

Processing Internal Hard Drives

Amy F. Brown

Ohio State

Chloë Edwards

Mississippi Department of Archives and History

Meg Eastwood

Northern Arizona University

Martha Tenney

Barnard College

Kevin O'Donnell

University of Texas at Austin

As archives receive born digital materials more and more frequently, the challenge of dealing with a variety of hardware and formats is becoming omnipresent. This paper outlines a case study that provides a practical, step-by-step guide to archiving files on legacy hard drives dating from the early 1990s to the mid-2000s. The project used a digital forensics approach to provide access to the contents of the hard drives without compromising the integrity of the files. Relying largely on open source software, the project imaged each hard drive in its entirety, then identified folders and individual files of potential high use for upload to the University of Texas Digital Repository. The project also experimented with data visualizations in order to provide researchers who would not have access to the full disk images—a sense of the contents and context of the full drives. The greatest challenge philosophically was answering the question of

whether scholars should be able to view deleted materials on the drives that donors may not have realized were accessible.

Introduction

Project introduction

This case study is the result of a project undertaken in Dr. Patricia Galloway's Digital Archiving and Preservation course at the University of Texas at Austin's School of Information in the spring semester of 2012. The authors were assigned the semester-long project of recovering the content of eleven internal hard drives in two collections held by the Dolph Briscoe Center for American History (BCAH), and preparing those materials for ingest into the BCAH's space on the University of Texas Digital Repository (UTDR). None of the hard drives had been previously examined by the BCAH's digital archivist.

Six hard drives came from the George Sanger Collection, and the remaining five from the Nuclear Control Institute Records. The hard drives from the Sanger collection form a small part of an extremely diverse collection comprised of video game audio files, paper records, and email backups from Sanger's work as a video game music creator. The Nuclear Control Institute (NCI) materials are also part of a larger collection consisting mainly of the Institute's and founder Paul Leventhal's paper records, but that also includes other digital material such as NCI's website. NCI, founded in 1981 by Leventhal, was a research and advocacy center for the prevention of nuclear proliferation and nuclear terrorism worldwide.

Project goals

The goal of the project was to provide researcher access to the archival content of the drives at three levels: the entire drive, the folder level, and the file level. Digital forensics methods were used to ensure that the integrity of the drives would be preserved at all times. Digital forensics, developed in law enforcement to examine digital evidence such as desktop and laptop computers and various storage mediaⁱ, was

particularly well-suited for this project given that it dealt with the same kind of material—internal hard drives—for which the field was developed. Another decision made early in the project was to use open source software whenever possible in order to both test the open source software that was available and to replicate the experience of working in a repository without extra funds for incidental software. Furthermore, open source software, by definition, makes source code available to users, thus improving transparency in the project and reducing the risk of the software doing something unknown to the files in question.

Archiving the contents of the drives at three levels meant that the project took place in three distinct phases. In the first phase, the team imaged each drive in its entirety for storage in the BCAH's dark archive, and then experimented with visualization software to provide researchers with a surrogate for the actual disk image, which would not be publicly available due to ethical concerns, detailed further below. Visualizations of the entire drives would also provide context for the material that was the focus of the second and third phases of the project. The second phase, providing access to individual folders, focused on finding the best method to create partial images of the folders determined to be of potentially high use. The third phase, aimed at providing access to individual files, consisted of preparing and running a test batch ingest to upload the 465 selected files to the UTDR.

The next section details the process of imaging the drives and experimenting with visualization software. Section three describes the search for an appropriate method of partial imaging for selected folders. Section four details the batch ingest process, followed by the conclusion. Section five concludes.

Phase 1: Providing drive-level access

Collection inventory

Before any disk imaging could occur, the drives were physically inventoried and assigned unique identifiers. The drives' storage size, physical size, manufacturer, model

or series name and number, date of manufacture (in some cases an exact date from the hard drive label, in some cases approximated from the type of drive), other identifying numbers such as product numbers, serial number, computer brand of origin (if known), operating system (if known), any creator labels, and any other parts that came with the drive, such as jumper shunts and parallel ATA cables, were all recorded in a spreadsheet ⁱⁱ.

