Преглед НЦД 4 (2004), 6–21

Kevin Kiernan
(University of Kentucky, Department of English)

Brent Seales and James Griffioen
(University of Kentucky, Department of Computer Science)

THE REAPPEARANCES OF ST. BASIL THE GREAT IN BRITISH LIBRARY MS COTTON OTHO B. X*

Abstract. We illustrate how digital technology can help restore greviously damaged, practically inaccessible, medieval manuscripts. Our example was an important version of the Old English *Life of St. Basil the Great* now clinging to existence in charred, illegible, disordered, and strangely dispersed fragments.

Key words: digital restoration; ultraviolet imaging; 3D imaging; editing tools

One of St. Basil the Great's great miracles was rendering a manuscript totally illegible. He was able to ruin all but one reading in this manuscript during his lifetime, and he obliterated the last stubborn reading as his first posthumous miracle. While he was still alive, he used "remote access" in another miracle to alienate a manuscript from its rightful owner (the devil, as it happens), then tore it up. It seems therefore devilishly ironic that the Old English versions of his life mostly survive in ruined bits and illegible pieces – including two halves of the folio describing the manuscript he ripped apart! In an interdisciplinary project called "The Digital Atheneum: new techniques for restoring, searching, and editing humanities collections," we are attempting to reverse some of the damage done to manuscripts in the infamous Cottonian Library fire of 1731, and to make them accessible again through electronic editions.[1] We chose to confine our research to heavily damaged manuscripts, because we would presumably master easy problems while striving to solve difficult ones, and we would make available some currently unusable material for new research. In this group of manuscripts one of the biggest challenges for Anglo-Saxon scholars and computer scientists alike is Cotton Otho B. x, once a large Old English miscellany dominated by saints' lives written by an eleventh-century monk named Ælfric. The first saint's life in this wreck of a manuscript is the *Life of St. Basil the Great*, which furnishes many opportunities to show how new computing techniques and electronic editions can help restore damaged manuscripts and provide easy access to formerly inaccessible texts [2].

The Anglo-Saxons of the tenth and eleventh centuries venerated the prolific Church Father, St. Basil the Great (330–379) [3]. To judge by the *Life* he wrote, Ælfric admired St. Basil as an activist bishop and monk who wrote a monastic rule, contributed to the Eastern ritual of the Mass, fought heresies, bravely stood up to emperors, and worked many miracles in the service of his flock. In the early tenth century King Athelstan (924–939), a spectacularly successful collector of saints' relics, had acquired among his hundreds of relics one of St. Basil's teeth as well as his bishop's crosier [4].

In addition to the extant reliquary lists, evidence of his cult in late Anglo-Saxon England comes from seven surviving liturgical calendars [5] as well as from three extant vernacular copies of his life [6]. Two of these manuscripts of his life are in fragmentary state, eerily reminiscent of that reliquary tooth King Athelstan acquired, but one full version survives in the handsomely written and well-preserved British Library MS Cotton Julius E. vii, the basis of our only modern edition [7]. The complete Julius version allows us to understand where the other surviving fragments once belonged in their respective manuscripts. The University of Toronto's Dictionary of Old English project has greatly simplified the task by providing an online version of Skeat's edition [8]. It is as if King Athelstan had St. Basil's dental records to go with his reliquary tooth.

Even with the aid of Skeat's edition of St. Basil, however, it is impossible for a reader today to sit down with Cotton Otho B. x and read the remaining fragments in their correct sequence. The manuscript was in such miserable condition after the fire of 1731 that it was considered unusable and was packed away and forgotten for well over a century in a garret in the British Museum. In 1837 Sir Frederic Madden rediscovered it and many other lost manuscripts in the garret, and later as Keeper of Manuscripts assigned a staff member named N.E.S.A. Hamilton to try to sort out the thoroughly disordered leaves in preparation for restoration binding. In 1863, when Otho B. x was ready for a new binding, Madden was not satisfied with the results and subsequently tried to order the leaves properly himself, but without great success.[9] For example, after all of Hamilton's and Madden's efforts, one must still today be ready to flip pages back and forth to read the first four folios of Cotton Otho B. x in their correct order: first, fol. 60, beginning with the verso; second, fol. 36, also beginning with the verso; third, fol. 49; and finally fol. 1.

To read the St. Basil portion in sequence is no easy task. One must begin with folio 3, skip to folio 5, and then move back to folio 4. After that, to stay in sequence, the reader must take a day-trip to Oxford and the Bodleian Library to study a stray leaf from Cotton Otho B. x that was taken there in November 1731, a couple of weeks after the fire [10]. The reader must stay alert, because the Bodleian Library, in the tradition of this unlucky manuscript, has mislabeled the leaf with the verso as the recto, and the recto as the verso. Finally (perhaps more taxing than the trip to Oxford), this diligent student must return to London and the manuscript in the British Library, and try to read line by line from fragment 50 to fragment 6 throughout the recto, and line by line from fragment 6 to fragment 50 throughout the verso, because "folios" 50 and 6 in the manuscript are actually two parts of the same folio in the wrong order. In short, the texts in Cotton Otho B. x are practically inaccessible and are accordingly ideal candidates for some radical restoration.

The official foliation is not only useless, but actually detrimental to a study of the manuscript, because its numbers lead readers in all the wrong directions. To facilitate study of this manuscript I have devised an "electronic foliation" that presents the leaves in their correct order, but also keeps track (in parentheses) of their official British Library foliation numbers and of the Bodleian Library pressmark for MS Rawlinson Q. e. 20, the leaf from St. Basil now alienated in Oxford. With all the surviving leaves of Cotton Otho B. x in order, St. Basil should begin on the fifth folio of the codex, not the third. Thus for the five (not really six) surviving leaves of St. Basil this e-foliation is folio 5(3), 6(5), 7(4), 8(RQe20v)r, 8(RQe20r)v, 9(50+6) for the recto, and 9(6+50) for the verso.

