

Hembrough Software Collection Project

Date Started: July 2014

Background:

The Hemborough software collection was donated by Georgine Hembrough, wife of late Gary “Coach” Hembrough. Gary was an educator and coach – teaching physical education and driver’s ed, and coaching football and basketball - for Champaign Unit 4 schools.

Gary maintained many interests. After retiring in 1994 he started G&H Computing Services, “focusing on self-taught computer skills he acquired when coaching football and track.” Gary’s passion for cooking lead to Coaches Cooking Team, a business and competitive BBQ team specialized in homemade food and BBQ. (See Obituary: <http://www.news-gazette.com/obituaries/2012-10-13/gary-hembrough.html>)

Project Overview:

I envision four stages in this project:

- 1.) Initial inventory of the floppy disks and documentation contained in boxes and cabinets in rm. 429
- 2.) Deduplication of redundant software; retaining the highest quality/most complete version of the software
 - a. Definition of highest quality/most complete
 - i. Professionally manufactured software
 - ii. Complete sets of software (no missing install disks)
 - iii. Packages that include manuals/packaging/instructions
 - iv. Software with serial numbers, licensing information.
- 3.) Evaluating value of retaining software
 - a. Comparing to NSLR; Cabrinety collection
 - b. What is the enduring value of the software
 - i. Is it proprietary?
 - ii. Is it difficult to find a usable version of the software?
 - iii. How many files do we have in our archival collections that require the software for data conversion?
- 4.) Transferring software from current media into digital repository
 - a. Test that software is readable
 - b. Determine where the software is usable:
 - i. what operating systems, computing hardware environments are required
 - ii. if the software is installed on a legacy computer do we have a way of importing/exporting the source and migrated files to/from the legacy computer (e.g. does the computer have a USB, networking or some other interface which will allow it to read data in need of migration?)

Project Plan:

Stage 1

1. Sabrina will review all of the software within the cabinets and box on the table, research software titles for complete information (referring first to the Cabrinety collection finding aid) and will enter this information into an Excel spreadsheet.
2. Software is first identified as belonging to one of two categories:
 - a. Manufactured software –software that is manufactured by a software publishing company or software developer. This software will often have professionally pressed labels. The label text should include the software title; it may include version, copyright or license information.
 - b. Copied software – this is software that we believe Gary has copied or backed up for his personal use. Copied software will often have a black and white label, low graphics or no graphics, or may be handwritten.
3. After identifying the whether the software is manufactured or copied software titles are identified and categorized using a locally created controlled list.
 - a. Note on controlled list: In consultation with a colleague at Stanford University, Charlotte Thai, who is creating metadata for the Steven M. Cabrinety collection of historic micro computing software in partnership with the National Institute of Standards and Technology (NIST), it was determined that no authoritative list of software categories is currently available. Charlotte indicated that NIST has developed a local controlled vocabulary which is in use in the Cabrinety finding aid (see: <http://www.oac.cdlib.org/findaid/ark:/13030/kt529018f2/>).
 - b. The UIUC controlled list of software categories is based on the Cabrinety/NIST categories as well as those established by Gary within his disk labeling schema. The list and examples are provided below:
 - i. Communication software: software that assists in communication
 1. Voice
 2. Text/Chat
 3. Email
 4. Web browsers
 5. Personal information management
 6. File / document sharing
 - ii. Educational software: Software intended to educate or assist in K-12 and college-level learning
 1. Dictionaries
 2. Encyclopedia
 3. Educational games
 4. Training/courseware including sports training
 5. Cooking software
 - iii. Entertainment software: Leisure or hobbyist software

1. Media Player applications
2. Image Viewer applications
3. Cooking and recipe applications
- iv. Graphics software: Software containing pre-created graphics. This category
 1. Clip art programs
 2. Pre-created illustrations which are not productivity suites
 3. Screen savers
- v. Games: Software that is typically interactive; has an element of strategy and may not be exclusively related to learning. Games are differentiated from entertainment primarily for the interest in preserving games within the digital preservation community.
 1. PC games
 2. Console games
- vi. Productivity software:
 1. Software which assist in the creation of documents, posters, flyers, reports, databases and drafting documents
 2. Office Applications – e.g. MS Office
 3. Database Applications – e.g. dBASE
 4. Graphics Production Applications – e.g. Adobe Illustrator
 5. Audio Production Applications
 6. Business productivity applications – e.g. Quicken
 7. Computer aided design applications – e.g. AutoCAD
 8. Font packs
- vii. Utilities software: Software relating to computer maintenance; e.g. computer troubleshooting; database and data recovery; anti-virus and anti-malware software such as Norton Antivirus
 1. Hardware drivers
 2. Antivirus/Malware
 3. Operating systems utilities
- viii. Magazines: Software that contains publications (added 4/15/2019)
 1. Magazettes or “magazine on diskette”
 2. Monthly publications
- ix. Development software: Computer software/programming development environments
 1. Integrated Development Environment (IDE)
 2. Programming language applications – e.g. C++, J++ developing
- x. Operating Systems (O/S) software: Software related to running the computing environment
 1. Operating systems applications – e.g. MS Windows
- xi. Other categories
 1. Other categories may be added to the list; however, software titles should be researched to ensure that a new category is required.

