2015

File Naming in Digital Media Research: Examples from the Humanities and Social Sciences

Jerome W. Crowder
University of Texas Medical Branch

Jonathan S. Marion
University of Arkansas

Michele Reilly
Central Washington University, reillym@cwu.edu

Recommended Citation

Crowder, J. W., Marion, J. S., & Reilly, M. (2015). File naming in digital media research: Examples from the humanities and social sciences. *Journal of Librarianship and Scholarly Communication, 3*(3), eP1260. DOI: 10.7710/2162-3309.1260

JLSC

ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | http://jlsc-pub.org

Volume 3, Issue 3 (2015)

File Naming in Digital Media Research: Examples from the Humanities and Social Sciences

Jerome W. Crowder, Jonathan S. Marion, Michele Reilly

Crowder, J. W., Marion, J. S., & Reilly, M. (2015). File Naming in Digital Media Research: Examples from the Humanities and Social Sciences. *Journal of Librarianship and Scholarly Communication,* 3(3), eP1260. http://dx.doi.org/10.7710/2162-3309.1260

ISSN 2162-3309 10.7710/2162-3309.1260

THEORY

File Naming in Digital Media Research: Examples from the Humanities and Social Sciences

Jerome W. Crowder*
Assistant Professor & Associate Director, Institute for Medical Humanities, University of Texas Medical Branch
Jonathan S. Marion
Assistant Professor, Anthropology, University of Arkansas
Michele Reilly
Associate Dean of Libraries, Central Washington University

This paper identifies organizational challenges faced by Social Science and Humanities (SSH) scholars when dealing with digital data and media, and suggests improved file naming practices in order to maximize organization, making files easier to find, more useable, and more easily shared. We argue that such skills are not formally discussed in the literature and therefore many scholars do not recognize the problem until they cannot locate a specific file or are sharing files with colleagues. We asked SSH scholars to share their file naming strategies (or lack thereof) and we use these narrative anecdotes to discuss common problems and suggest possible solutions for their general file naming needs.

IMPLICATIONS FOR PRACTICE

1. This article focuses on the crucial but often-overlooked role, and significance, of file naming practices in digital media research, storage, and archiving in the social sciences and humanities (SSH). We offer practical file naming solutions that librarians can introduce to SSH colleagues and scholars.
2. Narratives provided by SSH scholars detail their file naming inconsistencies, which lead to confusion, loss, or error. Librarians can address these issues and support scholars who find file naming challenging or are seeking new strategies.
3. Librarians can implement our suggestions to help SSH scholars develop their own file naming protocols so they can better find, retrieve, and share their data. These protocols can also serve as a foundation for creating data management plans increasingly required for federally funded grants.

Received: 05/11/2015 Accepted: 10/08/2015 ** Authors, listed in alpha order, all contributed equally*

Correspondence: Jerome Crowder, University of Texas Medical Branch, Institute for the Medical Humanities, 301 University Blvd, Primary Care Pavilion Ste 2.301, Galveston TX 77555-1311, jw.crowder@utmb.edu

PREFACE

A colleague shared this story with the authors when asked about his experience with naming files:

> Recently I team-taught a course with a colleague in a different department. His staff person uploaded all of the articles for our course to Blackboard so the students could access them. The titles of the files were the titles of the papers themselves, without a date, without an author, without any meaningful data to organize the file. When I asked the staff person about this convention she told me that it has always been done that way, as it makes it easier for the students to recognize it in Blackboard. But, I thought, "if students have a syllabus, with the citations in it, then they will know the author's name and publication date for the appropriate class day, listing the entire title seemed superfluous." It was a small consolation when the staff member later confided that just listing the author's name/date is much easier for her to input and it was not difficult for the students understand. Though she was reluctant to change the system for fear of faculty reprisal!

In response to this anecdote and our experience with other Social Sciences and Humanities (SSH) colleagues, we offer this paper to the library community as a description of research faculty behaviors and as an opportunity for academic librarians to have an immediate impact by helping SSH researchers better organize their files and work more efficiently. Such work is imperative given the requirements of federal grant funding in which researchers provide data management plans when they submit proposals to funding agencies (e.g. NSF, NIH, NEH).

While we recognize that professional librarians and archivists already adhere to various best practices for digital file naming in their own work,[1] they should not assume that colleagues across campus understand the need for such practices. This paper highlights common issues with file naming and digital organization faced by scholars so that librarians can assist faculty and graduate students, particularly associated with SSH, to develop better data management plans and personal information management practices. Improved personal organizational skills can increase the chances for competitive funding and enhance research and teaching functions.

[1] Because we are focused on digital content organization strategies for individual researchers, it is outside of the scope of this paper to discuss best practices for institutional digital archives or enterprise-level digital asset management systems. While personal practices are informed by similar approaches, this paper is not intended to suggest best practices for library and information professionals.

INTRODUCTION

This article focuses on the crucial but often-overlooked role and significance of file naming practices in digital media research, storage, and archiving in the social sciences and humanities (SSH). More than just labels, file names are a vital part of data and information management, and can serve as tremendous tools both for ongoing analysis and future research. While the current social science and humanities literature is replete with content on current research practice and theory (e.g. Mitchell, 2011; Pink, 2007; Postma & Crawford, 2006; Rose, 2007; Spencer, 2010; Strong & Wilder, 2009), there is a dearth of material on current data management theory and practices (see Jablonko, 1989; MacQueen & Milsten, 1999; or McLellan, MacQueen, & Neidig, 2003 for tangential discussions on the topic). Such issues are increasingly important amidst growing attention to, and emphasis on, data management plans (Adamick, Reznik-Zellen, & Sheridan., 2012).

