

## THE NYANGWE DIARY OF DAVID LIVINGSTONE: RESTORING THE TEXT

### *Enhancing the Humanities through Innovation*

This application is for a Digital Humanities Start-Up Grant, Level II. We propose a pilot project to:

- Build on previous spectral imaging of medieval palimpsests to devise and implement new processes and technologies to spectrally image the Nyangwe field diary of the Victorian medical missionary and explorer David Livingstone (1813-1873), and to process the resulting image data with open source software in order to recover the faint and often illegible text of the diary.
- Publish online a critical edition and image database so that Livingstone's unredacted diary is accessible to scholars for the first time, with both marked up transcription and processed images.

This pilot project will break new ground in the application of advanced digital imaging and data management in the humanities. It will:

- Extend the use of the spectral imaging from medieval palimpsests to faded nineteenth-century paper manuscripts, thereby challenging its scientific participants to identify optimal imaging techniques for a nineteenth century paper manuscript and educating scholars of the nineteenth century on the potential of the new technology for their field.
- Explore new frontiers for cross-disciplinary collaboration in the technical processing and scholarly study of digitized cultural objects: data will be marked up with all metadata needed for retrieval, scholarly and scientific analysis, *and* interoperability with related projects.

Livingstone wrote the field diary in 1871 during a five-month sojourn in the eastern Congo village of Nyangwe. The diary is of significant historical value because it describes a violent and unprovoked massacre of the African population by Arab slave traders. This event would become iconic for late-Victorian era abolitionists around the globe. Equally important, the diary reports the immediate circumstances leading up to Livingstone's famous meeting with Henry Morton Stanley, the correspondent for the *New York Herald*, in November 1871. Written with the only materials available to Livingstone—ink from a native plant and broad sheets of an 1869 edition of *The Standard*—the diary's text is now largely illegible. Spectral digital imaging techniques developed over the last decade to recover erased writing in medieval palimpsests open the possibility of recovering Livingstone's faded writing and so restoring a central text to the study of nineteenth-century imperialism and colonialism.

The project will be developed through a scholar-scientist-institution transatlantic collaboration. *The Early Manuscripts Electronic Library* (EMEL) of California, which recently worked with the scientific team from the Archimedes Palimpsest project ([www.archimedespalimpsest.org](http://www.archimedespalimpsest.org)) on a pilot project with palimpsests at St. Catherine's Monastery of the Sinai, will organize the same scientific team to spectrally image the Nyangwe diary. The David Livingstone Centre in Blantyre, Scotland, which holds the diary, will provide access to original materials. Two literary scholars (one of whom is Project Director) will prepare the critical edition for publication through *Livingstone Online*, which is based at the Wellcome Trust Centre for the History of Medicine at University College London.

The proposed pilot project is the first phase (May 2010 to October 2011) of a three-phase program to produce a dynamic, online critical edition of David Livingstone's final African exploration diaries (1869-1873), perhaps the most complete extant set of such diaries produced by any Victorian explorer of Africa. The second phase (2012) will focus on the remaining field diaries from 1869 to 1871, and the third phase (2013-2014) on the field diaries from the last two years of Livingstone's life, including those that record the meeting with Stanley. Experience gained from the pilot project will guide the execution of the entire program and provide a model for collaboration between imaging scientists and humanities scholars to produce comprehensive digital editions of literary manuscripts that treat such manuscripts as both texts *and* material and cultural objects.

### *Start-up Activities and Goals*

Given the historical value and technical challenges posed by the Livingstone Nyangwe field diary, we plan to pursue two sets of interrelated goals, the first scientific and the second scholarly.

The first set of goals involves spectrally imaging the field diary and processing the images to reveal Livingstone's faded handwriting over the printed text of *The Standard*, as well as any erasures or corrections. These results, which form the foundation of our project, are not guaranteed and so involve a significant amount of risk as per the "high risk/high reward" paradigm. Our work to image and process images of the pages of the field diary offers the potential to (1) separate Livingstone's writing from the printed undertext, and (2) reveal and clarify the overtext. We will build on previous work with the Archimedes Palimpsest ([www.archimedespalimpsest.org](http://www.archimedespalimpsest.org)). Based on work at the Library of Congress and forensics research, the imaging team will also quantify the ink and provide spectral properties and samples to ink databases in the Library of Congress and U.S. law enforcement community. If needed, contacts in the U.S. forensics community will be asked to provide advice and guidance on the processing and enhancement of the ink(s).

