1999). The master is a TIFF (Tagged Image File Format) file and access or derivative files may be GIF (Graphics Interchange Format), JPEG (Joint Photographic Experts Group), PDF (Portable Document Format), or TIFF files, depending on the context: default screen image, emailed version, downloadable access copy,

or highest resolution copy intended for print or off-line analysis.

EEBO presents works in the same sequential 'open book' layout as in *EEB*. Illustrations appear inline in a page and as separate downloadable images. The user can browse both page images and extracted illustrations in a forward–backward click through or perform a search. Alongside each brief citation in a list of search results are applicable icon links to a full bibliographic record, digitized content, available illustrations, and, for subscribing libraries, available encoded and searchable transcriptions compiled by *EEBO-TCP*.

A generation of devoted *EEB* users lobbied their institutions to acquire the more user-friendly online version of this indispensable source. Early on, more than 150 libraries subscribed to *EEBO*, a number that exceeded the number of libraries with a standing order for *EEB* over its history (Sandler, 2003). Like *EEB* before it, *EEBO* is now a 'must have' scholarly tool.

EEBO preserves the limitations of microfilm—the cropping, poor registration, and low microfilm image granularity—while adding the strengths and limitations unique to the digital medium—the convenience and added value of digital access as well as the erosion of detail. Scholars are won over, despite *EEBO*'s considerable reproduction flaws.

2.3 From digital facsimile to encoded ctext

Because image files are not directly searchable, an *EEBO* search is run on the bibliographic record alone. While relatively error-free searchable text can be automatically compiled using Optical character Recognition (OCR) software, not all text images make good candidates for OCR. Early English characters are particularly difficult to translate because of font variability in the print. The varying quality of *EEBO* digitized images only adds to the difficulty. When running current OCR technology against Early English works produced a high error rate, ProQuest decided that OCR was not a viable choice for *EEBO*. Even if OCR were possible, variations in Early English spelling and punctuation would make searching difficult and unreliable.

To resolve the problem of text searching, ProQuest collaborated with the TCP, a project of the University of Michigan, the University of Oxford, and the Council on Library and Information Resources (CLIR). The TCP creates fully searchable texts of early English and American published works. This collaboration established *EEBO-TCP*, a collection of selected *EEBO* works transcribed into searchable SGML/XML encoded standardized text. Of the targeted 25,000 works, approximately one-third were completed by the end of 2005. Supporting library partners contribute financially or operationally, and gain access to the viewable and searchable transcriptions within *EEBO*.

EEBO-TCP adds an additional production layer to *EEBO*. A human operator consults the digitized microfilm copy on screen, not the text in hand or the microfilm copy, transcribing and tagging the text with SGML/XML codes. While TCP transcription standards have been carefully established, there is room (some scholars might say—a lot of room) for subjective interpretation of content and the occasional human error. Keyboarding is outsourced offshore (JISC: The Joint Information Systems Committee, 2004). Transcribed text may stray from the original enough to produce artifacts or errors that influence the direction of scholarship. Since the transcriptions are based on a digitized microfilm image rather than the original printed work, the text may contain omissions and distortions related to the production of *EEBO*, and the process of remediation.

The transcribed text is browsable, searchable, and detached from its source. The transcription is not just another facsimile; it is a new work. Changes of font, typography, spelling and presentation intended to bring clarity to the text, and embedded coding to render the text searchable, create a new work with an ambivalent relation to the original. Early English authors did not write in XML mark-up code and their readers did not read standardized unadorned text! *EEBO-TCP* transcriptions are a modern academic creation. The transcribed text differs from the microform and digital image formats by several degrees of manipulation and bears the least resemblance or likeness to the

original print of the three Early English collections described in this article.

3 Through The Digital Veil

3.1 Surrogate or 'real thing'?

In a current online marketing brochure, ProQuest states that *EEBO* includes 'cover-to-cover full-page images that show the works *exactly as they appeared in their original printed editions*' (italics added) and that subscribing libraries can show users 'what the original readers saw, back when the Wars of the Roses still raged'.[7] Its promotional literature implies that *EEBO* contains clone-like copies of the original printed work. The student and scholar can therefore happily reside at home or their institution and conduct primary research, instead of traveling the world to libraries that still permit access to the original.

