Managing Data Sources

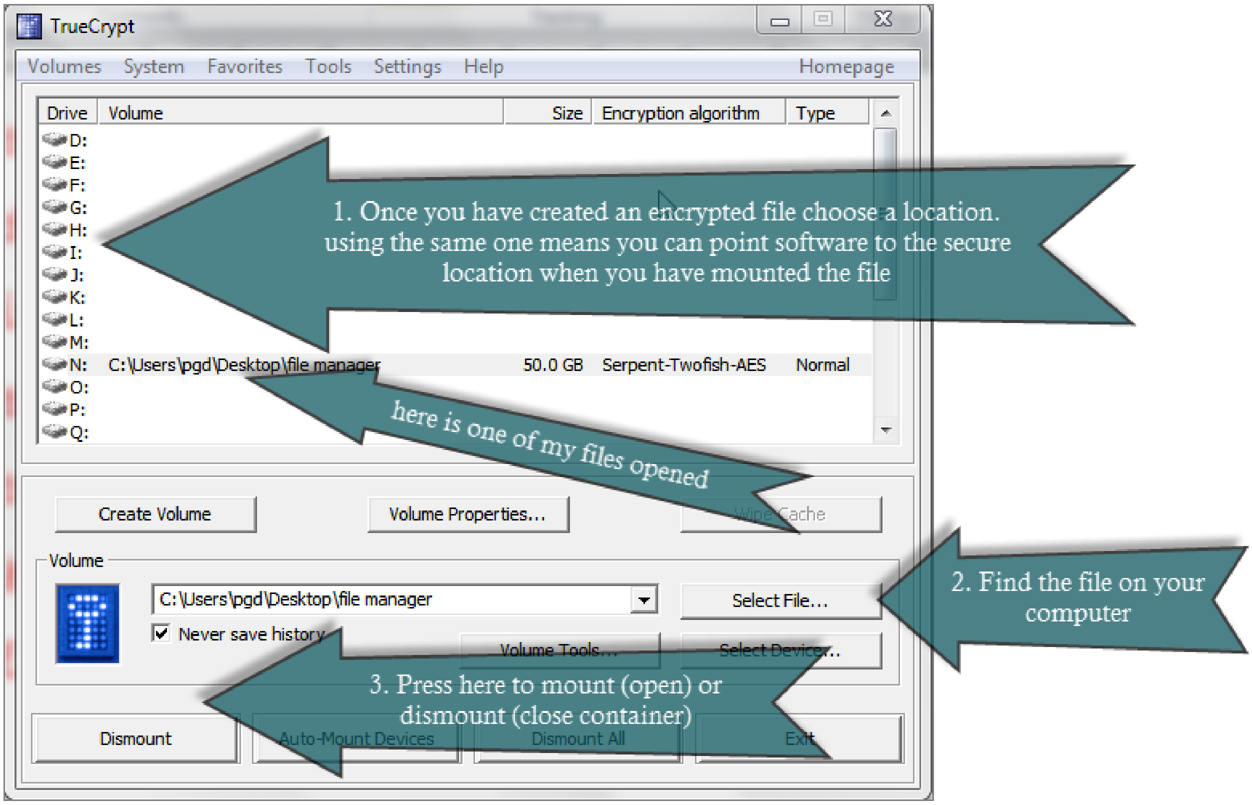
When dealing with datasets of any kind, there will always be security and ethical issues to consider. Being an ethical and reflexive researcher requires that you think carefully about the ways in which you will navigate the ethical dilemmas (both the expected and unexpected) that are bound to arise when generating small or large datasets. From the very moment you begin designing your research study, you must think about how you will protect, store, and eventually share your data.

Finding a way to systematically and efficiently manage your data will make your research process run more smoothly, safely, and transparently. We highlight some of the new policy initiatives that are shaping how researchers think about protecting data and sharing data for re-use.

Protecting Data

Well-designed research studies build in multiple ways by which to keep the data confidential and secure. Your ethical duty requires that you protect participants from harm, assuring that their identities are protected and that they are aware of the ways in which the data will be produced, stored, and shared. Institutional ethics boards typically have explicit requirements related to data storage, including when and how collected data should be stored and eventually destroyed. It is important to become familiar with your institution's guidelines, while also employing other protective strategies. In the UK you may also need to be aware of the data protection act. Institutions will have a dedicated officer who will be able to advise you.

