

Reading the Invisible Library: A Retrospective

W. Brent Seales*

*Professor and Chair
Computer Science Department
University of Kentucky

Abstract

On July 20, 2015 the Israel Antiquities Authority (IAA) made a dramatic announcement about an indistinct, badly damaged scroll from their collection. Without being opened, the severely burned and crumbling ancient document had been read. The IAA released to the world an image from inside the scroll of text that had been rendered purely from non-invasive imaging using a set of software algorithms. The words shown were from Leviticus, making the damaged fragment the oldest extant copy of the Hebrew Bible outside the Dead Sea Scrolls themselves. This paper describes a somewhat personal narrative of that discovery: the progression of the research team's technical work over two decades that led to the result; a review of the significance of the discovery itself; and the prospects for future work using non-invasive imaging and software analysis to recover text from obscure and badly damaged artifacts.

Introduction

The digital library was born in the 1990's, with several key technologies inspiring a new approach to manuscripts and library collections. The Internet, digital imaging, and at-scale systems for taking in information created a vision of a different kind of library--A Digital Renaissance [1]. This global movement to digitize and thereby preserve and make accessible the world's bevy of information continues today, with millions of books having been ingested [2] and hundreds of thousands of manuscripts now available digitally for scholarship.

Information, once made digital, can be organized, indexed, and disseminated on an enormous scale using computer networks and large digital repositories. But from the beginning, the problem of *damage* has presented a roadblock to progress. Damage is the natural result of aging. Manuscripts and books--the "technologies" used to record human experience for the past three thousand years--are not exempt from the ravages of time (Fig. 1). So even after crossing the chasm from analog to digital, that leap of digitization was inhibited, or even defeated, by the profound damage found in manuscripts.

The work we describe in this paper deals with damaged written material: not creating damage, but defeating it. Rewinding it. Removing it. Overcoming it. It describes a twenty-year effort that led to a software system and a vision for recovering information in the “invisible library” once thought lost forever. We give a brief review of the early work that motivated our line of thinking and finally led to a framework for digitally unwrapping materials that cannot be opened physically. We show the significance of the restoration of the text from within the scroll from En-Gedi [3]. Finally, we conclude with future prospects and what we hope will be an inspiration for others to continue using non-invasive techniques in a quest to reveal the invisible library.



FIGURE 1: Examples of damaged materials. Left: Unrestored damaged manuscript from the Cotton collection (British Library). Right: Egyptian papyrus scroll (British Museum).

Early Work

Early attempts at creating digital libraries focused solely on document imaging. 2D digital photos of pages were taken and then assembled as an electronic version of a book or document. But in just five years, our research team would move beyond merely capturing an electronic image of a document to actually restoring it, thereby creating versions of texts not even in existence.

The Electronic Beowulf

Our first project was the the *Beowulf* text, which had been imaged at the British Library in London. The manuscript was part of The Cotton Collection, the “most important collection of manuscripts ever assembled in Britain by a private individual,” according to the British Library. Sir Robert Bruce Cotton (1571-1631), along with his son and grandson, gathered during their lifetimes thousands of manuscripts, charters, rolls, and seals dating from approximately the 4th century to the 1600s [4]. Unfortunately, a fire in 1731 destroyed some of the documents and damaged many others. Among those that survived was *Beowulf*. But its pages were charred from the fire and cockled from the water used to save it. The damage and subsequent repair left each parchment page wrinkled, distorted, and impossible to support in its restored paper frames without obscuring and hiding important parts of the text.

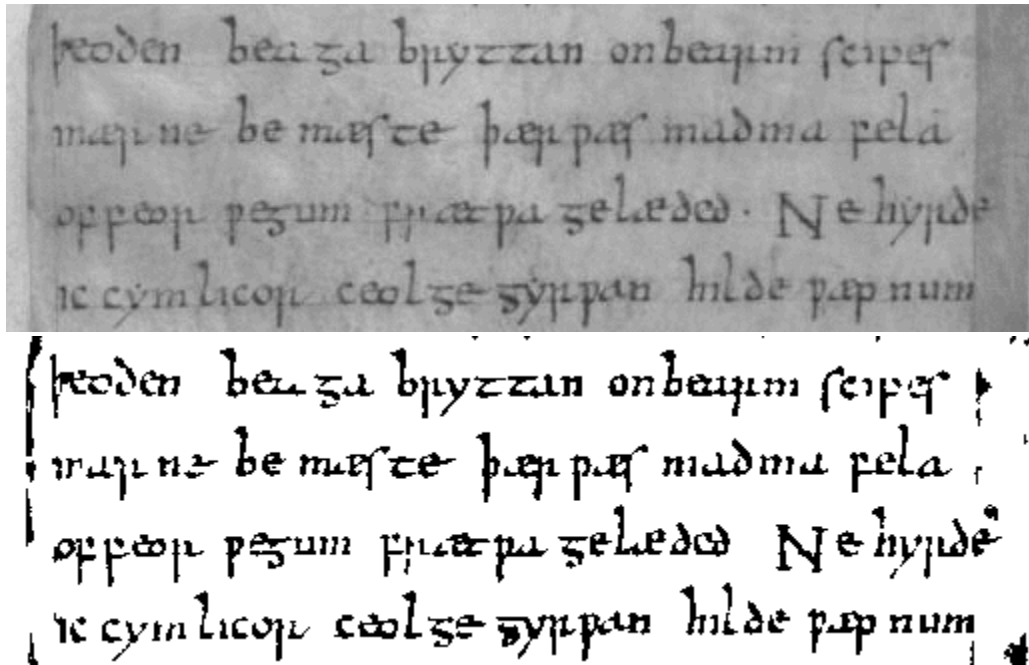


FIGURE 2: The digital edition required the processing of manuscript images in order to link and then visualize images, transcriptions, and translations. Shown is *Beowulf* folio 129 verso, demonstrating the first step in processing to find text and link text to transcript.

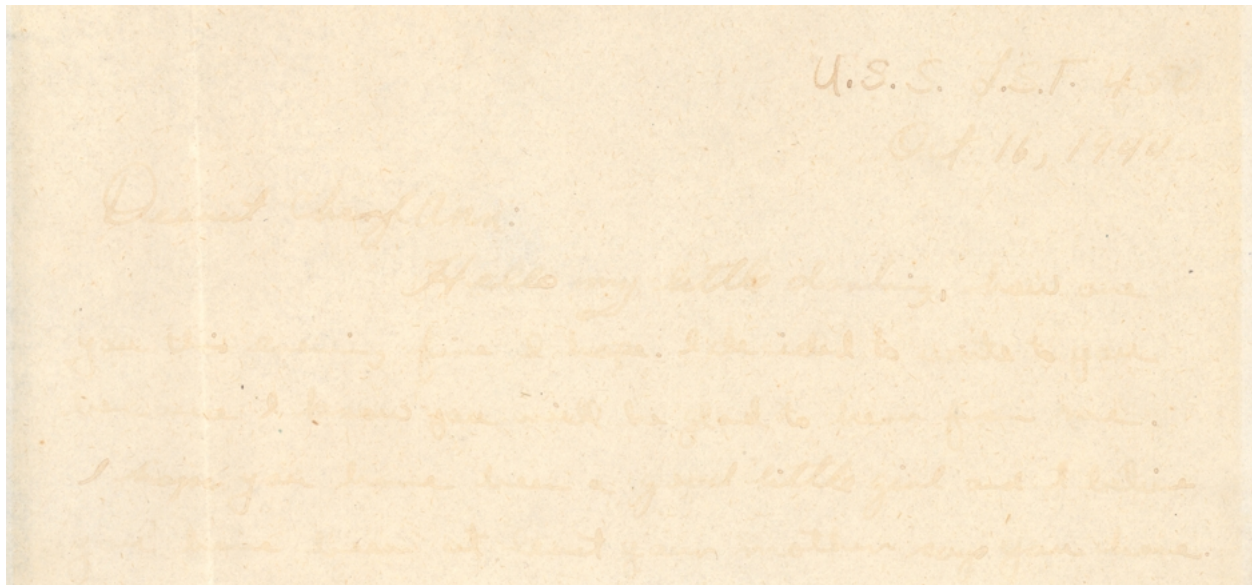
In 1995, we built software to organize the 2D photographs of these damaged pages (Fig. 2). At the very front end of the digital library movement, our 2D “digital edition” of *Beowulf* [5] was pioneering. It was, after all, one of the firsts. But given its implementation on the computer hardware of the time, it was also severely limited, and the results left us wanting. Yes, we had successfully created an electronic version of the manuscript, but we immediately realized that flat, 2D images were insufficient for truly creating exact digital replicas of texts. Ancient documents are not flat. Damage has occurred over time, and at a minimum there is wrinkling and buckling of pages. Often, as was the case with *Beowulf*, there is further destruction due to fire, water, or invasive restoration efforts. Therefore, if we wanted to create an exact facsimile of an aged manuscript, we needed a way to capture the undulations of warped pages. In other words, we needed 3D.

Another dissatisfaction with the *Beowulf* digital edition was that it did not sufficiently explicate the distortions resulting from such damage. Buckling and wrinkling change a manuscript. Words become hidden within wrinkles or smeared over undulations, for example. Although straightforward photography will show these occurrences, it cannot reveal their causes. In other words, with basic photography, it is impossible to discern what elements apparent in the digitized manuscript are original and which ones have been introduced by damage over time.