Enthusiastic scholars echo this notion of consulting the 'real thing'. John P. Schmitt, Regis University, describes the experience of accessing EEBO as 'a little like waking up in the British Library after closing time. The rare books of the British Library, Harvard, the Folger, the Huntington, and many others are suddenly accessible in their original appearance' (Schmitt, 2003). The enthusiasm engendered by the liberating experience of accessing facsimiles anywhere anytime coupled with the potential of online analysis undreamed of with print or microfilm, suspends the scholar's incredulity. The longer they look, the more the facsimile becomes the 'real thing'. The scholar rationalizes the only version of the work she will ever examine—the '*only* thing'—as the '*real* thing'.

The *EEBO* image is virtual. Although it manifests itself as a real object, it lacks the physicality of its microfilm or print predecessors. It is a representation of the original microfilm encoded in a matrix of binary 0's and 1's in a data file stored in computer memory (Technical Advisory Service for Images, 2006). It has multiplicity but no duration or place. The file is decoded as pixels refreshed on the screen at a rate that appears to the eye of the viewer as a physical reality. However, try to touch it, and you touch the computer screen. Try to reach in and pull the image out magically like a rabbit from a hat, and encounter failure. Each on-screen expression is an ephemeral array of pixels on a computer screen framed by the dual windows of the *EEBO* interface and computer monitor, and limited in duration to the last screen refreshment. The ambiguous materiality of the digital image is not unique. Photographs, for example, are a representation based on a matrix of dots. However, unlike the virtual screen image, photographs have duration and take up space.

While digitization gives unprecedented access to content, that content is distorted by virtue of its production, and the print work it purports to represent with exactness, while seeming so tantalizingly accessible, is illusive. The materiality of the microfilm and virtuality of the digitized object introduce new and conflicting layers of scholarly evidence related to remediation. In *EEBO-TCP*, content is transformed through the interpretation and re-engineering of the typographical evidence viewed on screen.

EEB is a surrogate of an Early English printed book (Tanselle, 2001)—itself a remediation of a handwritten manuscript. *EEBO* is a surrogate of a surrogate. However, substitution is not equivalence. Neither the microfilm nor the digital facsimile is equivalent to the original physical book, despite the enthusiastic descriptions of the publisher and scholars to the contrary. Remediation acts like a distorting lens or opaque veil through which the scholar 'sees' the mediated Early English book. Scholars may attempt to penetrate the opacity of the new host medium through imagination and scholarly grit, but in the end, the facsimile is not, and perhaps never can be 'the real thing' or equivalent of the original.

3.2 Suspension of disbelief

The willing suspension of disbelief, a concept encountered in literary theory, may account for the scholar's credulity before a digital facsimile. Like the reader of a work of fiction, or a member of a live theatre audience, the *EEBO* scholar is engaged with the surrogate as if it were 'the real thing', knowing very well that it is not. Claims for an identical relationship between a digital facsimile

and its original, and the transparency of the user-resource experience, ignore the myriad transformations that occur when producing a facsimile in another medium. The process of remediation lowers a digital veil of varying opacity between the scholar and the original work. The scholar, gazing through that veil, seduced by the transformations wrought by remediation, suspends disbelief in order to advance the study of the text. Successful remediation depends on the witting and unwitting complicity of the viewer.

4 Filters and Limits of Remediation

Digital facsimiles have changed the way the modern scholars do their research and exposed students and even non-academics to works they would rarely, if ever, have viewed on microfilm and in the original print. If the digital facsimile is to serve as a virtual and ubiquitous stand-in for the original work and undergo such wide scholarly scrutiny, then it is very desirable that its identity be as true as possible. What standard of identicalness should such a facsimile have? Should it match the likeness of two copies from the same print run? Despite small differences, the illusion of identicalness in print runs is compelling. The remediated work has a much lower standard of identity with the original. However earnest the effort and advanced the technology, the migration from original to copy, and from one medium to another results in mutation. Remediation launches a new artifact: a 'point of view' copy essentially transformed through its migration to a new medium. The following is a review of selected production and media filters and limits in *EEBO* and *EEB* that reveal how fundamental and transformative the process of remediation is, even for the facsimile, a relatively simple and straightforward remediation type.