Here is an overview of all the surviving folios:

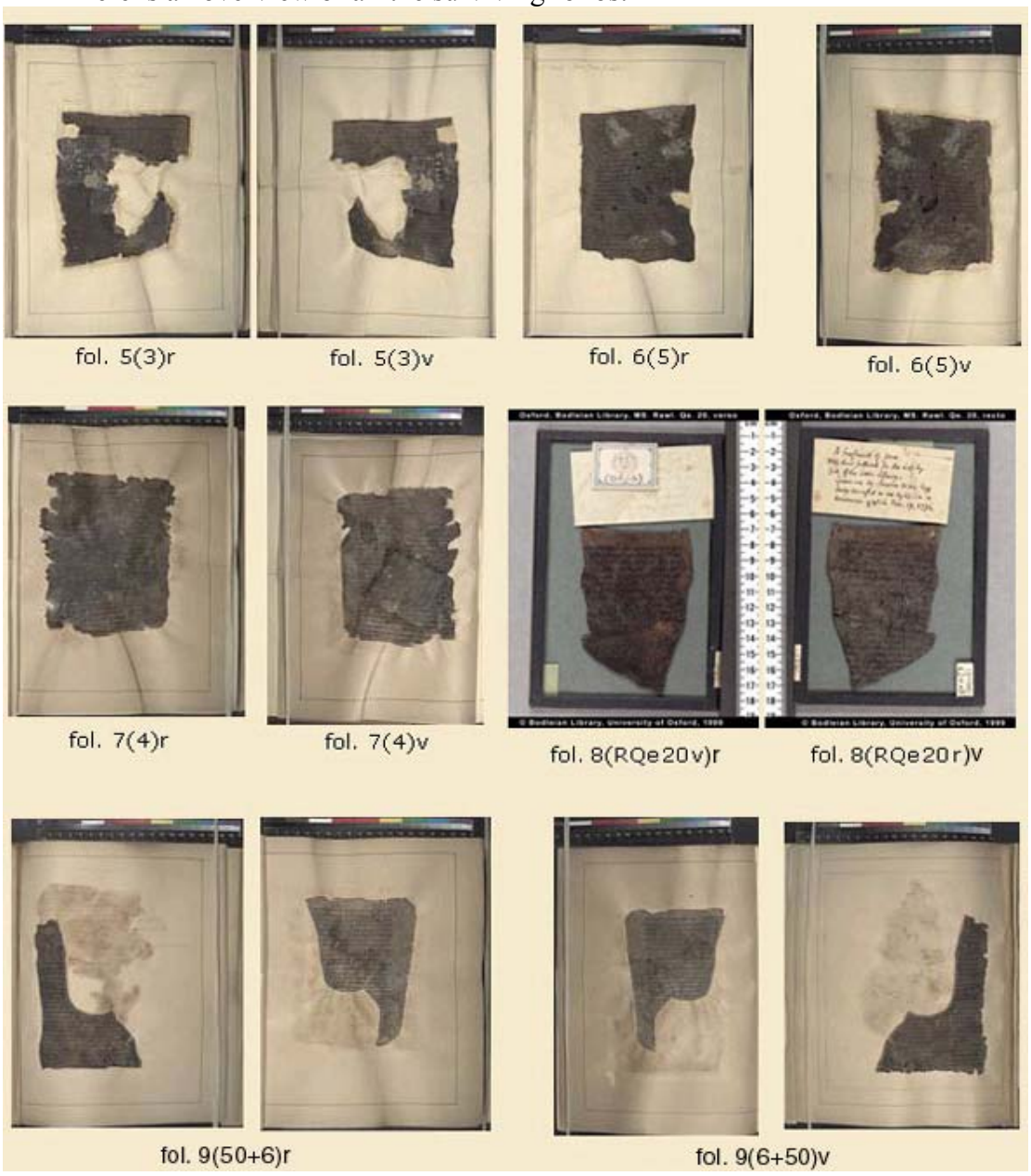


Fig. 1 Thumbnails

This useful e-foliation will never have more than a virtual reality, for two overpowering reasons: although they may easily rectify the mislabeling of the recto and verso, the Bodleian Library, Oxford, is not likely to repatriate the leaf it now owns to its original manuscript at the British Library; and the British Library, even if it does eventually put the leaves of Cotton Otho B. x into their proper order, is not likely to remove fragments 50 and 6, the misbound parts of a single folio, from their protective paper frames and rebind them together [11]. With the cooperation of the respective repositories, however, these highly improbable events are of course easy to accomplish in an electronic edition.

By putting together all of the extant fragments of the *Life of St. Basil the Great* in Otho B. x and fully collating them for the first time with the version that survives in

Julius E. vii, it is possible to see how much survives of the original manuscript. The best preserved leaves show that there were 29 lines of text per folio page. The fragment of a leaf at Oxford preserves only about 22 lines, but we have the text from Julius to show what was lost at the side and the bottom of the recto, and we can restore with relative certainty the amount that was similarly lost at the side and the bottom of the verso. Between the end of fol. 8(RQe20r) v and the beginning of the split leaf, fol. 9(50+6)r are 66 lines of verse, which would have fit on a single folio (58 MS lines). Thus nearly half of the original manuscript of the *Life of St. Basil the Great* in Cotton Otho B. x still survives today.

One manuscript leaf that might have been a fascinating research project is unfortunately beyond restoration, because St. Basil destroyed it. In the culminating episode of Ælfric's *Life of St. Basil the Great*, a rich widow who had been living "as a pig in muck" (*swa swin on meoxe* Skeat 528) determines to amend her life by writing down all her sins on a vellum leaf, sealing it with lead, and then asking St. Basil to obliterate the list of sins. Deciding to take on this innovative approach to penitence, St. Basil prays that as Christ's own deed blots out sins and as all our sins are written down with Him, Christ should help him deface the manuscript (Skeat 541–548). After a night of prayer, St. Basil successfully blots out all but the most mortal sin (Skeat 551–553). Claiming he is too sinful to eradicate this one, he sends her to a hermit in the wilderness to erase it. The long journey seems a ploy to get rid of her when St. Basil, having foreseen the time of his death, immediately prepares to die as soon as she leaves [12]. It is a curious excursion, because her pilgrimage turns out to be a waste of time. The hermit tells her that only St. Basil the Great can help her [13]! Rushing back to Caesarea, she is understandably upset to find St. Basil dead and lying on his bier. "She then threw the writing on the bier, and told the men about her misdeeds." One of the attending priests, curious about her remaining sin, retrieves the manuscript and sounds frustrated when he discovers that everything is obliterated. "Why are you so worked up, woman?," he yells at her; "This vellum is blotted out!" (Skeat 640–644).

Fig. 2 Fol. 5(3)r

What does this strange story have to do with restoring manuscripts? A modern-day miracle would satisfy our desire to know the past and make the writing legible again. Ultraviolet and digital image-

processing can to some extent gratify these mundane desires on many of the illegible leaves that have survived from *St. Basil*. The first remaining folio of Cotton Otho B. x, fol. 5(3)r, is a good example. It resembles a medieval Hell's mouth that has devoured the text the missing vellum was supposed to save. Collation with the surviving version in Cotton Julius E. vii reveals that both the righteous and the unrighteous have been lost in this seemingly ravenous maw, including Eubolus, St. Basil's teacher and subsequent disciple, other learned Athenian philosophers, and two distinguished classmates, St. Gregory the Theologian of Nazianzus, and the Roman and Byzantine emperor Julian the Apostate (360–363). Historians (Old English *wyrdwriteres*) and their books (*bocum*) have not fared well in this Hell's mouth, in addition to the philosophers, theologians, and emperors.

