

Tool K: Common Means of Getting Access to Large Data Files

File Transfers

If the state agency responsible for housing and sharing education data plans to send you large data files containing data on individual students, that agency must choose a means of transferring very large data files while ensuring the security of the data. Their methods of transfer could include:

EMAIL. Although less popular than a few years ago, email is still an option for transferring data files, particularly if those files are of moderate size. Every working professional has an account, and transferring files this way is almost free.

Yet many data files exceed data storage or transmission limits, and traditional email accounts are typically not secure. If your agreement stipulates that you destroy all data after your agreement expires, email will leave many loose ends to resolve. Files with protected information about individual students could remain in your inbox, sent email, mailbox backups and other places. You can contract with a secure email system if you find yourself working with a partner that can only email their data files.

SECURE DROP BOX. Drop boxes are external file storage spaces people can access if they have proper credentials. The files would be deposited there by one partner and retrieved by another. Many working professionals already have access to one or more drop boxes, and they are easy to use. It is very important to restrict access to the drop box to approved personnel and to destroy all files after they have been used.

WAREHOUSE OR DATAMART RETRIEVAL. Your data provider may have an established a data warehouse or datamart from which you can extract the data you need. A data warehouse is a primary data repository for the organization and is typically protected behind the organization's most secure firewall or security protocols. A datamart is often a smaller subset of a data warehouse, containing data on a specific topic area.

Data warehouses and datamarts can minimize problems of version control, because the organization analyzing data can retrieve updated files when it needs them. Access to data warehouses and datamarts is highly controlled and can be complex to negotiate.

Direct Access to Files

Some state agencies have systems for gaining direct access to their data files:

TRADITIONAL REMOTE ACCESS TO HOST SYSTEMS. In these situations, the organization provides the researcher or analyst credentials for accessing the data within their existing systems. They do not grant authorization to extract files from those systems. In rare cases, this can mean that the researcher must use the data provider's computer system at the data provider's offices to conduct analysis.



In some cases, researchers are authorized to connect via the internet to a computer on site and take control of that computer remotely. Lastly, researchers may be given rights to establish a Virtual Private Network (VPN) connection to the collector's network systems to access the required applications and data.

This mechanism can provide most immediate access to the data without intervention from the data provider's personnel. Yet such virtual connections can be slow and unstable, especially when researchers must analyze large data sets.

SPECIAL ACCESS TO SYSTEM DATA PROVIDED BY A CUSTOM HARDWARE OR SOFTWARE SYSTEM.

Some agencies have software and hardware systems specifically to provide access to researchers and other institutional partners. In those cases, researchers have access protocols that require either connecting to designated software systems over the internet or installing specific connectivity hardware at their local offices.