A good place to begin your data storage process is by encrypting anything that you would not be comfortable having read by the general public. In other words, assume that personal information, sensitive data, and stored files (e.g., transcripts) must be encrypted. There are a variety of encryption programs that can support you in this process. One free and easily usable encryption software package is TrueCrypt. It uses multiple security algorithms to protect files. Any size of folder or hard drive can be encrypted, and once this encrypted space is created data can be added to it as to any external or hard drive. The secure drive appears in Windows Explorer as a separate drive yet blends in with the rest of your computer drives. Whatever you do, do not lose your password. Allegedly, even the FBI cannot crack it.



Drive	Volume	Size	Encryption algorithm	Type
D:				
E:				
F:				
G:				
H:				
I:				
J:				
K:				
L:				
M:				
N:	C:\Users\pgd\Desktop\file manager	50.0 GB	Serpent-Twofish-AES	Normal
O:				
P:				
Q:				

Volume: C:\Users\pgd\Desktop\file manager

☑ Never save history

1. Once you have created an encrypted file choose a location. using the same one means you can point software to the secure location when you have mounted the file

here is one of my files opened

2. Find the file on your computer

3. Press here to mount (open) or dismount (close container)

It is also crucial to store your data on authorised, password-protected desktops, memory sticks, hard drives, and/or mobile devices. Anti-virus and firewall protection programs serve to add another layer of protection. Assume that all digital storage formats will potentially fail. Thus, save your data in more than one secure location. If you have paper-based data sources (e.g., field notes), prior to shredding them, scan them and place them within your digital database. If you use cloud-based storage systems, keep in mind that this is less secure, with some ethics committees not even allowing researchers to store data online (see Chapter 2 for a discussion of the pros/cons of cloud-based note-taking systems).

Preparing Data for Sharing and Reuse

In many countries, funding agencies are now requiring that researchers with funded research studies develop data sharing and management plans. For example, the National Science Foundation and National Institutes of Health in the US now require that researchers explain how they will share their data with the larger scientific community. In the UK the United Kingdom Data Archive (UKDA) curates many social science and humanity projects.

Several recent science policy initiatives in the UK and US have resulted in an increased number of scientific data collections (Cragin & Shankar, 2006). These may include raw and/or already analysed data, metadata, supplemental information, links to other databases, or biographic information. Librarians and archivists, as well as computer scientists and programmers (amongst others), are involved in the creation and maintenance of these databases. Many universities now house digital repositories in which their faculty and graduate students can store and make public their data collections. There are also discipline-specific repositories. Social science data archives (e.g., Roper Center for Public Opinion Research) and topic-specific data libraries and archives (e.g., Princeton's Cultural Policy and the Arts National Data Archive) are becoming more common.

The pervading assumption is that if data is easier to access it will result in reuse. However, the little research focused on data reuse has pointed to the challenges inherent to analysing qualitative data outside of the context(s) in which it was collected (e.g., Berg & Goorman, 1999; Zimmerman, 2008). At present, there is minimal research examining how local context will and should be communicated to re-users. Additionally, much of the discussion around data use and reuse has revolved around quantitative research, with the natural sciences leading many such conversations. Thus, it is important for you, as a qualitative researcher, to think carefully about the assumptions of your particular methodological and theoretical approach as you develop a data management plan.

Regardless, one way to ensure that your data is ready for auditing and/or reuse is to carefully document your data. The qualitative data analysis software packages explored in Chapter 7 can be used for this purpose. Ask yourself: What would someone using my data for the first time need to know? Think about the most useful ways to share your data collection process, outlining in great detail the tools used, participants included, and data sources generated. While many qualitative researchers may never share their data in digital repositories, documenting your data is still a good idea. In many ways, assuming that your data might be reused will push you to create an audit trail.

Once you have encrypted your data and securely stored it, you will at minimum need to share the data with other members of the research team. Chapter 3 discussed some of the tools that support the collaborative research process, including tools such as Dropbox. While there are a variety of new tools, issues around security of such sites remains a concern. Anything in the “cloud” will never be entirely secure. For instance, when sharing video data, in lieu of delivering password protected hard drives or sending videos via unsecure email or via Dropbox (which most ethics committees will likely not approve), you might consider exploring whether your institution has secure server that you and your colleagues can both access in order to upload and download data. The key is to always be working with encrypted data, which will create more flexibility for the ways in which you can safely and securely share your data.