The NCI portion of the collection consisted of five internal hard drives in the custody of the BCAH. All were extracted from founder Paul Leventhal's working computers. Two of the drives were dated confidently to 2002 and 2000 (NCI 1 and 5, respectively); exact dates for NCI 2-4 were undetermined, but creation dates for materials on the drives indicated a range of dates from the mid-1990s to the early 2000s. Hard drive size ranged from 53 megabytes to 40 gigabytes. Materials on the drives were extremely variable. Like any internal drive, much of the space was taken up by operating system files. The materials in the Sanger hard drive collection consisted of six internal hard drives from computers utilized by Sanger throughout his career dating from 1998 to 2004.

Imaging the drives

Having taken a complete physical inventory, the next step was to take a full image of each of the drives using digital forensics software. As stated above, using digital forensic methodology, disk imaging in particular, made sense for this project. Imaging creates an exact bit-for-bit copy of the item being imaged, whether an entire hard drive or a portable flash drive. After an image is taken, it is no longer necessary to work from the original hardware, thus easing wear and tear on fragile artifacts, and once verified using checksums or hash values, the image can be mounted as a read-only drive for in-depth examination of its contents with the assurance that no changes have been made to the data. The read-only image can also be used as a test-bed for whatever other procedures are required to fully process the collectionⁱⁱⁱ.

In creating its images, the group followed the basic capture workflow laid out by John in his 2008 article on archiving digital personal papers: ensuring “(i) audit trail; (ii) write-protection; and (iii) forensic ‘imaging’, with hash values created for disk and files...(iv) examination and consideration by curators (and originators), with filtering and searching; (v) export and replication of files; (vi) file conversion for interoperability; and (vii) indexing and metadata extraction and compilation ^{iv}.” The group did deviate slightly from this workflow in that the sheer number of file types present on the drives prevented the conversion of all the individual file conversions.

To create the images, the group used AccessData FTK Imager software (version 3.1.0.1514) on the Forensic Recovery of Evidence Device (FRED) suite held by the School of Information’s Digital Archaeology lab ^v. FTK Imager is a free download version of the proprietary forensic imaging software FTK Toolkit and comes loaded on FRED’s laptop. Although the Imager suite does not contain the same functionality of the Toolkit, we found that it was more than sufficient for our needs, as well as allowing us to capture disk images without the inconvenience of working through the command line. It is worth noting that since the completion of this project, significant progress has been made on the development of an open source software suite called BitCurator, which is aimed at facilitating digital forensics workflows specifically for archives^{vi}.

The use of FTK Imager catalyzed the project’s greatest philosophical challenge: the proper handling of deleted files. Because FTK Imager is forensic software, it is calibrated to show deleted files on the hard drive, which are denoted by a small red X on the folder icon. Although these deleted files were included on the full image by default, given that the donors were almost certainly unaware that we could access those files and by deleting them, showed their intent not to donate those files to the BCAH, the full images will not be released to the public but will remain as security copies in the Briscoe’s dark archive against damage to the original hardware. Kirschenbaum, Ovenden and Redwine discuss the ethics of handling deleted files in detail in their section three of their CLIR report on digital forensics^{vii}. Making sensitive materials such as passwords, Internet search queries, and health information available for public use could result in a

privacy invasion for donors. This concern is only increased for materials the donor may not even be aware are present on a given piece of hardware. Kirschenbaum, Ovenden and Redwine recommend that going forward, deleted files be addressed with the donor before any donation has been made, thus making the disposition of deleted files an explicit part of the donor agreement. Unfortunately, the project group had no such agreement and could not reach either of the donors for clarification. As a result, the group decided the potential damage of making these materials available far outweighed the benefits, and deleted files were not included in any publicly accessible materials.

Philosophical issues aside, the actual process of imaging was very simple, a matter of selecting “Capture Image” and filling in the appropriate metadata and destination for the completed image (in our case, a 500 GB external hard drive, also connected to FRED. Getting the drives to turn on, however, was not so simple, and on our first attempt, only two of the eleven hard drives, both from the NCI collection, were successfully viewed and imaged. A problem arose when our remaining nine drives were not detected by the write-blocker, which team member O’Donnell soon determined was due to incorrectly placed jumper shunts, plastic pieces whose placement indicates whether the drive is serving as a C: drive or a subsidiary in a multi-drive system. After moving the jumper shunt(s), to the Cable Select position, we had no trouble viewing and imaging the contents of the other nine drives.