The Seacrist Letter

Shortly after the work on *Beowulf*, a woman contacted me out of the blue. She had a letter her father had written to her during World War II that had become unreadable over the years, with the words fading to the point of invisibility. She wanted to know if we could do anything to fix it. Using digital image processing to enhance the text on the page and then applying various algorithms to separate and clean it, we were able to restore the faded writing (Fig. 3). The daughter could once again read her father's endearing closing words: "Oceans of Love, your Daddy, Frank."

To be able to recover a father's loving words to his daughter, words that he sent from an ocean away without knowing if he would ever see her again, felt wonderful. I became hooked on the idea of harnessing technology to rewind textual damage in all its forms. But, we needed a set of techniques for dealing with the ravages of time. In 1999, such tools did not exist. So we set out to build them.



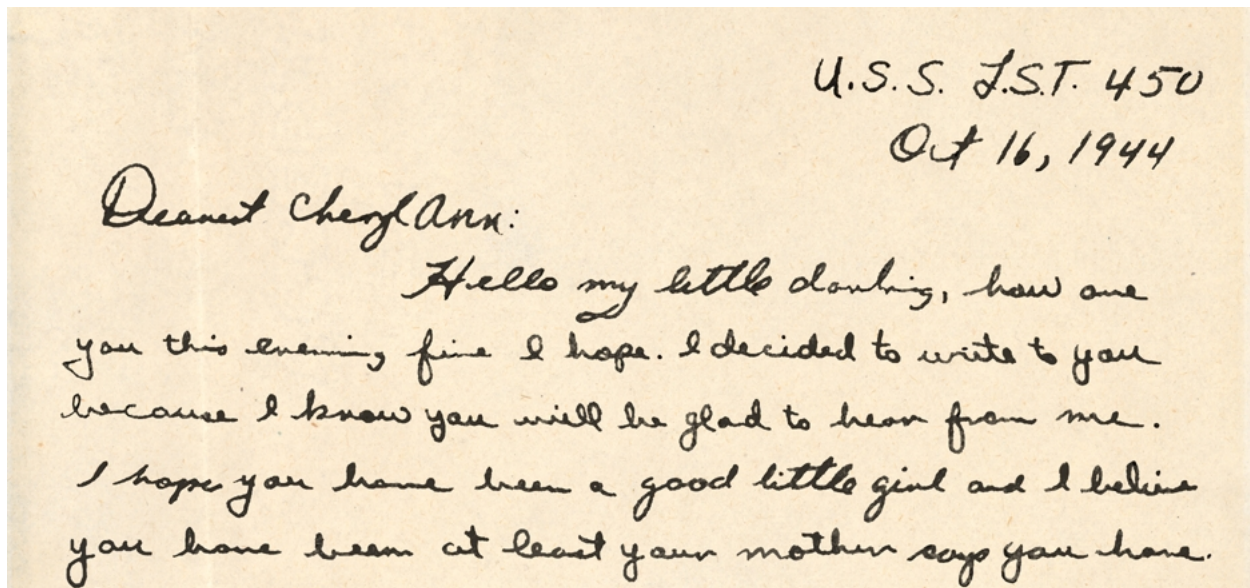


FIGURE 3: Seacrist letter and its restoration.

Otho B.x from The Cotton Collection: 3D imaging and Flattening

Our quest took us back to the beautiful but damaged manuscripts of the Cotton Collection, specifically the Otho B.x [11, 12, 13]. Our goal was to develop digital processes for flattening the warps and wrinkles in documents. The system we created for this also enabled, for the first time, the acquisition of accurate 3D shape information from documents.

The demands of modern conservation have rendered unthinkable the idea of using a physical process to flatten manuscripts. Full-contact restoration efforts have, over the years, inflicted so much additional damage that today's conservators have forbidden almost all human contact with these friable artifacts. Heightened protection measures therefore necessitated the creation of a non-contact system for acquiring a digital three-dimensional model of the page.

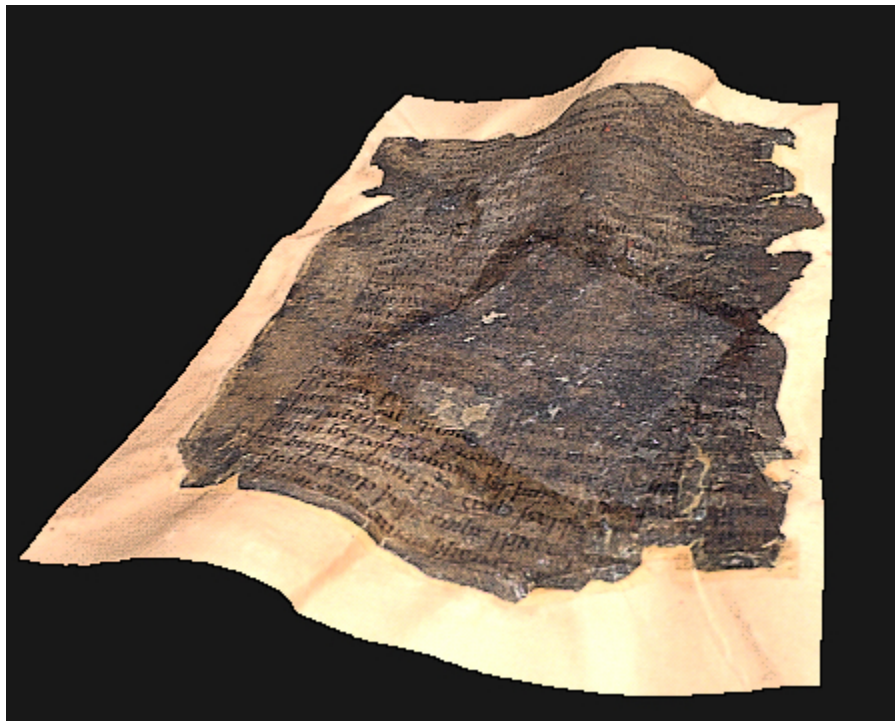
Using structured light with calibrated camera and projector systems, we built a laboratory prototype that allowed us to acquire both the 3D shape of a manuscript page and its high-resolution photograph, which were aligned to each other. The system relied on visible light and required no physical contact with the fragile manuscript. Although tedious, it was completely non-invasive and harmless.

We went back to the British Library in 2000 and used this system to image the Otho B.x. The shape information we acquired gave us terrifically realistic 3D renderings of the pages of the manuscript. The wrinkles were visible as an exact shape model and were viewable in 3D from all angles. The system's ability to simulate any viewpoint and light source direction allowed its interactive viewer to create interesting effects. One such outcome was a low angle of light relative to the surface shape, also known as a raking light effect, which emphasized the surface

detail. The combination of all these effects created a never-before-experienced immersive interaction with the digital pages. A patron could now call up 3D, metrically-exact models of the Ottho B.X's pages and engage with them in a way that proved extremely compelling.

While we were interested in this innovative display environment, high-tech document visualization was not our primary goal. Our main purpose for capturing 3D structures was to test our algorithm for restoring a wrinkled page to its original flattened state. We believed that if we could electronically simulate the laws of physics in our software--for example, program it to mimic the pull of gravity or the force of a heated iron--we could digitally process a wrinkled 3D page back to its unwrinkled state.

We had tested our algorithm in house by hand-crumpling a page of text, scanning it, applying our software to the data, and then comparing our result to the physical, unwrinkled page. Our algorithm proved quite accurate. While it did not address such damage as holes, tears, shrinking, or stretching, the algorithm was extremely effective at digitally smoothing the buckled page. By subsequently applying this algorithm to the entire set of Ottho B.X digital 3D images, we created a completely new edition containing only flat pages, a version of the manuscript that does not exist in reality. [6]



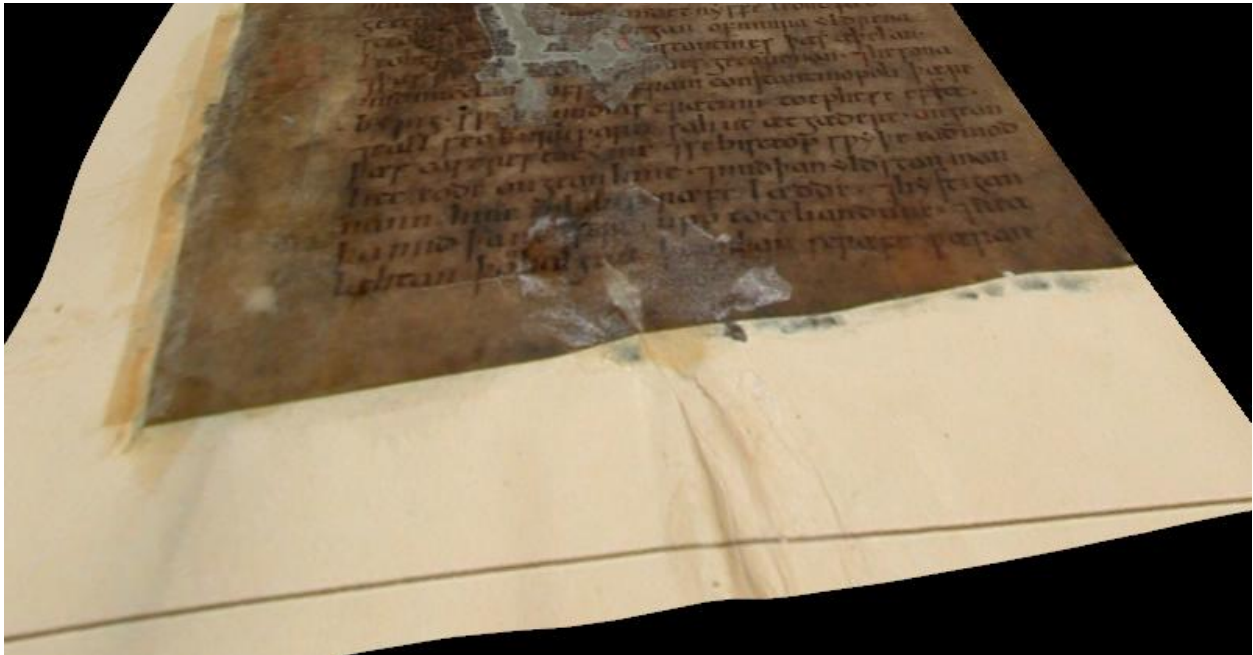


FIGURE 4: Digital flattening of the Otho B.x. First image reveals the 3D shape capture, followed by pre-flattening and then post flattening (Otho B.X 18 verso).

