Julia Josfeld

Internship in Textual Sciences
at the University of Rochester
(August & September 2017)

During my studies in medieval English literature, culture, and language, I have become increasingly fascinated with the discipline of Digital Humanities, particularly with regards to the digitization of medieval and early modern textual culture. When it comes to medieval texts, one of the persistent problems is the scarcity of sources and the challenges surrounding their preservation, edition, and dissemination. As technology advances, an increasing number of libraries and archives have begun to digitize their collections of manuscripts in order to make them more easily accessible to a wider audience of scholars and the interested public. This digitization can take many forms, from a straightforward reproduction of high-resolution photographs online, to highly elaborate electronic editions, and even collaborative transcription projects. One approach that caught my eye in particular, however, is the new field of "Textual Sciences."

About nine months before I started my internship, I came across a TED talk on YouTube.[1] In it, Professor Gregory Heyworth, currently working at the University of Rochester, talked about the possibilities for examining cultural heritage artefacts with new technologies, such as spectroscopy or multispectral imaging (more on this later). He explained that he was pioneering a new field of study that he and his colleagues are calling "Textual Sciences". It combines scientific methods and new technologies with more traditional modes of literary and cultural study to reach new insights into even well-known texts. I was intrigued by this innovative idea and contacted Professor Heyworth to ask how I could learn more about this field and its application to my particular interests. It was something of a surprise (and an enormous stroke of luck), when he suggested that I should come to Rochester for an internship in the summer which, of course, I was only too happy to do.

The Institution

Rochester is a city in upstate New York that has been a centre for imaging technology from the turn of the twentieth century. Both Kodak and Xerox were founded here and had their headquarters or main factories in the city for most of the 1900s. More recently, the Chester F. Carlson Centre for Imaging Sciences was founded at the Rochester Institute of Technology (RIT) and in 2016, Professor Heyworth and his research team moved from Mississippi to join the University of Rochester (UR). Together, researchers from these two institutions created the Rochester Cultural Heritage Imaging, Visualization and Education initiative, R-CHIVE for short, which has since been coordinating research and organizing conferences in this interdisciplinary field of study.[2] My internship in Rochester was primarily with one sub group of R-CHIVE, the Lazarus-Project.

The Lazarus-Project is "a multi-institutional international collaboration between imaging scientists, humanities faculty, graduate students, and independent consultants", as they describe themselves.[3] It is a mobile lab that uses different imaging methods, mainly multispectral photography,

---

[1] Gregory Heyworth, *Gregory Heyworth: How I'm Discovering the Secrets of Ancient Texts* (2015) <https://www.ted.com/talks/gregory_heyworth_how_i_m_discovering_the_secrets_of_ancient_texts> [accessed 6 March 2018].

[2] cf. Rochester Cultural Heritage Imaging, Visualization and Education, *R-CHIVE: Imaging a Future for the Past* (2018) <https://r-chive.org> [accessed 6 March 2018].

[3] cf. their website: Gregory Heyworth, *The Lazarus Project: A Multi-Spectral Imaging Project* (2017) <https://lazarusprojectimaging.com> [accessed 6 March 2018].

to recover lost or damaged text, inscriptions, images, and so forth from manuscripts or other artefacts in museums, libraries, and archives all around the world. Among their many projects have been the digitization of the Vercelli Book, an eleventh-century manuscript with great importance to scholars of Old English poetry, parts of the Dead Sea Scrolls, and most recently the Behaim Globe, the oldest extant globe of the world, currently situated in the Germanische Nationalmuseum in Nuremberg.[4] As technology progresses, Professor Heyworth and his team keep adding new and improved methods for digitisation and image processing to their roster, while also working to spread the knowledge of this relatively new field.

During my internship, I was able to participate in a range of research undertaken by the Lazarus project and R-CHIVE more generally. Most of that work centred on multispectral photography or the images obtained through it, so a short explanation of this process seems in order.

The Projects

Multispectral Imaging

At its core, multispectral imaging means that an object, e.g. a manuscript page with text that has become illegible due to fading or some other damage, is illuminated by different-coloured lights, ranging from infrared all the way to ultraviolet, and extremely high-resolution images are taken with a special lens. The images are then processed in different ways, using a variety of computer programmes, in order to make any lost text visible again. Sometimes it is as easy as looking at images taken under ultraviolet light, which has been used for decades by scholars trying to recover faded text, but there are also more intricate scenarios. Because different materials (parchment, ink, paint pigment, dirt, etc.) are chemically distinct, they can have different reactions to light of a specific wavelength. Iron gall ink, for example, one of the most common inks used in medieval manuscripts, starts to fade under infrared light of relatively short wavelengths, while other inks stay visible even under longer waves. So if a text in iron gall is written on top of, and thereby obscuring, older writing in different ink, using images taken under different infrared lights can be a way of making the undertext visible again. Of course, the situation is rarely this straightforward, but it provides a good example of the underlying principle.