4.1 Digitizing microfilm

Although archival quality is superior when created directly from print rather than microfilm, ProQuest's decision to digitize the microfilm facsimiles it owned rather than print copies it did not own, was a logical and shrewd one. The cost of using the latter would have been prohibitive. Even if the originals were still extant, would libraries permit access to a commercial enterprise with so much potential to harm their culturally important holdings? In the end, the choice was fortunate since the technology current at the time the project was launched would almost certainly have required the originals to be dis-bound—an alarming prospect given the relative rarity of Early English book copies. In the short term, scholars now find themselves with a resource, purchased at great expense by their institutions, that has equivocal value because it is not a complete scholarly tool. ProQuest's choice, driven by economics and practicality, places the remediated microfilm image at the center of Early English textual studies for years to come, at least until there is a renewed drive to digitize extant print copies directly using state of the art technology, thereby substantially surpassing *EEB* and *EEBO* in resolution and completeness. *EEBO* prolongs the influence of the microform facsimile and ensures that libraries continue to acquire and maintain *EEB* as a necessary supplement.

4.2 A single copy

EEB and *EEBO* typically include only a single copy of an edition of a work. This pragmatic decision is an unnecessarily limiting legacy in the digital environment where space and access issues are less significant than for print, and where online access provides scholars with an ideal setting for comparative analyses. For the bibliographer, one copy may be better than nothing if no other copy exists, but as many copies as possible is best.

4.3 The best available copy

In theory, the *EEB* project sought only the best available print copy to microfilm. In practice, the majority of selected copies came expediently from the nearby British Museum library or The British Library (founded in 1973 from the Museum's core collection). Improvements in communication and cooperation may make it easier to identify 'best' copies from around the world as the *EEB* project nears completion, but cost and expediency are still major selection factors. Even if a scholar were

to review all or even several print copies of a title located in libraries across the globe, the selection of the 'best' copy would be highly subjective. What is best for one purpose may be poor for another. For example, which is the best—an incomplete copy that includes an exceptional hand-painted frontispiece, or the copy without the painting but otherwise complete?

4.4 'Best quality' image

Microfilm and digital scanning operators have a lot of technical leeway to achieve the best quality image. They make choices to de-speckle, de-skew, color enhance, and edit out blemishes, annotations or shadow from the verso page. Their technical skill and subjective judgment affect the quality of the image, making the notion of 'best quality' relative. The quality of digital images depends on 'the quality of the initial scan, which depends on the equipment, its calibration, the judgment of the scanning operator in using its capabilities, and environmental characteristics, such as dust and lighting' (Arms, 1999).

4.5 Content amputation

While some content loss may be inevitable due to the production process, a facsimile is by definition lossless. If content is missing, it is also missing in the original. Is content loss in *EEB* and *EEBO* benign or a form of amputation—accidental or deliberate gaps created for reasons of expediency, subjective production decisions, or inherent qualities of the medium?

- *Cover-to-cover.* Despite UMI's intentions, microfilm and subsequently digitized microfilm, do not replicate the original cover to cover. Besides the inevitable pages missed in the filming or scanning process, there are other notable gaps. Works routinely start with the title page and end with the last page of text, excluding front and end pages, consecutive blanks, and special front matter, such as incunabula paintings. This practice results in the loss of valuable details about book production, illustration practices, and reading practices gleaned from page notes and inscriptions. Digital scanning adds more human and machine error into the mix. For example, partially scanned pages mean the missing section is lost forever unless rescanned.
- *Cropping.* By current microfilm and digital standards (The Library of Congress, 1997), an image must include a minimum margin surround of a target object. This standard, sometimes referred to as the '1-inch rule', guarantees that an object is captured in its entirety, including its edges. Especially in the early years, *EEB* scanning operators routinely cropped page margins to fit the narrow 35 mm film width or correct skew. By slicing away important material evidence such as marginalia or poor plate registration, systematic cropping misrepresents the physicality of the original work (Viscomi, 2002).
- *Binding evidence.* By systematically excluding images of binding covers and spines, *EEB* and *EEBO* contain no evidence for the study of binding techniques, or cover or spine design and lettering. Current digital facsimile projects, such as the British Library's *Turning the Pages*[8] project, routinely reproduce binding evidence.
- *Illustration extraction. EEBO* adds illustration findability to the un-indexed and caption-less illustrations in Early English books, by capturing them as separate numbered images. However, it does not capture decorative details, such as decorated initials, small illustrations or decorations, and marginalia. Automated extraction and the intervention of a scanning operator have considerable potential for error and distortion. For example, automated capture routinely mistakes unusual text layout or column formatting for an illustration or a table.