Scroll Prototypes: From Wrinkles to Unwrapping

Prior to this development, we had thought of manuscript restoration only in terms of taking 2D photographs of pages and trying to clean them up or enhance them, as we did with the Seacrist letter. But the power we saw in our flattening process led us to believe that we had not yet

reached the limits of digital surface restoration. Perhaps most, if not all, physical processes like wrinkling, shrinking, and folding could be digitally undone given the right starting point and the correct physics-based algorithms.

Our thoughts turned to the most badly damaged items we could imagine: the Herculaneum papyri, a set of approximately 1800 ancient scrolls buried in ash after the eruption of Mount Vesuvius in A.D. 79. It seemed unimaginable in 2000 that a rolled, carbonized, brittle papyrus manuscript could in any way be restored to a readable state. Physical efforts had failed miserably from the very first attempt in the 18th century, and yet such methods were still being employed as late as the 1980s. The result was absolutely fragmentary wreckage [7]. So the development of a completely non-invasive and digitally-based approach that required no hands-on manipulation became a sort of grand challenge for us.

It turned out that by the end of 2001, we had the technical machinery to achieve this sort of “virtual unwrapping.” We just didn’t realize it yet. Our mechanism for flattening a surface to remove a wrinkle was only one-half a dimension away from what we needed to perform a complete 3D unwrapping. And that one-half dimension was to be found at the very beginning of the process, at the image capture stage. As noted earlier, our standard method for digitizing damaged objects was visible photography. But when an object is completely closed and contains surfaces that can never be exposed to a camera lens, a different approach is required. The answer, we learned, was tomography.

We made this leap in thought thanks to other work that had introduced us to the power of surgical technology, specifically in minimally invasive foregut surgery. The first pre-operative step in these procedures was to image the patient with a tomographic device. This technology, known as Computed Tomography (CT) and for which Cormack and Hounsfield jointly received a Nobel Prize in 1979, was being used abundantly in the medical field at the time.

With CT, the entire inner structure of the human anatomy is exposed in a volumetric way through the use of x-rays. A CT acquisition is fully 3D and non-invasive, but also metric--the structures can be accurately measured in global units--unlike the other imaging methods we had been using for antiquities. We soon recognized the potential power of CT for our purposes, and set to work to test its possibilities.

For our first experiment, we created a scroll prototype by applying oil paint to art canvas, rolling it up, and imaging it using a commercially available medical-grade scanner (Fig. 5). It was a complete success. The CT images captured everything about the canvas and paint. We used our software to manipulate the image so that we could single out the part of the data that was of interest to us and mapped it onto a surface structure of a page. Not only were able to see the ink, but we also unrolled the surface virtually to reveal all of the writing hidden within the folds of the canvas.

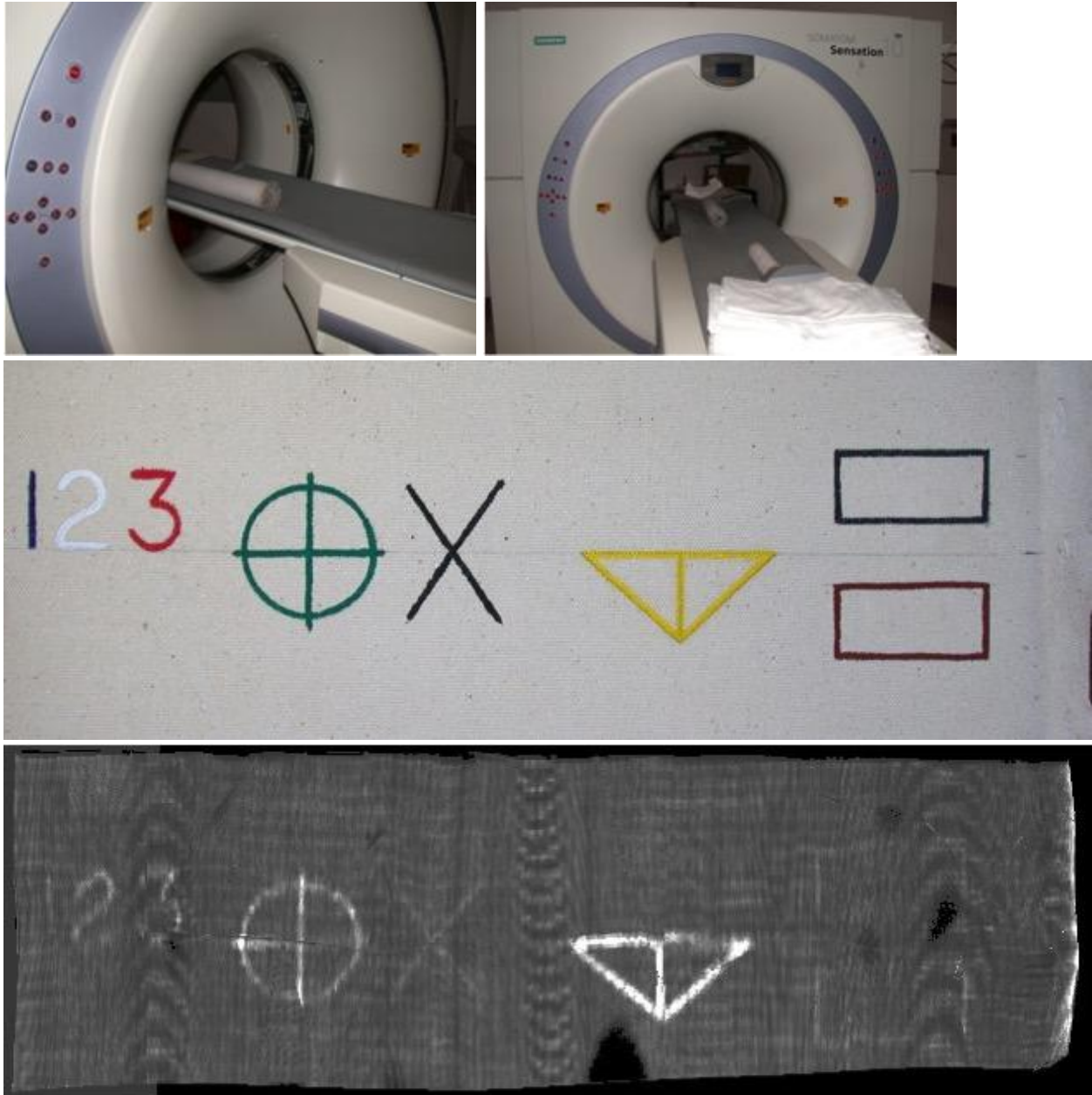


FIGURE 5: *Our first experiment using the medical scanner, circa 2001. The CT machine with our prototype inside; the reference photo of the physical document; the virtually unwrapped digital image.*

This first experiment, which was completed in 2001, led to the construction of a complete set of proxies that we scanned and analyzed [8]. Using this data, we built the step-by-step pipeline of algorithms necessary to convert scan data into a complete photographic image that reveals writing. We felt certain that the steps of our pipeline were correct and generally applicable; and the obvious proof we had in front of us--the ability to read text from real objects that we had constructed, wrapped up, and scanned--propelled us excitedly forward.

Ecclesiastes Book Binding: Complete Unwrapping

Our successes with these proxies gave us the justification we needed to request funding and access to real collections. The first authentic object to be pushed through the pipeline and completely unwrapped was a fragment from within a bookbinding the University of Michigan library [9, 10]. This binding contained approximately seven layers, all stuck together. The top layer was clearly visible and displayed Hebrew writing from the book of Ecclesiastes. We therefore postulated that the hidden layers of the binding might also contain text.

After scanning the binding and applying our algorithms to the data, the analysis revealed writing on the reverse side of the top layer. It appeared to be a continuation of the Hebrew Ecclesiastes text written in the same hand. We subsequently removed the top layer from the bookbinding by hand and could see that the physical item perfectly matched our digital version (Fig. 6). This small fragment thus confirmed our technical approach and opened up the possibility of real access to larger collections of damaged material [20].





FIGURE 6: *Original text shown on Ecclesiastes book binding, followed by the digitally rendered text.*

The Restoration Pipeline

Within a decade, we had conceived, prototyped, and crystallized a computerized process enabling the complete digital restoration of an object without physically opening it. This software pipeline, which we still use today, involves a five-step process: imaging the document, segmenting the layers, texturing the layers, flattening the images, and merging the flattened pieces into a single, virtually-unwrapped “master view.”

Imaging the Document

The first step is to acquire a digital image that comprehensively shows the internal structure and the contents of the object being restored. Volumetric, 3D imaging has come a long way since our camera-based work on the Cotton manuscripts (Fig. 7). Today, a number of choices exist for performing non-invasive, penetrative, volumetric scanning, and any of them can be used in our pipeline. We have relied on micro-computerized tomography (micro-CT) scanning for most of our work.