Information scientists identified the need for naming schemes for files and images early on (Arms, 1996; Eakins & Graham, 1999), and many libraries and state archives across the country maintain best practices for file naming (e.g. Colorado University Libraries, 2008; North Carolina Department of Cultural Resources, 2008; University of Illinois, 2013; University of Oregon Libraries, 2010). Yet while many institutions have policies in place to aid in data retrieval and utilization, individual scholars typically follow piecemeal practices. This article illustrates typical problems SSH faculty and graduate students have in file naming and provides concrete strategies and rationales for data and information practices which in turn set the stage for good archival practices. From file naming conventions to folder structures, this article considers the different perspectives and approaches of researchers creating files and suggests best practices to help maximize the utility and accessibility of research and records.

LITERATURE REVIEW

Information professionals have been organizing, sorting, and storing electronic files in a systematic way for many years but it has only been within the last ten years that research has been conducted in the field of Personal Information Management (PIM). William Jones (2007) describes PIM as the

> practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use, and distribute the information needed to complete tasks (work-related or not) and fulfill various roles and responsibilities (for example, as parent, employee, friend, or community member). (p. 453)

Despite this emerging research, the topic of file naming and how it relates to humanities and social science researchers is rarely addressed. A majority of the articles written on the topic of electronic file organization concentrate on personal information practices within the corporate workplace or the student context. For example, Khoo, Luyt, Ee, Osmana, Lim, and Yong (2006) carried out a small study on the file management behaviors of university students by harvesting filenames, folder names, and file structure from personal computers and notebooks. The authors sought to understand how participants organized, maintained, and managed their files; how they retrieved information and documents; and what the "relationships [were] between file structure, user behavior and cognition."

Not surprisingly, Khoo et al. (2006) found that a majority of users stored their files on their desktops and in folders on their hard drives. Subsequently, users tended to use their desktops for their *working* or temporary files and the folders on their hard drives for folders for storage, archives and working files. Most users in the study named files and folders based on project or work tasks; though those users with deep folder structure organization used "broader and more generic folder names at the first level" (Khoo et al., 2006). These findings suggest that users are familiar with managing levels of files, but do not suggest a systematic approach to naming and organization. While Khoo et al. discovered that there were a number of categories for folder naming conventions including "document type, organization function or structure, and miscellaneous or temporary" (Khoo et al., 2006) such findings reinforce the idea that file and folder naming practices need to be communicated to practitioners. This opens up an excellent opportunity for library and information professionals to begin the conversation with these file creators.

Where Khoo et al. (2006) focus on student practices, Peters (2002) discusses the data organization challenges faced by general computer users. The author reports on a qualitative study designed to determine the practices and awareness of personal information management in the popular sphere and explores "how people store and retrieve all types of information" (Peters, 2002, p. 2). The study observes three types of behaviors and gives them the following names" "Luddite Larry, Pragmatic Polly, and Gadgety Georgina" (Peters, 2002, p.3). Though users "want to build strategies and use tools that process information in a timely, speedy and reliable manner that is natural to them in whatever particular situation they might be," the author finds that users, for the most part, are not following a simple strategy (Peters, 2002, p.1). This finding was confirmed in a recent study by Zastrow (2014):

> In a survey of the personal archiving practices of 110 writers, most admitted to practicing 'benign neglect' as a records management strategy, and 80% said they would welcome instruction on digital preservation as they felt they lacked the technical ability and knowledge to back up and preserve their own files.
> (Zastrow, 2014)

As Zastrow indicates, there are opportunities for information professionals to instruct students, faculty, social scientists, humanists, and digital content creators in file naming and folder structure basics to aid in organizing data. In this paper, we present simple strategies in which librarians and other information professionals could train students, scholars, and other content creators.

LEFT-HAND SIDE OF THE DOT: WHAT YOU NEED TO KNOW

All digital files have file names. Computer documents, scans, voice recordings, videos, photographs, and more get downloaded and accessed by people using file names to choose which file they want to view, move, copy, or attach. Until and unless files are deliberately renamed, however, machines generate default names that can be uninformative and confusing (e.g., _IMG1163). Since different types of files—from different machines, by different manufacturers—use different default naming conventions, it is particularly valuable to assign names that are useful, i.e. that tell something about the data. While thumbnail images sometimes get used as visual cues, ultimately computer access is achieved via designated file names. Since this is already true for all digital files, why should researchers care about developing and maintaining specific file naming conventions? The following case is an example of a user who could benefit from good file management, and suggestions of practical strategies to overcome their organizational issues.

A historian working in an archive must keep close track of the many primary sources she uses in her research. She has found photographs, cassette tapes, note books, and ephemera that pertain to her question, but how will she use these in her work once she exits the archive? After digitizing these sources, she will need to organize the files in a logical manner so she can later retrieve and use the data when necessary. It is not enough to place the different file types in separate folders (e.g. “photos”, “audio”, or “stuff”) and scroll through them later expecting to find them. This researcher needs a system so she can differentiate between similar file types as well. Depending upon the type of work she is conducting, she could begin by using the surnames of each family as the root for the files pertaining to that family. The filenames of the scans of the Smith Family photos, audio recordings, and letters would begin with SMITH and could be followed by the date they were created: SMITH1957, SMITH1966. Following this root with a hyphen separates the kind of data from more specific identifiers, such as the type of media, Black & White (BW) or slide, followed by a number identifying the file’s place in the series of tapes or images. Everything on the right-hand side of the dot then designates the type of file it is, .jpg, .pdf, or .mp3. For example, a copy of the third audio cassette would be “SMITH1957-3.mp3” or a scan of the third black and white photo, “SMITH1957-BW-003.jpg”. Such strategies will help the researcher find and use her data more effectively.

As seen in this example, when different file types are all related to each other (i.e. they are all from the same project, subject, or date) they should all be named the same and use the different format types to distinguish them. For instance, if a social science researcher has created word files, images, and audio all from the same meeting, instead of naming them "documents about meeting 1", "photos of meeting 1", and "recordings of meeting 1", it makes more sense to name them all "meeting 1", and let the file extension designate the type of data (e.g. "meeting1.doc", "meeting1.jpeg", and "meeting1.wav"). With multiple files in the same format, simply differentiate between them using systematic designations such as sequential numbering or time codes. (More examples will be discussed later in the article; see Zhang (2007) for description of an entire project design).

