CLINICAL PSYCHOLOGY
SCIENCE AND PRACTICE

Archiving for Psychologists: Suggestions for Organizing, Documenting, Preserving, and Protecting Computer Files

Jamie DeCoster, Center for Advanced Study of Teaching and Learning, University of Virginia
Jamie O'Mally, Institute for Social Science Research, The University of Alabama
Anne-Marie R. Iselin, Duke University

Psychological researchers create a large number of files as part of their work, including study stimuli, assessment forms, data sets, analytic output, and manuscripts. We argue that it is fundamentally important that psychologists develop systematic ways of archiving these files. A well-designed file archive can greatly improve the efficiency of locating information, the security of stored files, the ability to recover from human and mechanical errors, the generation of future studies, and the sharing of knowledge with other psychologists. A survey of clinical psychologists demonstrated a need for greater knowledge and training in archiving. To address this issue, we describe the abstract demands that a file archive must meet and then provide concrete suggestions on how to meet these demands.

Key words: **computer files, data archiving, documentation, file archiving, organization, preservation, protection.** ***[Clin Psychol Sci Prac 18: 246–265, 2011]***

The purpose of psychological research is to generate knowledge about human thoughts and behaviors by describing them, explaining them with theories, and finding ways to alter them with interventions. As part of this process, psychologists commonly create participant consent forms, study materials, treatment protocols, assessment forms, data sets, analysis programs, and statistical output. They must also submit applications seeking approval of the study from local review boards, and many commonly submit grant applications to receive external funding. Once the work is completed, it is summarized and presented to others in the form of journal articles, book chapters, conference presentations, policy briefs, and institutional reports. Both the materials created while conducting research as well as the final reports provided to others are generally maintained as computer files. Efficiently handling the vast array of information contained in their computer files is therefore an important task for all psychologists.

Address correspondence to Jamie DeCoster, Center for Advanced Study of Teaching and Learning, University of Virginia, 350 Old Ivy Way, Suite 100, Charlottesville, VA 22903. E-mail: jamied@virginia.edu.

FILE ARCHIVING IN CLINICAL PSYCHOLOGY

Psychologists have been regularly working with computers since the 1980s. Given that many of the prominent research laboratories and clinics have been handling electronic files for decades, an important question is whether the field has already successfully dealt with the issue of archiving. Do psychologists need more information about how to maintain their files? What is the value of our proposed discussion of file archiving? To answer these questions, we surveyed members of the Society of Clinical Psychology (APA Division 12) to determine what file archiving practices clinical psychologists currently use, how these have affected their professional performance, and how important they felt it was for themselves, their students, and their subordinates to receive training in file

archiving. The items in the survey were selected to represent a variety of potential issues related to these topics.

Respondents were recruited through e-mails sent to the APA Division 12 mailing list on February 13, 2008, and May 14, 2008. A total of 146 individuals completed the survey. Respondents did not receive any compensation for their participation. Approximately 87% of the respondents (127) held doctorates, whereas 13% (19) had master's or bachelor's degrees. The survey was conducted online and was hosted on a secure server at The University of Alabama.

In the survey, we first asked respondents whether they had ever experienced a number of problems that commonly occur as a result of poor file archiving. The results, presented in Table 1, provide clear evidence that the respondents experienced many of the problems we described. A majority of the respondents had to redo work after losing a file, had to stop working on a document because a file they needed was in a different location, and had wasted time working on an out-of-date version of a file. These findings suggest that providing psychologists with information about file archiving techniques could greatly increase their productivity.

We then asked respondents to report their beliefs about the importance of file archiving and whether they felt a need for greater training on this topic. The results, presented in Table 2, suggest that although psychologists feel that training in file archiving is important, they have not typically received such training themselves, nor have they provided training for their students and subordinates. The majority of our respondents are currently unsatisfied with their knowledge of file archiving techniques, an issue we hope this article will begin to address.

Table 1. Respondents' experiences of problems resulting from poor file archiving

Question	Percent Responding "Yes"
Have you ever had to redo work because a saved file was lost or damaged?	75
Have you ever had to stop working on a document because you needed a file that you kept on a computer in a different location?	76
Have you ever lost information on a portable storage device (e.g., floppy disk, CD, USB memory stick, etc.) that wasn't backed up?	38
Have you ever wasted time editing an out-of-date version of a document or analyzing an out-of-date data set?	65
Have you ever re-analyzed a study and gotten different results without understanding why?	23

Table 2. Respondents' opinions about training in file archiving

Question	Percent Responding "Yes"
Have you received instruction on file archiving?	12
Have you taught your students/supervisees how to archive their files?	20
Do you believe that graduate students in your field should receive explicit training in file archiving?	83
Do you personally feel satisfied with the file archiving techniques that you use?	40

The primary goals of this article are to present a set of guidelines for creating a successful file archive and explain how following these guidelines can benefit psychologists. We will specifically suggest procedures that would prevent each of the file archiving problems presented in Table 1. Our discussion will focus on the archiving of computer files, although the general principles apply to archiving information in any medium. Almost all of the information currently used in social science research is, at some point, translated into an electronic format. Even tools that are primarily used in paper form, such as questionnaires, consent forms, and assessment protocols, are commonly stored as computer files and printed only when needed. Because so many of the materials psychologists create are stored electronically, appropriately archiving computer files is key to ensuring that they are kept available, usable, and secure.

The suggestions we provide below emphasize how to archive files related to research. We chose to not focus on archiving information related to clinical practice for three main reasons. First, the nature of materials archived in clinical practice varies greatly based on many factors, making it difficult to devise a single set of guidelines that could be applied broadly. American Psychological Association (2007) record-keeping guidelines note that the exact information included as part of a clinical record depends on the professional context in which the information was generated, the legal requirements of the jurisdiction and institution in which the psychologist is practicing, and case-specific

ethical concerns. Second, practicing clinicians do not necessarily want to retain all of the information they gather on a consumer. Recording clinical speculations about a consumer or qualitative thoughts about therapist–consumer interactions may be detrimental to clinical processes and could create legal difficulties in certain situations. It is therefore difficult to suggest general guidelines for archiving clinical information because there will be individual factors that will determine what information should be preserved and what information should be purged. Third, many institutions use commercial software programs to structure the way clinical files are saved, reducing the role of personalized file archiving systems. Finally, the field of health informatics specifically addresses archiving, accessing, and using clinical information, so we feel that this topic has already received coverage in the literature (Hovenga, Kidd, Garde, & Hullin Lucay Cossio, 2010).

An effective file archive must successfully meet four demands. First, it must provide a structure to organize how files are stored so that each file can be easily retrieved by anyone needing to access it. Second, the archive must contain documentation on the files so that the information in each file can be used appropriately. Third, the archive must keep backup copies of the files so that they can be repaired or replaced in the event that a file is damaged or lost. Finally, the archive must protect its files against unauthorized access. Below, we consider each of these demands in a separate section. For each demand, we first explain why it is important that the demand be met and what problems can occur if it is not. We then provide specific suggestions that psychologists can implement to help their archives meet the demand. While we focus on archiving research files, the general principles and some of our specific suggestions will apply to the archiving of information used in clinical practice. For example, the suggestions we provide can help clinicians meet American Psychological Association (2007) and Department of Health and Human Services (2002) guidelines for record keeping.

There are entire areas of study, such as bioinformatics, neuroinformatics, and biomedical informatics, addressing file archiving in fields outside of psychology. Many of these fields proliferated owing to the rapid accumulation of knowledge in biological and genomic research. Data sets in these fields are large and complex, frequently based on data from research collaborators with diverse backgrounds (e.g., molecular biologists, computer programmers, and data technicians; Achard, Vaysseix, & Barillot, 2001). Letovsky (1995) stressed the importance of keeping advances in file organization on par with expansions in scientific knowledge. We hope that the guidelines we set forth here similarly give psychologists useful ways to keep their archiving practices on par with the expansion of their scientific knowledge.

