

NEH Digital Humanities Advancement Grant**Project Title: Building a Decision Tree for Watermark Identification in Rembrandt's Etchings—The WIRE Project****Data Management Plan****Data to be Generated**

- **Computer code for interrogative watermarks decision tree.** The key data resulting from this project will be in the form of a complete, interrogatory computer decision tree, with coded branches for each of the 54 known types of Rembrandt watermarks and the approximately 500 individual subvariants within these types (with the addition of newly discovered watermarks as they arise). In addition to the decision tree's usefulness to researchers, the value added in data terms will be the individually observed verbal differentiations among these subvariants that identify them as unique, a set of data currently absent from the literature on this topic.
- **Image datasets.** Project participants are currently working with an image dataset supplied by Dr. Erik Hinterding, which includes one digitized radiograph image for every subvariant in the full decision tree, in a combination of TIFF and JPG formats. As new watermarks are discovered, they will be imaged by project staff or institutional partners, and added to the tree. The future addition of further images from museums and other archival sources is anticipated, but will not be enacted until this project is complete and image permissions are negotiated.
- **Algorithms.** Project collaborators Dr. Vikram Krishnamurthy, professor of electrical and computer engineering, Cornell Tech, and his PhD student Sujay Bhatt are developing algorithms in collaboration with project co-director Dr. C. Richard Johnson that will use decision tables extracted from the constructed decision trees to determine potentially alternative branch configurations of decision sequences to minimize the number of questions needed to reach the variant end-points. Algorithms will also be sought to 1) expand the decision trees to accommodate newly discovered watermark types, and 2) allow the tightest group classifications for images of watermark fragments. This research and the specific algorithms devised (typically programmed in Matlab) will be openly presented in the form of web postings (on the websites of the Johnson Museum and of the algorithm creators) and conference and journal papers, and will thus be accessible to all. Separate funding will be sought to accelerate algorithm development activities.
- **White paper.** Upon completion of the decision tree construction phase of the project. This white paper will be made available via the website of the Johnson Museum.
- **Final report to NEH.** When the project is complete, in compliance with NEH procedure. This report will also be made available via the website of the Johnson Museum.

Period of Data Retention

All computer code for the Rembrandt watermark decision tree for all 54 watermark types will be made freely and publicly available within six months of completion of the project end date, and will be maintained through Cornell University/the Johnson Museum through at least the year 2021. Dr. Johnson is now in exploratory discussions with the RKD (The Netherlands Institute for Art History) in The Hague about the possibility of an eventual transfer of archival management of the data, where it may fit most logically alongside the RKD's large store of related data on seventeenth-century Dutch art [see the example of the RKD's Rembrandt Research database: <https://rkd.nl/en/collections/other-web-services/rembrandt-db>].

Data Formats and Dissemination

All computer code (written in HTML, CSS, and Java) will be made freely and publicly available within six months of completion of the project via GitHub or equivalent third-party repository for open-source code. Also within six months following completion, a fully illustrated PDF instruction manual for the creation and coding of decision trees will be made available as a free download via the Johnson Museum website.

Data Management and Maintenance

All computer code, including embedded image files used for differentiation of watermark types, will be stored during the project period on the file server used by the Johnson Museum, which is maintained by the College of Arts and Sciences, Cornell University, and backed up daily by Cornell University Information Technologies. The code will be maintained and updated by project team staff and ultimately transferred to GitHub or an equivalent third-party repository.