The scientific team will bring equipment from the U.S. and image the diary in two phases, which at times will overlap: "Standard Imaging" and "Experimental Imaging." Standard Imaging uses proven hardware and techniques developed for the Archimedes Palimpsest to digitally capture images in 13 narrow bandwidths and with raking lights. Experimental Imaging exploits new optics and processing techniques that may offer the best potential to reveal the overtext on paper. The capture of multiple registered images across a range of spectral bands using different techniques enables digital processing of the images to reveal information not readily discernable to the naked eye. This requires not only a range of technologies, but also enhancements to the imaging and processing software. This is being developed with open source software to allow follow-up processing by scientists and scholars around the globe.

The second set of goals involves addressing metadata, access, and data management challenges in (1) transcribing the overtext of the Nyangwe field diary, (2) producing a multidisciplinary critical apparatus, (3) encoding the transcriptions and apparatus in XML, (4) developing the appropriate tools for the online display of the texts, and (5) making the product available to the public by July 15, 2011, the 140th anniversary of the Nyangwe massacre. The apparatus will benefit from the considerable expertise of the two participating literary scholars: Victorian colonial literature and geographical discourse, African history, and British imperial history and biography (Wisnicki); history of nineteenth-century western and tropical medicine, and medical humanities (Harrison). Much of our work in transcribing, editing, and encoding the texts will follow practices already developed for the Livingstone epistolary materials published on *Livingstone Online* (see [www.livingstoneonline.ucl.ac.uk/technical/policies.html](http://www.livingstoneonline.ucl.ac.uk/technical/policies.html) and the hyperlinks to "our approach" and "critical path").

The XML markup of the transcriptions will conform to guidelines established by the Text Encoding Initiative (TEI). Online display will allow readers to organize, sort, and retrieve the material according to their needs. The transcriptions will recreate the original texts as closely as possible and will be in diplomatic format, thereby giving access—at user discretion—to any cancellations or other modifications Livingstone may have introduced in the process of writing. Readers will also be able to (1) consult Livingstone's entries by document, date, or continuous chronological series, (2) compare the corresponding entries in the diary and Livingstone's "Unyanyembe Journal," and (3) read the transcriptions alongside the images. The texts will be fully searchable, and NINES-ready RDF (Resource Description Framework) metadata descriptors will ensure interoperability with the other nineteenth-century digital objects (see [www.nines.org](http://www.nines.org)).

### *Ultimate Results and Importance of Proposed Work*

Lessons learned in the development of new imaging techniques, work processes, and advanced digital scholarship will benefit, respectively, imaging and preservation scientists, data managers, and scholars of

culture, history, and medicine. Building on previous imaging of documents with multiple layers of text, the pilot project will develop imaging and image processing techniques that are most effective to reveal overtext. The pilot project will also standardize both work processes to support the efficient flow of data *and* formats for data delivery to scholars and the broader community with open source software. For long-term digital preservation, the images and transcriptions will be hosted on the Internet in “flat files” that are not dependent on any specific GUI. For immediate and follow-up studies, all data and metadata will be hosted on servers for digital analysis of the image products. By preserving Livingstone’s work in multiple locations, the pilot project will therefore make Livingstone’s field diary and journals freely accessible to the public, thus contributing to conservation by reducing the need for handling the original documents. Furthermore, the pilot project (phase one) will lay the ground work for phases two and three of a program to create digital editions of all of Livingstone’s final diaries, and will highlight the value of Livingstone’s unadulterated, first-stage field diaries as a unique primary source for a diverse range of disciplines. All phases of the project will also stimulate reinterpretation of the diaries themselves as well as Livingstone’s final years in Africa, analysis of which has hitherto been based only on highly edited versions of the texts. Finally, our work will establish that publication of exploration field diaries (of Livingstone and others) requires both transcription to make the texts easily legible and imaging because of the additional information available from the diaries as cultural and material objects.