ORGANIZATION

A well-organized archive allows its users to quickly navigate to the directory where the file they want is located and then easily identify that file within the directory. There is no benefit to storing information in an archive if it cannot be found and retrieved by those needing the information at a later point in time. Poorly organized archives have inconsistent naming conventions and have directory structures that vary between projects, vastly increasing the amount of time it takes to locate files. Files could be lost completely if they are stored in a disorganized archive, and the individuals who originally worked on the files are unavailable or forget where the files are located (Weise, 2009).

It is important that an archive prevents *information confusion*, which occurs when users of the archive search for a piece of information and can identify multiple files that might contain what they want. This most commonly happens when a user has trouble identifying which is the most recent version of a file. The confusion can occur either because the files are named in an unclear fashion or because different versions of a file are stored in different locations using the same name. It is also possible to have information confusion within a data set when users have trouble determining which variable contains the values they want to use in an analysis. Information confusion forces people to waste time examining the specific contents of files to identify the information they contain.

The specific organizational structure that works best will vary from archive to archive, but it is important that everyone using the archive organizes their directories in the same way. This allows users

to easily identify any file they need, even if they did not originally contribute it to the archive (Weise, 2009). Individuals joining a new research group should be introduced to the organizational guidelines as soon as possible so that they can efficiently locate files in the archive and to prevent them from adding files and directories in a way that would confuse other users.

Suggestions to Improve Organization

Write Down a Set of Archiving Guidelines. The easiest way to ensure consistency among individuals using the archive is to write down a set of guidelines that users of the archive are expected to follow. These guidelines should describe how the directories should be structured, how files should be named, what files are to be kept in the archive, how file backups can be accessed, and how the security of the archive will be maintained. These guidelines can be updated as needed and should always provide current information about how the archive should be used.

Make Each User Responsible for Maintaining His or Her Own Files Within the Archive. Although it is possible to assign one person the responsibility of archiving all of the files, this is problematic for several reasons. First, it introduces a lag between the time when work is completed and the time when the files become available in the archive. Second, if the individual responsible for the archive is not directly involved in a project, he or she will have a limited understanding of its files and may document or organize them poorly. Finally, making a single individual responsible for archiving limits the number of people who know where things are in the archive. If that individual should become unavailable, others using the archive could find it very difficult to locate the files they need.

The procedures we suggest do not require a great deal of computer literacy, but they do require that users know how to copy, rename, and move files, as well as how to create new directories. Most archive users will already know these skills, but the need to train those who do not represents one disadvantage to making everyone responsible for archiving their own files. Anyone needing to learn file and directory management skills in either Windows or Mac environments should investigate the free online educational technology tutorials provided by Michigan State University (2010).

Use a Hierarchical Structure. The simplest and most straightforward way to organize the directories in an archive is to use a hierarchy. We suggest the following hierarchical levels.

(a) Investigator. An investigator would be any individual who is responsible for a set of files in the archive. When multiple investigators collaborate on a project, the investigator who owns the rights to the data being collected should be responsible for that project's files. Student investigators will commonly have their own directories in the archive and should be responsible for files associated with their own research projects. Individuals who are only interested in archiving their own files can skip this level of the hierarchy.
(b) Project. Each investigator will be responsible for one or more projects. A project is a collection of studies that are all designed to investigate the same basic constructs, theories, and hypotheses.
(c) Study. Each project will be associated with one or more studies, each representing data collected at a specific time and location.

Using these levels provides a natural way to keep related files together. It also makes it easy to give people access to files from a particular project or study while leaving the remainder of the archive secure. Individuals entering data only need to have access to the directories containing the necessary data files. Collaborators, on the other hand, can be given access to an entire project directory.

A hierarchical directory structure must balance the number of levels in the hierarchy with the number of files in each directory. A hierarchy with a small number of levels leaves too many files in each directory, making it difficult to find the one you want. Increasing the number of levels will reduce the number of files in each directory, but requires you to spend extra time clicking through the different levels to get to the directory containing your file. An efficient hierarchical structure has the fewest levels necessary to allow the

contents of the directories to be easily searched. You should try to limit the number of files contained in your directories so that the full list can be viewed on a single screen. However, there are other considerations, such as maintaining a consistent directory structure throughout the archive, that may be more important and cause you to violate this guideline.

Use Descriptive Names for Directories and Files. We recommend that you make your file and directory names as descriptive as possible. Using unique file names is one of the easiest ways to improve your ability to navigate through your archive (Marshall, 2007b). You should use meaningful names that will allow you and others to quickly know what type of information is contained in the directory or file. The more information you can include in a file name, the easier it will be to later identify the file (Smith, Budzeika, Edwards, Johnson, & Bearse, 1986). For example, instead of naming a file ''stdy3irb.doc,'' it would be more informative to name it something descriptive like ''Attribution study IRB proposal 2007-01-12.doc.'' Many psychologists developed their naming habits on older computer systems that limited file names to eight or fewer characters. However, both Macintosh and Windows operating systems now allow you to have file names of any length, as long as the full path name of the file is fewer than 255 characters.

Limit the Use of Special Characters in File and Directory Names. File names on modern systems can include letters, numbers, and a number of different special characters. However, you should be careful when using special characters in your file names because they may cause unexpected results when referring to the files within computer programs. For example, many programming languages have users put quotes around the names of files when loading them, so including quotes in your file or directory names can confuse the program. This is even true when the quote is used as an apostrophe (such as in ''Mark's files''). You should also be careful when using commas or spaces, because sometimes programs use these to separate elements in a list.

Take Advantage of Alphabetization When Naming Directories and Files. Most computer systems automatically present lists of directories and files in alphabetical order. Keeping this in mind while naming your directories and files can make these lists easier to search. We recommend three specific naming conventions that take advantage of alphabetization.

(a) Add numbers to the beginning of directory or file names if you want them to be displayed at the top of the list (such as ''00 Uncompleted forms''). Numbers come before letters alphabetically, so a name starting with a number will always be listed before names starting with letters. Lower numbers come before higher numbers, so you can even specify the exact order of your files or directories by coordinating the numbers you add to the front of the names.
(b) Always use ''leading zeros'' when enumerating files. For example, if you have 25 different images for a study, you should name the first file ''image01.bmp'' instead of ''image1.bmp.'' This gives all file names the same number of characters and ensures that they will sort in numeric order when alphabetized. Without leading zeros, the files would not alphabetize in numerical order because the names ''image10.bmp'' to ''image19.bmp'' all come before ''image2.bmp,'' and the names ''image21.bmp'' to ''image25.bmp'' all come before ''image3.bmp.''
(c) Always reference dates using the format YYYY-MM-DD. For example, the date March 9, 2010, should be referenced as 2010-03-09. Dates written in this format will alphabetize from earliest date to latest date. Dates written in more typical formats, such as MM-DD-YY, do not alphabetize in any meaningful order. You should also avoid using the names of months or seasons in your file or directory names, because these will not sort in chronological order when the names are alphabetized.

Include the Revision Date in File Names. To help keep track of the different versions of a file, you should add the date that the file was last modified to the end of its file name. This makes it easy to identify the most

recent version of a file, because it will have the most recent date. It also makes it easy to retain older versions of files because they will have different names than the most recent version. If you want to save different versions of a file created on the same day, you can add a letter after the date (e.g., ''raw data 2009-04-02a.sav''). In this case, the version with the highest letter will be the most recent version. When adding the revision date to your file names, you should keep the root name of the file constant to make it easier to locate prior versions of the file. The date can be placed either in front of the root name (such as 2010-12-14 IRB form) if you want the files in the directory to be sorted by date or after the root name (such as IRB form 2010-12-14) if you want the files in the directory to be sorted by the root name.

Archive Files in a Single Location. Having a single location for your archive makes it easier to locate files because you do not have to switch computers or drives to find which one has the file you want. It simplifies the procedure for making backups, because all of the files can be easily copied at the same time. Storing the archive across multiple computers can also lead to information confusion if a file is updated in one location but not in another.