A primary imaging constraint is that the scanning method must provide fine enough resolution to see layers on which there is writing. A certain number of samples through the thickness of a page, for example, are necessary in order to visualize text. If the writing is very small—in some cases letters are only a few millimeters tall—the sampling rate must be high enough to capture it. The material and physical properties of the object are also important, since the elemental

composition of the materials determines how the x-rays of the tomographic imaging will reveal (or not reveal) its secrets. But today's range of techniques for volumetric scanning gives us the flexibility to take advantage of technological developments and the ensuing potential for improved images.

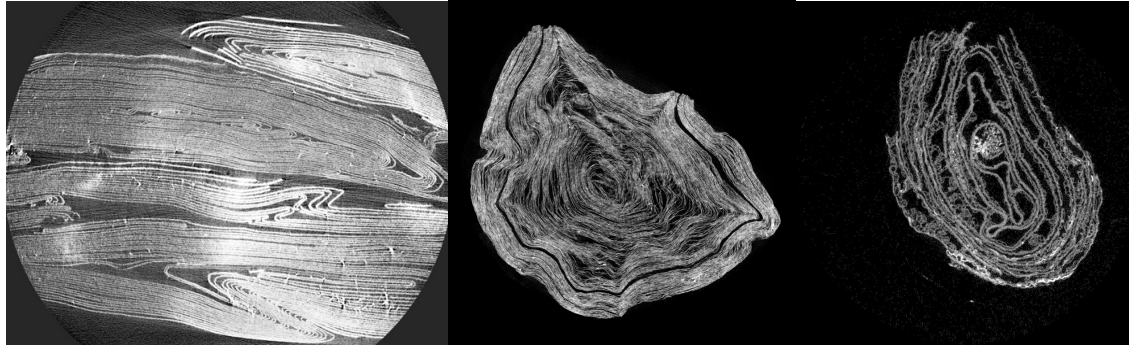


FIGURE 7: CT-scanning examples showing how the data appears. The scans are of the Franklin Papers, a Herculaneum scroll, and the scroll from En-Gedi.

Segmenting the Layers

To read a document, physical or digital, one must be able to separate and distinguish one page from another. This process becomes problematic in ancient documents. Pages often fuse together with the passage of time. In the case of writing surfaces made of animal skin, fat deposits therein can bubble, making it difficult for the computer to distinguish the true surface of a page from something that is just the back of a bubble. Furthermore, the layers in materials like papyrus are made from laminated fibers, which do not remain flat and intact when damaged.

Segmentation is the term denoting our process for addressing this wide range of potential problems. It's the first step after acquiring a 3D scan of a document. The evidence within the scan for surfaces on which there may be writing--layers of material--must be identified and modelled. We achieve this by treating the scan as a volume of 3D samples from the interior of the object to which we apply a variety of image processing and computer vision techniques. Specifically, as we have previously described [3], we perform this modeling using a triangulated surface mesh, which readily supports many operations that are algorithmically convenient: ray intersection, shape dynamics, texturing, and rendering. A surface mesh can vary in resolution as needed and forms a piecewise approximation of arbitrary surfaces on which there may be writing. The volumetric scan defines a world coordinate frame for the mesh model; thus, segmentation is the process of aligning a mesh with structures of interest within the volume. The goal of this alignment is to track as closely as possible to the undulations of the manuscript surface, and user adjustments are made along the way to ensure this happens.

In sum, the software works to extract a chosen layer, with the end result being a 3D, contoured representation of it, thereby digitally recreating the "page" that holds potential writing (Fig. 8).

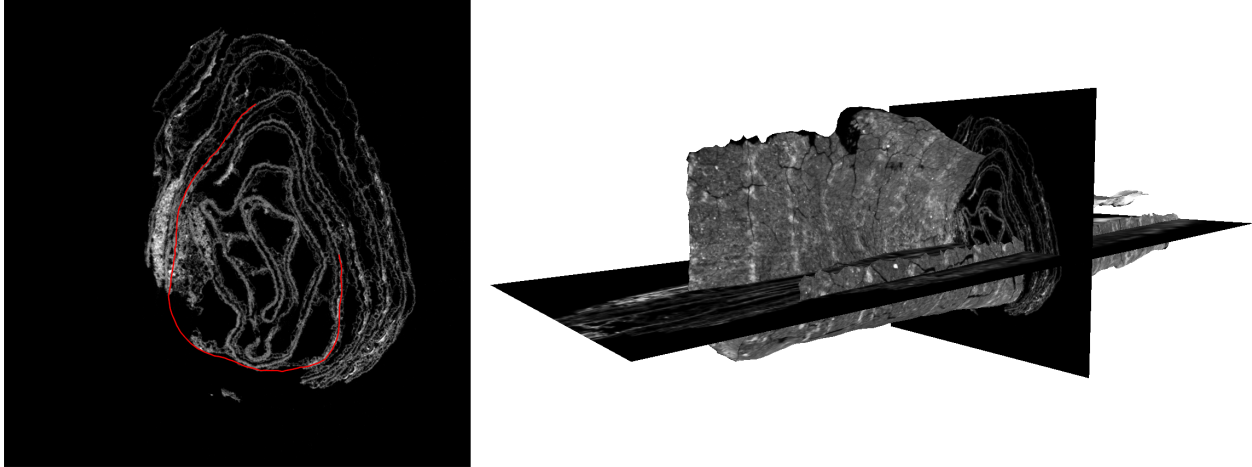


FIGURE 8: On the left is a slice from the CT-scan of a decayed 1st-century scroll [3]. One can see all of the tightly wound rolls that lie within. The red outline identifies the particular “wrap” being rendered in the image on the right, a process called “segmentation.”

Texturing the Layers

The next step is to extricate letters and words from the 3D scan and attach them to the recreated “page.” This process, called “texturing,” works by comparing intensity values in the scan. Each point on the reconstructed writing surface of the segmented mesh is assigned a brightness (intensity) value as it appears in the 3D volume. In the case of micro-CT, these intensities are related to density: brighter values indicate regions of denser material, while darker values suggest a less dense component. Comparing these intensity variations reveals the presence of ink. For example, since iron is a dense material, a coating of ink made from iron gall would appear brighter in micro-CT than the surrounding, less dense papyrus or animal skin surface on which it sits.

The texturing problem is one we started working on early in this project [9] as shown in Figure 9. We recognized that the assignment of intensities from the scan volume to the triangles of the segmented mesh could be done in a number of ways. Choosing wisely could help to overcome noise in the volumetric imaging, which contains noise in both the density estimates and the localization in 3D. But more than that, we found that noise in the segmentation--errors in localizing and estimating the shape of the layers--could be mitigated somewhat through the texturing process.

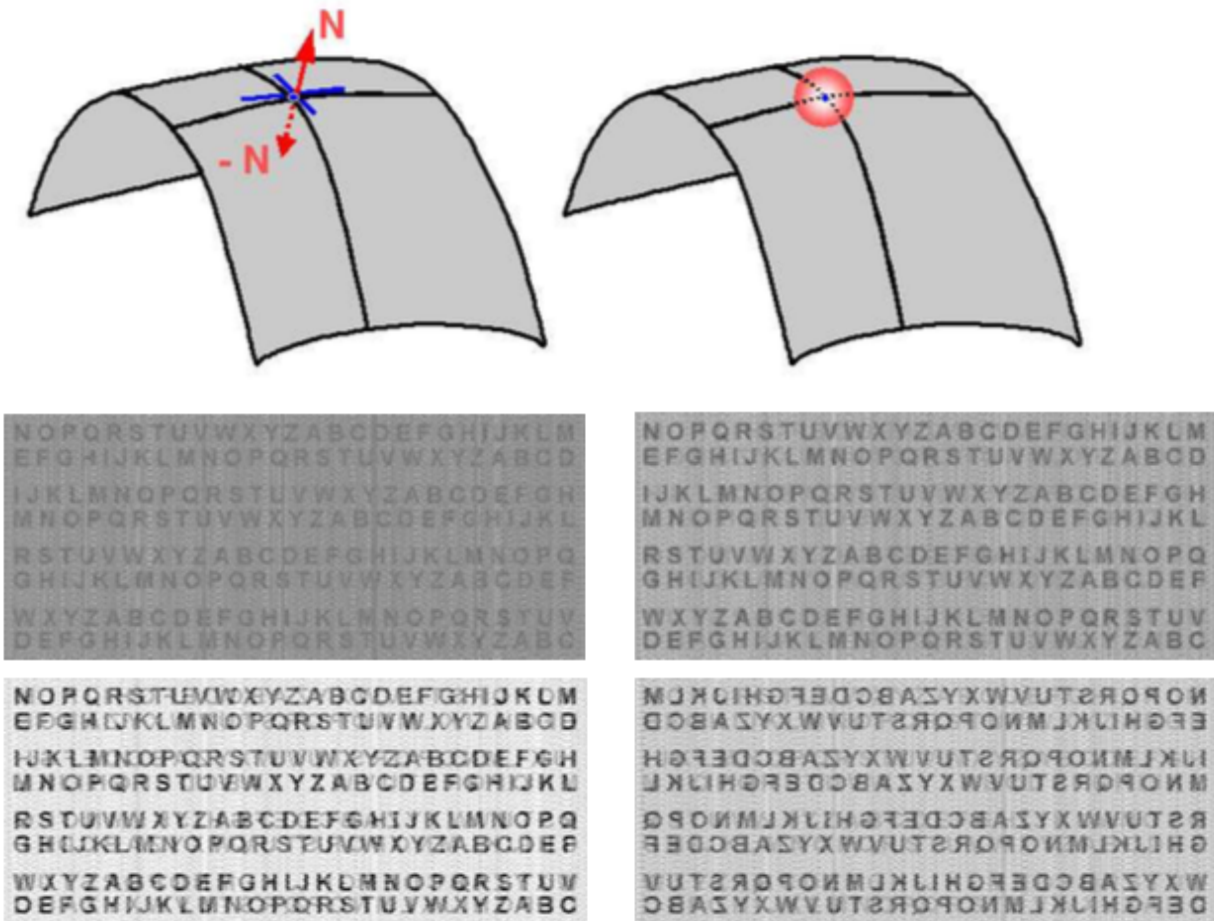


FIGURE 9: Early texturing [9]. The top sample shows texture geometry. The bottom reveals different textures resulting from the distance in both directions $+N$ or $-N$ from a simulation experiment.

