The Making and Knowing Project

Data Management Plan

This proposal has no significant digital products; nevertheless, data management is a responsibility upheld by all participants of the Research Workshop and overseen by the Making and Knowing Project (PI Smith). The methodologies, decisions, and data management strategies described below have been developed by the Making and Knowing Project in the creation of *Secrets of Craft and Nature in Renaissance France. A Digital Critical Edition of BnF Ms. Fr. 640*. These principles, influenced by the culture of "minimal computing" at Columbia University, guide all content creation and digital development. All decisions and modes of work are informed by two guiding principles: to generate and preserve all project data in sustainable and open formats while best facilitating the multifaceted collaborative research process; and to ensure commitment to open access and open source standards wherever possible. Data management will comply with Columbia policies for Research and Data Integrity as well as guidelines and standards for the social sciences and humanities, including the *Sustaining Digital Resources: A Briefing Paper for Leaders of Projects with Scholarly Content* and the *Socio-Technical Sustainability Roadmap*.

Because digital development has taken place in tandem with content development for the Making and Knowing Project, data is managed in the following contexts: 1) internal-facing active "working" documents/data, 2) publicly-accessible resources provided provisionally that may require maintenance, 3) stable and persistent (static) public resources, and 4) long-term archival and open access repositories.

1) Active "working" data: Many working documents and other assets are created and edited in proprietary platforms (Google Drive, Github) prior to being regularly transformed and exported in open standard formats to local archival backups as well as strictly-controlled cloud-based servers. Collaborative editing takes place largely via Google Drive and its application suite, chosen for their low entry-barrier to simultaneous, collaborative, and internationally distributed digital collaboration. This advantage offers sufficient benefits to counter the disadvantage of proprietary technology. Access is free for all contributors, with editing permissions and protocols controlled and regularly reviewed by the M&K Director, Digital Lead, and Assistant Director. Software development and online content is largely refined, compiled, and maintained in public Github repositories administered by this core team. Metadata, preliminary results of textual analysis, and project-created tools for data management and manipulation are also found in Github. Photographs, videos, and audio files are uploaded to Flickr, YouTube, and Vimeo. These public online storage sites also serve as tools for public engagement and dissemination. All content (from Google Drive, Flickr, Youtube, Vimeo, Github) is regularly downloaded for backup to local hard disk storage as well as deposited in Amazon Web Services S3 cloud storage. In order to avoid dependency on proprietary systems, M&K has developed procedures to routinely export and convert content to open standard formats (e.g., XML, TIFF, JSON, CSV), ensuring the data is fit for reuse. Exported and converted content constitutes a rich dataset that can be readily manipulated in the research process as well as in the final development of any public web-based representations.

2) Select public resources that require maintenance: Some digital scholarly content that is useful or compelling cannot adequately conform to minimal computing standards as they utilize

features beyond the limited corpus. Resources of this kind are regarded as "extended features," with an understanding that they may obsolesce without active maintenance (e.g., analytical workspaces using Voyant or Jupyter Notebooks). We are committed to limiting the number of resources requiring maintenance as well as maintaining the active ones for as long as possible. No core or critical work and data will be presented or preserved in this form. Any data underlying extended features will be stored and maintained for the long-term, and made publicly available (through Github, for example) so that similar work can be recreated.

3) Stable and persistent (static) public resources: Development decisions have prioritized the minimization of server-side technology in order to reduce maintenance overhead, security risks, and exposure to technological obsolescence. Resources rely upon a common and well-established technology stack: HTML/HTML5, with CSS and JavaScript served statically through a webserver. This simplified technology stack will more likely resist obsolescence, and reduce the need for resource expenditure on active maintenance to address issues such as security vulnerabilities or version upgrades of components and their dependencies, thereby ensuring its long-term usability. Until public release, infrastructure and content assets are developed and maintained using Apache HTTP, React JavaScript Library, GitHub, Amazon Web Services CloudFront and S3, and DigitalOcean virtual servers. All core work, including infrastructure and content assets and numerous public-facing core datasets, resources, raw data, and additional fundamental resources, are (or will be) available in public Github repositories, and, where appropriate, uploaded to repositories such as [Academic Commons](#), [Internet Archive](#), and [Zenodo](#).

4) Long-term archival and open-access repositories: The final and most complete forms of the research data, metadata, and media assets will be deposited for long-term archiving in Columbia University Libraries Fedora-based digital repository. The repository will manage these archival forms of the files stored in the Libraries' locally administered storage infrastructure (with content replication at a data center in Syracuse and tape storage at Indiana University) and in consortium-based preservation repositories such as the Academic Preservation Trust. All publication assets will be deposited and maintained by the Libraries under University and scholarly guidelines for long-term preservation. The Libraries have committed to hosting, serving, archiving, and maintaining this work.

Accessibility
The core team strives to ensure that all products, platforms, and methodology are accessible to individuals with disabilities. Columbia websites follow [WCAG 2.0 AA](#) standards, which together provide accommodations for people with cognitive or physical disabilities, impaired sight, and impaired hearing. Products and cloud services will be developed using their guidelines for increasing visual accessibility and compatibility with screen readers or other tools for hearing and sight impairments. Columbia offers a range of [guidelines and disability services](#), including disability liaisons for enrolled students, and accommodations for physical disabilities, impaired hearing and sight at workshops and events held at Columbia.