4. Disk Labeling and Organization:
 - b. **Professionally manufactured** disks will not be labeled with Post-it Notes
 - c. Disks that are **not professionally manufactured** will be labeled with Post-it Notes, in order of their placement in the cabinets:\
 - i. Disks that appear to contain copied software will be labeled *Soft_0001* and so forth with Post-Its.
 1. All Post-It note label numbers will be copied into the spreadsheet on the “Copied Software” page, in the “PreservationDiskLabel” field
 2. Copied software may have multiple software titles included on a single disk. If this is the case, list each software title individually tying the software back to the source disk by referencing the “PreservationDiskLabel” information.
 - a. Disks that appear to contain **personal material** or that have **no labels, illegible labels, or indecipherable labels** will be pulled from the cabinet and set aside for evaluation and/or discarding.

5. Software Research:
 - a. Sabrina will review and research each software title to complete the fields within the Excel spreadsheet. First point of reference: the Cabrinety finding aid. This finding aid may have complete information about the software. If the software title is found in the Cabrinety finding aid indicate this on the Excel spreadsheet. (e.g. InCabrinetyCollection field enter yes.)
 - b. If the software is not found in the Cabrinety finding aid, use Google to research the title. Although some information may not be available for a particular software title please be as complete as possible and reference several sources.

6. Data Entry:
 - a. Enter data into the 2014_softwareCollectionInitialTriage.xlsx spreadsheet. This spreadsheet should be centrally located in the Box.com cloud service: <https://uofi.app.box.com/files/0/f/1398762752>
 - b. Download the spreadsheet from Box each time you are editing.
 - c. 2014_softwareCollectionInitialTriage.xlsx field reference
 - i. **ApplicationName: Required.** Title of the software or application (e.g. MS Office 2013). If multiple software titles are on a disk enter each application name on a new line in the spreadsheet linking the titles back to a single disk through the **PreservationDiskLabel field** (e.g. disk label includes software1 and software2:)

Software1	soft_0001
Software2	soft_0002
 - ii. **PreservationDiskLabel: Required.** Post-It note applied to disk by Preservation Unit for internal identification.
 - iii. **Number of Disks: Required.** Total number of media included in the application package/software install package. Look for signifiers such as disk

1 or disk 3 of 4; some indicator that there are multiple pieces of media in a series of disks. This may be more difficult with the hand-labeled disks. Please review hand-labeled disks for potential inclusion in a series before entering data.

- iv. **Software Developer: Not Required.** The manufacturer or individual that developed the software code; this may be the same as the publisher.
- v. **Software Publisher: Not Required.** A **software publisher** is a [publishing company](#) in the [software industry](#) between the [developer](#) and the [distributor](#). In some companies, two or all three of these roles may be combined (and indeed, may reside in a single person, especially in the case of [shareware](#)).

Software publishers often license software from the original author-developers with specific limitations, such as a time limit or geographical region for a royalty consideration. The terms of licensing vary enormously, and are typically secret.(from:

http://en.wikipedia.org/wiki/Software_publisher)

- vi. **Software Version: Not Required but highly encouraged.** Number or name representing a new state of software within the same product family (e.g. MicroSoft Windows XP, MicroSoft Windows 8).
- vii. **Copyright Year: Not Required.** The year the software was published and registered with the US Copyright office.

viii. **TO BE CONTINUED**

- d. Once data is entered into the spreadsheet upload it back to Box.

7. Work Notes:

- a. Sabrina will keep a daily worklog in the Box.com account indicating work done and challenges encountered.

Background Reading and Reference:

- b. Read about the Cabrinety project at Stanford:
<http://web.stanford.edu/group/htgg/cgi-bin/drupal/>
- c. National Software Reference Library: <http://www.nsrll.nist.gov/new.html>
- d. Software Preservation Society FAQ:
- e. <http://www.softpres.org/faq>
- f. Video Game preservation at scale An Interview with Henry Lowood:
<http://blogs.loc.gov/digitalpreservation/2013/02/video-game-preservation-at-scale-an-interview-with-henry-lowood/>
- g. National Software Reference Library: <http://www.nsrll.nist.gov/new.html>