Appropriate file naming provides consistency, uniqueness, and meaning, and is an aid to access and preservation. If files are organized by date (or by the name of significant people, events, or topics), and *named accordingly*, it becomes far more obvious how different files relate to each other and where to look for related data. As such, instituting a standard file naming scheme early—or going back and batch renaming existing files—and then building this ongoing file naming process into the workflow will minimize the time and effort needed to search through materials in the future, and facilitate an organized expansion of digital data materials. Standardized naming also expedites various forms of collaborative and team research. Because everyone on the team understands how to name their files, the different pieces of the project all fit together when all the files are aggregated, and each member of the team knows how to access, make sense of, and use the data in their research. For various conventions and strategies concerning best file naming practices, see Alberta Government (2005); Controlled Vocabulary (ND); Kuhn (1995/2004), Minnesota Historical Society (2004); and North Carolina Department of Cultural Resources (2008).

Folder and sub-folder structures

Just as important as how to name files is how to structure their organization relative to other files. This may not seem problematic when there are only a few research files to deal with, but it matters increasingly more as one's research holdings continue to expand. As such, it is best to design robust and meaningful file naming conventions, directory systems, and directory structures from the outset. Already having strong file naming conventions in place allows one to simply mimic this structure in a way that is logical, e.g. putting files named for significant events in a folder for events. Additionally, adding the date to the already determined directory names links the data in another logical manner and makes the materials that much more searchable, e.g. searching by name (event/person) and/or date is now viable. While such added specificity is especially useful, researchers must be equally careful not to push things too far by adding too many additional file or directory

descriptors. Taken to the extreme (i.e. including every viable search term) will result in far too many files for the search to be useful. Consider file naming as a parallel to coding any research data: using multiple appropriate codes is far more versatile than artificially insisting on using only one, yet using too many dilutes the utility of coding in the first place. In a very similar way, the key to creating a strong and useful file/folder structure is finding an appropriate point of balance between creating a file naming and file/folder structure that is robust enough to allow for expansion and evolution, but not overly-general, complicated, or onerous to apply.

Good file naming should reflect the work itself and how the researcher expects to use the work in the future. For example, when Jerome Crowder (2013) took photographs for his fieldwork in Bolivia his primary folder for all files was called "bolivia-13", which reflected the geographical location of where all of the data within that folder were collected and the year they were made. Sub-folders were named, for example, "bolivia-13_audio", "bolivia -13_video", "bolivia-13_scans", "bolivia-13_notes"...Similarly, photos taken were placed in the appropriate folder and then named accordingly: "bolivia-13_001", "bolivia-13_002" … "bolivia-13_1234". Note: since the file type is already designated by the suffix (e.g. jpg or TIFF for a photo, .wav for a sound file), it does not need to be repeated in the file name. In his research on competitive ballroom dancing, Jonathan Marion (2008; 2010; 2013; 2014) has followed a similar labeling strategy based on event names and their associated year, e.g. "Blackpool_2013_0001", "Blackpool_2013_0002", and so forth.

This is only one approach. A digital archivist, in contrast, might begin archiving projects with a designated 'year' folder, such that all scanning projects in the year 2012 go in a folder titled "2012". The first project folder in that year could be designated with a "1" preceding all other names. Here, using a simple number system keeps all of the project files in order. For example: "2012_001_anderson", "2013_002_thompson", "2013_003_jasper", etc. In this instance, it is not the alphabetical order that is important, but the project number that keeps everything organized and straightforward.[2]

[2] This example assumes a preference for meaningful (i.e. intuitive) file and folder naming, which is the approach recommended in this paper. Digital archivists may also take a different approach to naming conventions depending on the scale of the data and the storage/retrieval system being used, as noted by the U.S National Archives and Records Administration (Puglia, Reed, & Rhodes, 2004): "Meaningful file names contain metadata that is self-referencing; non-descriptive file names are associated with metadata stored elsewhere that serves to identify the file. In general, smaller-scale projects may design descriptive file names that facilitate browsing and retrieval; large-scale projects may use machine-generated names and rely on the database for sophisticated searching and retrieval of associated metadata" (p. 60).

Files and versioning

At the most basic level, it is important to label each version of a file to help keep track of changes and updates. Beyond knowing which version is most up to date, such practices also document procedures. This is a very simple process, and can entail something as basic as labeling files with "-v" or "-v2" to designate and differentiate between the first and second versions of a file. Ostensibly the "last modified" data field should provide similar information, but this can be easily confused as files are moved, copied, or pasted, especially between different devices. Even if the automatically generated use-data becomes rewritten amidst various machines and uses, appropriate labeling ensures knowing the relationship between different versions of a file, for example, that version 2.2 of a file came before version 2.3 of that same file. While many word processing programs maintain version numbers in their metadata, incorporating version numbers in the title allows for quick recognition. There are no hard and fast rules for how to do this, but figuring out a system that makes logical sense—and using it consistently—is the key to success.

For instance, in writing *Visual Research: A Concise Introduction to Thinking Visually*, Marion and Crowder (2013b) used different version numbers, e.g. "-v1", "-v2", "-v3", to reflect the introduction of additional materials and used different version sub-numbers, "-v3.1", "-v3.2", "-v3.3", to reflect editing and revisions to extant materials. While this system was appropriate for that project, it may be just as effective to use a different system. What is crucial is that the system makes sense to both researchers and their collaborators and, more than anything else, that it is used consistently.

There is no need to version different types of media files differently. Rather, what is most important is that the file naming practices for related items—across various media types — should be the same. It is also advisable to make a key/code for the procedures chosen. While it is easy to memorize a system, over time it is easy to forget or misremember especially after an extended duration away from the activity. A written key may also be shared directly with collaborators or as part of a larger data management strategy. An example of what such a key might look like, in this case for research images, is that seen here in Figure 1 (following page).