Access Your Archive Remotely. Instead of transferring files to work on them from multiple locations, you might consider using the Internet to access your files remotely. This is one element of ''cloud computing,'' where files and programs are not tied to specific computers, but instead are more broadly accessible (Armbrust et al., 2009). Most commonly used operating systems have programs that allow you to remotely access other computers through an Internet connection[1]. Remote access provides several benefits over copying files between computers. First, remote users do not have to spend time making copies of their files before moving to a new location. Second, remote users will never have to stop working on a project because they failed to copy a particular file that they needed. Third, using files remotely reduces the opportunities for information confusion, because all of the files are always saved to the same remote location, no matter what computer is used to access them. Finally, remote access encourages users to work directly within the archive, which helps ensure that the archive contains the most recent version of all of its files.

Use Virtual Drives. Another way to take advantage of cloud computing is to have your files stored on a virtual network drive instead of on one of your own computers. A virtual drive is a section of a remote computer that can be accessed (by those with the appropriate permissions) as a disk drive over a local area network or over the Internet. The virtual drive can be simultaneously accessed by multiple people at multiple computers. This eliminates the need to transfer files between collaborators working on the same project, because everyone can directly add or modify files on the same virtual drive. The computer staff at most institutions can set up virtual drives with little difficulty. Once established, virtual drives require minimal effort to maintain on the part of the psychologist. Most institutions back up their mainframes on a daily basis and store copies of the backups in remote locations, providing superior protection for the archive.

Those who do not work at an institution providing virtual drives might choose to maintain their own web server. However, this can be challenging and, if not done correctly, can leave your files vulnerable to hackers. A somewhat simpler and cheaper option would be to have users connect to a network-attached storage device. Those wishing to avoid the responsibility of managing their own storage can rent online disk space from a web-hosting company. Files stored on a web host can be accessed by multiple people at multiple computers, duplicating many of the functions of a virtual drive. Many times files saved on a web host cannot be accessed directly, but must instead be accessed through a web browser. However, programs such as Webdrive (2009) allow you to access files on a web host as if they were on a normal disk drive. Some web hosts, like Dropbox (2009), work by installing a program that automatically allows you to access the web drive like a normal directory on your computer. This makes accessing the files much easier and prevents difficulties associated with forgetting to upload or download needed files from the web host.

Use Shortcuts Instead of Copying Files. Sometimes you may want to access the exact same file from multiple

locations in the archive. Instead of copying the file to each location, you should store the file in one location and add ''shortcuts'' (called ''aliases'' on Macintosh systems) to the file in the other locations. A shortcut is not an actual copy of a file. Instead, a shortcut contains a link to a file in another directory. When you click on a shortcut, the computer opens the linked file in its original location. When that file is saved, it is saved in the original location rather than in the location of the shortcut. This helps prevent information confusion because you are only working with a single copy of the file, even if you have shortcuts in a number of different directories. This can be particularly useful when a single file is used by multiple people or multiple projects, because it ensures that everyone is always working on the same updated version of the file. If you instead had copies of a file in multiple locations, you might update the file in one directory and then forget to update the file in other directories. An additional benefit to using shortcuts is that they take up very little space, so creating shortcuts in multiple locations will use less disk memory than copying the actual file to each of those locations.

Create Common-Access Directories. Sometimes researchers may want to add directories to their archive outside of the standard hierarchical structure. These would be used to contain files that would be of use to people working on different projects. The exact directories you might want to create would depend on who is using the archive, but below are some examples.

- *Organizational files*. Researchers might create a directory to hold a copy of the archiving guidelines and other organizational documents important to all archive users, such as schedules, phone lists, and research administrator guidelines.
- *Final products*. Copies of the final versions of papers, presentations, or grants may be kept together in a publicly accessible directory so that they can be readily referenced by collaborators and easily provided to other researchers or clinicians asking for reprints.
- *Templates*. The archive may include a directory to store templates of any frequently used documents or forms. Some examples would be letterheads, posters, consent forms, IRB applications, and journal submission letters.
- *Materials*. The projects managed within a research group typically share a common theme and have similar methodologies. Any materials that are likely to be used in multiple projects should be kept in a directory that all archive users can access. Different types of materials can be kept in different directories for ease of use. Some examples of such materials would be scales, images, assessment protocols, audio files, video files, and technical manuals for equipment.

DOCUMENTATION

A file archive should include descriptions and explanations of its files so that the information they contain can be verified and used accurately. Properly documenting all aspects of a research study makes it much easier to write up its methods and results for publication. Good documentation is even more helpful when a psychologist wants to reuse components of a research study in future work. Research studies contain many details that are difficult to recall after the project is completed. If psychologists wish to reuse a component that has not been properly documented, they must typically invest a large amount of time to understand the original methodology before they can make use of the old materials. However, when the original study has been accurately and thoroughly documented, a few clicks through an organized archive could easily retrieve all of the necessary details (Weise, 2009).

Having well-documented projects makes it much easier to share relevant information with other investigators. Although it is possible to directly contact individuals interested in your work and personally explain your projects, it is much more efficient to document aspects of your projects as the work is performed and then provide this documentation to interested colleagues. This is particularly important when sharing data collected from a research study. The data management guidelines of the Food and Drug Administration and the National Institutes of Health state that researchers should retain detailed information about how a study was conducted and how variables in data sets are to be interpreted to facilitate the exchange of

scientific information (McCartney, Burchinal, & Bub, 2006).

Maintaining accurate documentation is especially important for psychologists who are simultaneously overseeing multiple projects. Researchers commonly run several experiments examining the same underlying concepts. As they develop their research paradigms, psychologists might experience confusion about which components were used in each study. Documenting this information will make it easier for psychologists to determine which components affected which outcomes. This can help researchers design future studies that will be more likely to produce significant findings and treatments that will be more likely to benefit their participants.

General Suggestions to Improve Documentation

Document the Source, Use, and Coding of Measures. You should maintain full documentation for any measures you collect as part of a study. Purchased assessments will typically come with manuals, but you will need to provide your own documentation for measures you create or obtain by other means. You should record the full citation of the measure as well as instructions for its administration and scoring. You may additionally include copies of any programs that automatically score the measure, information about the psychometric properties of the measure, and notes from your experiences using the measure, such as how long it takes to administer and whether respondents find particular items confusing. For scales, you should specifically note the presence of reverse-coded items and the scoring of subscales.

Write Up the Study Method as Soon as It Is Designed. There is no better time to write up the methodology of a study than when it is fresh in memory. You should therefore create a document reporting the basic purpose of the study and the details of the procedure immediately after its design has been finalized. Any alterations made to the procedures during the course of running the study should be reflected in this document. The description of the method can be more concise and less formal than would be necessary for a journal article, but it is still advantageous to write in a clear and structured format so that this document can later be easily transformed into the method section of a manuscript.

Write Up Study Analyses as Soon as They Are Performed. The analyses of any major research question should be documented as soon as they are conducted. This documentation should report who performed the analyses, when the analyses were performed, which data set was analyzed, and what question was being examined. It should contain a copy of the syntax program used to analyze the data. Finally, it should verbally describe the results in a form similar to what would be found in a journal article. The document should contain relevant tables and graphs from the analysis output, but irrelevant parts of the output should be excluded to simplify the presentation. If there are many analyses, you might consider adding a table of contents to make locating specific results easier. Most word processors can generate tables of contents that will automatically update as you change the contents of the document.

Document the Results of a Study Even If It Fails to Support Your Hypotheses. Sadly but surely, some studies will not provide significant findings. Although you will likely not be able to publish such studies, it is still important to document their methods and results. This information can help you determine which procedures work and which do not, preventing you from unnecessarily replicating unsuccessful studies or repeating uninformative analyses.