4.6 Page distortion

By digitally preserving *EEB* page distortions, *EEBO* promotes the distortions as an integral part of a work.

- *Page curvature.* Pages curve outward from a distinct V-shape at the center of the top edge of the 'open book' in *EEB* and *EEBO*. Lines of text spread outward and downward like flying buttresses from the darkness of the centerfold. The curvature is often so extreme and the gutters so dark that nearby text is unreadable. We might

therefore forgive a novice reader for concluding that printers of the period adopted a curved line standard, with fonts of diminishing size and progressive 'squashed' appearance towards a black center strip bleeding deeply into the adjacent white space and sometimes into the text.

- *Page flattening*. Before cradle technology and the application of more rigorous microfilm production standards, technicians often attempted to correct page curvature by using manual page-flattening techniques. *EEB* has many examples of technicians using their fingers to pin pages back in a futile attempt to correct curvature. For example, STC/864:09[9] captures two thumbs pinning down the edges of the lower corners of the facing pages. *EEBO* preserves the image faithfully for the edification of digital archaeologists and bibliographers alike.
- *Page skew*. Print registration is the process of aligning text blocks to the edge of a sheet of paper, or graphics or tables within a text. When the alignment is 'out of register', the text block, graphic, or table appears to be off the square, or 'skewed'. There are many examples in *EEB* and *EEBO* of registration skew correction or 'de-skewing' by cropping and re-orientation of the page when microfilming or digitizing, and the subsequent corruption of bibliographic evidence.

4.7 Microfilm deterioration

At the start of its *EEB* project, UMI selected cellulose acetate base film, also known as 'safety film', and later the new more stable and rip-resistant polyester base film after the mid-1980s. Although acetate base film was a major improvement over the explosive nitrate base film that preceded it, it has its own problems. Acetate is subject to deterioration from off-gassing caused by acetic acid, sometimes referred to as the 'vinegar syndrome', named after the pungent smell associated with aging acetate microfilm collections. Over time, it can shrink, bubble, tear, and stick together (Baker, 2001). The soft layer of binding emulsion (gelatin or albumin) and silver halide compounds used to create the image are susceptible to both physical and bio-deterioration (Cappitelli and Sorline, 2005) that corrupt image information. Even with the best of physical and bio-controls, *EEB* master copies on acetate film would have suffered the same assaults on their integrity as library distribution copies. UMI's microfilm vaults contained similar nasty surprises and the distinct odor of vinegar. New microfilm masters generated to remedy such problems as off-gassing, physical deterioration, and the scourges of light, humidity and microorganisms, preserved evidence of past deterioration, that was later preserved digitally in *EEBO*.

4.8 Low resolution and loss of detail

Cornell University adopted a minimum of 600 PPI, 1-bit depth standard, as early as 1990 for its digital and Computer Output Microfilm (COM) project to preserve its nineteenth and twentieth century brittle books. It concluded this standard 'adequately captures the fine detail, elaborate serifed script, italics, and small body heights' that characterize fonts used in printing between 1850 and 1950. COM created from digitally scanned images using this standard met or exceeded the ANSI and Association for Information and Image Management (ANSI/AIIM) microfilm standards for image quality and permanence (Kenney, 1996). The *EEBO* standard of 400 PPI involves a considerable loss of detail more appropriate for derivative or access facsimiles of standard text rather than the fine detail of fonts and illustrations in Early English books.