There are a number of labs around the world that have multispectral capabilities, though the community is still very small as multispectral imaging has only been used for recovering cultural heritage artefacts for just over a decade. Starting in 2006, a team of interdisciplinary scholars (among them Professor Roger Easton, a colleague of Professor Heyworth's from RIT) used multispectral photography to image the Archimedes Palimpsest, a manuscript from the 10th century containing otherwise unknown treatises of Archimedes of Syracuse, that had been erased and overwritten with Christian texts in the 13th century.[5] Through their innovative techniques, the team was able to recover the lost text of Archimedes and since then, the approach has become more well-known. The specific application of this science is, however, still in its infancy and so far, the Lazarus Project is the only team using a *mobile* imaging lab to collaborate with institutions all over the globe.

I got the chance to observe the imaging process towards the end of my stay. Professor Heyworth became aware of an object that had been acquired by the United States Holocaust Memorial Museum in Washington D.C., which would offer an ideal subject for multispectral digitization. After

---

[4] A report on the precursor of this digitization effort can be found at Germanisches Nationalmuseum, *Digitalfotografische Erfassung und Erstellung eines 3D-Computermodells des Behaim-Globus* (2012) <https://www.gnm.de/forschung/archiv-forschungsprojekte/digitalisierung-behaim-globus/> [accessed 6 March 2018].

[5] For a complete account and more background cf. for example Reviel Netz and William Noel, *The Archimedes codex: How a Medieval Prayer Book is Revealing the True Genius of Antiquity's Greatest Scientist* (Da Capo Press, 2007).

some negotiations, he and the rest of the Lazarus team travelled to Washington for two days to image the object and provide the museum with both high resolution "beauty shots" (standard photographs) and processed images that recovered the faded text written on the object. Fortunately, it was possible for me to accompany the team on this expedition and to see first-hand the process of multispectral imaging.

Image Processing

My main occupation during the internship was the processing of images the Lazarus team captured in an expedition to the Museum of the Bible in Oklahoma City a few months before my arrival. ENVI, the software used for this purpose by Professor Heyworth and his graduate students, with whom I worked closely, was originally developed to process satellite images for climate research and law enforcement purposes.[6] However, its capabilities have also proven extremely useful for the multispectral analysis of cultural heritage objects and since the field is only just being established, more specialized tools have yet to be developed. The mathematical foundations of how the programme does what it does and why it works still elude me, but fortunately, Professor Easton (RIT) had developed a manual that made the interface easy enough to use and I was able to generate quite a few useful results.

Over the weeks, I even took up my own project. While the Lazarus team has been on many imaging expeditions, not all of the images they captured have been processed to an equal degree yet. To my delight, the images of the Vercelli Book were as yet less explored. In my studies and especially in my work as a student assistant, I have studied and worked with the texts from the Vercelli Book quite extensively; through a course with the Göttingen English Medieval Studies Department, I was even able to go on an excursion to Vercelli and handle the ancient manuscript myself. I was very excited, therefore, when I found out that only the most heavily damaged portions of the manuscript had so far been processed by Professor Heyworth's team and I was allowed to use images of some of the less damaged folios for practice material. Through this work, I was able to expand on the electronic transcription of some of the texts that I and my fellow student assistants had produced the ECHOE research project back in Göttingen.

RAM

In addition to the processing work, I also got to sit in on meetings between Professor Heyworth's team and researchers from University of Rochester's Data Science Department, who were collaborating on a project provisionally called RAM, Reading Ancient Manuscripts. The main aim of RAM (which was only started in 2017 and is currently ongoing) is to create a computer algorithm capable of transcribing medieval manuscripts accurately. While such optical character recognition (OCR) already exists for printed texts (with varying degrees of accuracy), the idiosyncratic handwriting of the Middle Ages has proven to be a far greater challenge. There are several groups of experts all over the world currently working on OCR for medieval manuscripts, but the Rochester team's approach seems to be yielding some promising early results.