Appraisal

The contents of the drives were appraised using a file directory exported as a .csv file by FTK. An examination of these files indicated that most of the Sanger drives did not contain any material that would expand the BCAH’s already extensive collection of Sanger materials, so the team’s priorities switched to the NCI hard drives. The highest priorities for ingest were determined to be drives NCI 1, 3 and 4, whose contents consisted largely of .doc files relating to Leventhal’s activities as a lobbyist and activist for nuclear awareness and disarmament, including research materials, reports, event

and conference planning, and correspondence. From these three drives, the individual folders that were to be ingested for the second level of access were selected.

The FTK Imager software is also capable of mounting the full image as a local disk. To perform operations such as virus scans or visualizations of an image, we mounted the disk image as a physical and logical image via the “Block Device/Read Only” mount method. This could be performed on any computer. Using this method, virus scans, an essential step prior to uploading any files to a digital repository, were performed using AVG Anti-Virus Free Edition 2012. Viruses were detected on several of the drives (most often in the temporary Internet files), and the results of the virus scan for each drive were ultimately uploaded to UTDR in each hard drive’s respective collection. No attempt was made to disinfect the files as the materials selected for ingest were not infected^{viii}.

Experimenting with metadata and visualization tools

Having completed the imaging process, copies of the full disk images of each drive were turned over to the BCAH for storage as security copies in its dark archives. The full disk images would not be made accessible to researchers due both to the ethical concerns about the inclusion of deleted files in the images discussed above, and more practically, to the images’ prohibitively large sizes. Since the full drive structure would not be publicly visible, the group elected to provide another representation of the full drives as surrogates for the disk images that could not be included with the individual folders and files selected for ingest. A similar problem in terms of providing context for individual folders and files has been addressed at Emory University, which provides access to Salman Rushdie’s digital archives via an emulator so that digital files are not seen solely as file paths^{ix}.

Initially, the group attempted to use archival open source software to determine the file types present on the drives, intending to upload the resulting reports to the UTDR. This proved problematic. Although the process of file format discovery and normalization has been automated in programs like Archivematica, the sheer number of

different file formats present on these hard drives—too many for Archivematica to recognize and migrate—precluded its use for this project. An attempt at using DROID (Digital Record Object Identification, a software tool developed by the UK National Archives) to identify our file formats fizzled because the report was too granular to be of any practical use given the number of file types present.

As a result of these difficulties, the group turned to visualization software that would graphically display the file formats within the directory structure. Visualizations would be able to show the percentage of each file format on the entire hard drive or in a single folder from the mounted image, as well as build file distribution tree maps and pie charts showing types of files. They would also provide more information about the context of each drive and provide insight into the way each creator used his drives. For example, a large amount of .doc files would imply that the creator was relying heavily on his computer's word processing capabilities. One note of caution: all of the visualization tools described here are real-time tools. Because of this, the software may recognize file extensions as coming from later programs than those in which they were actually created. For instance, the software will associate all files with the .doc extension with Microsoft Word, in spite of the fact that .doc was first used as a file extension by Word Perfect in the 1980s. Fortunately, these drives were recent enough that this did not cause any problems.

Test driving software: WinDirStat, SpaceSniffer and TreeSize Professional ^x

Maintaining the group's commitment to using open source software, the first visualization program tested was WinDirStat (version 1.1.2.80). In its main window, WinDirStat displays the percentage of the disk each folder occupied (this percentage can be calculated for the entire disk including free space, or only for the portion of the disk that is in use). The percentages window also displays the size of the folder in megabytes, the number of items, files, and subdirectories in each folder, plus the last date changed. Users can navigate down to the item level in each folder to compare files sizes for each individual file. WinDirStat also includes a window that shows the total number of files of

each file extension type, and estimates the total number of megabytes and percentage of the disk occupied by that file type—for instance, the program provides the percentage of files with a .doc extension. The different file extensions are color-coded for representation on a tree map. While WinDirStat is a useful tool for real-time visualization of data, the only way to export any data is via a screenshot. Taking a screenshot of the window that analyzes the directory structure or the file types would be difficult, however, because there is too much data to fit on the screen at one time, meaning that a successful picture would require piecing together multiple screenshot images to save the data. The tree map can be captured in a single screenshot, but the tree map itself contains no explanatory text. Therefore, the map only makes sense when viewed in conjunction with the file type window that shows which colors on the tree map correspond to which file type. For these reasons, the group decided not to use WinDirStat.

