Bentley Historical Library Digital Processing Note

Version 2.0: Effective from December 12, 2013

In processing digitized and born-digital materials, archivists at the Bentley Historical Library adhere to professional best practices to ensure that content is prepared for long-term preservation and that end-users have full assurance of the authenticity and integrity of digital files. These procedures include the following:

- **Virus scan**: all content is scanned for viruses or malware with the most recently available antivirus definitions.
- **Extraction of content from archive files (*.zip, *.tar, etc.)**: the identification and extraction of content from archive files ensures that this material will be appraised and subjected to the procedures as listed in this document.
- **Identification of missing file extensions**: files with missing or user-defined extensions are identified and an attempt is made to identify correct extensions to facilitate patron access to materials. No changes will be made if the results of file format analysis are ambiguous or inconclusive.
- **File format conversion**: files in proprietary or at-risk formats are identified and an additional copy is made in a sustainable preservation format.
    - Preservation copies created by the library are identified by a suffix that includes the letters "bhl" and a CRC32 checksum derived from the original file (i.e. concert_recording-bhl_2j72ea91.wav).
    - Preservation copies are stored alongside the original versions of files.
    - In some cases, access copies of material will be made so that patrons may more easily download and use content.
    - More information on at-risk and sustainable formats as well as Bentley Library conversion strategies are available at http://hdl.handle.net/2027.42/93307.
- **Personally Identifiable Information (PII) scan**: a best-faith effort is made to scan all material with Identify Finder software for credit card numbers, bank account numbers social security numbers, and passwords.
    - Archivists will redact this information from files or, if redaction is not possible, place the material under an appropriate restriction.
    - This procedure aligns with the University's guideline "Sensitive Regulated Data: Permitted and Restricted Uses" available at http://cio.umich.edu/policy/sensitive-regulated-data.php.
- **Technical metadata extraction**: the UK National Archives' DROID (Digital Record Object Identification) software is employed to extract technical metadata about all files in a deposit .
- **Packaging of content for long-term storage**: to streamline the storage, management, and access of digital content, archivists may package content in uncompressed ZIP files based upon the provenance, original order, and/or intellectual arrangement of the materials.
- **Checksum calculation**: MD5 checksums are generated upon initial ingest of content. Additional MD5 checksums are calculated at the completion of digital processing procedures for all individual files as well as ZIP files created by the library. These

checksums document the state of digital materials and will be used to verify the integrity of content over time. For more information on checksums, see http://en.wikipedia.org/wiki/Checksum.

The Bentley Historical Library maintains complete log files of all operations and records all significant events in a PREMIS (PREservation Metadata: Implementation Strategies) document. More information on the library's policies and procedures related to the arrangement, description, storage of and access to digital content may be found on the Digital Curation home page (http://bentley.umich.edu/dchome/index.php).

Disclaimer: Procedures are subject to change as new tools and standards are introduced.