4.9 Bi-tonal reduction

The resolution used for the *EEB* project was too low for effective grayscale capture, resulting in a bi-tonal black and white reproduction. According to Gabriel Egan and John Jowett, digitization reduced the quality of the evidence further. *EEBO* scanned the microfilm as bi-tonal digital images, capturing only black microfilm dots and stripping out all vestiges of grayscale with subsequent visible loss of detail (Jowett and Egan, 2001). While not particularly noted for color, Early English books do share basic color characteristics such as ink hue, paper color and tonal variation, and especially in the incunabula period, hand-painted frontispieces and illustrations, and rubrication. Rubrics, or text in red ink, such as capital letters, signal breaks or headings, or highlight text. They often mark the beginning

or end of a section, acting something like a paragraph indent for the modern reader. Because *EEB* and *EEBO* capture little tonal variation, they preserve little evidence of color use. Their bi-tonal black and white images make rubrics difficult to identify. Rubrified letters or markings, especially hand-painted ones appear muddy and indistinct. Bi-tonal digital reduction of microfilm removes or distorts evidence of microfilm deterioration, such as blemishes. A microfilm blemish may be mistaken for a black blob of unknown origin. Is it a printer's error or a blot from an annotator's pen? The multiple layers of remediation and deterioration merge into a single undifferentiated layer, making it difficult for the digital scholar to distinguish 'noise' from information, the influence of the medium, from 'real' evidence. To illustrate how bi-tonal black and white digital images capture the look of a printed work differently than color, we can compare two copies of the same page. The *EEBO* copy is in Ranulphus Higden's *Polycronicon*, published by William Caxton in 1482.[10] The other is a single leaf[11] posted publicly on the Internet from a small, digitized collection of incunabula leaves at the National Diet Library (NDL) in Japan. The copies appear to be from the same print run. Of the two copies, the NDL facsimile has more detail and verisimilitude. It captures the look and feel of an actual page better than the *EEBO* version, partly due to the leaf's truer alignment and appearance as a flat printed page, and to the absence of font, line, and gutter distortions. More important, the NDL facsimile is in color. The absence of color in *EEBO* makes the text less readable and alters the way the modern reader experiences the work compared with the Early English reader. Microfilm (and by extension digitized microfilm) is not a medium designed 'for natural, realistic tonal capture but for optimal legibility' (California Digital Library, 2001). Microfilm is primarily useful as a medium for the preservation of text.

5 Conclusion

5.1 Pouring old media into new

Delivered with exhilarating speed and ease of access, the accretion of media layers in *EEBO* may seem deceptively light. Bolter and Grusin describe how some digitized products are presented 'as if the content of the older media could simply be poured into the new one'. The digital medium strives to 'erase itself' in the process. The transparent online layer introduces no filters or barriers to the old medium (Bolter and Grusin, 2000). The publisher represents *EEBO* to the scholar as if the medium makes little difference. However, can old media simply be poured into a new communication matrix, and re-emerge as itself only better? Can 'being digital' faithfully clone 'being analog'? In the process of becoming digital, is the analog original transformed in fundamental ways that transcend changes attributable to speed and convenience alone?

Marshall McLuhan mused, 'no medium has its meaning or existence alone, but only in constant interplay with other media' (McLuhan, 2001). The research for this article suggests that new media *absorbs* old media in the process of remediation. In *EEBO*, the old communication technologies of microfilm and print are absorbed *within* the new online medium. They do not exist *alongside* or *separately* as perfect facsimiles in a new medium. There is a measurable loss of qualities unique to the original medium and a simultaneous gain of qualities unique to digital media. The process of remediation transforms content as it filters through a new medium. The media layers work as distancing and distorting agents. The scholar, several times removed from the Early English book when viewing an image in *EEB*, increases that distance in *EEBO*, and again in *EEBO-TCP*.