As we have previously reported [3], the ideal case would allow each surface point to be mapped directly to its 3D volume position, generating the best possible replication of text. In practice, however, errors in the segmentation process combine with artifacts in the scan to create the need for a filtering approach that can overcome these sources of noise. Therefore, as an improvement over earlier work [9], we implement a neighborhood-based directional filtering method, which gives parametric control over the texturing (Fig. 10). The texture intensity is calculated from a filter applied to the set of voxels within each surface point's local neighborhood. We make use of the surface normal, which is particularly important when attempting to recover text from dual-sided materials, such as books. In such cases, a single segmented surface can be used to generate both the recto and verso sides of the page [3].

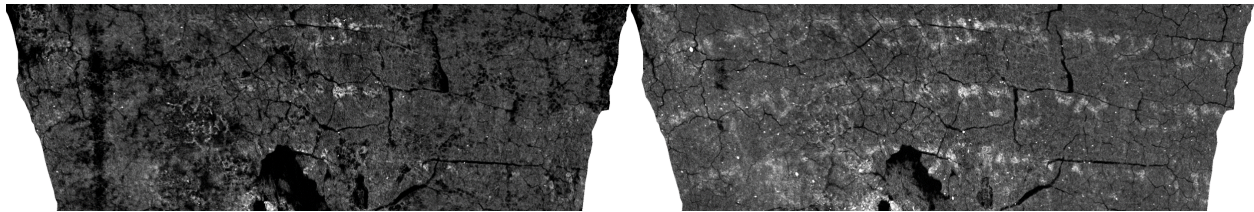
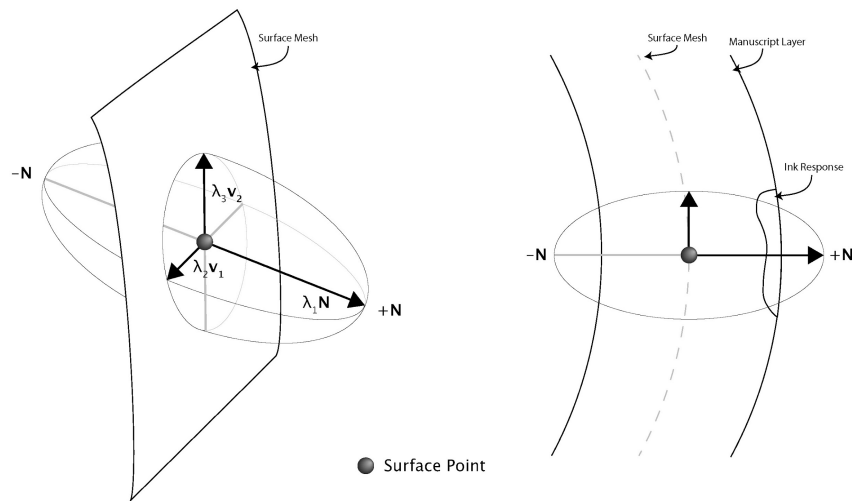


FIGURE 10: Improved texturing process [3] as applied to the scroll from En-Gedi.

Flattening the Images

Moving to a fully 3D scanning methodology gave us complete access to every wrap, layer, and turn within the interior of a book or scroll. The structure and shape of those undulating, curved 3D layers makes the digitized text difficult to read while it appears rolled up, because it's possible to view only one aspect of the page at a time. We needed a digital process to push the discovered 3D layers into flattened images for study and easier analysis. These methods are not unlike the early digital flattening techniques used on the Cotton collection [5].

We explored a number of methods for mapping a 3D surface onto a plane (Fig. 11). Our first attempt was an extension of the physics-based flattening methods for wrinkled manuscripts. We used a material simulation approach, mimicking the cloth modeling from the rendering community. In this model, the mesh is represented as a mass-spring system, where each vertex of the mesh is given a mass and the connections between vertices are treated as springs with associated stiffness coefficients. The mesh is relaxed to a plane through a balanced selection of appropriate forces and parameters. This process mimics the material properties of isometric deformation, which is analogous to the physical act of unwrapping.

We start with the 3D surface and pretend it is a piece of cloth to be flattened. The software simulates the unravelling process by stretching the corners out to calculate target positions. Pulling the corners is a first approximation to a cloth-like unwrapping operation. The software then pushes the surface onto a collision plane so that we can smooth out wrinkles. These

simulations are performed using an open-source physics engine, Bullet Physics. While most commonly used to produce visual effects in video games and movies, Bullet Physics simulates collision detection using soft and rigid body dynamics and thus provides the tools necessary for our conformal mapping method.

A major advantage of a simulation-based approach is the wide range of configurations that are possible under the framework. Parameters and forces can be applied per vertex or per spring. This precise control allows for modeling of not only the geometric properties of a surface, but also of the physical properties of that surface. For example, materials with higher physical elasticity can be represented as such within the same simulation.

We have also explored a hybrid approach that begins with existing parameterization methods, such as least squares conformal mapping (LSCM) [15] and angle-based flattening (ABF) [16], followed by a physics-based model. The purely geometric approaches of LSCM and ABF produce excellent parameterizations but have no natural way to capture additional constraints arising from the mesh as a physical object. By tracking the physical state of the mesh during a parameterization via LSCM or ABF, a secondary correction step using the simulation method could then be applied to account for the mesh's physical properties.

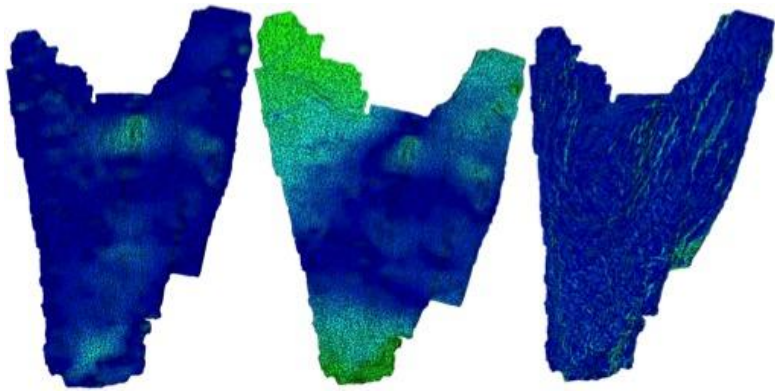
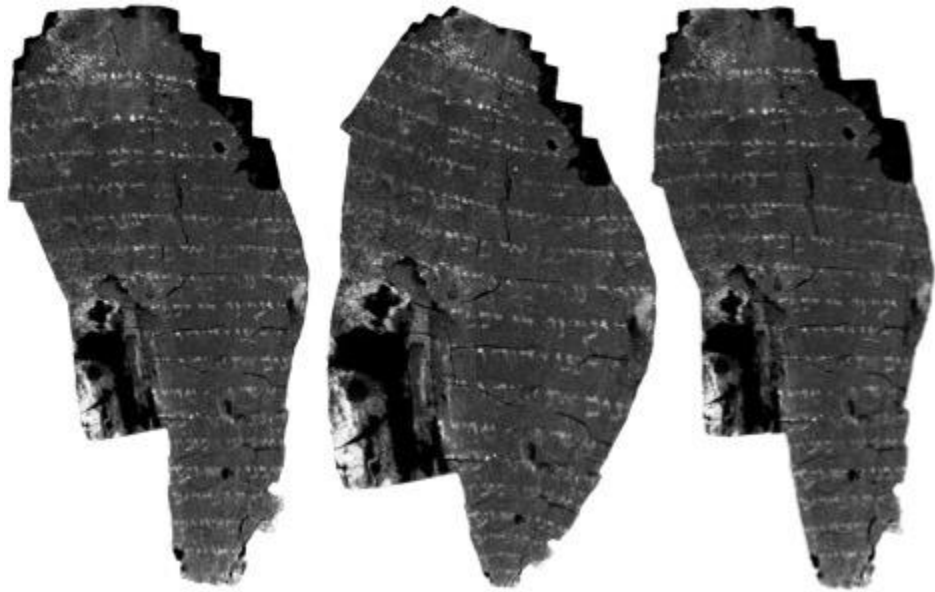


FIGURE 11: *Flattening samples [14]. The top image demonstrates the difficulty in reading a 3D wrap without any flattening, because the back side of the roll is not visible.*

Next, results from three flattening methods--physics-based, LSCM, and ABF--are compared. The bottom photo shows a heat, or error, map for the three approaches applied to a single fragment.

Merging the Flattened Pieces

Perhaps it is apropos that even a digital restoration approach to these badly damaged materials must work on pieces--fragments--and then reassemble them. Applying the framework in a piecewise fashion is necessary to accurately localize a scroll's highly irregular geometry. We have found no more exact method than to work on sections at a time, pushing them through the computational pipeline to identify and restore the writing, flatten it, and then merge those sections into a larger composite. We termed the final composite the "master view," which is the merged union of all the sections that are extracted and restored from within a volume.