While Marion and Crowder (2013b, p. 102-104) have previously suggested that special characters can be useful in this capacity, e.g. using "#" to designate cropping, or "!" for a fully edited and publication-ready image, it is important to realize that this is only advisable for working files. For archiving, it is best to use a standardized set of nomenclature (Kuhn, 1995/2004) so that data are widely searchable. Using standard characters, i.e. not special ones such as "!" or "+", make research materials more accessible by making them machine readable.

Key code	Definition
-it	Image treatment. Image has been cropped and color corrected. It is ready to display
-ar	Archival master image
-pr	Print ready

Figure 1. Sample versioning key for research image files

When establishing a system, it is also important that it is scalable. This will ensure that it will endure after the particular project has ended or if it becomes part of a long-term or broader group project. For example, using a two digit numbering sequence is limited to 99 projects without further modification. Instead, choose a scheme from the outset that will ensure that file names will not become duplicated. Therefore, when numbering similar types of files or sequences, anticipate the maximum number that are likely to be needed or used for additional versions or sequential steps. It is fairly easy to conceive going over 9 but what about going over 99 or 999?

The key here, as with folder naming and structuring, is to balance identification against access, providing enough characters, alphanumeric and numeric, to ensure uniqueness, but not so many that identifiers become overly complicated or that machine readability is compromised. Many computer operating systems restrict the number of characters in file names, disallowing anything beyond a certain length, so while a long, descriptive file name might aid in identification, it concomitantly makes the file largely inaccessible since it cannot be searched for by the operating system. As a general rule of thumb, 31 characters is usually the maximum accepted by most computer systems. Also note that another problem with long file names arises with the need to create a folder/sub-folder/file structure, as the ensuing path names become too unwieldy and complicated when transferring files between drives and machines.

Along these same lines, best practices involve using lowercase letters, numbers, and dashes when naming files, and not using other (special) character options. Some information professionals encourage the use of underscores to separate filename components (e.g. "2007_reillym"). Since the issue is simultaneously one of identification and access, it is important to recognize that underscores can produce a visual acuity problem given the common usability practice of underlining hyperlinks. One argument against hyphens is that some programs misrecognize them as minus signs, such as when used in spreadsheets

like Excel. Similarly, avoid using spaces in naming and versioning practices, as spaces can cause problems when included as hyperlinks and some systems may not accept them. No matter the particulars, however, always aim to create file/folder names that are logical, meaningful to all users, simple to read, and relevant. If a file is obscurely named, it can be almost impossible to locate in the future.

Working with existing files

Once an appropriate naming convention is determined, consider renaming existing files. If possible, this is the ideal solution; it brings all of the existing data into a unifying structure, and allows new data to be added in a systematic manner. The easiest and most efficient way to do this is by using a batch file (re)naming tool, such as that found in Adobe Bridge. The actual software process will only take seconds and will result in consistent file naming from all devices, including camera, video, audio recorder, or scanner. Going forward, develop a workflow: a standard set of operations (and an associated key) to the unique system (Marion & Crowder, 2013a; 2013b). Using the system on a consistent basis and in a consistent manner increases the likelihood of finding those files. It makes them more accessible for specific research needs, as well as more navigable to others while preserving them for later or long-term use. The authors agree with Zastrow (2014), "Once you have your basic file management framework set up, it's a piece of cake to add new folders and branches to the hierarchy." Indeed, a simple framework will provide the consistency to build a robust project no matter the discipline or file type.

Keywords and metadata

As mentioned in the story from the preface about file naming conventions, keywords and metadata are very important to keeping organized and finding files both quickly and efficiently. However, it is advisable that they not be placed in the filename itself. Rather, right-click and choose properties within the new window (Windows) (Figure 2, following page) or command-click (Mac) (Figure 3, following page) on a file icon in order to open a details section for that unique file.

Here there are a number of fields, depending upon the file type, which can be populated with the relevant data, including full titles, keywords, authors' full names, or perhaps even notes made on the item itself. What is important to understand here is that the operating system (OS) will search these fields, along with the file name, when a search is performed for specific files. That is, while the filename is visible to the user when organizing the files, the OS 'sees' much more than that. As such, there is no need to place all of the important information in a filename. Rather, simply be systematic in thinking about the folder organization and key in appropriate keywords and metadata to make it easier to find files later.

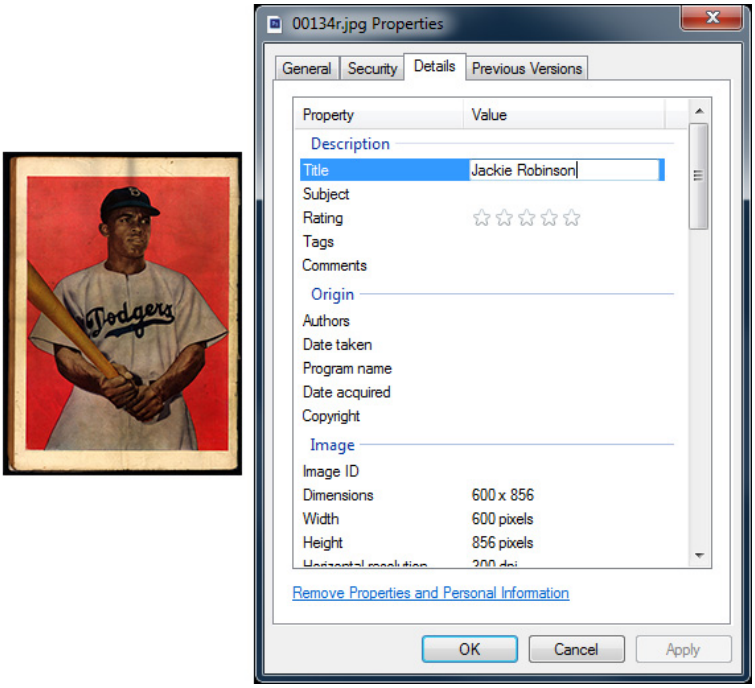

Figure 2. Metadata entry within details section of file properties (Windows)