5.2 Text preservation and access

EEBO's primary importance may be to provide online access to Early English texts and make them broadly accessible. It brings primary texts to the desktop of both scholars and students from an ever-widening range of disciplines and transforms the way in which Early English scholars do their research and teach. Where once only a few scholars and graduate students had access to the microfilm, or more rarely, to print copies, access is now broadly available, even to undergraduates. Possibilities for cross-disciplinary synergies

are boundless. Online analysis permits new kinds of research never possible in print or microfilm. In an instant, a scholar can sift through a wealth of content that previously might have taken a lifetime of painstaking, costly effort. *EEBO* expands the influence of Early English books far beyond remote, rare book collections or the narrow confines of microfilm reading rooms. As Peter White describes it, *EEBO* liberates rare texts from the restrictive cultural framework of 'the library' and 'learning' (White, 2004).

5.3 Replicating the book

The primary goal of a facsimile is to replicate the look and feel of the original in addition to the text. How successful are *EEB* and *EEBO* in replicating the Early English print book, for example, its 'dimensions, thickness, page form, and general design style?' (Landoni *et al.*, 2000). They preserve the text, but little of the book as a physical object. They present ambivalent information about key physical characteristics, such as size, presence, typography, and context.

5.3.1 *Dimensions*

Neither *EEB* nor *EEBO* portray physical size accurately. Standards for the presentation of scale or reduction ratios are absent or inconsistently applied through the long history of *EEB* microfilm production. The metadata excludes book dimensions and reduction ratios and since rulers are often absent, dimensions cannot be determined through either consulting the image or the bibliographic record. The absence of reliable dimensions is a serious barrier to bibliographic scholarship, where size matters.

5.3.2 *Physicality*

Details of how pages are gathered, or what sheet format (folio, quarto, octavo) was used, are not described in the metadata. Because of low resolution and cropping, such information is absent or difficult to garner from visual evidence. There are no watermark details for the paper and paper mill scholar. There are no descriptive or visual details about binding. Both microfilm and digital products have characteristics that emulate but do not capture the physicality of print. The distinct physicality of accessing microfilm bears little resemblance to opening a print book. For the *EEBO* user, the computer, computer monitor, keyboard, and mouse are the dominant physical presence. The works themselves are illusive virtual objects, matrices of light and dark on a screen. It is impossible to handle them physically or view them three-dimensionally.

5.3.3 *Gauging context*

In a print book, the reader can gauge context in relation to the whole work using visual queues. Fore-edge dimensions (edges opposite the spine) convey document length and position relative to the beginning or end, and therefore progress through the work. Abstract devices such as page numbers verify and specify the visual and physical sense of position. Equipped with a page number, or a numeric position within a reel, and an acquired sense of the time it takes to slow or fast forward to a targeted page, the experience of negotiating a work in *EEB* bears little resemblance to riffling back and forth through a print book. In *EEB*, the absence of fore-edges and other reliable physical queues makes the relation of the page to the whole, ambiguous. The navigation of a work in *EEBO* bears even less resemblance to the physicality of a book. Hidden algorithms, parse a matrix of binary digits, '0' or '1', and control access. Icons and other navigation links orient access. Sequential browsing imitates paging through a work, but the context is abstract and relies on the reader's memory. Alternatively, the reader may skip blithely from one image to another, using abstract and historical references such as page number, STC number, and reel number, rather than physical queues. Without context, the images materialize like objects adrift on a fogged in great digital sea.

5.4 Single source or matrix?

According to a ProQuest spokesperson, the ultimate goal of *EEBO* is 'to provide scholars with a single source for research on Early English Books, including bibliographic citations, full-page representations of all images, and ASCII-encoded text' (McLean, 2001). While *EEBO* provides the latter options and more, it stops far short of the goal of providing a 'one-stop' research environment and a single scholarly source for Early English books.

At the same time as it promotes *EEBO* as a single scholarly source for Early English books (McLean, 2001), ProQuest directs the scholar to also consult the *EEB* microfilm facsimile for its higher resolution images and *EEBO-TCP* for its full text transcript and searching capability. ProQuest describes *EBBO* as 'a complement, not a replacement for the higher resolution images that exist on microfilm' (Pack, 1999). The microfilm images have greater fidelity to the original printed works. Libraries, hoping to provide a scholarly collection of Early English publications must purchase or retain existing *EEB* subscriptions and associated preservation quality storage, and participate in *EEBO-TCP*. Early English scholars will insist on it. Even then, the quest for completeness is illusive without access to the original print and manuscript copies.

