Secure Data Handling: An Essential Competence for Evaluators

Konrad Czechowski, John Sylvestre, and Katherine Moreau
University of Ottawa

Abstract: Since it is paramount that the rights and welfare of evaluation participants and stakeholders be respected, we argue that the abilities and knowledge necessary to appropriately safeguard data ought to be considered an essential competence for evaluators. Building from past contributions, and in consultation with research ethics and data security experts from our home institution, recommended practices in the collection, handling, and storage of evaluation data were identified. A three-dimensional framework for secure data handling was developed, considering type of information handled, harm posed by a potential confidentiality breach, and corresponding steps to securing confidential information.

Keywords: competencies, data management, data security, program evaluation

In The Program Evaluation Standards, Yarbrough, Shulha, Hopson, and Caruthers (2011) identify a variety of standards and competencies for evaluators. Although some standards emphasize the importance of information accuracy, relevancy, and respect of privacy (e.g., standards A5 Information Management, U5 Relevant Information, and P3 Human Rights and Respect), few standards relate to...
evaluators’ roles and responsibilities in ensuring that they appropriately and securely handle confidential information collected during the course of evaluations. For example, standard A5, Information Management, includes a section on storing information, emphasizing the need for accurate and quality information storage, without invoking the need for secure storage. Since it is essential to respect and protect the rights and welfare of participants and communities in evaluations (Yarbrough et al., 2011), we believe that secure data handling and management of confidential information should be considered a core evaluation competence.

The Centre for Research on Educational and Community Services (CRECS) is an evaluation and applied research centre in the Faculty of Social Sciences at the University of Ottawa. In 2008, CRECS developed a manual to inform its evaluators and researchers on guidelines to collect, safeguard, and work with confidential data (Olson, Aubry, & Morier, 2008). In the almost 10 years since, the capacity to collect, share, and store data electronically has grown greatly. At the same time, there have been many stories of failures to properly safeguard confidential data by government agencies and by some of the largest private companies. These stories may lead to some uncertainty among evaluators, researchers, and students regarding the proper steps to follow when collecting, sharing, and storing data. This topic is particularly important in the context of our centre, which has a history of research and evaluation involving marginalized and vulnerable people. In the absence of clear institutional guidelines, CRECS undertook to develop guidelines for our researchers and evaluators. While there have been recent calls for researchers to enhance their research data-management skills to account for rapidly evolving data-sharing environments (Corti & Van den Eynden, 2015), we extend this call by encouraging evaluators to develop their skills in data management too.

For the purpose of this article, we define “evaluator” as a person who is collecting primary data or using secondary data for the purposes of conducting an evaluation. This can extend to any person who is part of an evaluation team or who will come into contact with data. “Data” can include a range of information collected on and by individuals or about program processes or outcomes. In this article, our primary concern is with confidential data with a particular focus on electronic data. However, most of the principles we outline can easily be extended to other forms of data.

In this paper, we present a three-dimensional framework to assist evaluators in making decisions regarding the handling of their data. We employed the following approach in developing our framework and producing recommendations for secure data handling:

1. we conducted an online search for best practices using terms such as “policy data security,” “privacy policy,” and “data security”;
2. we reviewed relevant ethical codes, both nationally and by discipline;
3. we examined the websites of the Privacy Commissioners of Ontario and Canada;
4. we consulted librarians specializing in data management;
5. we consulted the University of Ottawa data security architect;
6. we presented and obtained feedback on draft findings from the University of Ottawa evaluators, researchers, students, and research ethics protocol officers.

We believe it is essential that organizations have clear guidelines to equip evaluators with the knowledge, abilities, and resources to appropriately safeguard data and minimize the risk of a confidentiality breach. Further, we argue that clear institutional-level guidelines must accompany access to appropriate resources (e.g., relevant software) as well as training to use such resources. For those working alone, or in a private evaluation practice, we view it equally important to develop consistent practices and provide sufficient resources to protect confidential data. We hope that the information we provide in this article may inform decisions that all evaluators may adopt in their practices.