Moreover, charring, discoloration, new layers of tape, and relatively recent attempts at preservation by applying gauze over glue, which has turned opaque over the years, have made it very difficult to read much of the text that has survived around the gaping hole. Fortunately, ultraviolet counteracts some of these features obscuring the text. By digitizing the ultraviolet effects and then processing the images, we can clearly restore, for example, the badly faded rubric and even penetrate the gauze and glue.

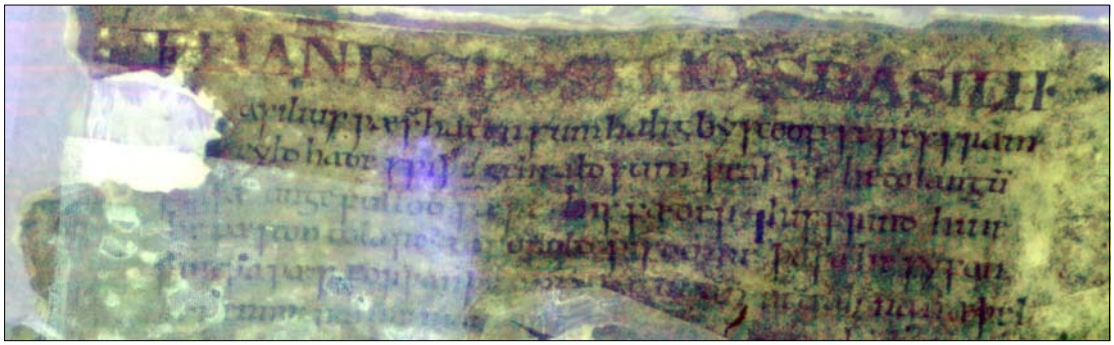
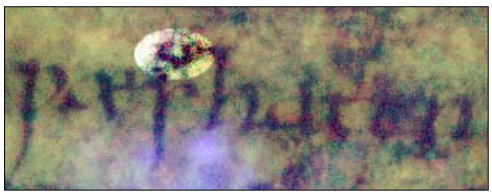


Fig. 3 Ultraviolet of 5(3)r

Even in strong light the leaf is so severely charred that it is difficult to read anything, but ultraviolet and contrast enhancement show that the rubric reads K[*a*]l[*end*] IAN[*uarii*] DEPOSITIO S[*ancti*] BASILII, attesting to Ælfric's eccentric dating in his sanctorale of St. Basil's feast day on 1 January, the date of his burial, instead of 14 June. Beneath the gauze to the left in line 5, it is now possible to see the normal genitive plural reading *wintra*, instead of the variant spelling *wintre* that appears in the Cotton Julius E. vii manuscript. Ultraviolet and image processing also disclose the scribe's superscript correction of *haten* to *gehaten* in the first line, a reading that scholars after the fire have been unable to see through the scorched vellum [14]. Here the scribe's superscript *ge* is digitally highlighted between *wæs* and *haten* [See Fig. 4]. Similar digital techniques reveal scores of variant readings Skeat was unable to detect and record in his published collations. Skeat says that he gives all variations that he is able to decipher,[15] but it is clear from studying the ultraviolet images that he was unable to see the great majority of variations. For example, in one twelve-line stretch of text on fol. 7(4)v3-15 (lines 172–185) Skeat finds only one variant reading, *segene* for *sægne*, "a statement, saying," whereas the ultraviolet image reveals



over two-dozen variants.

Another episode in St. Basil's career may be used to introduce the topic of remote access. The story is told on fol. 8(RQe20vr) rv, the leaf now in the Bodleian Library, Oxford (Skeat 204–264). After an arrogant altercation with St. Basil stemming from their schooldays in Athens, Julian the Apostate, who is on his way to fight the Persians, promises to lay waste to Caesarea when he returns. St. Basil warns the citizens of Julian's rage, and advises them to raise a tribute to placate him, advice that would have resonated with an Anglo-Saxon audience during the Viking incursions. Enjoying his own remote access, however, St. Basil is visited that night by the Virgin Mary and a heavenly host. They promptly mobilize the martyred St. Mercurius, whose relics and armor lie in St. Basil's church, to execute Julian in his camp for denying Christ and generally speaking pompously [16]. The Bodleian Library, Oxford, has recently made this leaf from Cotton Otho B. x accessible by remote access through its splendid website of manuscript images [17]. It is not apparent from the online images that the leaf is preserved in its own reliquary, sealed between glass plates, and kept in a small, book-like, solander made for the purpose. What can be seen, sealed with the fragment, is a note in the handwriting of the eighteenth-century Oxford antiquary, Thomas Hearne, explaining how it got into Oxford: "A Fragment of some MS. that suffered in the Loss by fire of the Cotton Library. Given me by Browne Willis, Esq. being brought to me by his son a commoner of Xt ch [i.e., Christ Church] Nov. 15. 1731."

In the context of digital restoration, it is interesting to compare the different ways that the British Museum and the Bodleian Library restored these fragments from the Life of St. Basil the Great in the nineteenth century. At the British Museum Henry Gough had perfected the process of inlaying leaves in paper frames, a process well-known from the Beowulf manuscript, and one that was used with varying success with the two damaged manuscripts containing the Life of St. Basil the Great. Ironically, Gough began working at the Bodleian, but was hired by Sir Frederic Madden to undertake the massive task of restoring the most ruined Cottonian manuscripts [18]. Gough's method was to trace each leaf on heavy construction paper, cut out the center leaving a retaining edge, and then paste the vellum leaf into the open space. One disastrous problem in the case of Otho B. x is that he used the new acidic paper for the frames, which have begun to crumble and have leached out stains from the leaves. A.S. Napier records the method of restoration for the Bodleian leaf in his 1887 note, "A Fragment of Ælfric's Lives of Saints," published a few months after the leaf was found by a librarian in a drawer in the Bodleian Library. Napier relates that "it was wrapped up in a piece of paper" containing the previously mentioned note by Thomas Hearne (378). According to Napier, "the fragment itself was so shriveled up and blackened by the heat that it was quite impossible to decipher it until it had been soaked in water and carefully stretched." The smooth edges of the fragment indicate that some trimming, whether on purpose or by accident, was done before the fragment was, as Napier says, "laid between two pieces of glass." The piece of paper that kept it safe from 1731 to 1887 was certainly trimmed, preserving only Hearne's note. As the Bodleian Library, Oxford, does not permit the use of its images on any website other than its own, I provide a sketch of the verso here.