5.5 The quest for an early english digital codex

Although improvements in communication technologies promise to bring the reality of the perfect 'clone' facsimile ever closer, is it achievable? Is it desirable? The cost would be prohibitive. The remediated digital facsimile cannot overcome its surrogacy or the incompleteness of absorption of the old medium. The question is not whether it needs to do this (does print have to justify that it is not a manuscript?), but whether publishers and scholars will openly acknowledge the limits of remediation, guard against publisher claims of authenticity and avoid scholarly artifact creation and misreading. Scholars may be keenly aware of the limits of remediated primary sources, but it is a learned awareness. Each generation must rediscover for itself the limits of new media and interpret how it affects their research.

EEBO has considerable scholarly value, but it does not contain identical copies of the original microfilmed by UMI and certainly not identical copies of the one pulled from the printing press, bound, and read when first published. Like the modern connoisseur of Renaissance paintings who prefers the darkened and faded colors created by time and environmental assaults rather than the vividly colored restoration, the student and scholar of Early English books runs the risk of revering the digital image in all its surrogate glory, and preferring it to the print book it is replicating.

The value of *EEB*, *EEBO*, and *EEBO-TCP* cannot be understated. Are they the foundation for a true Early English digital codex, accurately reflecting not only the text, but the physicality of the original? To create that paragon, it would be necessary to return to the original print copy, use new technology that will not damage or violate the print copy, and produce 3D and virtual world facsimiles. The examples of recent digitization projects at The British Library show progress towards achieving similitude. According to Oliver Grau, it is just a matter of years 'before computing power is available to realize high-definition spaces of illusion' (Grau, 2003). The digital codex will become a reality eventually, while the microfilm and digitized images will continue to have value as 'point of view' images and as highly processed historical artifacts.

References

Arms, C. R. (1999). Getting the picture: observations from the library of congress on providing online access to pictorial images. *Library Trends*, **48**(2): 379–409.

Baker, N. (2001). *Double Fold: Libraries and the Assault on Paper*. NY: Random House.

Bolter, J. D. and Grusin, R. (2000). *Remediation: Understanding New Media*. Cambridge, MA: The MIT Press.

California Digital Library Technical Architecture and Standards Workgroup (2001). Best practices for image capture. http://www.cdlib.org/news/pdf/BestPracticeImageCapture.pdf

Cappitelli, F. and Sorline, C. (2005). From Papyrus to compact disc: the microbial deterioration of documentary heritage. *Critical Reviews in Microbiology*, **31**(1): 1–10.

Grau, O. (2003). *Virtual Art: From Illusion to Immersion*. Cambridge, MA: The MIT Press.

JISC: The Joint Information Systems Committee (2004). *Early English books online: the holy grail of online resources?* http://www.jisc.ac.uk/index.cfm?name=news_eebo

Jowett, J. and Egan, G. (2001). Review of the Early English Books Online (EEBO). *Interactive Early Modern Literary Studies*, January: 1–13. http://www.shu.ac.uk/emls/iemls/reviews/jowetteebo.htm.

Landoni, M., Wilson, R, and Gibb, F. (2000). From the visual book to the WEB book: the importance of design. *The Electronic Library*, **18**(6): 407–19.

Kenney, A. R. (1996). *Digital to Microfilm Conversion: A Demonstration Project: 1994-1996*, Final Report to the National Endowment for the Humanities. Ithaca, NY: Cornell University Library, Department of Preservation and Conservation. http://www.library.cornell.edu/preservation/com/comfin.html

McLean, A. J. (2001). Early British printing meets the electronic age: a large-scale digitization case study. *Microform & Imaging Review*, **30**(4): 127–134.

McLuhan, M. (2001). *Understanding Media: The Extensions of Man*. Cambridge, MA: The MIT Press.