The second open source program tested was SpaceSniffer (version 1.1.4.0). SpaceSniffer's main display window is an interactive tree map that allows the viewer to zoom in to any section of the tree map to see file names and sizes; additionally, mousing over a file in the map generates a pop-up display that shows the file's creation, last modified and last accessed dates. The tree map display can also be filtered using a series of commands typed into the "Filter" bar. For example, by typing ">10MB" and hitting enter, the tree map will display only those files that are larger than 10MB; similarly, typing "*.doc" will display only .doc files. SpaceSniffer provides the user with powerful visualization tools, but a novice user might prefer an interface that displays all the possible options as icons rather than having to learn the full set of filtering commands. Also, while the tree map in SpaceSniffer is very informative in real time, once again, the only way to export the map is as a screenshot, and much of the information SpaceSniffer provides is lost in this static environment. In terms of export functions, SpaceSniffer will allow you to export a file report in .txt format that shows the file names and sizes groups under their parent directory. This report was useful, but it did not include as much information as the file directories exported from FTK. Additionally,

the file directories from FTK were in a .csv format that can be opened in Microsoft Excel and sorted based on multiple features. Therefore, the group decided to upload the file directory from FTK to DSpace rather than using the file report from SpaceSniffer.

Ultimately, a program called TreeSize Professional (version 5.5.4) was used to provide graphical visualizations. While TreeSize Professional is not an open source program, by using a thirty-day free trial and experimenting on the image of NCI 3, the smallest drive and one whose contents were to eventually be ingested at the item level, the group created:

- A pie chart showing the disk space allocation by file type, exported as a .png file;
- A tree map showing the disk space allocation by file type exported as a .png file;
- A full report from TreeSize Professional showing all the different types of file formats, arranged in the same hierarchy as the file directory and with the ability to drill down through the folders, exported as an .xml file;
- A date-of-last-change bar chart showing the percentage of files changed a range of number of years ago, exported as a .png file; and
- A spreadsheet of file format extensions, exported as an .xlsx file.

Unfortunately, after trying to create visualizations of the remaining NCI drives, a number of drawbacks to this program emerged. First, individual files in the .xml full report generated by TreeSize could not be seen, even when active content was enabled in the browser; individual files are represented only by an icon that reads “[files].” Secondly, the date reports only allowed for “number of years ago” as opposed to the actual year of creation, which meant these visualizations would be out of date as soon as the calendar year passed. The team opted to upload the full report anyway for the broader information it contained. Because the FTK .csv file also contained creation and modification dates at the file level, the date visualization was excluded.

Creation of the DSpace sub-community structure

Upon completion of the visualizations, the initial sub-community structure was created in the UTDR, which runs on a DSpace repository installation. The group determined that each hard drive would be its own collection under its creator's (NCI or George Sanger) sub-community heading, which also enabled the group to avoid over-processing the materials and users to see exactly how the selected directories of each drive originally existed. Under each collection, every directory-level .tar file (see the following section for details on .tar file creation) requested would be a separate item, and the visualizations of the full hard drive image would be a group of bitstreams in a single item. Any individual files uploaded via batch ingest (discussed in section V) would appear as individual items.

Corresponding to the three levels of access the group aimed to provide, three different levels of metadata were created for the materials in our UTDR collections, corresponding to entire drives, images of individual directories, and individual files. The broadest level—of entire drives—consisted of the collections (folders) in the UTDR and the metadata concerning the visualizations and other drive representations. As pointed out by Kirschenbaum, Ovenden and Redwine, the folder level is most representative of the archival concept of the fonds, so this was the logical selection for the DSpace “collection” level^{xi}. The names of the collections matched the naming schema used in the hard drive inventories. The collections also include top-level metadata about the materials on the drives, the visualizations of the drives, and a photograph of the original hard drive. The visualizations and textual representations of the entire drives were assigned metadata about their date of creation and descriptive metadata about their contents. All materials, including visualizations and .tar files were assigned metadata about the Briscoe and their Briscoe accession number.