This front view of the leaf is, as we have seen, mislabeled as the recto, rather than the verso. From what can now be deciphered in strong daylight or in the digital image, it appears that the text, especially on the recto, is no longer as legible as it was when Napier transcribed it. Based on my experience with the other Cotton Otho B. x leaves,

the text would almost certainly respond well to ultraviolet, but whether ultraviolet would penetrate the glass is perhaps less likely. The glass casing is not a good way to preserve the natural suppleness of vellum, and the Bodleian Library might consider using modern conservation techniques for the leaf, as the British Library is doing for the rest of the manuscript [19].

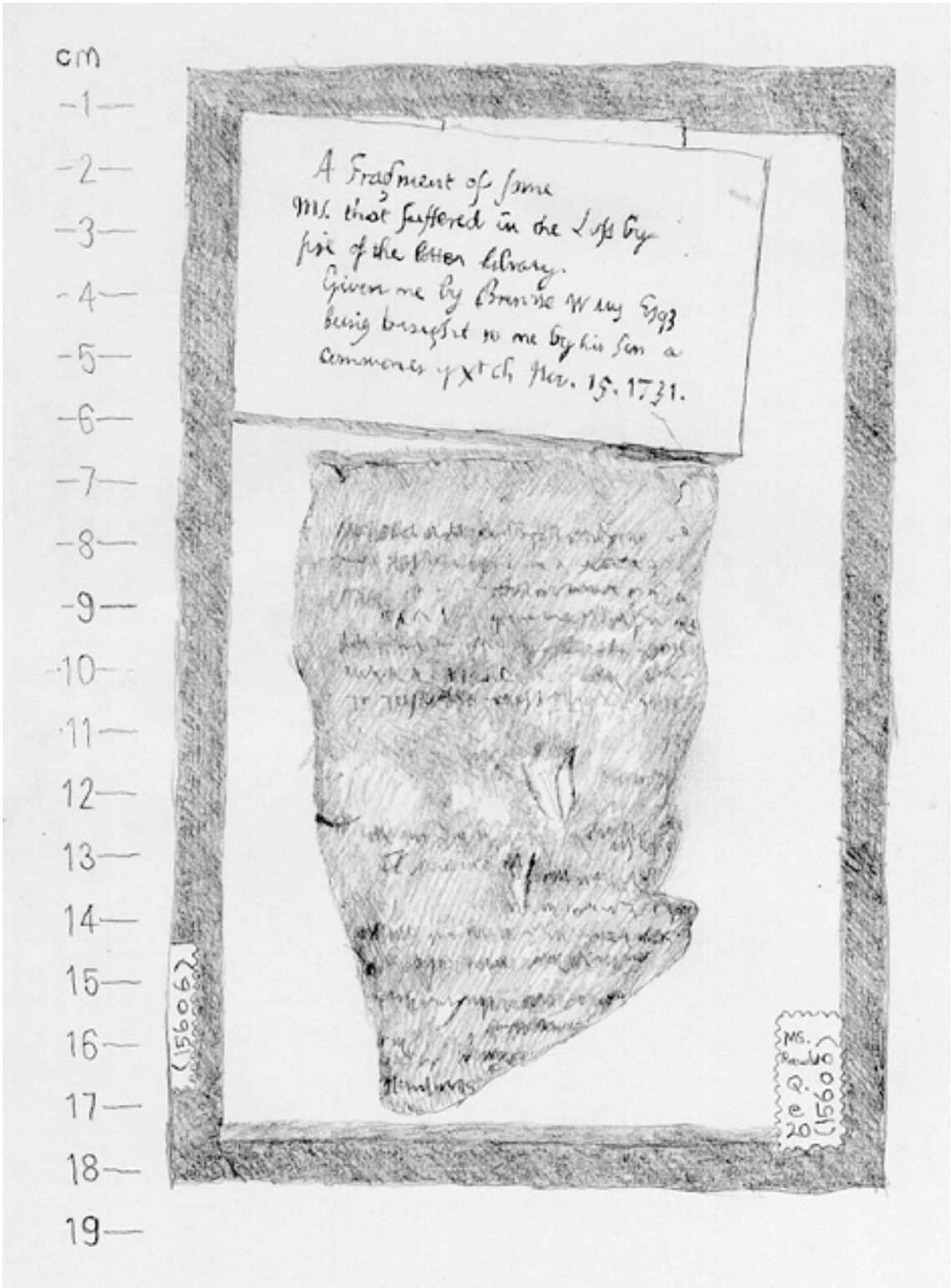


Fig. 5 8(RQe20r) verso

The next surviving leaf is now mounted as two separate folios in Cotton Otho B. x, the widely displaced fragment 50 and its other half, the correctly situated fragment 6, which I have digitally rejoined as one leaf and named 9(50+6)r and 9(6+50)v. The projecting vellum at the bottom of fragment 6 has shrunk inward, which prevents bringing the two fragments more closely together without obscuring some text on the adjacent fragment. It is perhaps significant that the text missing from the Bodleian leaf, which we can restore from Cotton Julius E. vii, has the same L-shaped form as fragment 50. The text on fragment 6, moreover, breaks off in the same area as the Bodleian leaf, around line 22. It may be that these leaves were cut for some unknown reason at the same time, perhaps during efforts to extinguish the fire.

Fig. 6 Merging 9(50+6)r

By strange coincidence, St. Basil himself tore apart a manuscript the devil acquired on this same leaf, fol. 9(6+50)v14-18 (Skeat 379–383). The Faustian story on this reunited folio is about a young man who makes a pact with the devil in order to marry the girl he sinfully loves. The devil dictates to the youth a contract renouncing Christ and his baptism, assuring the devil, he thinks, of the boy's company on Doomsday. After the boy and girl marry, in a part of the Life preserved only in Cotton Julius E. vii, St. Basil fights with the devil on behalf of the repentant young man to retrieve the autograph manuscript. St. Basil is unimpressed with the legality of the document, and as a result of his prayers the contract falls from the ether into his hands. After he confirms with the youth that it is indeed his handwriting, St. Basil promptly tears it up (Skeat 458). Skeat has memorably illustrated how difficult it is to read fol. 9(50+6)r and 9(6+50)v with the two parts so widely separated. He realized that the text came from about the same place in the story, but mistakenly concluded that fragment 50 must be from a different manuscript of the same text as fragment 6: "As noted at p. 70," he says, "one of the leaves in this MS. (leaf 50) does not belong to the MS. at all, so that the collations are here marked with the symbol O2." According to Skeat, "It is easy to see whence the leaf came, viz. from the other much burnt Cotton MS. with similar contents, i.e. from MS. V. (Vitellius D. 17)" (p. xvi) [20].