Document Changes Made to Your Files. You should maintain a ''version history'' for any files that you will be changing regularly, such as manuscripts, reports, analyses, and computer programs. To document the version history, we suggest that you include a table at the top of the file with one column indicating the version name or date and a second column indicating what changes were made to the file in that version. This will make it much easier for you to later find older versions containing sections that you deleted and undo changes to the file when needed.

Suggestions to Improve Documentation in Data Analysis

Use Descriptive Names for Variables in Data Sets. Using descriptive variable names makes it easier to select

which variables to use in an analysis and also makes it easier to read the statistical output. Consider using variable labels if you are using a statistical software package that supports them.

Coordinate Variable Names Across Data Sets You May Eventually Merge Together. When preparing the data for a study that makes use of multiple data files, you should try to name your variables in a way that makes it easiest to combine the files. If you will be merging data sets that have information on the same measures but different subjects (such as data sets from the same study conducted at different sites), then you should be sure that the data sets all use the same names for any variables they have in common. If you will be merging data sets that have information on the same subjects but represent different measurements (such as data sets representing assessments made on the same subjects at multiple time points), you will need an identification variable (such as the subject number) that uniquely identifies each subject. The identification variable must appear in each data set and have the same name across all of the data sets so that observations from the same subject can be linked together. Other variables should have unique names so that they can be added to the merged data set without conflicting with each other. You cannot have two variables with the same name in the same data set, so merging will be easiest if the non-identifier variable names are all unique. If the data sets being merged represent the same assessments over time, you might consider adding a suffix to the variable names indicating the time point to make them unique (e.g., MMSE_T1).

Maintain Accurate Codebooks for Raw Data Sets. We recommend creating codebooks for all raw data sets containing detailed information about each variable. Given access to the codebook, a naïve researcher should be able to analyze the data without checking an external source. If any of the variables in the data sets were collected from survey items, the codebook should report the exact questions and response options in the survey. We suggest giving the codebook a file name that is similar to the data set it explains so that they are listed next to each other in the directory.

Use Summary Data Sets for Analysis. It can be very difficult to find specific variables in the raw data sets if a study used a large number of measures. Assuming each scale item was recorded as its own variable, a moderately sized survey can easily result in a data set containing hundreds of variables. Navigating a data set of this size is a cumbersome task, even with an accurate codebook. As a solution to this problem, you should consider creating summary data sets that only contain the variables needed for analysis. Most investigators do not analyze the responses to individual trials or items, but instead only look at composite measures. A summary data set would contain these composites, but would drop all of the variables corresponding to specific items or assessment trials. Summary data sets should be saved as separate files, because it is important to retain the raw data sets in case you want to recalculate a composite measure. Researchers may choose to create multiple summary data sets from a single raw data set if they are interested in different aspects of the data.

Create Data Dictionaries for Summary Data Sets. Data dictionaries are to summary data sets as codebooks are to raw data sets. Data dictionaries should focus on the information an investigator would need to accurately analyze the data set. The data dictionary should provide the name of the data set to which it refers, as well as a short description of the study from which the data were collected. The remainder of the dictionary describes the variables in the data set. For each variable, the data dictionary should report its name and provide an explanation of the values it can take. For categorical variables, this requires that you report which values correspond to each possible category. For variables derived from scales, this requires that you report the range of possible values, and what high and low values mean theoretically. For naturally continuous variables, this requires you to report the unit of measurement. Every variable should be represented in the data dictionary; however, sets of similar variables can be described together in a single entry.

Apply Corrections to Raw Data Sets. If you discover an error in your data, you should correct it in the raw data set and then re-create any summary data sets that would be affected by the error. If you only correct the

error in a summary data set, you may forget to apply the correction the next time you use the raw data to derive a new summary data set.

Analyze Data Using Syntax. Most statistical programs give you the option of either selecting your analyses through pull-down menus or typing out a set of commands in a way that looks like a computer program. The latter method is referred to as ''syntax'' and is better than using pull-down menus for several reasons. The primary advantage is that the syntax files can be saved and rerun at a later point in time. The first time you do an analysis, it will likely take you longer to write a syntax program than it would to perform the analysis using the menus. However, it will take much less time if you ever have to run the analysis again. Modifications that require only small changes to a syntax program would require you to completely redo your analysis if you were using the pull-down menus. In addition, you can sometimes reuse parts of an earlier syntax program if you perform similar analyses on a different data set. Creating syntax files to analyze your data also provides you with a permanent record of the details of your analyses, making it easier to interpret the analytic output.

When using syntax, we suggest that you make your programs as complete as possible, so that you can run them on their own without having any human steps in the process. Ideally, your syntax should start by loading the raw data file and end with the completed analyses, performing all of the necessary steps in between. There are two main advantages of this. First, it gives you a complete record of all of the data preparation that is needed for the analyses, making it simple to describe your analytic procedure. Second, it makes it very easy to rerun the analyses should you decide to add or remove steps to your data preparation, or add or remove cases from your data set.

Include Comments in Programs. It is beneficial to embed detailed comments throughout programs (including syntax for analyses) to make it easier for you and others to later determine how the program functions. You should insert a comment at the start of the program to provide a general description of its purpose. Add comment headings at the beginning of each major section of the program that clearly explain what the program is doing at that point. Whenever you are creating a new variable, you should include a comment describing what the variable represents and how its values are assigned and interpreted.

PRESERVATION

An archive must accurately store information so that it can be retrieved and used at a later point in time. The archive should include methods to easily recover from both human and mechanical errors (Sprehe, 2005). There are two major types of threats to the integrity of files in an archive. First, the equipment on which the archive is stored can suffer catastrophic damage, destroying all of the files it contains. The archive must therefore contain procedures to back up its files and store them in a secure location so that the archive can be re-created in the event of a disaster. Second, individual files may become corrupted through user error, equipment failure, or viruses, so that the file still exists but does not contain the appropriate information. The archive must therefore also contain procedures to save prior versions of the files so that the information can be salvaged if the most current version is damaged. A successful archive must use separate procedures to protect against these threats; saving prior versions of your file in the archive will not help if the entire archive is destroyed, and the backup of a corrupted file will be just as unusable as the original.

Suggestions to Improve Preservation

Regularly Back Up the Archive. The need to replace or repair files is an eventuality, not merely a possibility. Backups should be made on a regular basis (McCartney et al., 2006) and should be automated when possible. How often you should back up your files depends on how often information changes in the archive and what technology you are using, but we suggest that you perform backups at least once a month. Multiple copies should be made of each backup, and copies of earlier backups should be kept for a period of time, even after more recent backups have been made. While it is not necessary that all users of the archive have direct access to the backups, everyone should know who to contact in case they need to retrieve their files. Backup files

should not be edited directly. Instead, they should only be used to replace or repair files in the archive.

Keep Backups in a Separate Location. At least one backup copy of the archive should be kept off-site. Fires, floods, and tornadoes can not only damage the equipment storing your archive but can destroy everything around it as well.

Make Sure the Most Current Version of Each File Is Saved to the Archive. The archive is not of much value if it only contains outdated files. Users should either work directly from the archive or otherwise ensure that any new or modified files are immediately copied to the archive.

Maintain Your Own Backups of the Archive. It is important to maintain your own backups even when using a virtual drive or web host. The individuals maintaining these systems typically back up the files daily, so you can usually rely on them to correct short-term problems. However, the institution or host will not always retain these backups for long periods of time. Some institutions delete backups older than a month or two to limit their liability if someone should happen to compromise the security of their servers. When using a virtual drive, you should find out how long backups are retained by the institution or host and then plan your own long-term backups accordingly.

It is particularly important to maintain your own backups when working with a web-hosting company. If you ever have a conflict with your provider, they can choose to cut off your access to the archive and its backups. While you may be able to address the situation through negotiation or legal means, losing access to your files during this period could cause many problems.