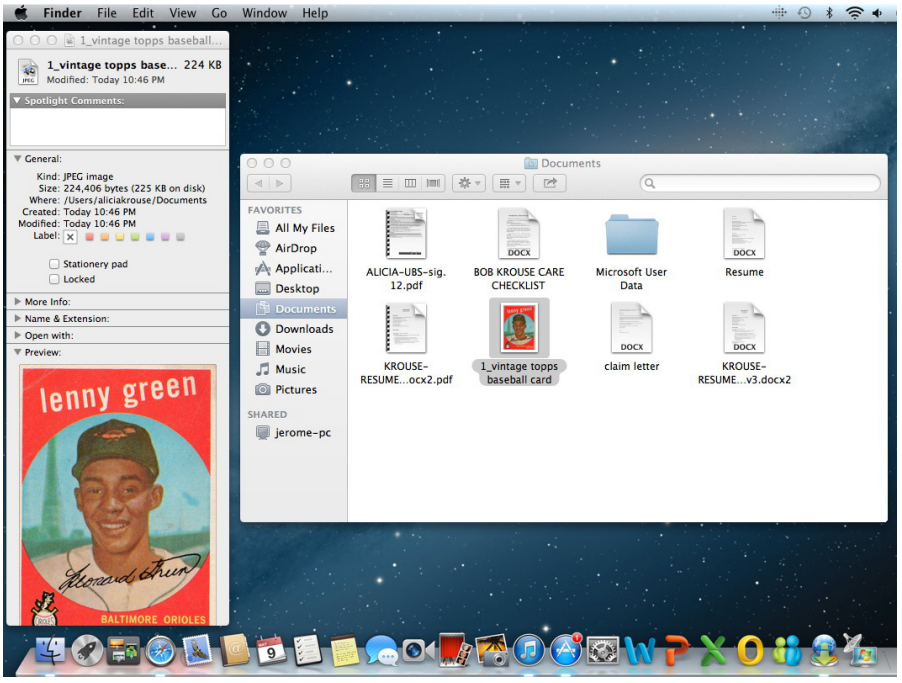

Figure 3. Metadata entry within details section of file properties (Mac)

Data organization structure

Like file naming, placement of files is just as much about organization as it is about understanding the architecture of the computer. File names on both MACs and Windows-based machines are limited to 260 characters of path name. It is unadvisable to 'bury' files deep in the data organization structure, but rather keep them near the surface where most working files exist. Several key considerations here include differentiating between local/use files, backup files, and the procedures implemented (if any) to synchronize between them.

Local/Use Files

Files are usually organized in 'trees', the root of which is commonly the computer name or drive name, such as C\: or D\:, where most general files are placed, e.g. "Research", "Professional", "Photos", "Teaching", "University", and "Personal".Next, place more specific files within those files (this is called 'nesting'). Within the "Research" folder, for instance, there may be sub-folders/files labeled by grant name or award date, the name of the project, or even "current" and "past", or under the "Professional" file resides all presentations and associated documents in a folder called "Meetings", within which each individual meeting is organized by chronological order. Using this example, a potential data path looks like this:

C:\desktop\professional\meetings\13AAA\13digihuman_final.doc

In this example, 36 characters (of the 260 maximum characters per data path) have already been used before getting to the filename itself. Considering that the majority of users keep their working files on their desktops (Khoo et al., 2006), some people might argue that this example is rather deeply 'buried' in the system, but it could be much worse. Here, at least, there is a logical organization whereby that document was filed in that folder. Not to belabor the point, but the importance of systematic organization cannot be overstated. Indeed, users are already starting the process of 'losing' their files when they cannot figure out where to put them. This pitfall may be easily avoided, however, by building a *systematic* structure, starting with the desktop, or root directory, which enables (relatively) easy navigation up and down the trunk and access to the themes that make the most sense. Having thought through this process from the outset (or gone back and established such a system after the fact), thereafter one can usually quickly and easily file documents in the manner that makes sense.

Backup files

For purposes of this paper, the term 'backup' signifies copying files with intentional redundancy onto external media. External storage may be portable, like a flash drive, or

more substantial like an external hard drive, or even an optical disk (CD-ROM, DVD-ROM, Blu-Ray). Today, the most popular and accessible means for backing up is actually "in the cloud", i.e. uploading and storing files via online server space. Whatever the preferred method of backing up data, it is best to maintain the same filing system for the backup files as for the local/use files. In fact, when setting up the 'back up' for the first time, copy the entire file structure, so that all folders and files for a project are included in the back-up. Using the example suggested above, there would be folders on the backup system that mirror the exact same folder configuration that appears on the desktop: Research, Professional, Photos, Teaching, Universal, and Personal. All of the sub-folders would then be included in the external copy (backup), so when updating the individual files, the sub-folders will already be there as well. When making new sub-folders on the local use drive, simply copy, whether by "drag and drop" or "copy and paste", those over to the backup drive, since both MAC and WINDOWS operating systems will notify you if you a given file already exists in the destination folder.

On a related front, realize that having redundant copies can be supremely important to protect against data loss, but that it is crucial to keep these copies synchronized. One does not have to synchronize all of the files simultaneously, but establishing a synchronization routine will guarantee that you have copies of all of your important files, even if you never need them. Important files should be backed up regularly; a minimum would be once per week, depending upon the data. There are applications that will schedule automatic backups or manual backups, which may take longer; these are also advisable. The problem of not having the data, of course, occurs when a drive crashes or becomes corrupted, or a computer dies or is stolen. It is only at those times when one realizes the importance of properly backing up and knowing what all has been protected through such a backup.

Downloading from the Internet

One of the things discovered with experience, and highlighted in the following SSH case studies, is the importance of keeping articles used for research and teaching just as well organized as the papers and articles produced for others to read. This raises the question: when accessing documents via the Internet, from J-STOR or a similar digital database, how does one download and deal with these files? Academic articles accessed online are commonly formatted as PDF documents, typically with files names which may make little obvious sense (e.g. "11204572", "XXx-RTJ", "ATW-3132B"). Typically what the file name designates can be discerned, but the important question to ask is: will the file be recognizable via the filename on the desktop at some indeterminate point in the future? Chances are slim that any general user will recognize it; instead, one will need to open each specific file in order to determine the content and where it should be placed.