APPLICABLE ETHICAL CODES AND LAWS

Evaluators work in different contexts and therefore may have different statutes or legislation dictating rules of compliance. They may be members of different associations, each of which may have its own guidelines or regulations concerning the handling of confidential information. This section provides a brief overview of applicable ethical codes and laws, starting with guidelines provided by the Canadian Evaluation Society, which are relevant to all Canadian evaluators. Next, we provide an overview of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2), and the Personal Information Protection and Electronic Documents Act (PIPEDA), each of which would apply to those working in particular institutional settings or to evaluators using previously collected research data for evaluation purposes.

Canadian Evaluation Society

The Canadian Evaluation Society provides ethical guidelines for its members, divided into three sections: competence, integrity, and accountability. Their guidelines emphasize that evaluations should be designed and conducted in a manner that protects the rights (including the right to privacy) and welfare of all participants, but they offer no further specifics for the protection of confidential data. They stress that evaluators should act with integrity and confer with clients on contractual decisions such as confidentiality, privacy, and ownership of findings and reports (Canadian Evaluation Society, 2014).

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2)

The Tri-Council Policy Statement is a joint policy initiated by the presidents of Canada’s three federal research agencies (Canadian Institutes of Health Research [CIHR], Natural Sciences and Engineering Research Council [NSERC], and...
The objective of the policy statement is to express the agencies’ commitment to the promotion of the ethical conduct of research involving humans in Canada. The agency occasionally updates these guidelines. This review uses the version of the Tri-Council Policy Statement published in December of 2014 (TCPS 2, 2014). The policy statement asserts that individuals should design research with the inclusion of safeguards to ensure the privacy of participants and measures that protect confidentiality. Although the policy does emphasize the importance of taking necessary precautions to address privacy and confidentiality issues, it does not specify further what those precautions ought to entail and under what circumstances.

According to the TCPS 2, individuals must outline any exceptions to privacy and confidentiality, whether legal or ethical, in the process of free and informed consent and be approved by a Research Ethics Board (REB) prior to beginning data collection, unless the information is publicly available. According to Article 2.5 of the policy, when conducting non-research activities such as program evaluation, quality assurance and quality improvement studies, or performance reviews, approval by a REB is not necessary. However, it is common that individuals may subsequently use data originally collected for such activities later for research purposes. According to the policy, this is considered secondary use of information not originally intended for research and may require REB review (TCPS 2, 2014)

Evaluators who rely exclusively on secondary use of non-identifiable information are not required to seek participant consent but are required to seek REB review. Identifiability can be context specific (e.g., use of coded information is considered non-identifiable only if a researcher or evaluator does not have access to the key) and consent from participants for secondary use of identifiable information is not always necessary if there is REB approval (for more about secondary use of information see Articles 5.5 and 5.6 of the TCPS 2, 2014). Tri-Council requirements can be useful to evaluators since there is often a lot of overlap between research and evaluation activities. Much of what the policy statement covers is useful in an evaluative context, especially related to secondary use of research data or use of evaluation data for research.

**Personal Information Protection and Electronic Documents Act (PIPEDA)**

PIPEDA is a five-part federal statute covering the management of personal information for private-sector organizations during the course of their commercial activity; the statute and its provincial counterparts may apply to certain evaluation projects. The statute focuses on the protection of personal information in the private sector and handling of electronic documents. PIPEDA sets out 10 “fair information principles” outlining basic privacy obligations under the law, which include accountability, identifying purposes, consent, limiting collection, limiting use, accuracy, safeguards, openness, individual access, and recourse (Office of the Privacy Commissioner of Canada, 2015). The statute outlines broad
responsibilities related to protection of electronic data but does not go further in specifying appropriate means to safeguard electronic data.