Retain Older Versions of Files. People sometimes accidentally delete important parts of a file that need to be retrieved later, or they make incorrect changes to a document that need to be undone (Marshall, 2007a). Having older versions of a file will help you reconstruct information that was incorrectly altered and help you re-create a file if the current version becomes corrupted. It is easier to re-create a document from an earlier version than to re-create it completely from scratch. Retaining old versions also allows you to undo changes that you made to a file if you decide that an approach you took in an earlier version is better. Sections that you choose to delete from the final version of one document might later be useful in other documents, so retaining the old versions could increase the efficiency of your future projects. Finally, retaining older versions of your files provides a record of your work on a project, should the originality of your ideas ever be called into question.

We suggest that you create a subdirectory to store the older versions of files in any directory containing files that change regularly. Every time a file gets updated, the older version should be moved from the current directory to the subdirectory rather than being overwritten or deleted. By storing previous versions in a subdirectory, the main directory will only hold the most current version of any file. This provides two benefits. First, it reduces the number of files you need to examine when searching through the directory. Second, it removes any questions regarding which copy of a file is the most recent version. This is particularly useful when multiple people are editing the same document.

Keeping older versions of your files should not have any adverse effects on your archive. As long as these files are retained in a subdirectory, they should have little impact on the daily use of the archive. Moving the files to a subdirectory is just as easy as deleting the files, so retaining them does not require any extra effort. As long as the files are not cluttering the main directories, they should also not interfere with your ability to locate files. You typically do not need to worry about having many old files in the same directory because you will only rarely need to access these files. Retaining these files does require disk space, but this is a minimal concern given the sizes of current storage devices. For example, the word-processing file for this article was 205 kilobytes, whereas the typical hard drive on a new computer today holds approximately 500 gigabytes. This means that one standard hard drive would be able to hold over 2,400,000 word-processing files. Retaining old versions of documents is therefore unlikely to cause storage problems.

Preserve the Original Raw Data Sets. The raw data from a study are typically modified multiple times before they are subjected to analysis. These modifications may remove invalid cases, transform or recode existing variables, or create new composite scores. Each time a change is made to a data set, however, there is a chance that an error may accidentally be introduced. It is therefore best to save a copy of the original raw data file without any modifications so that the data sets can be corrected in case an error is found. Those using the archive should be prevented from making unwanted changes to these files.

Never Work Directly From a Portable Storage Device. A portable storage device (such as a USB memory stick) should never be the only place that the most recent version of a file is stored. These devices are small enough to be easily lost and stolen, and so the information on them should never be thought of as secure. Memory sticks are also highly susceptible to power surges and are easy to damage even through careful everyday use. Although the chips used inside of memory sticks have been tested to survive over 10,000 rewrites (Lewis, 2005), the industry has never intended these devices for permanent storage, so the physical components are much less durable. The following is an anecdote posted online by someone who had her memory stick (which she refers to as a ''jump drive'') fail.

> Good way to lose all of your data!! I've had two… jump drives which have both failed after less than 1 year of use. In both cases, I lost ALL data on each. Before they failed, I thought they were wonderful—in fact I carried my ENTIRE professional data and document base in one of them. They were wonderful until they failed. [The company] offers no method to salvage the data. They offered to send me replacements if I send the drives back—drives which still contain sensitive and irreplaceable data. What a bargain, I get a replacement after losing everything. If you insist on jump drives, back them up daily!! They can fail without warning. (Stevens, 2006)

Floppy disks are an even less dependable alternative to memory sticks. They are made inexpensively, resulting in compromised read/write quality. Files saved on them should be expected to last no more than a couple of years (Slater, 2007). While CDs and DVDs are more reliable than floppy disks, they still have the potential to fade, making the files they contain unreadable. There is evidence that standard-quality CDs are only reliable for about 2 years, and higher-quality CDs are only reliable for about 5 years (Deacon, 2007).

If you decide to use a portable storage device to transfer files between computers, you should always copy the file from the device to the new computer before working on it. You should then save the updated version to the computer before copying it back to the portable storage device. This will ensure that there is always a copy of the most recent version of your file on a more permanent storage drive.

PROTECTION

Researchers need to be able to control who has access to the information contained in their files. Review boards and the federal government are paying greater attention than ever to the security of information collected as part of research studies and the provision of clinical services. Recent problems with unauthorized individuals gaining access to personal data have increased governmental (e.g., Johnson, 2006), professional (e.g., FDIC, 2007), and academic (e.g., Rutgers, 2006) concerns about how files are stored and accessed. An archive must protect its files from unauthorized access or modification both by those who regularly access the archive as well as from outsiders. However, it is important to balance the ability to protect your information with the ability of authorized users to access the information (Clark & Wilson, 1987). The usefulness of an archive can be severely diminished if the procedures used to protect its information prevent those using the archive from being able to access their files when they need them. Researchers must decide what security procedures they want to employ that will best balance the protection of their files with the ability to easily access the information they contain.

There are several different reasons that researchers may add protection to their files. Sometimes researchers will want to limit the ability of people to read files because they contain sensitive information. Other times they will want to allow users to view a file, but will want to prevent anyone from changing its contents. It

is also important to protect the archive as a whole from outside security threats such as viruses (which can sabotage the contents of files), keyloggers (which can record passwords entered into a computer), and sniffers (which can intercept information sent over a network). Finally, individuals should be careful about the physical security of their computers and storage devices to prevent them from being lost or stolen.

Suggestions to Improve Protection

Use Password Protection. Users should be required to provide a login and password before they can access files in the archive. This helps to prevent unauthorized users from viewing files. Each user should have a unique login, so that different users can be given access to different files.

Use Strong Passwords. A good password uses a combination of lowercase letters, capital letters, numbers, and symbols. You should not use any meaningful numbers (like years, phone numbers, or zip codes) because these can be easily discovered. You should also avoid using real words (in English or other languages), because those attempting to crack passwords commonly use programs that repeatedly try words listed in the dictionary. However, you might consider interlacing words with other characters (e.g., h2o7u3s^e) to create a strong password that is easy to remember. One of the best ways to generate a strong password is to make an acronym of a sentence that would only be meaningful to the user, like ''ilrbotfe'' for ''I like reading books on the fire escape.'' This can then be combined with numbers to make a very secure password.

Limit File Access Whenever Possible. As a general rule, you should not give people access to files that they have no reason to use. Those interested in file security often follow the ''principle of least privilege,'' where users are given the lowest level of access that is needed for them to perform their required tasks (Saltzer & Schroeder, 1975). This minimizes the damage that can result from mistakes and minimizes the threat to the archive that can result if an account is compromised.

Separate ''Read'' and ''Write'' Access to Directories. Most operating systems allow you to separately determine who can view and who can edit each file. Giving more people the ability to modify a file increases the chance that errors may be introduced into the file (Weise, 2009). You should therefore only give individuals ''write'' access when you specifically expect them to be editing the file. If you have only a small number of collaborators, it may be fine to allow everyone full access to the archive. However, it becomes more important to control access as the number of individuals using the archive increases. The details of how to set these permissions depend on the operating system and are beyond the scope of this article. However, the technical support branch of your institution will typically know how to help you establish the permissions for your archive.

Separate Access to Identified and De-identified Files. As part of the Health Insurance Portability and Accountability Act (HIPAA) of 1996, the U.S. Department of Health and Human Services was required to develop a standard to protect the identifying information of individuals receiving health care. The final version of these guidelines (DHHS, 2002) suggests that organizations and individuals make use of ''de-identified'' files to help protect the privacy of people whose data is contained in computer files. De-identified files exclude information that could be used to determine the identities of research participants or clinical consumers, such as the individual's ''name, street address, telephone and fax numbers, e-mail address, social security number, certificate/license numbers, vehicle identifiers and serial numbers, URLs and IP addresses, and full face photos and any other comparable images'' (DHHS, 2002, p. 53234). Those working with a data set who do not need this personal information can be given access to de-identified files, while the original identified files are preserved in a secure location. While many psychological research projects fall outside the legal boundaries of HIPPA, it is still good practice for researchers to restrict access to identified files to prevent identity theft and unauthorized access to private information.