#### *Environmental Scan*

Our work builds on previous projects to recover information from significant historical artifacts, such as fragments of the Dead Sea Scrolls, the Archimedes Palimpsest, the Sarvamoola Granthas in Udipi, India, the colophon of a 15th-century Florentine Siddur, and palimpsests at St. Catherine’s Monastery, among others. Our team has a successful track record using advanced imaging technologies to recover illegible writing and has experience integrating images with spatially mapped transcriptions of text. The proposed project will further refine imaging and processing techniques that were pioneered during the Archimedes Palimpsest project and supplement those established techniques with experimental techniques. Applying a relatively new combination of imaging techniques to nineteenth century paper, printed text, and ink of unknown composition likewise poses a high risk/high reward challenge.

Our project supports an emerging critical trend in the digitization of nineteenth-century literary and cultural materials to produce rigorous, XML-based, interoperable projects focused on a single author (see [www.nines.org/about/scholarship/scholarlyProjects.html](http://www.nines.org/about/scholarship/scholarlyProjects.html)). Our project bears most affinity with those projects that attempt to provide multiple versions of a single text or that image and transcribe at least a portion of an author’s extant manuscripts, including: the Rossetti, Walt Whitman, and William Blake archives; *Herman Melville’s Typee*; and the *Dickinson Electronic Archives*, the latter of which, like our project, places manuscript materials at the fore and uses high-quality color images to highlight the additional information embedded in Dickinson’s manuscripts as material and cultural objects.

Finally, the project fills a significant gap in critical editions of Livingstone’s work. There have been few if any editions of the diaries, notebooks, and journals from the final expedition (1866-1873). Most critical endeavor related to this period has focused on Livingstone’s letters. In posthumously publishing *The Last Journals of David Livingstone* (1874), Waller did use Livingstone’s “Unyanyembe Journal” and a few of Livingstone’s field diaries (though not the Nyangwe field diary, despite a claim to the contrary [see 1874:114]), but, as Dorothy Helly has shown conclusively, Waller sought to portray Livingstone as a national hero and so edited, censored, and tampered with the data available in the original manuscripts.

#### *History and Duration of the Project*

During the last four years, the project director (Wisnicki) has been researching and writing a critical, book-length study of Victorian travel and exploration literature centered on Africa, provisionally titled *Fieldwork of Empire: Intercultural Dynamics in the Production of Victorian Colonial Literature*. Several articles drawn from this project have been published or are forthcoming in top journals such as *Journal of*

*Colonialism and Colonial History* (2007), *Victorian Studies* (2008), *History in Africa* (2008), *Victorian Literature and Culture* (2009), and *Studies in Travel Writing* (2010). His work on this monograph has also resulted in an invitation to become a research consultant for *Livingstone Online*.

Wisnicki's research, in particular his need to use the Nyangwe field diary, occasioned a query to L-SHARP, the listserv of the Society for the History of Authorship, Reading, and Publishing, and eventually resulted in a series of discussions with the Archimedes Palimpsest imaging team. The team took interest in the diary because of the opportunity to apply techniques developed with older, parchment-based undertexts to a more modern, paper-based overtext. The team brings extensive experience in imaging science, imaging technologies, technical program management, and data management in support of advanced imaging programs. The team has always provided fully integrated solutions to ensure all stakeholders, including scholars, conservators, and funding organizations receive the expected deliverables in a timely fashion.

Specifically, the team's effort in developing a pilot capability to image the Livingstone diary will build on our advanced imaging work to support the studies of manuscripts and documents in cultural institutions such as the Library of Congress, the Walters Art Museum in Baltimore, and St. Catherine's Monastery at Mount Sinai, Egypt. This includes development and application of spectral imaging and processing techniques to image the parchment Archimedes Palimpsest beginning in 1999; the paper Waldseemüller 1507 World Map at the Library of Congress in 2007; and a range of historic manuscripts, maps and prints including, in 2009, drafts of the Gettysburg Address and the Declaration of Independence on paper.