As we have previously noted, [3] the creation of the master view is an important final step, because each of the prior steps in our pipeline involves a series of algorithmic decisions and approximations. Since textual identification is the primary goal, we tolerate some mathematical and geometric error along the way to ensure that we extract the best possible images of text. Therefore, the final merging and visualization step becomes significant not only for composing small sections into a single master view, but also for checking the correctness and relative alignments of individual regions. As such, it is crucial to preserve the complete transformation pipeline that maps voxels in the scan volume to final pixels in the unwrapped master view so that any claim of extracted text can be independently verified.

This merging process presents its own challenges, mainly involving workflow, in terms of keeping track of where all the pieces should go. In addition, errors can occur when lining up all of the edges. We have considered two key approaches for the merging stage: quick or resampled.

The quick merge is the alignment of texture images from small segmentations to generate a composite master view. As such, it favors a fast track to the master view. However, because each section of geometry is flattened independently, the merge produces distortions that are acceptable as an efficiently computed draft view, but must be improved to become a definitive result for the scholarly study of text.

The resampled merge favors accuracy and minimal errors in the merged sections as they are aligned. It offers a more precise merge step because it relies on the piecewise meshes which generate a high-quality master view. After all segmentation work is complete, individual mesh segmentations are merged in 3D to create a single surface that represents the complete geometry of the segmented scroll. The mesh from this newly merged segmentation is then flattened and textured to produce a final master view image. Since mesh merging is computationally expensive compared to texture merging, it is not ideal for the progressive feedback required during segmentation of a scan volume. However, as the performance of

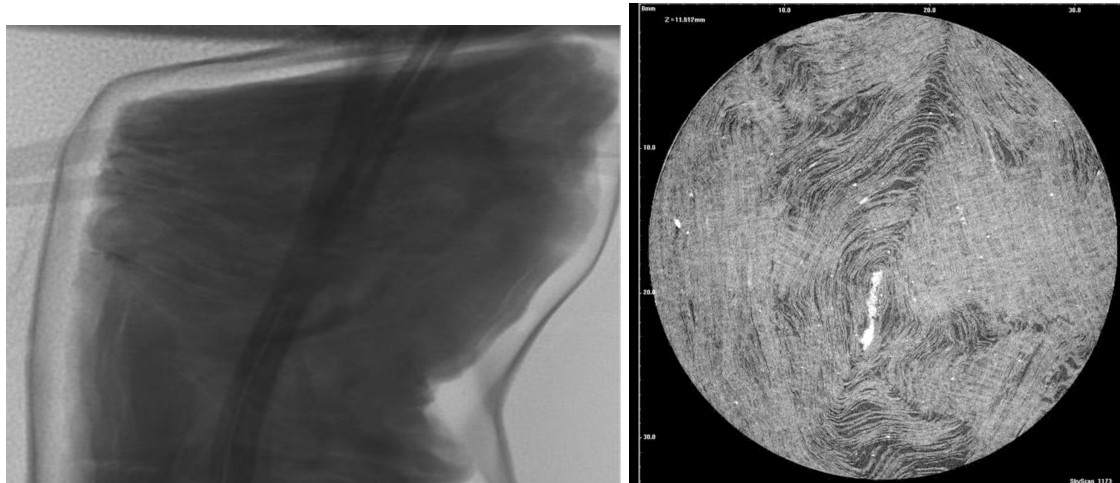
algorithms improves and larger segmented surfaces become practical, it is likely that mesh merging will become viable as a user cue during the segmentation process [3].

Once merging is complete, the end-to-end pipeline from the volumetric scan to the master view is preserved in the data and can support interesting visualizations of how the scanning captured the writing.

Toward Herculaneum

Armed with our fully developed pipeline, we were ready to tackle the scrolls from Herculaneum. In 2008 the Institut de France granted us permission to study a Herculaneum fragment [17, 18, 19]. Our investigation led to the discovery of lead and other trace elements in the ink and represented the first ever instance of reported heavy metals in Herculaneum writings. The complete results were published in 2009 in the proceedings of the Academy's annual program [21, 22]. Although the quantities were very small, the presence of metal-based ink, a surprising discovery that we were first to report, gave us hope that imaging methods based on absorption, such as our favored micro-CT scans, could produce contrast at the site of the writing. As a result of the knowledge we gained from the fragment study, the Institut de France granted permission for two complete and intact Herculaneum scrolls to be scanned on site using a micro-CT machine.

We conducted the scanning during the month of July 2009. This became the first ever volumetric scan of a Herculaneum scroll, and it revealed the complexity of the internal structure and the utter scale of the task facing us (Fig. 12). We immediately saw that reading the papyrus layers in these tortured and damaged scrolls was going to be more challenging than expected. By 2009 standards, our prototype software, which by then was a few years old, was ill-prepared to handle both the data sizes produced by the newer scanning technology and the complexity of the scrolls' internal structures. We were now dealing with terabytes of data, versus gigabytes, and the scrolls were so carbonized and tightly wrapped that in many places the scans showed little to no separation between layers.



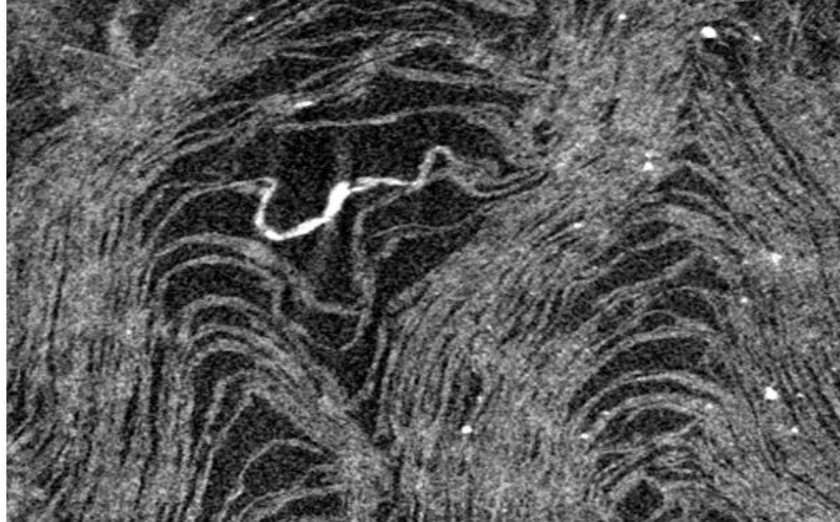


FIGURE 12: Left is the first radiograph of a Herculaneum scroll (by us in 2009). Right is a typical slice after the micro-CT reconstruction. Bottom is a detail shot showing the complexity of the scrolls' internal structure [21,22].

In trying to process the images, we learned that segmentation of Herculaneum scroll strata, as well as that of other ancient manuscripts, presents a number of unique problems; and that the naive application of existing segmentation techniques can cause poor or completely incorrect results. Depending on the type of material being scanned, scroll strata can appear fuzzy or almost indistinguishable due to time-induced distortion, disintegration, or other deformation of the writing medium. Commonly used writing surfaces, such as papyrus, can easily fray or suffer fire damage, while fat deposits in animal hide can bubble, producing local defects in the stratum surface that make it difficult to visually follow across CT slices. Furthermore, tracking a single stratum through an entire scan is made more difficult by the fact that undulations in the scroll strata can cause the separations between layers to disappear and reappear at random. In effect, layers merge together and then separate later on. By nature of the CT-scanning process, strata have the same relative density. When they merge together, they become nearly indistinguishable.

In addition to these material-specific problems, micro-CT introduces its own anomalies and deformities into scan results, such as beam hardening, streaking, and ring artifacts [23], that must be compensated for in the final raw CT slices. These extra artifacts can make image processing much more difficult on scroll CT data and are heavily dependent upon the parameters selected at the time of the scan.

We looked at these challenges and realized that our software pipeline was not yet ready to solve them.

Stymied, But Not Stalled

The brick wall encountered with the Herculaneum scroll [24] turned out to be a blessing in disguise, because it gave us some time to think about the problems and to experiment with how to best address them. While working on two primary challenges (the complex segmentation of layers and the production of ink contrast) we turned to a refinement of earlier techniques and field-deployed a full manuscript alignment system.

The *Illiad* and the Chad Gospels: Aligning Digital Copies

The oldest complete copy of the *Illiad*, the Venetus A, had been photographed in 1901 but had never been digitized. In 2007, we joined together with the Center for Hellenic Studies (Harvard University) and made a complete digitization of the manuscript on site at the Marciana Library in Venice [25]. Each page was captured in 3D, and some were imaged under ultraviolet light.

Through a series of experiments based on practical necessities [26] we developed a full image-to-image registration system that allowed the 1901 photographic facsimile to be aligned with the newly captured digital images. The result allowed a stunning view of the differences between the two images. The text aligned perfectly, clearly revealing the differences between the 2007 digitization and the 1901 photographic facsimile. Surprisingly, in some cases the 1901 facsimile was actually the “better” image of the two, because text had disappeared from the document during the intervening 100 years.

This image-based alignment operation was a variation of the surface modeling approach that we had been using for segmentation and virtual flattening as part of the restoration pipeline. By treating an image as a set of elastically-deformable points, one could be stretched and fitted exactly onto another in such a way that the text of both matched perfectly. This alignment, with the common surface model as a technical implementation, produced dramatic ways to visualize changes in a manuscript over time. We began to think about the possibility of aligning more than just two instances of a manuscript. What if many images had been taken over the lifetime of a manuscript?