Evaluators should be familiar with its principles and be aware of any similar provincial statutes. Several provincial statutes have been deemed “substantially similar” to PIPEDA, including those of British Columbia, Alberta, Quebec, Ontario, New Brunswick, Nova Scotia, and Newfoundland and Labrador. Organizations subject to provincial legislation deemed substantially similar to PIPEDA are exempt from PIPEDA with respect to the collection, use, and disclosure of personal information within their province (Office of the Privacy Commissioner of Canada, 2013).

**PRESCRIBING APPROPRIATE DATA-HANDLING PRACTICES: A MODERN CHALLENGE**

While institutions and associations may have general guidelines regarding secure data handling, it is up to evaluators to exercise their judgement in deciding what practices to implement and when to implement them. Evaluators may feel ill-equipped for this decision making if they have not received training or do not have access to resources to make decisions regarding their data and its safeguarding. For example, guidelines may instruct an evaluator to “password-protect sensitive data,” leaving them to decide for themselves what constitutes sufficiently “sensitive” data to warrant password protection. Moreover, an evaluator may be unsure about whether, and at what point, they should encrypt a document (and, of course, how to encrypt it and what software to use). This illustrates how a seemingly simple directive, to password-protect sensitive data, requires that an evaluator make many decisions about how exactly and in what context to execute the directive.

**A THREE-DIMENSIONAL FRAMEWORK FOR SECURE DATA HANDLING**

We argue that the abilities and knowledge necessary to make decisions related to appropriately safeguarding data ought to be an essential competence for evaluators. This would mean that such abilities and knowledge ought to be part of an evaluator’s basic training. Evaluators work across a broad range of contexts with different types of data, serving a diverse range of clients. We have developed a framework to aid in the understanding of the nature of one’s data to ensure that evaluators and program stakeholders take proper steps to safeguard their evaluative data.

Having consulted data handling experts, and reviewed relevant literature, we have come up with a three-dimensional framework for data handling. First, we consider the type of information handled (the extent to which data is identifying), next, the harm associated with a potential breach of confidentiality, and finally, the extent to which an evaluator should secure their data. We believe that our
three-dimensional framework is applicable at an individual level, when an evaluator is working on a specific project, as well as at an institutional level if they are developing their own procedures for secure data handling.

**Understanding and identifying types of information**

When an evaluator handles personal information, its source may have a reasonable expectation of privacy. When acquiring or working with data, it is important to determine the type of data before considering steps to safeguard them. The TCPS 2 (2014) has identified the following five categories to assess the extent to which researchers can use information to identify a participant, all of which we believe are also valuable for evaluators:

Directly identifying information—the information identifies a specific individual through direct identifiers (e.g., name, social insurance number, personal health number).

Indirectly identifying information—the information can reasonably be expected to identify an individual through a combination of indirect identifiers (e.g., date of birth, place of residence, or unique personal characteristic).

Coded information—direct identifiers are removed from the information and replaced with a code. Depending on access to the code, it may be possible to re-identify specific participants (e.g., the principal investigator retains a list that links the participants’ code names with their actual name so data can be re-linked if necessary).

Anonymized information—the information is irrevocably stripped of direct identifiers, a code is not kept to allow future re-linkage, and risk of re-identification of individuals from remaining indirect identifiers is low or very low.

Anonymous information—the information never had identifiers associated with it (e.g., anonymous surveys) and risk of identification of individuals is low or very low. (p. 59)

Ethical concerns regarding privacy decrease as data move on the continuum from directly identifiable information toward anonymous information (TCPS 2). When possible, it is best to collect data as low on the above list as possible; anonymous or anonymized data will keep the risk of re-identification of data low. Challenges in anonymizing data are amplified in sensitive contexts where balancing the integrity of the data with participant anonymity can be challenging (Saunders, Kitzinger, & Kitzinger, 2015). Indeed, evaluators often keep their data in an identifiable form, or have no choice but to collect identifiable data (e.g., if identifiable characteristics are directly evaluated or if follow-up with participants is necessary). In instances where this is the case, an evaluator ought to endeavour to de-identify or anonymize their data as soon as possible.
Assessing the risk of a possible confidentiality breach