In fact it is easy to see from the script and layout that Otho B. x and Vitellius D. xvii are quite different manuscripts, and N.R. Ker readily saw that fragments 50 and 6 were two parts of the same leaf from Otho B. x [21]. Skeat's embarrassing blunder nonetheless shows how difficult it is to read the two parts of the same folio when they are separated by 44 folios. Even if they are ever rebound in the correct order, these two parts will most likely remain in their paper frames, because removing them might cause further damage. A digital restoration in a single image, on the other hand, is easily accomplished, once or as many times as desired, without any possible harm to the manuscript.

The fragments are reasonably legible, even in ordinary light, and bringing them together and enhancing the contrast render them the most easily read of all the surviving leaves. A "free transform" procedure in Photoshop makes it possible to join the fragments more closely, by moving to the right the bit of shrunken vellum on fragment 6 that would otherwise cover some text on fragment 50. The same free transform function could bring the two fragments quite close together. It is important to keep in mind, however, that these fragments are not really planar. The digital camera produces a flat, two-dimensional result, which is an accurate facsimile only if the original object is also flat. In fact, each manuscript leaf in Cotton Otho B. x has its own three-dimensional properties, which appear as ambiguous distortions in the two-dimensional digital photograph. These three-dimensional properties in the objects themselves were not intentionally created but were caused by the way they bend when the manuscript is opened, by their different texture caused in part by the action of fire, water, and shrinkage, and by the way each has reacted to the individual paper frames that hold them [22]. Computer scientists Brent Seales and Michael Brown are currently experimenting with three-dimensional modeling to see if the fragments can be more accurately rejoined by taking into account the constantly changing three-dimensional properties of each leaf. In order to record the structure of a three-dimensional object, Seales and Brown have built an inexpensive, portable device from commonly available off-the-shelf hardware, a digital light, or LCD (liquid crystal display), projector. With this device, one can capture millions of three-dimensional sample points which together produce a very fine reconstruction of the shape of the surface of the fragments. These points form the basis for a mesh of triangles that approximates to a very fine degree the shape of the fragment, and onto which the high-resolution texture from the digital photograph is rendered to give an accurate, metric rendition of the exact shape and color of the object. Of course, with complete shape information together with high-resolution digital photography, it is possible to view the mesh as a textured image looking very much like a leaf in the manuscript, and also to view it as a "wire-frame" mesh, the structure of triangles from the points recording the three-dimensional properties of the object.

A two-dimensional rendering does not do justice to the three-dimensional modeling which is best represented in video. While there are many available 3D acquisition technologies, there are a number of design considerations that have led us to the system we have developed for scanning manuscripts.

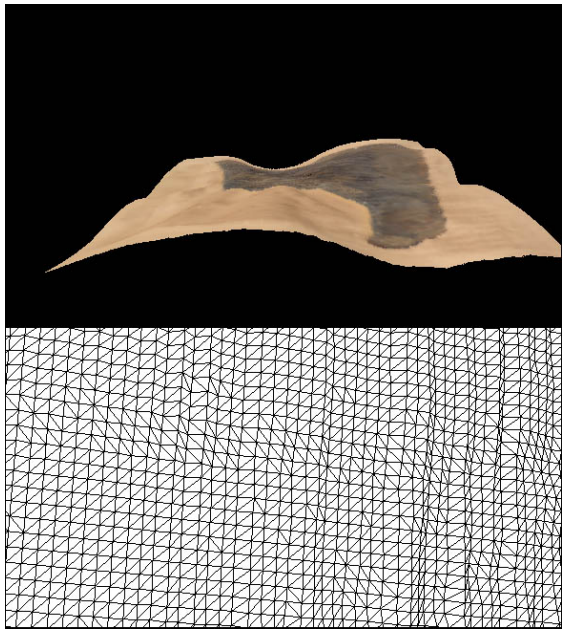


Fig. 8 2D v and

Challenges to address include:

- Specialized acquisition hardware is expensive
- Most existing systems are not well-suited for imaging the same object under a variety lighting conditions
- Highly collimated light-source (laser) may not be desirable for sensitive materials
- Most systems are not easily integrated into current setups

One important design and deployment goal is to be able to acquire 3D shape representations without expensive alterations of the existing hardware setup. Also, it is very important to allow the recovered 3D data to be easily registered with imagery under different lighting conditions, since this is a desirable restoration practice.

Furthermore, not all artifacts require 3D scanning. If the curator or librarian in charge of digitization feels that the object is flat enough and does not warrant 3D imaging, then 2D imaging alone can be performed. Considering these goals, a laser scanner or similar devices were not appropriate. Instead, we opted to use structured light techniques based on an off-the-shelf light projector as a controllable light-emitting device.

By using a light projector, we can directly convert an existing 2D acquisition system into a 3D acquisition system. The projector is introduced into the digitization setup and is used to illuminate points on the surface of an object. The projector turns on its pixel P(x, y) and illuminates a 3D point, M, on the surface of an object. The camera observes this illuminated point as a bright spot in its image at coordinate C(u, v). When the exact geometries of the camera and projector are known, this device-to-device correspondence can be used to reconstruct the illuminated 3D point, M. We recover the required geometries during a simple calibration step, completed before the digital scan commences. By repeating the projection and detection steps for each projector pixel P(x, y), a dense set of 3D points can be recovered on the object's surface. This dense set of 3D points is used to create an accurate 3D model of the scanned object.

The set of points recovered in the scan must be converted to a structured representation so that the digital image, or texture, can be mapped and displayed coherently. This conversion is done by connecting the individual points together into a triangulated mesh. The mesh captures the space in between each sample point and provides a large set of 3D "faces" onto which the digit image can be rendered. Other conversions and internal representations, such as height maps, are also used to facilitate a rapid and interactive display for the end-user. The result of the scan is a coherent

representation of the surface of the manuscript. We have estimated the accuracy of the scanning process to be on the order of 0.5 millimeters variation in depth. At that level of accuracy it is possible to capture very small variations in the surface of the vellum, and to make accurate measurements of the size and volume of the various features on the surface of the manuscript.