Use Software to Protect Your Computer From Internet Attacks. You should always have virus protection and an Internet firewall active on your computers. Virus protection software, such as that provided by AVG (2010), McAfee (2010), and Norton (2010), will

prevent you from accidentally downloading or running infected programs. Firewalls block unauthorized users and programs from accessing your computer over a network. The Macintosh and Windows operating systems both come with firewalls, although others can be purchased or downloaded from the Internet. It is important that you keep both virus protection and firewall software up to date, because hackers are constantly generating new ways to break into computers. Most commercial products automatically download updates as they become available.

Do Not Open E-mail Attachments With Unfamiliar File Extensions. Hackers will often try to get access to computers by sending ''Trojan horse'' programs as e-mail attachments. These are files that, when opened, will change the settings of your computer to make it vulnerable to Internet attacks. You should therefore never open any attachments with file extensions that you do not recognize. There are many different types of programs that can make unwanted changes to your system, so you should consider most file types to be potentially harmful. Usually picture files, video files, and audio files are safe, but you should be wary of anything else. You should even be careful with attachments sent from e-mail accounts you know, because hackers can steal contact lists and pretend to be from a familiar account. If someone sends you a file that you were not expecting, contact them first to see if they really sent it before opening the attachment.

Only Download Software From Reputable Web Sites. Trojan horse programs are sometimes posted on the Internet instead of being sent through e-mails. You should therefore only download programs from web sites that you trust. Two reputable web sites that provide software are http://www.shareware.com and http://www.download.com.

Protect Yourself Against Keyloggers. Keyloggers are memory-resident programs that keep track of the key presses made on a computer. They are commonly used to spy on computer activity and to steal passwords. People will sometimes install keyloggers on publicly available computers, so you should always be careful about logging into any online accounts when you are not in your home or office. To help protect against keyloggers on your own systems, you can install key scrambling software such as Antilogger (Zemana, 2010) and KeyScrambler (QFX Software, 2010), which can encrypt your keystrokes so that keyloggers cannot interpret them. If you do happen to find a keylogger on your machine, you should probably reformat it because there is a good chance that someone has stolen an administrator password, which would compromise the entire system.

Encrypt Your Sensitive Files. One great way to protect your sensitive data is to use file encryption on your computers and storage devices. A file encryption program, such as TrueCrypt (2010), encodes the contents of a folder or drive so that it can only be accessed in its original form using a decryption key. This prevents the files from being opened even if they are stolen. It is particularly important to encrypt files on laptops, because these are most vulnerable to theft. However, it is important that you remember your decryption key. Good encryption software will not provide a way to access the files without the key, because this would present a security loophole.

Encrypt Sensitive E-mails. By default, e-mails are sent through the Internet without any encryption. Any computer along the message path can potentially read or distort the contents of the e-mail. People with access to the computers forwarding e-mails will sometimes use ''sniffer programs'' to scan the contents of messages in an attempt to find sensitive information in e-mails. It is therefore important that you always e-mail confidential information in an encrypted form. You can use programs such as iSafeguard (MXC Software, 2010) or PGP (PGP Corporation, 2010) to encrypt your e-mails so that they can only be read by the intended recipient.

Provide Physical Security for Any Devices Containing Files. Much attention has been given to protect computers against electronic threats, but many important security breaches have occurred because people have lost their computers or had them stolen (Privacy Rights Clearinghouse, 2010). You should avoid keeping sensitive information on portable storage devices and laptop

computers because these are often targeted by thieves. These devices should be kept out of sight, preferably in a locked drawer or cabinet, when they are not being used. When traveling, always keep laptop computers and other storage devices within your sight to discourage theft. Do not save passwords for web sites you visit when using a laptop computer. You should habitually lock the rooms containing desktop computers in your home or office. You should also consider chaining your computer components to the supporting furniture to make them more difficult to steal.

Erase Any Storage Devices Before Disposal. You should make sure that there is no sensitive information on a storage device before you throw it away. Otherwise, someone else might obtain the device and be able to extract the information it contains. Keep in mind that deleting a file does not actually remove it from your computer—it can still be retrieved if someone has the appropriate software. You might consider physically destroying hard drives by drilling holes in them and placing them near high-powered magnets to prevent anyone from being able to access their contents. If you do not want to destroy the device, consider using a permanent erasure program such as Active@ Killdisk (2010) or Wipedrive (White Canyon Software, 2010) to ensure that the original files cannot be retrieved.

One thing that many people overlook is that all modern photocopiers contain hard drives. These drives store images of documents scanned, faxed, or copied on the machine. If your office plans to sell, return, or retire a photocopier, you should make sure that the hard drive inside the photocopier is erased first. Some people intentionally purchase used photocopiers to find sensitive documents on the hard drives, which can then be used for fraud or blackmail (CBS News, 2010).

A SAMPLE ARCHIVE

Figure 1 illustrates the directory structure of a file archive for a fictional laboratory focused on autism research. We have designed this archive so that it embodies the principles suggested in the previous sections.

The ''00 Common Files'' directory contains files that would be used by all members of the archive. Inside of this directory, there are subdirectories for organizational documents, scales, technical manuals, and templates. The ''01 Final Products'' directory contains final versions of grants, papers, and presentations that were authored by members of the autism laboratory. This directory would not hold working copies of such products—only versions that are in a form that would be appropriate to distribute to people outside of the laboratory. When alphabetized, the numbers included at the beginning of their file names cause these two directories to be listed before the directories for individual psychologists.

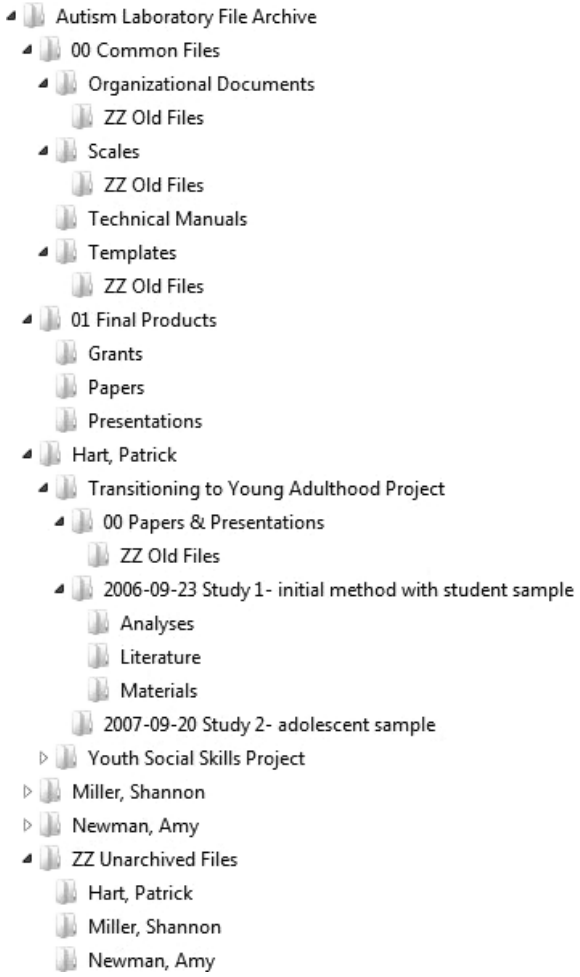

Figure 1. Example of an archive.

Next, there is a subdirectory for each psychologist in the laboratory. Underneath each of these directories is a subdirectory for each research project coordinated by the psychologist. Underneath each research project

is a subdirectory for papers and presentations and a subdirectory for each study associated with the project. Each study directory has subdirectories containing the identified data sets, de-identified data sets, analyses, computer programs, resources, and study materials. The data directories would include codebooks for each raw data set and data dictionaries for each analysis data set. The analyses directories would contain syntax programs used to analyze the study, outputs from these analyses, and analysis write-ups. Computer programs used as part of the study would be stored in their own directory. The literature directory would contain articles and other reference materials that were acquired to help develop or write up the study. Finally, the materials directory would include copies or shortcuts to documents used to conduct the study.