Observing this interdisciplinary collaboration was inspiring on several levels. For one, the emerging technology has the potential to cause a paradigm shift in the humanities. Depending on how finely the data scientists can tune their algorithm, it might be able not just to transcribe texts more accurately than a human scholar could, but also to detect similarities and relations between styles of handwriting. If this functionality can be implemented successfully, it could potentially revolutionize the discipline of palaeography by making it much possible to calculate the likelihood with which two

---

[6] For more information ENVI, cf. Harris Geospatial Solutions, *ENVI: The Leading Geospatial Analytics Software* (2018) <http://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx> [accessed 6 March 2018].

texts were written by the same hand. Extrapolating from this, it might be possible to date and localize new manuscripts based on comparison with an existing database of medieval hands.

Beyond the academic significance of the project itself, though, it was enlightening to witness how much can be gained when researchers from different disciplines come together to exchange information and pool resources. For the data scientists, RAM started out as something of an intellectual exercise, since OCR was developed decades ago and cutting-edge Computer Science research has moved on to other topics. They only realized what impact their project might have when they presented it to us, a group of medievalists, who were overwhelmed by the possible implications such a programme would have for our field. On the other hand, Professor Heyworth and his team are able to provide important input and expertise with regards to medieval handwriting, which can help the programmers avoid pitfalls and find the most suitable subjects for their study. Clearly, this kind of interdisciplinary exchange and collaboration can open new avenues of research and new possibilities for the Humanities as well as Computer Sciences, the scope and ramifications of which would be very hard to gauge, let alone reach in isolation.

Reflection

Being involved in such diverse projects and seeing Professor Heyworth and his team in action has taught me a great many things. Beyond the technical skills for image processing and the theoretical knowledge of multispectral imaging, I gained insight into the state of Digital Humanities in North America and also learned more about effective project management than I ever had before. Watching Professor Heyworth cold call a museum curator because he was interested in imaging one of their artefacts, and convincing her to let the team come in and take images was truly instructional. I was also able to help Professor Heyworth by translating his correspondence with the Germanische Nationalmuseum and other members of the research team in Germany about the Behaim Globe imaging project, which is slated for summer 2018. Getting a feel for his polite but straightforward style of communication, so unlike the often convoluted phrasing used by many German academics, has helped me to improve my own writing and communication immensely.

On a grander scale, this internship has once again proven to me how important the integration of IT into the different fields of the Humanities is going to be in the near future. From what I have been able to discern, this realization has been much more widely accepted in the US than in e.g. Germany, where the Digital Humanities are still facing a lot of resistance. In the US, technology is integrated very widely at the administrative level as a matter of course. Applications for scholarships, the distribution of course materials, and submission of papers all happen electronically, which sets a stark contrast to the more paper-based bureaucracy still pervasive in many areas of German academic life. Digital communication also seems to play a larger role; every one of the many meetings I attended in Rochester, for instance, was joined by at least one person via Skype or Google Hangout. While I have been part of a few international projects during my studies in Germany, important meetings have usually been scheduled so that all participants could be present in person and the suggestion that external team members could naturally skype in has never been raised. This might of course be a peculiarity of the particular projects I have been involved in but the use of Skype for regular meetings stood out to me as an uncommon but extremely useful (and somewhat obvious) practice.

On the academic side, Digital Humanities as a discipline seems to be gaining more traction across the Atlantic than it has so far in the European context. Again, this manifests in different areas. Many academic job vacancies in the US now ask applicants for at least some basic knowledge of, if not concrete research in DH. Likewise, scholarships and funding opportunities for specifically digital projects are more common in North America. Germany and other European countries are starting to see an increase in Digital Humanities courses, conferences, and funding, but there is still much more

pushback against the emerging discipline (starting from the discussion about whether it should even be considered a discipline in its own right).

Some of this reluctance is of course justified. There are important discussions ongoing, or yet to be had about the place of digital methods and subjects in established areas of the Humanities. Most DH scholars are not blind techno-optimists that believe computers will solve all problems and make more traditional approaches to e.g. History or Literary Studies obsolete, but they can sometimes be dismissive of the value of these approaches or too uncritical of their own. As an example, OCR may never reach the sophistication needed to aid palaeographers in their research, let alone replace their expertise entirely. But at the same time, its improvement and systematic implementation could become an important milestone in the digitization of historical archival material on a scale never seen before. It is important, therefore, to find a balance between naïve optimism and technophobia, and after my internship, I am convinced that this balance can be found through greater interdisciplinary exchange and collaboration.