In fact, the Chad Gospels--an early Insular Gospel book--had been imaged at at least five or six different times over the years. We continued developing the piecewise, non-rigid warping and registration prototype, and in 2010 we applied it to a new set of Chad Gospels images [27]. Mapping every prior photograph of the manuscript to the most recent digitization, the team created a virtual time lapse of every page. With perfect alignment of the words on every page from every image, flipping through them was like marching through time.

Valuable in its own right for diachronic alignment, the advancement and refinement of the non-rigid warping and registration piece of the pipeline was coming into place for a return to the full unwrapping problem. The team had solved a number of nagging problems, including the management of very large image sizes and the automatic matching and warping needed to

deform one surface model into another. These innovations proved crucial to the segmentation, flattening, and merging required for the full restoration pipeline.

Phase Contrast Tomography

At about this time, I attended a professional conference in Helsinki and met Uwe Bergmann, a physicist from Stanford University [18, 19]. After conversations and technical explanations, it became clear that there was a new tomographic approach in use that could provide more sensitivity to the ink than was possible before. As John Seabrook recalls it [31], after I explained the problem we were having with regular tomography, Uwe “didn’t bat an eye, but simply responded with three words: ‘Phase contrast, man.’”

His comment launched a multi-year quest for access, both to machinery for applying phase contrast tomography and to the intact scrolls themselves. Because phase contrast tomography relies on a coherent x-ray beam that can only be produced at a synchrotron, it is not possible to conduct the imaging on site within the confines of a museum or library. This raised the bar considerably in receiving permission to work on material.

Although we had completed the first virtual unwrapping on the University of Michigan book binding fragment as early as 2007, and had subsequently scanned and unwrapped sections of actual Herculaneum scrolls, the setbacks in processing data from the first-ever Herculaneum tomography in 2009 resulted in a completely redesigned software pipeline over the next several years.

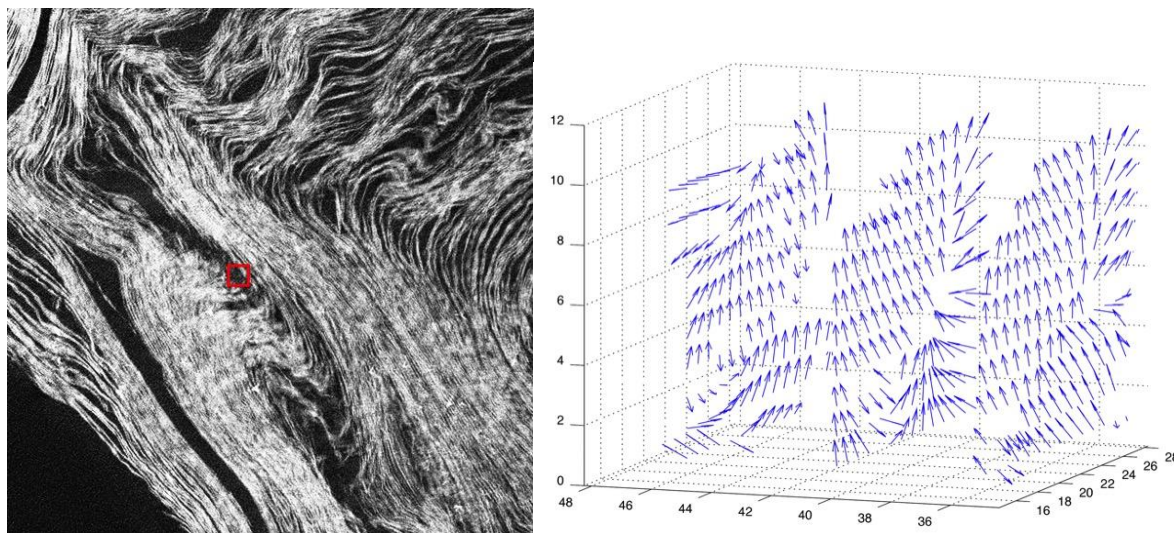


FIGURE 13: *During my time as a Visiting Scientist at Google, we developed new methods using the “structure tensor” to inform the segmentation process. This progress improved the software suite in automating the virtual unwrapping pipeline. The example shows a slice from the Herculaneum scroll and the estimated slice directions for a region of interest.*

While working toward phase contrast tomography and permissions to access material, we continued developing the software suite, maturing it to the place where we were once again ready to attempt to read the most damaged material anyone could imagine (Figure 13). We thought it would be from Herculaneum. But instead, it came from a desert in Israel.

The Scroll from En-Gedi

En-Gedi, Israel is the desert oasis where David hid from King Saul in the biblical account of 1 Samuel. But in 1970, it became the site of an exciting discovery. Right there on the shore of the Dead Sea, an Israeli archeologist pulled a blackened, 3-inch, cigar-shaped stick out of the ground. He was excavating the ruins of an 8th century BCE synagogue, and the ground where he was standing was actually the site of the ancient temple's holy ark.

This piece of charcoal, therefore, represented a dramatic discovery, as it was almost certainly a sacred scroll. But, burned and charred from a fire in the 6th century AD, then further damaged by 1500 ensuing years of deterioration, it was impossible to unroll and verify the crumbling scroll's contents without completely destroying it. So, despite the archaeologist's hunch that he had found something incredibly significant, the artifact was shelved and then eventually locked away in a vault at the Israel Antiquities Authority. There it remained untouched and unread for almost half a century.

In 2014, Pnina Shor, curator and director of the Dead Sea Scrolls Project at the Israel Antiquities Authority, contacted us and wanted to know if we could take a look at some data she had acquired from a volumetric scan of the scroll. We agreed, she gave us a hard drive containing the CT scan data, and in a few short months we achieved the impossible. Using our process of virtual unwrapping that we had worked for 15 years to develop, we revealed the scroll to be part of the Bible, the first chapter of Leviticus to be exact, and we did it without ever touching, opening, or even seeing the scroll.



FIGURE 14: *The scroll from En-Gedi on the left, and Pnina Shor and IAA colleagues at the press conference revealing its contents.*

When we sent Shor our preliminary results, she immediately called a press conference for the following week (Figure 14). She told the press, “When we saw the results we almost fainted. We had been certain it was just a shot in the dark.”

Shor’s shot in the dark—when pushed through our virtual unwrapping software pipeline—turned out to be the oldest Hebrew Bible ever found other than the Dead Sea scrolls and the only one ever uncovered in a Jewish synagogue. As such, it is one of the most significant biblical findings of the 21st century.

A Complete Restoration

Using our pipeline, we restored and revealed the Hebrew text on five complete wraps of the En-Gedi scroll and made a complete textual critique of the writing possible. Thanks in part to the remarkable spatial resolution now possible with micro-CT, our resulting master image equals the best photographic images available in the 21st century, with an effective resolution of 1500 dots per inch. The high quality of our final result enabled Hebrew and biblical scholars to arrive at dramatic conclusions regarding the scroll’s significance.

As noted by these scholars [3, 28], one can clearly see in the master view the remains of two distinct columns of Hebrew writing. These columns contain legible and countable lines, words, letters, and spacing (Fig. 15). Clearly restored is part of one sheet of a scripture scroll that contains 35 lines, of which 18 have been preserved and another 17 have been reconstructed. The lines contain 33- to 34 letters and spaces between letters; spaces between the words are indicated, but are sometimes minimal. The two columns extracted also exhibit an inter-columnar blank space, as well as a large blank space before the first column that is larger than the column of text. This large blank space leaves no doubt that what is preserved is the beginning of a scroll.

Armed with the extraction of this readable text and its historical context discerned from carbon dating and other related archeological evidence, the scholars were able to accurately place the En-Gedi writings in the canonical timeline of biblical text. The dating of the En-Gedi scroll to the third or fourth century CE falls soon after the period of the biblical Dead Sea Scrolls (third century BCE – to second century CE) and several centuries before the medieval biblical fragments found in the Cairo Genizah, which date from the ninth century CE onwards. As such, the En-Gedi scroll provides an important extension to the evidence of the Dead Sea Scrolls and offers a glimpse into the almost 800 years of near silence in the history of the biblical text.

Scholars also noted that, based on their knowledge of the development of the Hebrew text, the En-Gedi Hebrew text is not vocalized, there are no indications of verses, and the script resembles other documents from the late Dead Sea Scrolls. The text deciphered is completely identical with the consonantal framework of the medieval text of the Hebrew Bible, traditionally named the Masoretic Text and which is the text presented in most printed editions of the Hebrew Bible. On the other hand, one to two centuries earlier, the so-called proto-Masoretic text, as reflected in the Judean Desert texts from the first centuries of the

Common Era, still witnesses some textual fluidity. In addition, the En-Gedi scan revealed columns similar in length to those evidenced among the Dead Sea Scrolls.