At the bottom of the archive is a directory named ''ZZ Unarchived Files'' to separate files contained in the archive from those outside the archive. The expectation is that all files stored within the archive will follow strict organizational and naming conventions, but no such expectations are held for files outside the archive. This provides a place to store files created before the archiving guidelines were implemented. The ''ZZ'' included at the beginning of the name ensures that this directory always appears at the bottom of the list.

You will notice that the archive uses the hierarchical structure we suggested in our section on organization. This makes it easy to allow collaborators access to files they need while restricting access to files they do not have permission to view. A simple way to assign permission levels parallels the hierarchical structure, so that individuals can be granted access to all of the files for an investigator, a project, or a study. All investigators would be able to read the files in the common directories, although the ability to edit the contents of these directories may be restricted to a single individual to prevent inadvertent modification of these files. There are separate directories for identified and de-identified data sets, making it easy to restrict access to participants' private data.

A specific naming convention would be applied to all files in the archive. First, the project name and study name would be included at the front of all of the file names so that it is easy to identify where each file came from and to help ensure that the files contained in the archive have unique file names. The file names would be long and descriptive to make it easy to identify what is contained in each file. Any files that might be updated would have the date of the last revision included at the end of the file name, making it easier to identify the most recent version of that file. Older versions of files would be moved to subdirectories to reduce the number of files contained in the main directories.

DISCUSSION

In this article, we have discussed the importance of establishing systematic guidelines for organizing, documenting, preserving, and protecting the computer files created by psychologists. Organization is important to make it easier to find information in the archive. Documentation is important to ensure that the information in the files is interpreted correctly and to facilitate the sharing of files with other psychologists. Preservation is important to keep the information being archived available and uncorrupted over time. Protection is important to keep unauthorized individuals from accessing sensitive information. We provided a number of specific suggestions to improve the way that psychologists organize, document, preserve, and protect files in their own archives.

We presented a specific example of a file archive that embodies our suggestions. Psychologists wishing to organize their own files might consider using our example as a starting point. Of course, the best archive is one that is tailored to the specific characteristics of the information it contains and that develops as the needs of the individuals using it become better understood. Our example might be used when psychologists are initially developing an archive, later modifying it as they see fit. The most important thing is to explicitly define the archiving guidelines (Weise, 2009). This way, any aspect of the archive that creates problems can be easily identified and corrected.

Those developing file archiving guidelines for their organization or research laboratory should train other users on the recommended procedures specifically and on file archiving more generally. The first priority should be to instruct users what files they should be working with, where those files are located in the

archive, and how the files they create should be named. New users should also be made aware of any security issues likely to directly impact their work, such as logon and password procedures, what types of programs are allowed on laboratory/organizational computers, and which files should not be moved off-site. Other aspects of archive management, such as the creation and retrieval of backups, the theory behind the organization of the directories, and the possibility of remotely accessing the archive, can be explained as users gain more experience.

Our discussion of file archiving has focused on the topics and issues that are most relevant to clinical psychologists, but we would like to suggest that many other types of researchers can benefit from these ideas. We feel that most of our suggestions can easily apply to those working in any of the social sciences. The specific nature of the research and collaborative environment may require some changes, but the organization, documentation, preservation, and protection of computer files are important concerns for all modern researchers, no matter what their specific content areas may be.

Working from an organized file archive can have benefits for research above and beyond increasing the efficiency of accessing information (Humphrey, Estabrooks, Norris, Smith, & Hesketh, 2000). By maintaining a more complete record of the procedures and results associated with a project, psychologists can make more accurate inferences about the topics being studied. Archiving also forces psychologists to thoroughly describe the characteristics of their methods, helping them to identify and correct any shortcomings. Finally, systematic file archiving requires psychologists to describe their projects in ways that can be easily understood by others. Taking this broader perspective can lead psychologists to devise methods and materials that are applicable to a wider range of topics, increasing the scientific impact of their projects.

It can be difficult to introduce a systematic file archiving approach to a laboratory that has already been collecting information for years in an unsystematic way. The prospect of renaming and reorganizing all of the old files to meet the new guidelines is usually daunting enough to halt any such transition in its tracks. We therefore suggest that those wishing to introduce new file archiving procedures select a point in time after which any new projects must be archived according to the new guidelines, but not attempt to correct older files. Adhering to archiving guidelines will be less disruptive when the procedures are in place at the start of a project (Humphrey et al., 2000). In addition, those working in the archive will more readily see the value of systematically archiving active projects than inactive projects, so focusing on the archiving of new files will increase adherence to the guidelines (Weise, 2009). It would be best if projects following the new guidelines were stored in a separate location, such as on a different computer or in a different directory, to help ensure the integrity of the new archive. This makes it clear that all of the projects stored in the new location must follow the new archiving guidelines.

We would like to note that individuals do not have to commit to following a complicated system to benefit from systematic file archiving. Many of our suggestions, such as using a hierarchical directory structure or adding dates to the end of file names, require little or no additional effort to implement once the guidelines have been established. Files must be stored in directories and must be given names, so this may as well be done using an efficient methodology. Other suggestions, such as creating data dictionaries or writing up the results from unsuccessful studies, do create additional work beyond what most psychologists currently put into archiving their files. The initial implementation of the archive will take some careful thought, and time must be spent developing the guidelines that will be followed (Sprehe, 2002). When archiving procedures are set in place, however, using the new procedures will become second nature. Systematic archiving methods can even increase productivity by preventing archive users from individually spending time deciding how to organize, document, preserve, and protect their files (Weise, 2009). Our personal experience has been that the cost of systematic file archiving will be notably less than the savings it provides by increasing efficiency. However, psychologists need to make their own decisions regarding what procedures are needed within their own file archives. While we believe that all of these procedures would benefit psychologists, we understand that individuals may be apprehensive about

making large changes to their current practices. In this case, psychologists can begin by incorporating our simplest suggestions, adding procedures that require more effort only when they have personally experienced the need for them.

The development of a systematic file archive takes dedication, care, and patience. It takes effort to set up the initial guidelines and to see that those using the archive consistently follow them. The benefits of establishing a systematic archive accrue over time, as users have more opportunity to retrieve, review, and share the files it contains. Psychologists who now use systematic file archiving practices often had to suffer from many setbacks owing to poor file archiving before they became convinced of the need for these procedures. It is evident from the data we collected that psychologists feel that training in data archiving is important. Unfortunately, this contrasts with the lack of scholarly information available on data archiving for psychologists. Although other disciplines, such as library sciences and business enterprise (e.g., Sprehe, 2002), offer general guidelines for data management, they have rarely provided detailed suggestions. Most sources stressing the importance of archiving suggest that individual enterprises devise guidelines specific to their needs (e.g., Sprehe, 2002). We have endeavored to provide such suggestions for the field of psychology. We hope that our discussion helps convince new psychologists and those who train psychologists of the value of employing good file archiving techniques from the very start of one's career. We also hope that those who already believe in the importance of systematic file archiving can benefit from the specific suggestions we provide on how to organize, document, preserve, and protect computer files.

NOTE

1. Microsoft Windows comes with a program called *Remote Desktop* that can be easily configured to remotely access other computers running under Windows. A version of this program is also available that allows you to access Windows computers from Macintosh computers (Microsoft, 2009). The program *TeamViewer* can also be used to remotely access Windows computers, but with a secure connection (TeamViewer, 2007). Apple provides an application called *Apple Remote Desktop* (Apple, 2007) that can similarly be used to remotely access Macintosh computers, although this software does not come with the operating system. The program *RealVNC* enables you to set up a system to receive remote connections across multiple operating systems, including Windows, Mac OS, and Linux (RealVNC, 2007).