This process can be made much easier, making the materials being collected that much more accessible for current and future use, by renaming files as they are being downloaded. When downloading files from the Internet, one is given options to designate both where to save the files locally and under what file name. If possible, this is the ideal time to rename the file and place it in the appropriate folder, such as for a specific research topic or set of class readings. Recognizing that there is not always sufficient time, one useful practice is to create a document repository where all downloaded files are saved until there is time to go back to them for proper renaming and filing. (Note: a repository file is also a good idea for scanning documents or images.)

Reference management applications—such as Endnote (paid, endnote.com), RefWorks (paid, refworks.com), Zotero (public, zotero.org), and Mendeley (public, mendeley.com)—can also help a researcher keep track of downloaded files. For example, once the .pdf file has been downloaded from the server, and saved to a specific folder, one can then associate that file with the reference in Endnote by simply dragging the paper's icon to the particular reference. Later, when the references are opened, the icon for the paper will appear in the attachments field. Double-clicking on the icon will then open the paper. When conducting research with the downloaded papers, the same applications can be used to search for the particular paper, which it will then link to for easy access.

Even when using such programs, appropriately (re)naming the files will save time in the long run and be advantageous overall. For instance, when copying a file for use in a class, or consolidating research materials for a specific project into one large file, or sharing a specific article with a colleague, one could use a reference management application, but properly named files will be more accessible. In essence, properly naming and organizing downloaded files can be as advantageous to organization and productivity as the care one takes concerning the materials one creates oneself.

EXAMPLES FROM OUR COLLEAGUES' FILES

As librarians are tasked with helping faculty develop data management plans, it is important to understand how social science and humanities scholars name their files. The following sample serves as an example of scholars who may or may not need help with creating such plans. When asked, these SSH scholars' responses were highly informative, spanning the continuum from very thoughtful naming conventions to not having a system. Here, selections from those responses illustrate some of the problems they have (which may not be unique to SSH faculty and graduate students) and the significance of establishing and maintaining robust file naming conventions. First, the scholars explained how they deal with versioning of chapters and articles, identification of letters, grant proposals and downloaded content:

- *I have a hard time finding files that I saved a good while ago and had forgotten what I called them.*
- *I have trouble when I don't provide that initial classification [on my file names]*
- *I don't really bother naming images because it is easier to identify them from an icon image.*
- *I have trouble finding files if the first word in the file is not a word or name that I would initially think of in searching for the file.*
- *The real issue is about what folder things are in, and of course, what computer they are on, and which USB drive I've saved it to.*
- *I don't use folders, I keep everything on my desktop in alphabetical order, I just scroll down my desktop until I find the file I'm looking for.*

As the scholars further explained, they have each developed systems that are now familiar but not necessarily satisfying for organizing digital files. Fortunately, there is no need to completely overhaul file naming conventions to become better organized. Rather, implementing a few strategic tweaks can significantly improve the 'hit rate' or efficiency in finding files on regularly used machines (e.g. laptop, desktop, tablet, flash drive, cloud drive). Using examples from these scholars to model how such changes can be easily implemented, here are some suggested strategies for improved organizing and file naming practices. (We have changed all names to 'Smith' and the year to '2015' for both privacy and clarity of presentation).

Example 1

> *I am a product of early computer socialization back in the DOS days, when filenames could only be 8 characters. I have an outdated tendency to use shorter filenames and as a result I organize things by folder name and truncate filenames. For a manuscript in preparation I use a fragment of the proposed title of the article and then a digit to indicate what draft, e.g. 'Futility 3.' If I am writing with a co-author I lead off with his/her last name, 'Smith-Futility3.' For articles [pdf or downloaded] I use author-year-words from title, 'Smith 2004 Determining Futility.'*

In this example, the scholar has applied several good conventions, including: keeping filenames short; using the last name of the author(s); and including the date. Nonetheless, with a few minor adjustments this system could be even better. How so? One suggestion would be expanding file naming length to include dates for this work, using lower case letters, and placing hyphens between the important data categories (e.g. "14-smith-futility-v1").

Using underscores is fine for local filenames to avoid using spaces, but substituting hyphens for web use prevents confusion with underlined hyperlinked documentation (Colorado University Libraries, 2008; Puglia, Reed, & Rhodes, 2004). Similarly, while the authors understand the historicity wherein filename length was an issue in the 'DOS days' (because of the operating system's bit rate), in today's 64-bit world filename length is no longer a technical issue. However, keeping folders neat and organized is still important, and longer filenames become cumbersome after about 20 characters. Making these few simple adjustments, would enable this scholar to immediately sort self-produced work based upon any one of those pieces of data. Similarly, when sharing these files with co-authors or editors, they too will immediately have a better sense of what file they are dealing with, the version number, and the year the project commenced.

Example 2

> *Image files I handle more casually. It's more a point to get them in the right folder than to name them correctly, though I often open the information file and add important metadata there.*

As previously noted, file types are immediately identified by what appears on the right-hand-side of the dot ("."), with the suffixes ".jpg," ".tiff," ".mov," and ".wav," standing for image, video, or audio files. In this case, this scholar relies on this alone to help her identify her project files. Suggested improvements might be to use a keyword or simple name to identify the entire project, and then use that across all of her files, along with the creation date, to improve the 'findability' of the files. Batch renaming of files is quite straightforward and easy to do, so the authors suggest renaming this project "smith-14-#.*" where the # is the file number and * stands for the format type (Marion & Crowder, 2013b). Remember, while dates and times are saved automatically by the digital machine when creating an image or audio file, specifying these in the filename facilitates immediate recognition, as with a file named: "16jan14-smith.wav".