Some Drawbacks

Of course not all experiences in these two months could be positive. One stumbling block I encountered throughout my time with Lazarus resulted from working at the cutting edge of a brand new discipline: there was no standard literature or reference material yet, covering the theory of multispectral imaging or the basics of image processing. In my explorations, I was therefore limited to whatever instruction the team members had time to give me and if I encountered a problem, I often had to wait for assistance because I had no real manual to refer to. In a similar vein, I found it occasionally difficult to understand the science behind multispectral photography or especially image processing, from just the explanations given by Professor Heyworth and his colleagues. While they were very patient in explaining the more baffling concepts multiple times, it still would have been helpful to have written reference material, which I could then have dissected in my own time with the help of google and an introductory textbook to Physics.

Additionally, the projects I was involved in over the summer were all on a rather long time scale. While I was able to participate in all individual steps from taking the multispectral images, to processing them, to helping a doctoral student prepare some of her final results for presentation at a conference, I was not able to observe a single project from start to finish. My work with the Vercelli images came closest to a complete task, as I was able to set my own goals and follow through on them, but eight weeks were not enough time to get a grasp of any of the Lazarus team's many projects in full detail. This was a particular shame with regard to RAM because I found its interdisciplinary endeavour highly compelling and would have enjoyed being part of its further development. Still, the two months provided me a very good overview of the work done by the Lazarus Project and gave me a solid idea of what research in this specific academic context looks like.

Personal Summary

During my time studying English, History, and Medieval Studies at Göttingen and Oxford (where I did a one year Master of Studies as a year abroad), I have rarely felt as much part of cutting edge research as I did in my two months in Rochester. Many of the topics my studies have focused on previously were in areas that are either well-explored and only offer room for small innovations, or have fallen out of favour with current trends of research and are therefore not part of mainstream discussion. The opportunity to experience the beginning of a brand new field of study was invaluable and has proven to me the potential for innovative thinking and research inherent in the humanities – both on the well-trodden paths and off.

Partially thanks to this insight, I am increasingly sure that I see my academic future in the Digital Humanities. My focus will likely shift away from the Middle Ages more towards the Early Modern period, but the internship in Rochester has shown me how much I enjoy implementing technological

solutions to old questions. Especially the vision of OCR for historical documents offered by RAM and similar initiatives intrigues me and I plan to search out specialists in this area when it comes time to find a PhD programme. As mentioned above, I am wary of excessive techno-optimism, but I am truly convinced that we have not yet exhausted the possibilities offered by new developments in AI and supercomputing and after getting a glimpse of what the future *might* hold, I am determined to keep exploring this area.

All in all, I am incredibly grateful to have had this opportunity, and consider myself extremely lucky for all the support that made it possible, from Professor Heyworth's generous invitation to being awarded a PROMOS scholarship, to my amazing hosts in Rochester. I know that I will not return to the University of Rochester for my PhD and likely will not pursue multispectral imaging as a focus of my future academic career, but all the insights and experiences I gained through this internship have been invaluable and I would not want to miss them for the world.

References

Germanisches Nationalmuseum, *Digitalfotografische Erfassung und Erstellung eines 3D-Computermodells des Behaim-Globus* (2012) <https://www.gnm.de/forschung/archiv-forschungsprojekte/digitalisierung-behaim-globus/> [accessed 6 March 2018]

Harris Geospatial Solutions, *ENVI: The Leading Geospatial Analytics Software* (2018) <http://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx> [accessed 6 March 2018]

Heyworth, Gregory, *Gregory Heyworth: How I'm Discovering the Secrets of Ancient Texts* (2015) <https://www.ted.com/talks/gregory_heyworth_how_i_m_discovering_the_secrets_of_ancient_texts> [accessed 6 March 2018]

—— *The Lazarus Project: A Multi-Spectral Imaging Project* (2017) <https://lazarusprojectimaging.com> [accessed 6 March 2018]

Netz, Reviel, and William Noel, *The Archimedes codex: How a Medieval Prayer Book is Revealing the True Genius of Antiquity's Greatest Scientist* (Da Capo Press, 2007)

Rochester Cultural Heritage Imaging, Visualization and Education, *R-CHIVE: Imaging a Future for the Past* (2018) <https://r-chive.org> [accessed 6 March 2018]