This and related research into three-dimensional imaging promises to aid in the reconstruction of damaged manuscripts, in this case with a seamless rejoining of fragments 6 and 50. A process called "mosaicing" can stitch together digital images from different regions, even when the objects are digitized separately with potentially varying scales, to form a seamless global portrait. The same process of mosaicing will allow us to fuse images acquired at different times under different lighting conditions (for example, with bright light, ultraviolet, and fiber-optic backlighting) to achieve a more complete and legible result. Because the manuscript fragments are often quite distorted by heat, water, bending, or other factors, the computer scientists will adapt a process called "image warping," or digital stretching, to return the three-dimensional shape of the manuscript page to a planar object. They will, in other words, use the geometry of the three-dimensional image as the basis of a "post-warping" to undo the damage that caused the warping. These processes are of course ideal occasions for close collaboration between the computer scientists and the humanities scholar, because the latter must decipher the text in the damaged regions and assess the accuracy of the digital restoration.

In addition to developing methodologies for restoration, a chief aim of the project is to furnish a toolkit for accessing, searching, and editing damaged manuscripts. The Digital Atheneum team is now working on developing a system with innovative capabilities and functionality along with interfaces to access this system. Some major capabilities and interfaces we are developing include:

- three-dimensional Manipulation to allow the editor to control modeled images of manuscript pages in meaningful ways
- a Restoration function to apply image processing techniques to enhance and restore damaged manuscript pages
- Remote Access for network access to the database of images, text, and glossary entries (with full support for all the above features).
- a Tag-linking capability to permit the editor to coordinate regions of interest in image and text for analysis, annotation, or editorial reconstruction
- a Search capability to provide options for searching glossaries, texts, and images, or a combination of these components of an electronic edition
- an Import-Export capability designed to import data into the database for storage, search, and retrieval; and to export the same data from the database in multiple-option formats

Access to a distributed database and continual integration of new data and algorithms provide the backbone of the project. This research into new techniques for accessing the humanities collections is led by James Griffioen and his students in Computer Science.

At the heart of the system is a multimedia database capable of storing a wide range of data formats, including images of damaged manuscripts. The key to the database is its ability to store, search, and modify metadata, auxiliary information entered along with the data or extracted from the data. Because the amount of metadata can exceed the amount of original data by several orders of magnitude and is often as

important, if not more important, than the original data, being able to store and search the metadata efficiently and quickly is a primary goal. Another key to our work is the ability to process images to identify image objects (e.g., structures or letterforms) and record this extracted information as metadata for future content-based searches. Because the metadata can be searched using fast select and join operations, searching large amounts of metadata (image/manuscript content) can be done very quickly.

Several related digital library projects have taken a web-oriented approach for their storage needs. These systems are typically based on some restricted flavor of the SGML markup language; variations of HTML and XML being some of the favorites. One example used by scholars who work with textual materials is the Text Encoding Initiative (TEI). TEI is tailored to the markup of text-based documents via the use of hierarchically structured "tags". Clearly, TEI is not well-suited for non-textual data such as images, but more importantly, the storage format enforces a structure on the tagging mechanism that is often inappropriate or overly restrictive (e.g., the ability to support overlapping tags). Although our database is capable of storing metadata, like tags in TEI, it does not impose any of the unwanted restrictions of the SGML markup syntax TEI is built on. The primary task of the database is to store massive amounts of data and metadata and search it quickly. Decoupling the storage system from the way in which data is entered or retrieved allows the storage system to focus on efficient storage and fast retrieval while all metadata parsing and syntax restrictions, if any, can be enforced by a domain or content specific interface. If an SGML/XML based interface is desired and the syntax restrictions are enforced by the interface, converting between such standards to our database format are straightforward, which means the database storage system can be used at the heart of any digital library, including those based on web-based standards.

Another key feature of our database is that the system can be accessed by any number of interface programs run on any computer in the network. This means that the data storage system is shared by all applications that need access to the data and can be accessed by any machine anywhere in the world. Moreover, the data can be shared between users in different locations, facilitating collaboration between editors. The database is built using the Java Database interface (JDBC) and Java's Remote Method Invocation (RMI). Applications that wish to access the data stored in the DB simply issues queries on java objects to obtain result sets containing the desired information. Because JDBC is based on SQL, powerful searches can be specified using standard SQL queries to any JDBC-compliant SQL database.

Another feature and advantage of our database model is that it is able to record correlations between different stored data elements, even when the data have dissimilar formats. All data, as well as the correspondences between the data are stored in a single database and can be quickly retrieved and displayed by an application. With conventional approaches, one transcript of a manuscript might be stored in one file, the edited version in another file, and the associated images in other files - with no way to link or search the various pieces of the digital collection. By storing all the data, along with correspondence identifiers in the database, applications can easily establish connections between distinct data elements and quickly search for and retrieve related pieces of information.

The third aspect of our project revolves around the editing of these collections. We are now developing interfaces, tailored to the requirements of humanities editors, to

facilitate primary editorial tasks, critical to the editorial process of assembling and annotating the data:

- an Edition Production tool to document and coordinate all the editorial functions required to compile, edit, proofread, organize, and display all the components of a complex electronic edition
- a Content-tagging tool to assist the editor in structuring a transcript or edition and providing metadata (tags) to enable complex searches
- a Sorting tool to generate indices, wordlists, and tables of contents, and to organize and arrange individual texts and groups of texts (identifying and indexing manuscript folios and quires, for example)
- a Glossary tool, using customized templates for all parts of speech, to help the editor compile comprehensive glossaries from wordlists generated by the Sorting tool

In one of his miracles lost to both Otho B. x and Vitellius D. xvii, Saint Basil helps one of his disciples who wants to learn Greek the easy way, without studying it (ll. 512–523). Saint Basil tells him the request is beyond his powers, but after praying together the student is in fact able to speak Greek without studying it. One of the aims of this project is to provide comprehensive glossaries that will help students translate Old English without too much study. The glossaries will be displayed in HTML for convenient reading and browsing, but they will also be searchable in a myriad of user-defined ways through the database. For editors we are providing a facility for generating a wordlist from a properly prepared transcript or edition that can include folio and edition line numbers, with accompanying templates for editing the wordlist into a completely tagged comprehensive glossary.

In this example the left frame is an alphabetical list of all words occurring in the text (note that the unstressed prefix *ge-* is ignored in the alphabetization). The tool that sorts the list of words also gives all line numbers, but these are not displayed unless the button beside the word is clicked (in this case *cyldhade* was clicked and so displays the two line numbers where the word occurs). The highlighted number contains both the folio-line and the edition-line, because a damaged leaf survives; if a listing does not contain a folio-line, the manuscript is lost for this section, and we must depend on Cotton Julius E. vii to restore it.

