

6. Data Management Plan

We are committed to releasing the software and data resulting from this research freely to the community. As described in our proposal, we will launch a cyberenvironment based on the current PARAGON (<http://www.cdh.sc.edu/Paragon>) site, in which we have released a prototype of PARAGON for image collation resulting from Sapheos project. With the progress of this project, we plan to release a new updated version of PARAGON, including source code and installation guide, at least once a year in this cyberenvironment. At the end of this project, we will also release all our source code and documents under an open source license through a page on SourceForge, encouraging other researchers to use our work, benefit from our investment of resources, and alter our code to extend the benefits of our research. We will release our software in versions for different operating systems, including Linux/Unix, Windows and Mac OS.

The core testing datasets used in this project will be provided by our project partners, including the Spenser Archive and Shakespeare Quartos. They provide online access to a large collection of the high-resolution digital scans of the early manuscripts and prints. We will design convenient and compatible interfaces in our PARAGON software to easily and efficiently access, retrieve and process these online collections. Such an interface can be re-used by other researchers to evaluate their own image-collation tools on these data collections.

We will also archive and release the image-collation results from the test data. Specifically, we will develop and implement a metric to quantitatively evaluate the performance on a large collection of image data. We will release the source code of this metric in the PARAGON cyberenvironment so that other researchers can evaluate their own methods quantitatively on a same data set and conduct an apples-to-apples performance comparison to justify the strength and weakness of each method.

We are committed to maintain and update this cyberenvironment after the completion of this project. As mentioned earlier, updated information such as papers submitted and project development plan will be included in the portal. We will also embed modules in the portal that

allow activities tracking (provenance data), collaboration support (including blogs, chat, and wikis), and networking support.