In addition to the present project (phase one), it is anticipated that our scholar-scientist team will collaborate on phases two and three of the larger program to complete a digital critical edition of Livingstone's final diaries. Because the project represents a transatlantic collaboration, we will apply for supplemental funding in both the US and UK. In the US, we will apply to the NEH for a JISC/NEH Transatlantic Digitization Collaboration Grant and/or a Scholarly Editions and Translations Grant. In the UK, we plan to apply for one or more of the following: Wellcome Trust Pilot Grant, Wellcome Trust Programme Grant, British Academy Small Research Grant, British Academy Research Development Award, and Leverhulme Trust Research Project Grant.

#### *Work Plan*

The work will be carried out over an eighteen-month period and will explore two viable options to recover Livingstone's handwritten overtext:

1. Spectral imaging and processing to use the spectral differences to suppress the undertext and enhance the overtext. This is the most mature technique with known risk and opportunity for timely results.
2. Computer separation of the handwritten over text by subtracting the printed undertext with a clean copy of the particular issue of *The Standard*. A key challenge will be registration of the clean copy and the overwritten copy. This procedure will require quality scans of *The Standard* and the labor-intensive warping and stretching of the two images (the diary and *The Standard*) so that fiduciary points match up. This technique could require significant new technology integration, with higher risk. The imaging team will collaborate with forensics experts.

Image Collection. The imaging team has developed a state-of-the-art image capture system that maximizes the information recorded with minimum impact on the historical documents. The pages are illuminated by a number (typically 11-13) of narrow bands of wavelengths of light from the ultraviolet (UV) through the visible and the infrared (IR), each band generated by light-emitting diodes (LEDs). A primary advantage of this system is the LEDs do not generate heat that can damage the fragile pages. Depending on the materials on the pages, light at short wavelengths (UV and blue) may be absorbed and re-emitted by the process of fluorescence. When captured by the monochrome camera, this fluorescence

conveys significant useful information about the text. Images collected using light at longer wavelengths primarily show reflectance from the page that varies with wavelength, depending on the spectral response of the inks and substrate.

In addition to the Standard Imaging, up to six experimental imaging techniques may also be used:

1. Fluorescence emission imaging with spectral illumination from above and a color filter wheel.
2. UV reflectance with spectral illumination from above and a UV transparent lens and UV filter.
3. Transmissive imaging with normal spectral illumination from above and reflective substrates.
4. Transmissive imaging with spectral illumination from below through a transparent table.
5. Raking light or Polynomial Text Mapping (PTM) using spectral illumination at multiple points around the object, supporting development of a new technique with this system.
6. Nonimaging Spectrometry using an Ocean Optics spectrometer.

Image Processing. The imaging team proposes to render pseudocolor images from the spectral images, following an initial survey of the spectral responses of the inks. The pseudocolor images will be constructed by a process developed during the Archimedes Palimpsest project and since refined. The processed image is a combination of the UV fluorescence image and the image from a red illumination. Since the over and under texts comprise different inks with different spectral responses, this difference can be used to render each spectral response in contrasting colors. We also propose to apply principal component analysis (PCA) to these images, which maximizes the data variance in the images to enhance contrast between principal components. Transmissive imaging and PTM may also offer new techniques for separating texts in different inks on paper.

Core image processing operations will make use of the open source software package ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, <http://rsb.info.nih.gov/ij/>, 1997-2009). ImageJ is widely used by scientists throughout the research community on all major operating systems (Unix, Windows, Macintosh, Linux). Written in Java, ImageJ provides an extensive range of tools for image acquisition, processing, display, and mathematical manipulation. With an extensible plugin architecture, users can add functionality using low (Java plugin) and high (script languages such as Perl and Javascript) levels.

ImageJ supports all major file formats and major metadata and archival standards with extensive support for existing metadata and spectral imaging standards. Our team uses ImageJ extensively in imaging of major cultural heritage materials, drawing on existing tools for core operations. Specialized tools that we have implemented to streamline our particular workflows, operations, and quantitative analyses include:

- File manipulation routines to handle the key formats (multispectral TIFF and DNG image cubes), precisions (8-, 16-, and 32-bit integer and floating point), and large file sizes (up to 1.3 Gigabyte);
- Fundamental data operations to perform flat fielding, local normalization, color reconstruction and display, basic calibration, and advanced quantitative processing;
- Highly specialized computational algorithms for quantitative image processing operations (pseudocolor, spatial filtering, PCA, local and kernel operators, logical and linear image algebra);
- Metadata creation and inspection, and archival manipulation tools, which work in conjunction with tools and structures furnished by our data manager, Doug Emery.