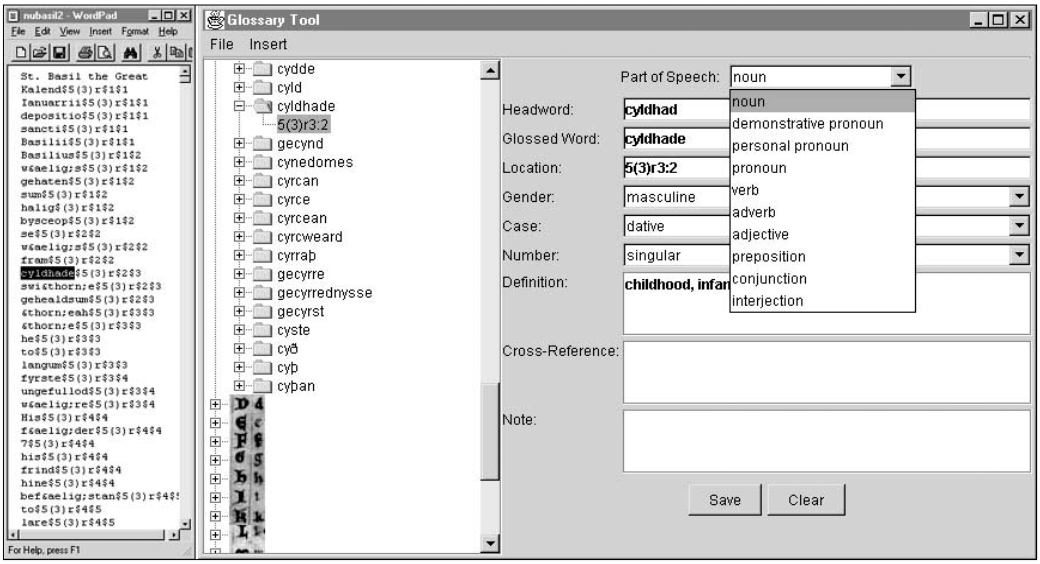
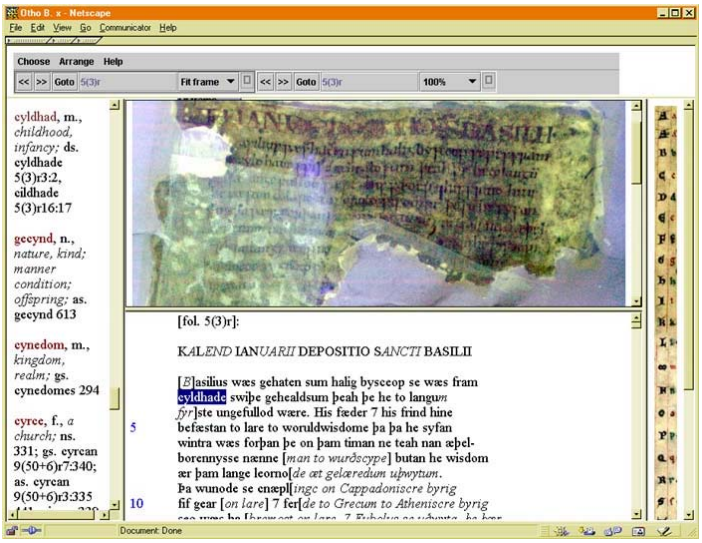


Fig. 9 Glossary Tool template

After filling in the remaining information, the user clicks Save and the entry, properly tagged, is saved for storage in the database. All the entries, collated by headword, are simultaneously tagged for HTML display (in the expectation of moving to XML) as shown in the left frame below.



The costs and difficulties of inaugurating a digital library of complex electronic editions have to be balanced with the benefits of widely disseminating relatively inaccessible cultural treasures while developing new technological capabilities. It has to be recognized, too, that the collaboration of computer scientists and humanities scholars is a highly unusual situation in academia today, and it would require profound changes in university structures for this kind of interdisciplinary work to occur on a normal basis. None the less, most research libraries, if motivated, can absorb some significant costs of such a project by using equipment already available and by reassigning staff to digitize documents. All institutions should begin nurturing and honoring programming skills and building a programming culture across the disciplines, particularly for the humanities, where the cultural resources are rich but the programming facilities are typically poor or non-existent. Creating a digital archive is not difficult, however. Capturing images with a digital camera is much easier and faster than traditional photography, the results are immediately apparent, and the saved images may be used and reused (and sold and resold) without further cost if properly archived and backed up. There is another valuable resource that is usually free, the expertise of research scholars in the humanities. As Andrew Prescott has argued, the collaboration of librarians and curators with research scholars with intimate knowledge of the collections, who are all dedicated to making the resources more widely available, is one of the most effective ways for defining a project and starting a successful digital library. As for the Digital Atheneum, we intend that at least one result will be to save subsequent projects some of the costs and difficulties of developing new techniques for restoring, searching, and editing humanities collections [23].

References

* The first version of this article was given by Kiernan as "Creating Electronic Editions from Medieval Manuscripts" for the International Foundation of St. Cyril and St. Methodius in Sofia, Bulgaria, on 1 August 2000, at the invitation of the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. It was first published as "The Reappearances of St. Basil the Great in British Library MS Cotton Otho B. x," *Computers and the Humanities* 36:1 (February 2002), 7–26, and is reprinted with revisions by permission of the publisher.

[1] Supported by the National Science Foundation's Digital Library program, IBM's Shared University Research grant, the British Library, and the University of Kentucky, the project combines the expertise of computer scientists and humanities scholars in an effort to make newly

accessible some of the most badly damaged manuscripts from the Cottonian collection in the British Library. The principal investigators are Kevin Kiernan, Brent Seales, and James Griffioen, with the assistance of Linda Cantara, C.J. Yuan, Katherine Wenger, Michael Brown, Michael Rogers, Demorah Hayes, Kenneth Hawley, and Ashwin Gokhale at the University of Kentucky, and David French at the British Library.

[2] The codex is in such desperately poor condition that the British Library has taken it out of circulation and the conservation laboratory is reviewing various extreme methods to halt its progressive deterioration. One problem is that the fragments were unfortunately framed in acidic paper in the nineteenth century, and the paper frames are both disintegrating around the manuscript leaves and staining them. The brittle vellum leaves have themselves crumbled in places, and previous conservators have sometimes used ill-advised methods to hold them together. One such method was gluing pieces of gauze over brittle vellum, which rendered these passages unreadable as the glue aged and became opaque.