Phase 2: Providing folder-level access

Having created and uploaded a set of visualizations that would provide context and serve as surrogates for the entire drives, the group turned to the second phase of the project: creating partial images of the specific folders selected for ingest to the UTDR. Because files and folders exported from FTK can only be viewed in FTK, the group elected to create .tar files, an open standard that captures folder structure while simultaneously excluding deleted files and allowing researchers to navigate and open individual files. Additionally, at the time of this project, a GUI version of FTK was not available for Mac computers. Although a command line interface was available in a beta version, the group felt this was a significant access issue for the anticipated user base.

A number of methodological options for creating .tar files for the images were considered. The standard method is to mount an image in the Linux operating system and use the command line interface to create the desired images. Because this is not a particularly user-friendly method for people who are not familiar with Linux and because of the size and complexity of our drives, the group decided to investigate alternative tools and successfully created .tar files using the open source file archiving program 7-Zip (version 9.20).

Mounting the disk images through FTK in order to ensure they were still write-blocked, .tar files of the folders selected for ingest were created. The group then discovered (after making a large number of .tar files) that using FTK's export option would often export deleted files. This made it necessary to create the .tar file directly from the mounted image, which purposefully did not include the deleted files. Another discovery stemmed from the creation of a .tar file of a directory that included multiple folders. Opening the .tar file in WinZip, a common file-compression software, caused the directory structure to flatten and all the files to display alphabetically rather than in their original hierarchical order. Opening the .tar file in 7-Zip does preserve the directory structure and for this reason, the group recommended to the BCAH that its research

terminals be loaded with 7-Zip or another program that supports the creation of .tar files so that researchers may open the files and see the folders as they were originally assembled.

Once the .tar files were uploaded to the UTDR, they were assigned metadata based on the materials in the file. All materials on the drives were attributed to the creator of the drives. Full inclusive dates, which consisted of the earliest “last modified date” through the latest “last modified date” of all the files in the directory, were included for each of the .tar files. Creation dates, however, were not visible. The name of the directory was used for the title metadata (e.g., \My Documents\), and the description of the item consisted of the file path (e.g., Partition 1\DISK2PART01 [FAT16]\[root]\My Documents\).

Phase 3: Providing file-level access

Preparation

At this point, the group entered phase 3 of the project, namely, providing file-level access to selected individual files from the NCI 3 drive, which required the determination of a method of ingest. Four hundred and sixty-five files had been selected for individual preservation, well above the threshold for manual ingest. For this reason, the group opted for a batch ingest, a method used successfully by groups working in prior iterations of Galloway’s course^{xii}. Although the group was unable to obtain permission to perform batch uploads to the UTDR, we felt it was important to ensure that our materials would upload properly and therefore, a practice batch ingest was performed on an iSchool server and the files prepared for later ingest into the UTDR. Due to the large number of files slated for ingest, the group opted to use Perl scripts to automate as many of these operations as possible.

Metadata and ingest package

Running the New Zealand Metadata Extractor (version 3.2) produced unique .xml files for each item, but these files could not be used for ingest into DSpace or the

UTDR because they did not follow Dublin Core standards. Using a modified version of a Perl script originally created by Sarah Kim^{xiii}, a previous student in the course, each .xml file was converted to a Qualified Dublin Core .xml and then moved into a newly created file. After downloading Active Perl, the script simply needs to be placed in the collection folder at the level of the .xml folders, then double-clicked.

The next step was to complete the ingest package. We ran additional scripts to create the contents files and renumber the files sequentially, both of which were requirements for batch ingest into DSpace. Finally, a command line was prepared to ingest all 465 files into DSpace. At the conclusion of the project, the authors did not have permission to perform said ingest, but the practice ingest was successful.

Conclusion

This project illuminated a number of valuable insights into working with internal hard drives. Using digital forensics methods to image the hard drives allowed the group to produce an identical, bit-for-bit copy of each hard drive with minimum difficulty and maximum confidence that the copy was both exact and created without compromising the original bitstreams. Once the images had been created, FTK's ability to mount the drives as read-only images provided the group with a crucial test bed for experimenting with visualizations and partial imaging techniques, as well as peace of mind that the original bitstream would remain unaltered.