ACKNOWLEDGMENTS

Support for this article was provided by NIH/NIDA grant R01DA018920 and NIH/NIMH grant T32MH018269. The file archiving techniques we describe were developed as part of the first author's work with the Center for Mental Health and Aging and the Institute for Social Science Research at The University of Alabama, and we would like to thank the members of both these groups for their help in refining our ideas. We would like to thank Joan Barth, Debra McCallum, and Cindy Roth for comments made on an earlier version of this article. We would also like to thank Lynn Snow and Donald Cervino for contributing suggestions.

REFERENCES

Achard, F., Vaysseix, G., & Barillot, E. (2001). XML, bioinformatics and data integration. *Bioinformatics Review*, *17*, 115–125.

Active@ Killdisk. (2010). *Active@ Kill Disk hard drive eraser. Low level format*. Retrieved August 26, 2010, from http://www.killdisk.com/

American Psychological Association. (2007). Record keeping guidelines. *American Psychologist*, *62*, 993–1004.

Apple. (2007). *Apple remote desktop 3*. Retrieved March 5, 2009, from http://www.apple.com/remotedesktop/

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., et al. (2009). *Above the clouds: A Berkeley view of cloud computing*. Tech. Rep. UCB/EECS-2009-28. Berkeley: EECS Department, University of California.

AVG. (2010). *AVG—Anti-virus*. Retrieved August 26, 2010, from http://www.avg.com/

CBS News. (2010). *Copy machines, a security risk?* Retrieved September 17, 2010, from http://www.cbsnews.com/video/watch/?id=6412572n

Clark, D. D., & Wilson, D. R. (1987). A comparison of commercial and military computer security policies. In *Proceedings 1987 IEEE Symposium on Security and Privacy* (pp. 184–194). Oakland, CA: IEEE Computer Society Press.

Deacon, N. (2007). *Using CDs as an archive medium*. Retrieved March 5, 2009, from http://web.ukonline.co.uk/suttonelms/articles9.html

Department of Health and Human Services. (2002). Standards for privacy of individually identifiable health information; final rule (45 CFR parts 160 and 164). *Federal Register*, *67*, 53182–53273.

Dropbox. (2009). *Dropbox—Secure backup, sync, and sharing made easy*. Retrieved September 1, 2009, from http://www.getdropbox.com/

FDIC. (2007). *FIL-32-2007: FDIC's supervisory policy on identity theft*. Retrieved March 5, 2009, from http://www.fdic.gov/news/news/financial/2007/fil07032.htmlhttp://psycprints.ecs.soton.ac.uk/archive/00000279/-html

Hovenga, E. J. S., Kidd, M. R., Garde, S., & Hullin Lucay Cossio, C. (2010). *Health informatics: An overview*. Amsterdam, The Netherlands: IOS Press.

Humphrey, C. K., Estabrooks, C. A., Norris, J. R., Smith, J. E., & Hesketh, K. L. (2000). Archivist on board: Contributions to the research team [19 paragraphs]. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* [Online Journal], *1*(3). Retrieved March 5, 2009, from http://www.qualitative-research.net/fqs-texte/3-00/3-00humphreyetal-e.htm

Johnson, C. (2006). *Protection of sensitive agency information*. Retrieved March 5, 2009, from http://www.whitehouse.gov/omb/memoranda/fy2006/m06-16.pdf

Letovsky, S. (1995). Beyond the information maze. *Journal of Computational Biology*, *2*, 539–546.

Lewis, B. K. (2005). *USB flash drives. Originally printed in the Sarasota PC Monitor*. Retrieved March 5, 2009, from http://www.spcug.org/reviews/bl0511.htm

Marshall, J. (2007a). *Five ways to prevent data loss*. Retrieved March 5, 2009, from http://wordprocessing.about.com/od/troubleshooting/a/dataloss.htm

Marshall, J. (2007b). *Six ways to keep your documents organized in Word*. Retrieved March 5, 2009, from http://wordprocessing.about.com/od/wordprocessingsoftware/a/6organize.htm

McAfee. (2010). *Malware detection, virus protection, Internet security, total protection*. Retrieved August 26, 2010, from http://www.mcafee-home.com

McCartney, K., Burchinal, M. R., & Bub, K. L. (2006). Best practices in quantitative methods for developmentalists. *Monographs for the Society of Research in Child Development*, 71. 9–23.

Michigan State University. (2010). *Educational technology tutorials*. Retrieved December 20, 2010, from http://edutech.msu.edu/online/index.htm

Microsoft. (2009). *Connect across platforms with Remote Desktop Connection*. Retrieved May 11, 2009, from http://wwwmicrosoft.com/mac/products/remote-desktop/default.mspx

MXC Software. (2010). *iSafeguard: Signing and encrypting files and emails made easy!* Retrieved August 26, 2010, from http://www.mxcsoft.com/

Norton. (2010). *Norton AntiVirus 2010, Norton Internet Security 2010 and Norton 360 Version 4.0*. Retrieved August 26, 2010, from http://antivirus.norton.com/

PGP Corporation. (2010). *The leader in file encryption software, hard drive encryption, and enterprise security*. Retrieved August 23, 2010, from http://www.pgp.com/

Privacy Rights Clearinghouse. (2010). *Chronology of data breaches*. Retrieved August 26, 2010, from http://www.privacyrights.org/sites/default/files/static/Chronology-of-Data-Breaches_-_Privacy-Rights-Clearinghouse.pdf

QFX Software. (2010). *QFX Software—Anti-keylogging software and more*. Retrieved August 26, 2010, from http://www.qfxsoftware.com/

RealVNC. (2007). *RealVNC remote control software*. Retrieved March 5, 2009, from http://www.realvnc.com/

Rutgers. (2006). *Identity theft compliance policy*. Retrieved March 5, 2009, from http://policies.rutgers.edu/PDF/Section50/50.3.9-current.pdf

Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE*, *63*, 1278–1308.

Slater, J. C. (2007). *How to archive your thesis/dissertation/project data*. Retrieved March 5, 2009, from http://www.cs.wright.edu/~jslater/archivingdata.pdf

Smith, P. C., Budzeika, K. A., Edwards, N. A., Johnson, S. M., & Bearse, L. N. (1986). Guidelines for clean data: Detection of common mistakes. *Journal of Applied Psychology*, *71*, 457–460.

Sprehe, J. T. (2002). *Enterprise records management: Strategies and solutions. A white paper prepared for Hummingbird LTD*. Retrieved August 25, 2009, from http://www.hummingbird.com/alt_content/binary/pdf/collateral/wp/rmstrategies.pdf

Sprehe, J. T. (2005). The positive benefits of electronic records management in the context of enterprise content management. *Government Information Quarterly*, *22*, 297–303.

Stevens, K. (2006). Great way to lose all your data!! *Amazon.com: Reviews for Lexar Media 128MB JumpDrive Sport JDSP128-231: Electronics*. Retrieved March 5, 2009, from http://www.amazon.com/Lexar-Media-128MB-JumpDrive-JDSP128-231/dp/customer-reviews/B0001CFSW2

TeamViewer. (2007). *TeamViewer instant desktop sharing*. Retrieved March 5, 2009, from http://teamviewer.com/

TrueCrypt. (2010). *TrueCrypt—Free open-source on-the-fly disk encryption software for Windows 7/Vista/XP, Mac OS X*. Retrieved August 26, 2010, from http://www.truecrypt.org/

Webdrive. (2009). *Webdrive managed file transfer, document collaboration software, secure FTP*. Retrieved August 17, 2009, from http://www.webdrive.com/

Weise, C. (2009). Best practices for electronic records management. *Infonomics*, *23*(2), 48–50.

White Canyon Software. (2010). *WhiteCanyon: Tools to erase hard drive data and prevent identity theft*. Retrieved August 26, 2010, from http://www.whitecanyon.com

Zemana. (2010). *Secure your Internet banking and make your online shopping safe*. Retrieved August 26, 2010, from http://www.zemana.com/

Received April 13, 2011; accepted April 17, 2011.