Example 3

> *For articles from books and journals saved to my machine, I label them as 'author-short title' if it is written by an author whose name I know well. For an article by authors whose name I will not remember, I just label it by its short title, but I make the first word of the short title the major subject of the article so it is easily searchable among all of my digital files. For my own documents, I use that same short title method, sometimes augmented by dates if there are multiple versions, or name of the class that it is associated with, etc.*

This scholar has introduced a number of variables into his file naming conventions, which may serve in the short-term, but can become quite cumbersome in the long-term. Full article titles can be placed in the metadata, so those should not be included in the file name. Similarly, if using a reference manager (like Endnote or Mendeley), the PDF files can be attached to the reference record, so one never need search a title in the filename again, but instead can use the reference manager to find files instead. For this example, the authors recommend one keyword from the working title preceded by the specific date of the version or the year followed by the keyword and then the version number (all separated by hyphens).

Example 4

> *I use old FORTRAN conventions: projectname.DAT for data files. And for output files I use *.LIS or often *.OUT. For text files, they are usually *.doc or *.txt. Also, extra extension in text document files like the date. For project names I usually use a short mnemonic for the project: ojo for evil eye, but add initials for subsamples: ojoMXL (Mexican lay sample), ojoUSD (US dr sample).*

As in the first example, this scholar is accustomed to abbreviated filenames based in experience with the programming language Fortran. Early standards for this command-line driven application only allowed six-character-long filenames, requiring users to become creative in their naming conventions (thus the mnemonics). The authors would suggest that this user venture beyond such long-held conventions to include dates and/or version numbers. Just because the mnemonic works for this scholar does not mean that others will understand it, and in today's digital environment one must at least consider how filenames will work when shared with students, colleagues, and collaborators. When working on a team, for instance, agree upon a project name and establish a system for how and when filenames will be updated depending upon what counts as a new version number. This is neither complicated nor difficult, yet is often overlooked and can be of critical importance (especially when there are multiple copies of files floating around on various machines or on the Internet).

Example 5

> *My files are a mess. I named them something I think I'll remember later but never do. Later when I try to find my files I usually have forgotten what I named them. I search by what I think I might have named them but rarely do I find them. I consistently have to redo work because of this.*

This scholar provides an exemplary case of why developing and utilizing a consistent file naming practice is vitally important. This researcher's practice could already be vastly improved with the simple addition of the year date (2015) before each folder and file name. Just by including this basic information, she would no longer need to know the whole file name when searching; she would just need to remember the year in which he saved the file to start narrowing down her search. She could further refine this by adding the month (e.g. "2015-02"). Alternatively, if she chose not to add this information to each file, another way to tackle the problem would be to develop a system wherein she started a new folder each month titled by date; better yet would be to title each folder by date and subject (e.g. "2014-02-planning").

Example 6

> *Depends on what I'm doing. For chapter drafts I either use chapter 1.draft 1.doc or agriculture chapter.draft 1.doc. This lets me trace back each iteration of the chapter if I have questions about why I did what I did. Whether I start with chapter number or topic depends on whether I've figured out the organization of the book. I use similar schemes for work related items, i.e. survey syllabus.spring 2013.doc or Fall 2013 TA List.doc or Schuster rec ltr. Hoover grant.doc.*

This scholar has an excellent grasp on how to name her files for easy searching and retrieval. She applies sound versioning practices, and her folder structure ensures that associated items are grouped together. The only suggestion the authors might have is that she stops using periods (.) and spaces to define separate versions. Spaces introduce unnecessary characters and periods tell computer operating systems to look for a file type. In some instances this could affect search results, file sharing capabilities, and publishing to the web. Instead the authors would advocate using hyphens or underscores as discussed earlier.

CONCLUSION

The importance of sound file naming strategies to maximize the use and access of data files cannot be overstated. Academics create many files every day; if not organized, over time it becomes increasingly difficult to easily and quickly access specific files. Published in 2001, the Library of Congress 'Naming Conventions for Digital Resources' established the standard by which information scientists have subsequently based their own conventions (Colorado University Libraries, 2008; Minnesota State Archives, 2012; Puglia, Reed, & Rhodes, 2004). The social sciences and humanities have responded very differently,

however, with little explicit attention paid to the topic whatsoever.[3] Specific institutions may have their own policies to facilitate data retrieval and collaboration, but it is becoming increasingly more important for individual scholars to develop sound personal practices—especially amidst rapidly proliferating data files, and growing emphasis on the need for sound data management plans. Adhering to some basic best practices for organizing digital materials will maximize materials' utility and accessibility. From file naming conventions to folder structures to systematic backups, if research materials are appropriately and systematically organized and named, one will always know what each file is and where to find it when needed.

While this paper explores examples from humanities and social sciences scholars in particular, the available literature about personal information management illustrates that these dilemmas are not isolated to social science and humanities professionals. Rather, the examples provided here concern basic issues encountered by many scholars on a daily basis. Fortunately, none of the obstacles noted are difficult to overcome. They are all easily remedied and are worth addressing in order to save time, energy, and frustration whenever there is a need to access a file on a computer.

The authors acknowledge that the systems suggested here are not the only ones or even the best in any given circumstance. The examples and suggestions provided from the authors' own work and that of colleagues' are meant to help illustrate how we think about these issues—and why having any system is so important. It is especially important that scholars involved in shared projects develop a system and that they all know how to version (or at least identify and appropriately reference different versions of) the paper or project under revision. As co-authors who have collaborated on a variety of papers, presentations, and publications—both with each other, and with other colleagues on different projects—the authors cannot overstate the importance of establishing common practices on joint projects. Establishing shared file naming conventions will facilitate content-based interactions as time can be dedicated to thinking about and working on the project itself instead of time being wasted making sure that everyone is looking at the same sets of materials and versions of the paper.

[3] If ever mentioned, file naming appears in methods-centered articles that concern analysis of qualitative data.