Data Management. The Nyangwe field diary project will provide a complete package of images with documentation and full metadata. The resulting archive data set will be based on the archive and metadata model used for the Archimedes Palimpsest. The product will be a completely self-documenting and autonomous data set. The collection of the appropriate metadata will require planning before, collection during, and organization after the imaging sessions. Metadata collection will be conducted as part of the imaging with the MegaVision PhotoShoot software that controls the image collection. After imaging and processing is completed, the imaging logs will be collated with other required metadata to generate complete metadata records for raw and processed images, and to assemble the final data archive. This will

be a full digital archive of images that meets library and archival standards, as we have produced for previous projects. The archive will be hosted by *Livingstone Online*, while backup data hosting will be facilitated by *The Early Manuscripts Electronic Library*.

XML Markup and the Digital Product. The Livingstone Journal Digital Product will include XML tagged and marked up images and transcriptions produced as follows.

1. All images will be converted to TIFF, with all metadata included in the header of the TIFF images, including spectral and spatial metadata.
2. Wisnicki and Harrison will transcribe the text from the images and identify and mark up significant content elements (e.g., names, places, key terms).
3. The transcriptions will be marked up with metadata in XML format. This will include spatial data, with tags linking each line of text to the line on the image for the appropriate text.
4. With metadata in the header of the TIFF images, we will produce XML tagged TIFF image products of the journal images for all imaging modalities, registered as much as possible.

### *Staff*

As project director, Wisnicki will devote approximately one-third of his time to producing and marking up the transcriptions. He will submit the completed project to NINES for peer review, and ensure that the results of the project are properly integrated with other relevant materials on *Livingstone Online*. Harrison will devote approximately one-fourth of her time the project and assist in producing and marking up the transcriptions of Livingstone's text. Both Wisnicki and Harrison will develop the appropriate scholarly apparatus. Both will manage the imaging effort and the team of imaging scientists (Christens-Barry, Easton, Knox) who will conduct the imaging and image processing during a 12-month period. Emery will provide overall data management and ensure metadata and data meet TEI and other applicable standards. Phelps will serve as grant administrator.

### *Final Product and Dissemination*

Project results will be disseminated through a combination of media. The full set of data produced by the present project will be made available through *Livingstone Online* and, through the assistance of *The Early Manuscripts Electronic Library*, in on-line backups and mirror sites. This combination will ensure the continued availability of the data product. Formal publication of the marked up transcription and image database on *Livingstone Online* will reach both non-specialists and a broad scholarly audience with interests ranging from textual scholarship and the preservation of historical materials to Victorian exploration and geography, African history, British imperial history and biography, and the medical humanities. Interoperability with NINES will ensure that the data can be regularly consulted by scholars with interests in the digitization of nineteenth-century manuscript materials. The successful completion of each phase of the project will be announced in professional newsletters. Finally, the imaging scientists and scholars will also present on their findings at professional conferences.

To reach the public, we will distribute press releases at key project milestones, and we will hold a formal event on July 15, 2011, the 140th anniversary of the Nyangwe massacre, at the David Livingstone Centre. Finally, given the historical significance of Livingstone's Nyangwe diary, the unique scholar-scientist-institution nature of our collaboration, and the combination of textual and visual material that the project will produce, our work also offers the potential for educational museum exhibits in the US and the UK. These exhibits can detail how Livingstone's efforts as an explorer and abolitionist impacted the history of colonialism and slave trading in Africa, and how our unique cross-disciplinary collaboration helped recover and preserve the texts of Livingstone's final exploration diaries.

The project white paper will report innovations in image capture and image processing, data management, and workflow processes, and describe lessons-learned from scholar-scientist collaboration, integration of cross-disciplinary data, and dissemination through a broad range of media.