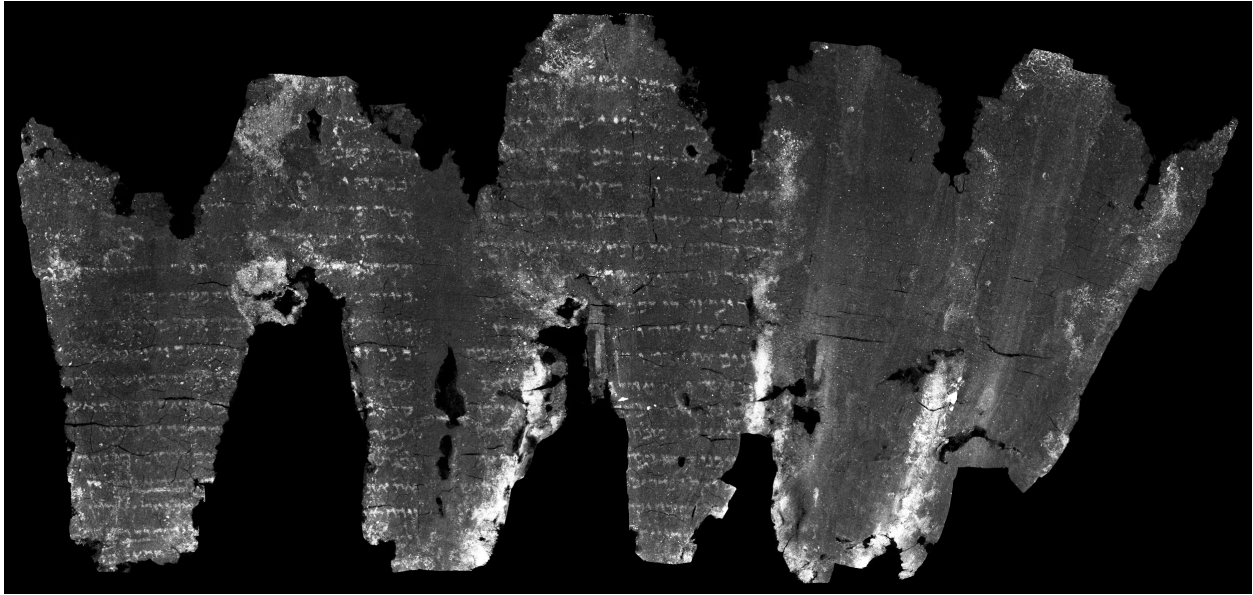


FIGURE 15: *The fully unwrapped En-Gedi scroll master view [28].*

Conclusion and Future Work

After more than two decades of work, we had finally achieved the impossible. We had pushed a hopelessly damaged, extremely friable, permanently closed scroll all the way through our software pipeline, from high resolution micro-CT digitization to complete virtual unwrapping and reading. Without our computational pipeline and the textual analysis it enables, the En-Gedi text would be totally lost for scholarship, its value left unknown.

This incredible achievement represents the first of what is sure to be many successes in making the invisible text visible through virtual unwrapping. Badly damaged artifacts sit on shelves in museum and library archives throughout the world, collecting dust while their potentially valuable contents remain locked away by their wrecked conditions. Mayan manuscripts; burned books and handwritten items; Native American copper plates; amulets; damaged scroll fragments at many locations; cartonnage; and fragments of earlier manuscripts bound within book bindings are just a few examples (Figure 16). And of course, the scrolls from Herculaneum still wait to be read, despite a few hints at letter forms [29, 30]. Once the first complete and intact Herculaneum scroll is read, there will be immense pressure to read others and then to excavate the remains of the library. The discovery, digital unwrapping, and reading of more scrolls, were it to happen, would represent the largest trove of classical materials ever revealed.



FIGURE 16: Franklin papers; Dynastic material (Asia); damaged manuscript from the cathedral at Chartres.

Our dedication to perfecting the technical pieces of our pipeline, coupled with our perseverance in the very human endeavor of building trust and gaining access to invaluable collections, has finally converged to create a solution for restoring some of the world's most profoundly damaged materials. From the fused and buckled pages of disintegrating books to the inner wraps of charred and carbonized scrolls, the world's vast invisible library can finally be made visible in a completely non-invasive, damage-free way.

Acknowledgements

Many students have contributed to the development of the decades of work summarized here by the authors, including Ryan Baumann, Michael Brown, Abigail Coleman, Matt Field, Sean Karlage, Yun Lin, Seth Parker, and CJ Yuan. A number of colleagues have also partnered in various positive ways, including Chris Blackwell, Duncan Clarke, Steve Crossan, Daniel Delattre, David Jacobs, Richard Janko, Dirk Obbink, Ross Scaife, and Pnina Shor. I gratefully acknowledge the editorial support of Christy Chapman in the preparation of this manuscript. And finally I am indebted to the institutional support that has given me access to world class material: The British Library, The British Museum, Center for Hellenic Studies, The Escorial, The Google Cultural Institute, The Institut de France, The Israel Antiquities Authority, Lichfield Cathedral, and the Marciana Library.

References

1. W. Brent Seales, Executive Producer. "Imaging the Iliad: A Digital Renaissance." vimeo.com/21758332. 2007.
2. Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science*, 2010.
3. William Brent Seales, Clifford Seth Parker, Michael Segal, Emanuel Tov, Pnina Shor, Yosef Porath, "From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi," *Science Advances* 21 Sep 2016 : e1601247.
4. The British Library, <http://www.bl.uk/reshelp/findhelprestype/manuscripts/cottonmss/cottonmss.html>.

5. Kevin Kiernan, "The Electronic Beowulf." <http://ebeowulf.uky.edu/>.
6. Michael S. Brown, "New Techniques and Algorithms for Acquiring, Restoring, and Displaying Digital Collections." PhD Dissertation. UK Department of Computer Science. 2001. Director, Brent Seales.
7. David Sider, "The Library of the Villa Dei Papiri at Herculaneum", Getty Publications, 2005.
8. Y. Lin and W. B. Seales, "Opaque Document Imaging: Building Images of Inaccessible Texts," International Conf. on Computer Vision (ICCV), pp 662-669, Vol 1, Beijing, 2005.
9. Y. Lin, Physically-Based Digital Restoration Using Volumetric Scanning (University of Kentucky, Lexington, 2007).
10. W. B. Seales and Y. Lin, "Digital Restoration Using Volumetric Scanning," Proc. 4th ACM/IEEE Joint Conf. on Digital Libraries, pp. 117-124, Tucson, AZ, 2004.
11. M. Brown and W. B. Seales, "Document Restoration using 3D Shape," International Conference on Computer Vision, pp 367-375, July 2001.
12. M. Brown and W. B. Seales, "Image Restoration of Arbitrarily Warped Documents," IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), pp 1295-1306, Vol 26, No 10, October 2004.
13. M. Brown, W. B. Seales, J. Griffioen, and K. Kiernan, "3D Acquisition and Restoration of Medieval Manuscripts," Communications of the ACM, pp 58-59, Vol 44, No 5, May 2001.
14. W. Brent Seales, C. Seth Parker, USA, Pnina Shor, Israel, USA, "[Virtual Unwrapping: Quantitative Distortion Analysis of Flattening Applied to the Scroll from Ein-Ged](#)", 2nd International Conference on Art & Archaeology, December 11-14, 2016, Jerusalem, Israel.
15. B. Lévy, S. Petitjean, N. Ray, J. Maillot, "Least squares conformal maps for automatic texture atlas generation," ACM Trans. Graph. 21, 362–371 (2002).
16. A. Sheffer, B. Lévy, M. Mogilnitsky, A. Bogomyakov. "ABF++: Fast and robust angle based flattening." ACM Trans. Graph. 24, 311–330 (2005).
17. R. Baumann, W.B. Seales, R. Scaife, "Recent work in the EDUCE Project", Digital Humanities 2008, Oulu, Finland.
18. W.B. Seales, J. Griffioen, and D. Jacobs, "Virtual Conservation: Experience with Micro-CT and Manuscripts", EIKONOPOIIA: Digital Imaging of Ancient Textual Heritage, Proceedings of the International Conference, Helsinki, 28–29 November, 2010.
19. W. B. Seales, J. Griffioen, and D. Jacobs, "Virtual Conservation: Experience with Micro-CT and Manuscripts," Finnish Society of Sciences and Letters: EIKONOPOIIA. Digital Imaging of Ancient Textual Heritage. Commentationes Humanarum Litterarum, Edited by V. Vahtikari, M. Hakkarainen, and A. Nurminen, Vol 129 p.81-88, 2011.
20. W. Brent Seales, Executive Producer, "Reading the Unreadable." <https://www.youtube.com/watch?v=wQ8v69wJ3K4>.
21. W. B. Seales and D. Delattre, "Virtual Unrolling of Carbonized Herculaneum Scrolls: Research Status (2007-2012)", Cronache Ecolanesi, Vol 43 pp. 191-208, 2013.
22. W. B. Seales, "Lire Sans Detruire Les Papyrus Carbonises d'Herculanum", Communication, l'Institut de France, 2010.
23. G. T. Herman, Fundamentals of Computerized Tomography: Image Reconstruction From Projections (Springer Science & Business Media, Berlin, 2009).

24. Jim Warren, "UK scientists stymied in effort to read ancient scrolls", Lexington Herald Leader, May 24, 2010.
25. R. Baumann, R. Scaife, W.B. Seales, "Imaging the Venetus A Manuscript for the Homer Multitext," Digital Humanities 2008, Oulu, Finland.
26. R. Baumann, D. C. Porter, W. B. Seales, The use of micro-CT in the study of archaeological artifacts, in 9th International Conference on NDT of Art, Jerusalem, Israel, 25 to 30 May 2008.
27. Stephen Parsons, Seth Parker, W Brent Seales, "The St. Chad Gospels: Diachronic Manuscript Registration and Visualization," Manuscript Studies, University of Pennsylvania Press (to appear), 2017.
28. M. Segal, E. Tov, W. B. Seales, S. Parker, P. Shor, and Y. Porath, "An Early Leviticus Scroll from En-Gedi: Preliminary Publication", Textus Vol 26, 2016.
29. V. Mocella et al., "Revealing letters in rolled Herculaneum papyri by X-ray phase-contrast imaging," Nat. Commun. 6, 5895 (2015).
30. O. Samko et al., "Virtual unrolling and information recovery from scanned scrolled historical documents," Pattern Recognit. 47, 248–259 (2014).
31. John Seabrook, "The Invisible Library", The New Yorker, November 16, 2015.