Title: Digital Humanities Integration and Management Challenges in Advanced Imaging across Institutions and Technologies

Abstract: (998 words)

This rapid development and testing project brought together international partners, scholars and collections in an exploratory, pilot effort from November 2015 to March 2017. The international, multidisciplinary team demonstrated that nondestructive digital imaging techniques and technologies have potential to make texts visible in Egyptian Ptolemaic papyrus mummy mask cartonnages. A major challenge in working across the different technologies, disciplines and institutions was integrating data from diverse technical imaging systems and work processes, requiring new and proven digital humanities data management capabilities.

Before this project, other scholars destroyed the masks to access the papyri, denying future researcher access to the primary historical artefacts. It capitalized on digital humanities skills and data management techniques in assessing the integration of non-destructive digital imaging technologies to make texts visible in layers of papyrus in mummy cartonnages for open research and analysis. Intermediate goals, such as detecting the presence of text, also proved valuable in highlighting the destructive techniques used to study mummy masks and offering scientifically valid approaches for documenting the initial state of objects and their production for future research.

A global team pulled together expertise from science and the humanities, including: digital humanities, Egyptology and papyrology, medicine, dentistry, particle physics, imaging science, data and project management, and systems engineering. Team members rapidly implemented a phased and agile approach at multiple institutions to develop and apply increasingly complex imaging, processing and data integration techniques to penetrate the paint and papyrus layers in mummy cartonnage and host all data online.

**Data Integration**

Project data integration was dependent on common data and metadata standards for ease of image correlation and integration, as well as effective data and project management across disciplines, technologies and institutions. All the different imaging modalities (Multispectral imaging, X-ray fluorescence, Optical Coherence Tomography, X-ray microCT, Terahertz and others) yielded very different data sets from each technology and institution. Integration of images from multiple imaging sources offered potential to apply the strengths of multiple imaging techniques for ease of visualization by scholars and curators. Integrating data from a variety of equipment required significant planning and collaboration across institutions and disciplines. This required streamlined standardization processes and/or more time and resources to devote to this part of a program.

The integration of data and work processes from a variety of scientific tools, disciplines and institutions required storage, dissemination, and searchable access to data from instruments that provide output in different formats, some of which were unique to the research methods and disciplines. While common standards and processes across institutions were encouraged, this was difficult with data and standards from technologies as diverse as nuclear synchrotrons and optical cameras. In addition, many contributors to this project volunteered their time and equipment for imaging and basic processing, but had limited time to spare from their day-to-day

responsibilities – ranging from medical personnel preventing blindness to particle physicists studying elemental changes in bone formation.

## Data Storage and Management

The approximately 300 Gb of data products– including images, individual reports, captured and processed data sets, analytical data and metadata– are now freely available online at https://www.ucl.ac.uk/dh/projects/deepimaging/data. This data set comprises a core content set of digital images, analytical data and technical reports on the imaging and analysis of mummy mask cartonnage and modern surrogates from the multiple imaging institutions. UCLDH also established a project website to host the project information and data at https://www.ucl.ac.uk/dh/projects/deepimaging for scholars, scientists and the public.

Collecting, organizing and hosting data with appropriate metadata from multiple institutions and systems around the globe proved to be a complex problem. This included providing access to and sharing of timely, complete, and relevant data during the project. This was due to both different data collection standards and the wide range of output from proprietary equipment. A key strength of this program was all institutions agreed to make all data freely available under Creative Commons license. This allowed the free exchange of all data for digital processing, analysis and research.

The data structures of the Archimedes and Galen Palimpsests and the University of Pennsylvania's OPenn served as models, but had to be adapted to include the various types of data sets for each image and data collection modality. To support scientific data integration, the team also used the Library of Congress CLASS-D data model. Some adjustments were needed to previous flat file access protocols to make the data product more accessible to users and future researchers. As an example, the large captured multispectral data sets were put in separate folders from the processed images, with the former available for follow-on digital processing and research, and the latter available for immediate visualization of our findings produced with current processing tools.

The need for quality assurance to verify and validate the data proved important. Once the data was integrated, some type of feedback mechanism was needed to validate and check the data against other data in collaboration with the collector as part of collaborative research. This highlighted the value of the data in conjunction with other data, with feedback on the efficiency and quality of the data and its reproducibility as initially structured and standardized. This significantly improved data sharing and preservation across the research team.

## Conclusions

Effective data management, integration and technical support are critical enablers in any broad digital research program to ensure data availability for follow-on research, even those (like this one) with a limited budget. The ability of imaging equipment to produce a standard data output with relative ease of use by the operator and researcher is important to the visualization, storage of and access to the data. Standardized procedures and data output better allow independent imaging of the same object with multiple technologies, with subsequent integration of data to leverage the strengths of each technology and technique.

Additional research could optimize various imaging technologies to help identify new imaging methods and support their development into useful and efficient tools for future imaging applications. In addition to text detection, the portability, availability and cost of the imaging system must be considered in determining its utility.