Although some technical challenges arose during the imaging process, as well as in the group's initial attempts to handle the wide variety of file formats present on the drives, ultimately, the most pressing philosophical issue was the ethical question of how to deal with deleted files. The group settled this issue by creating visualizations of the drives to allow researchers access to a surrogate of the full disk image, while the actual full disk images were sent to the BCAH to be stored on tape in a dark archive. Although there were multiple deleted files that the group's client at the Briscoe felt would be valuable additions to the collections, those files were not pulled out for public access.

In line with these observations, the details of avoiding capture of deleted files proved to be a significant ongoing issue, perhaps exacerbated by the group's use of digital forensics software. Repeated testing is essential to ensure that only those files the archivist intends to capture are actually being captured. Providing access to the contents of the drives at three levels of granularity allows researchers to access visual surrogates for the entire drive, thus providing context for the images of individual folders and specific files also available on the UTDR.

The group's reliance on open-source software tools also proved to be a mixed blessing. While open-source tools carry the advantage of being free, many of the tools tested here, particularly for the production of visualizations, did not perform up to the group's expectations, especially in terms of the exportability of reports, the key factor behind the decision to use the propriety TreeSize Professional. One reason for the dearth of suitable software may be that all the visualization tools the group tested were intended for personal use on working computers and storage, rather than for archival use on legacy hardware. This suggests an avenue for future archival software development. Ultimately, this project provides valuable information for the future of working with internal hard drives using digital forensics methods and open source software.

**Appendix: Comparison of tree maps generated by WinDirStat,
SpaceSniffer, and TreeSize Professional**

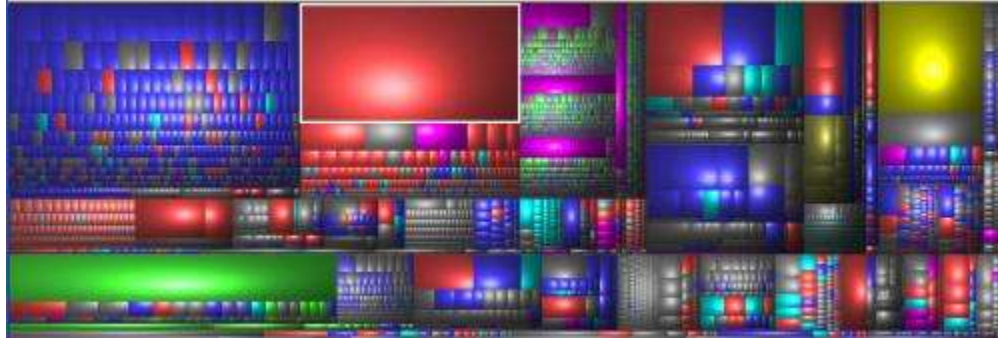


Figure 1: Tree map of NCI 3 generated by WinDirStat, exported via a screenshot. The color coding of the image indicates the types of files found on the hard drive, but there is no effective way to export the list of file types that serves as a key to the color coding of tree map.

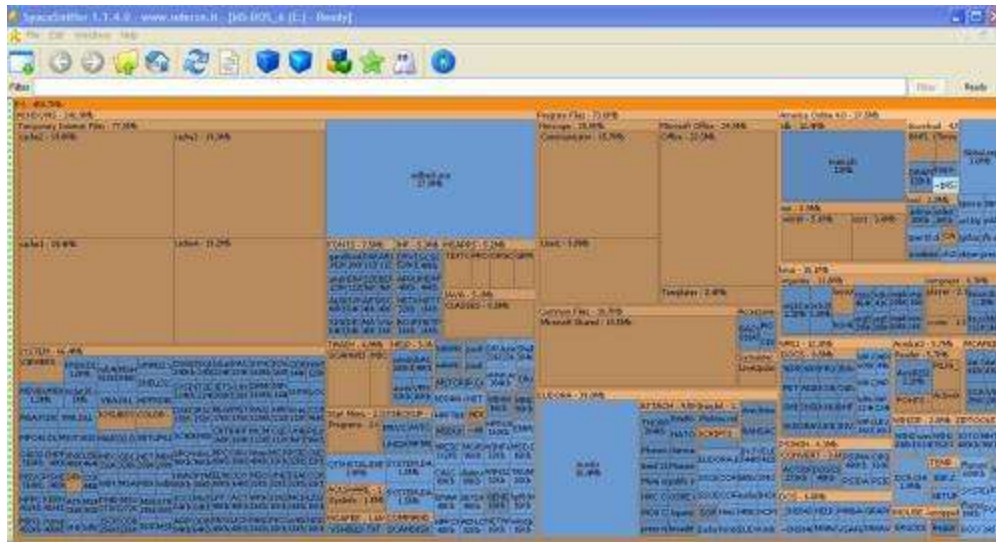


Figure 2: Tree map of NCI 3 generated by SpaceSniffer, exported via a screenshot. The program allows the user to filter the tree map using a large array of filtering commands, but this interactive capability is lost when an archivist needs to upload a static image to a DSpace collection.