[3] For the standard introduction to St. Basil's life and works, see Paul J. Fedwick, "A Chronology of the Life and Works of Basil of Caesarea," *Basil of Caesarea: Christian, Humanist, Ascetic*, vol. 1, ed. Paul J. Fedwick (Toronto: Pontifical Institute of Mediaeval Studies, 1981), 3–19.

[4] Max Förster, *Zur Geschichte des Reliquienkultus in Altengland, Sitzungsberichte der Bayerischen Akademie der Wissenschaften*, Phil.-Hist. Abt., Jahrgang 1943, 8 (Munich), 63–80. See also Patrick Conner, *Anglo-Saxon Exeter: A Tenth-Century Cultural History* (Woodbridge, Suffolk, UK; Rochester, NY: Boydell Press, 1993).

[5] Francis Wormald, ed., *English Kalendars Before A.D. 1100*, Henry Bradshaw Society 72 (London 1934). According to Michael Lapidge, "St. Basil is commemorated in a large number of Anglo-Saxon calendars - seven - but always on June 14. Four of the calendars in question are from Winchester. In commemorating St. Basil on January 1, therefore, Ælfric was not following Winchester use" (p. 123), "Ælfric's Sanctorale" 115-129, in Paul Szarmach, ed., *Holy Men and Holy Women: Old English Prose Saints' Lives and Their Contexts* (Albany: SUNY Press, 1996); but as Conner points out (p. 28, note 36), two of these calendars also give the date as January 1, one of which comes from New Minster, Winchester (Cambridge, Trinity College MS R.15.32, p. 15).

[6] The manuscripts, with the exception of one alienated leaf, are all in the Cotton collection of the British Library: Cotton Julius E. vii, Cotton Vitellius D. xvii, and Cotton Otho B. x. A fragment of a Latin *vita* paleographically dated in the early tenth century was found in a binding in Exeter; see Conner, 28–29.

[7] *Ælfric's Lives of Saints ... edited from British Museum Cott. MS. Julius E. vii with variants from other manuscripts*, ed. Walter W. Skeat, Vol. 1, EETS OS 76 & 82 (London, 1881). The contents and organization of Cotton Otho B. x is significantly different from the manuscript Skeat uses for this edition. The other manuscript is Cotton Vitellius D. xvii.

[8] *The Complete Corpus of Old English in Electronic Form*, ed. Antonette di Paolo Healey, with Richard Venezky and Peter Mielke (Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto, January 2000).

[9] See Andrew Prescott, "'Their Present Miserable State of Cremation': the Restoration of the Cotton Library," *Sir Robert Cotton as Collector: Essays on an Early Stuart Courtier and His Legacy*, ed. C. J. Wright (London: British Library Publications, 1997), 391–454.

[10] Not all of the missing Cotton manuscripts were destroyed in the fire. Prescott has recently informed us that the wife of David Casley, deputy librarian of both the Royal and Cotton libraries, on at least one occasion gave a visitor to the library "two bundles of MSS" from the damaged Cottonian collection! (Prescott, p. 442, note 51). The Bodleian leaf allows one to imagine the dismal scene outside Ashburnham house in the days following the disastrous fire.

[11] Given this situation, perhaps a more virtually real e-foliation would leave the Bodleian, Oxford, leaf out of the numbering and name BL folio 50 "fol. 8(50)" and BL folio 6 "fol. 9(6)." The problems with this solution are that a reader will not know where the Bodleian leaf belongs, while at the British Library the sequence 8(50)+9(6) for the recto would be reversed on the verso as 9(6)+8(50).

[12] Skeat was even suspicious of textual corruption. Observing that "there is an abrupt transition here," but pointing to line 633 (where the story of the sinful woman suddenly resumes), he concludes that "nothing is lost" (p. 83, note 1).

[13] The hermit's ineffectiveness probably reflects Ælfric's attitude that active, socially engaged monks were more admirable than solitary, contemplative hermits. See Mary Clayton, "Hermits and the Contemplative Life in Anglo-Saxon England," in Szarmach, 147–175.

[14] Although Humphrey Wanley before the fire correctly transcribed *gehaten* (p. 191), Skeat says, "I read it *haten*, as noted on p. 50" (p. 545, note 1); and N.R. Ker in his *Catalogue of Manuscripts Containing Anglo-Saxon* (Oxford: Oxford University Press, 1957; reissued 1990) agrees with Skeat: "Beg. 'Basilius wæs haten'" (p. 225).

[15] "Of this homily there are two other copies, viz. in MSS. O. and V., both of which are much burnt. I give such variations as I could decipher" (p. 545).

[16] See 8(RQe20r) v11-12:249.

[17] <http://image.ox.ac.uk/show?collection=bodleian&manuscript=msrawlqe20>

[18] See Prescott, pp. 410-415.

[19] An apparently successful method of restoring the suppleness of vellum is discussed in I. K. Belaya, "Softening and Restoration of Parchment in Manuscripts and Bookbindings," and "Instructions for the Softening of Parchment Manuscripts and Bookbindings," *Restaurateur* 1.1 (1969), 20–48 and 49-51.

[20] Vitellius D. xvii is also rendered more legible with magnification, UV, and image processing; the script makes it immediately evident, however, that the leaf from Otho B. x does not belong to it.

[21] The spine has "Homilies for Saints' Days, Brit. Mus., Cotton Ms. Vitellius D XVII;" see Wanley in George Hickes, *Linguarum Vett.* (Menston: Scolar Press, 1970), vol. 2, p. 206, for its description before the fire. The manuscript leaves are now much tinier than surviving fragments of Otho B. x, but they were more skillfully or carefully inlaid in non-acidic paper; even the unfortunate gauze reinforcement was well done, for the glue does not obscure the writing beneath, as it often does with Otho B. x. Many of the leaves of Vitellius D. xvii are out of order and reversed, however, suggesting that Madden had difficulty reading it too. St. Basil is on fols. 79v-83r21.

[22] A vivid example even in 2D is the image of fol. 5(3) r above, which clearly shows how the paper frame has buckled in reaction to the shifting shape of the supple vellum it is supposed to hold in place.

[23] For recent projects advancing the goals discussed in this article, see the website for our collaboratory for Research in Computing for Humanities at <http://www.rch.uky.edu/>.

Kevin Kiernan
University of Kentucky, Department of English, Lexington, KY 40506
mailto:kiernan@uky.edu
Brent Seales
University of Kentucky, Department of Computer Science, Lexington, KY 40506
mailto:seales@netlab.uky.edu
James Griffioen
University of Kentucky, Department of Computer Science, Lexington, KY 40506
mailto:griff@netlab.uky.edu