REFERENCES

Adamick, J., Reznik-Zellen, R, & Sheridan, M.. (2012) Data management training for graduate students at a large research university. *Journal of eScience Librarianship, 1*(3), 180-188. http://dx.doi.org/10.7191/jeslib.2012.1022

Alberta Government. 2005. *Naming conventions for electronic documents*. Retrieved from http://www.rimp.gov.ab.ca/documents/publications/DocumentNamingConventions.pdf

Arms, C. R. (1996). Historical collections for the National Digital Library: Lessons and challenges at the Library of Congress. *D-Lib Magazine*, (April). http://dx.doi.org/10.1045/april96-c.arms

Colorado University Libraries. (2008). *Guidelines on file naming conventions for digital collections*. Retrieved from http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

Crowder, J. (2013). Becoming Luis: A photo essay on growing up in Bolivia. *Visual Anthropology Review*, *29*(2), 107-122. http://dx.doi.org/10.1111/var.12008

Eakins, J. P., & Graham, M. E. (1999). *Content-based image retrieval*. JISC Technology Applications Programme, Report 39.

Jablonko, A. (1989). Filing and retrieval of visual documentation. *Visual Sociology*, *4*(2), 141-150. http://dx.doi.org/10.1080/14725868908583647

Jones, W. (2007). Personal information management. *Annual Review of Information Science and Technology*, *41*(1), 453-504. http://dx.doi.org/10.1002/aris.2007.1440410117

Khoo, C., Luyt, B., Ee, C., Osmana, J., Lim, H. H., & Yong, S. (2006). How users organize electronic files on their workstations in the office environment: A preliminary study of personal information organization behavior. *Information Research*, *12*(2), paper 293. Retrieved from http://www.informationr.net/ir/12-2/paper293.html

Kuhn, M. (1995/2004). *A summary of the international standard date and time notation*. Retrieved from http://www.cl.cam.ac.uk/~mgk25/iso-time.html

Library of Congress. (2001). *Naming conventions for digital resources*. Network Development and MARC Standards office. Retrieved from http://www.loc.gov/marc/naming.html

MacQueen, K. & Milsten, B. (1999). A system approach to qualitative data management and analysis. *Field Methods*, *11*(1),27-39. http://dx.doi.org/10.1177/1525822X9901100103

McLellan, E., MacQueen, K. M., & Neidig, J. L. (2003). Beyond the qualitative interview: Data preparation and transcription. *Field Methods*, *15*(1), 63-84. http://dx.doi.org/10.1177/1525822X02239573

Marion, J. S. (2008). *Ballroom: Culture and costume in competitive dance*. Oxford: Berg.

Marion, J. S. 2010. Photography as ethnographic passport. *Visual Anthropology Review*, *26*(1), 24-30. http://dx.doi.org/10.1111/j.1548-7458.2010.01045.x

Marion, J. S. (2013). Competitive ballroom dance. Berg encyclopedia of world dress and fashion. *Berg Fashion Library Online*, Volume 10, Online Exclusives. Retrieved from http://www.bergfashionlibrary.com/view/bewdf/BEWDF-v10/EDch10714.xml

Marion, J. S. (2014). *Ballroom dance and glamour*. London: Bloomsbury.

Marion, J. S. & Crowder, J. W. (2013a). Maximizing workflow and metadata. *Anthropology News*, *54*(3-4), 20-22.

Marion, J. S. & Crowder, J. W. (2013b). *Visual research: A concise introduction to thinking visually*. London: Bloomsbury.

Minnesota Historical Society. (2004). *File naming*. Retrieved from http://www.mnhs.org/preserve/records/electronicrecords/docs_pdfs/erfnaming.pdf

Minnesota State Archives. (2012). *Electronic records management guidelines*. Retrieved from http://www.mnhs.org/preserve/records/electronicrecords/erfnaming.html

Mitchell, C. (2011). *Doing visual research*. Thousand Oaks, CA: Sage.

North Carolina Department of Cultural Resources. (2008). *Best practices for file naming*. Retrieved from http://www.ncdcr.gov/Portals/26/PDF/guidelines/file naming.pdf

Peters, R. E. (2002). Organize this! Investigating personal information management practices. In *Proceedings of the conference on designing interactive systems*. New York: ACM Press. Retrieved from http://www.cc.gatech.edu/gvu/ii/persona/disPIMpaper.doc

Pink, S. (2007). *Doing visual ethnography* (2nd ed.). Thousand Oaks, CA: Sage.

Postma, M., & Crawford, P. I. (Eds.) (2006). *Reflecting visual ethnography: Using the camera in anthropological research*. Leiden: CNWS Publications.

Puglia, S., Reed, J., & Rhodes, E. (2004). *Technical guidelines for digitizing archival materials for electronic access: Creation of production master files – Raster images*. Retrieved from http://www.archives.gov/preservation/technical/guidelines.pdf

Rose, G. (2007). *Visual methodologies: An introduction to the interpretation of visual materials* (2nd ed.). Thousand Oaks, CA: Sage.

Spencer, S. (2010). *Visual research methods in the social sciences: Awakening visions*. New York: Routledge.

Strong, M., & Wilder,L. (Eds.) (2009). *Viewpoints: Visual anthropologists at work*. Austin: University of Texas Press.

University of Illinois. (2013). *Library digital content creations > 2.0 best practices for file naming*. Retrieved from http://www.library.illinois.edu/dcc/bestpractices/chapter_02_file naming.html

University of Oregon Libraries. (2010). *File naming and tracking changes (Version control)*. Retrieved from http://library.uoregon.edu/datamanagement/file naming.html

Zastrow, J. (2014). The digital archivist, PIM 101: Personal information management. *Computers in Libraries*, *34*(2). Retrieved from http://www.infotoday.com/cilmag/mar14/Zastrow--PIM-101--Personal-Information-Management.shtml

Zhang, A. (2007). Creating online historical scrapbooks with a user-friendly interface: A case study. *D-Lib Magazine*, *13*(11/12). http://dx.doi.org/10.1045/november2007-zhang