Figure 3: Tree map of NCI 3 generated by TreeSize Professional, exported by the program as a .png image. This tree map is an example of TreeSize's "3D" mode; there is also an option to export the tree map from the "2D" mode.

About the Authors

Amy F. Brown received her MSIS and BA in English from the University of Texas at Austin. She is currently employed at the Ohio State University Library as the Special Collections Cataloging Coordinator, providing access to Ohio State's rare materials.

Chloë Edwards holds a BA in International Relations from Oglethorpe University and an MSIS from the University of Texas at Austin. She is currently employed as a government records archivist at the Mississippi Department of Archives and History.

Meg Eastwood received her BA in Biology from Grinnell College and her MS in Information Studies from The University of Texas at Austin. She currently works as an Academic Programs Librarian at Cline Library (Northern Arizona University), providing research and course support to the students and faculty of the College of Engineering, Forestry, and Natural Sciences.

Martha Tenney received her MSIS from the University of Texas at Austin, where she also worked at the Human Rights Documentation Initiative. She is currently the Digital Archivist and personal librarian for the American Studies Program at Barnard College in New York City.

Kevin O'Donnell is an information professional living in Austin, TX. He holds an M.S. in Information Studies from the University of Texas at Austin and a B.A. in Political Science from Baylor University.

ⁱ John, Jeremy Leighton. "Adapting existing technologies for digitally archiving personal lives." In Ponencia presentada en las actas de Fifth International Conference on Preservation of Digital Objects, pp. 48-55. 2008. Accessed 10/24/13 at http://www.bl.uk/ipres2008/presentations_day1/09_John.pdf.

ⁱⁱ Full inventories containing all of these details can be accessed at <https://pacer.ischool.utexas.edu/handle/2081/23529> (NCI drives) and <https://pacer.ischool.utexas.edu/handle/2081/23518> (Sanger drives).

ⁱⁱⁱ Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending digital repository architectures to support disk image preservation and access." In Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries, pp. 57-66. ACM, 2011. Accessed 10/24/13 at <http://dx.doi.org/10.1145/1998076.1998088>.

^{iv} John, 2008, p. 47

^v Although FRED, which consists of a laptop plus several different hardware write-blockers and their associated cables, was a useful tool for us, its functionality can easily be replicated with a laptop and the appropriate external write-blocker.

^{vi} Lee, Christopher, Matthew G. Kirschenbaum, Alexandra Chassanoff, Porter Olsen, and Kam Woods. "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions." D-Lib Magazine 18, no. 5 (2012): 3. Accessed 10/25/13 at <http://www.dlib.org/dlib/may12/lee/05lee.html>.

^{vii} Kirschenbaum, Matthew G., Richard Ovenden, Gabriela Redwine, and Rachel Donahue. Digital Forensics and Born-digital Content in Cultural Heritage Collections.

Council on Library and Information Resources, 2010. Accessed 10/24/13 at

<http://www.clir.org/pubs/reports/pub149/reports/pub149/pub149.pdf>

^{viii} To access these and other files referenced as being uploaded to the UTDR as part of this project, see <http://repositories.lib.utexas.edu/handle/2152/7020>.

^{ix} Kirschenbaum, Ovenden and Redwine, 2010.

^x See the appendix for comparative screenshots of the imaging software tested.

^{xi} Kirschenbaum, Ovenden and Redwine, 2010

^{xii} Kim, Sarah, Lorrie Dong, and Megan Durden. "Automated Batch Archival Processing: Preserving Arnold Wesker's Digital Manuscripts." (2009). Accessed 10/25/13 at <http://repositories2.lib.utexas.edu/handle/2152/6003>.

^{xiii} The original Perl script is available at <https://pacer.ischool.utexas.edu/handle/2081/10676>.