Pack, T. (1999). Bringing literature alive: Early English Books Online reshape research opportunities. *EContent*, **22**(6): 26–30.

Review - Nicholson Baker, *Double Fold: Libraries and the Assault on Paper* (2001). *Microform & Imaging Review*, 30(3): 82–98.

Sandler, M. (2003). Academic and commercial roles in building "The Digital Library". *Collections Management*, **28**(1/2): 107–119.

Schmitt, J. P. (2003). Early English Books Online. *The Charleston ADVISOR*, **4**(4), http://www.charlestonco.com/review.cfm?id=141

Tanselle, G. T. (2001). Not the real thing. *The Times Literary Supplement*, **5134**: 14.

Technical Advisory Service for Images (TASI) (2006). *Do Digital Images Exist in the Real World?* http://www.tasi.ac.uk/advice/creating/digitalimage.html.

The Library of Congress [Memory Project] (1997). *Request for proposal for digital images of pictorial materials*. RFP 97-9. http://memory.loc.gov/ammem/prpsal9/rfp9c.html

Viscomi, J. (2002). Digital facsimiles: reading the william blake archive. *Computers and the Humanities*, **36**(1): 27–48.

White, P. (2004). *Early English Books Online (EEBO)*. Presentation at the European University Institute, Florence. http://www.iue.it/LIB/Guides/History/IUE-EEBO_files/frame.htm#slide0002.htm

Notes

1 Available at: http://en.wikipedia.org

2 In a ProQuest Company brochure, *Early English Books Online*. Available: http://www.proquest.com/products/pdf/eebo.pdf

3 STC or *Short-Title Catalogue of English Books* is comprised of: *A Short-Title Catalogue of Books Printed in England, Scotland, & Ireland and of English Books Printed Abroad, 1475-1640*, compiled by Alfred W. Pollard and G. R. Redgrave (STC I, Pollard & Redgrave) (London: Bibliographical Society, 1926; 2nd edn, 1976–91), and the *Short-title Catalogue of Books Printed in England, Scotland, Ireland, Wales, and British America and of English Books Printed in Other Countries, 1641-1700*, compiled by Donald Goddard Wing (STC II, Wing) (Modern Language Association of America, 2nd edn, 1982–98).

4 *Thomason Tracts* is based on a separate catalog of over 22,000 pamphlets, periodicals, and broadsides collected by George Thomason, a London publisher and bookseller, during the English Civil War and Interregnum, 1640–61. The *Early English Books Tract Supplement* includes small publications such as broadsides and pamphlets, from sixteenth and seventeenth century Britain, organized and collected in 'scrapbooks' or tract volumes located primarily in The British Library.

5 As of February 2003, *EEB* included only about 40% of the titles in the *English Short Title Catalogue* (*ESTC*), a comprehensive index of Early English works published in Great Britain and its dependencies between 1473 and 1800. *ESTC* is a union catalog covering the Early English collections of over 1,600 libraries worldwide. It includes shelf marks and location, and identifies the *EEB* copy (normally only one).

6 Bit depth, sometimes called 'brightness resolution', defines the palette range of tones or colors of a pixel. Higher bit depth results in improved preservation quality with more potential research value over the long term.

7 In the ProQuest Company brochure, *Early English Books Online*. Available at: http://www.umi.com/products/pdf/eebo.pdf

8 Available at: http://www.bl.uk/onlinegallery/ttp/ttpbooks.html

9 In Kinki Abenezrah's, *An Everlasting Prognostication of the Change of Weather: Collected and Compiled for the Common Vse and Profit of all Country Men*, published in London in 1625 STC (2nd ed.)/62 (microfilm), and included in EEBO.

10 Ranulf [sic] Higden's *Prolicionycion* [sic], printed by William Caxton in 1482 (STC (2nd ed./13438, Image 312) in *EEBO*.

11 The same page in another copy of Ranulphus Higden's *Polycronicon* digitized for the *Incunabula Collection* of the National Diet Library in Japan. http://www.ndl.go.jp/incunabula/e/collection/incu_65.html