It is important that evaluators assess the extent to which data carry a risk of harm to a participant if anyone were to disclose their data to an unauthorized party, thereby breaching their confidentiality. Not all information of the same type carries the same risk associated with a breach. For example, indirectly identifying information such as the address of a worker at a fast-food restaurant participating in an evaluation of the restaurant’s recycling practices is very different from the address of a resident participating in the evaluation of a low-income housing program who revealed they sell drugs to pay their rent. Although in both instances the information collected is indirectly identifying, the consequences of a confidentiality breach of the former would be far less damaging than the latter, where the resident may experience serious consequences such as loss of housing or criminal liability. Therefore, in addition to considering the type of data an evaluator is working with, it is imperative that they also consider the risk of harm associated with a data breach. Based on the Harvard Information Security Policy’s five data security levels (Harvard Information Security, n.d.), we have developed four levels of risk associated with evaluation data.

1. **Low-risk confidential information.** This level refers to information that in its present form would not likely cause harm to an individual or group if disclosed, which evaluators have nevertheless decided to keep confidential. Information at this level may include anonymous information, anonymized information, or coded information that if de-coded and disclosed would not cause serious harm to an individual or group. Examples include anonymous survey data, completely anonymized data, unpublished intellectual property (e.g., manuscript drafts).

2. **Sensitive confidential information.** This level refers to information that if disclosed in its present form can reasonably be expected to cause some damage to an individual’s reputation, or cause embarrassment. Information at this level may include low-risk data that has not yet been anonymized and coded information that if de-coded and disclosed could cause harm to an individual or group. Examples include bank account numbers, student records (e.g., transcript), and information supplied in confidence.

3. **Information that would likely cause harm if disclosed.** Information at this level includes information that if disclosed in its present form could create a risk of social, psychological, reputational, financial, legal, or other harm to an individual or group. This level may include sensitive data that has not yet been anonymized or cannot be anonymized because confidential information is needed for analysis but that, if disclosed, may cause harm to an individual or group. Examples include health card information, diagnosis of mental illness, and credit card numbers.

4. **Information that would cause severe harm if disclosed.** Information at this level includes information that if disclosed in its present form
could create a risk of criminal liability, loss of employment, or severe harm to an individual or group. This level is for data of the most sensitive nature, and evaluators should anonymize these data as soon as possible. Highly confidential information that cannot be fully anonymized because it is needed for analysis must be handled with extreme care. Examples include social insurance numbers and information about illegal activity.

Importantly, we note that examples corresponding to each level above may move up or down the scale depending on context. For example, student records may remain at the second level if they belong to students registered in a regular class, while they may move up to a level three or even four if they were to belong to students who self-identified as having cheated in an exam and who have provided this information in confidence to an evaluator.

We emphasize that the best way to avoid a breach of confidentiality is to not collect identifying data in the first place; we emphasize that the collection of identifying information not relevant to evaluation objectives may unnecessarily lead to a higher level of risk associated with a breach. It is therefore imperative that evaluators collect only data that are absolutely necessary to reach the goals of their projects.

**Practical steps to securing data**

After an evaluator has had the opportunity to assess the type of data they are working with and reflect on the harm associated with a potential confidentiality breach, they can consider what steps they should take to appropriately secure their data. Table 1 provides an overview of steps one can take to secure one's electronic data when considering the level of risk associated with a breach of confidentiality (and, by extension, the extent to which data are identifying). Since all projects are different, this framework is intended to help evaluators decide where their data may fit and what steps they should take to secure their data. The purpose of this framework is to encourage evaluators to think of their data in this three-dimensional way, in considering the type of data they are working with, the risk associated with a confidentiality breach, and steps to securing their data based on type and risk. The same principles in the first two dimensions (type of data and risk) can apply to non-electronic data as well. The third, steps to securing data, will not specifically apply to non-electronic data, but evaluators could nevertheless use it as a guide to consider how they can best keep their data secure.

Table 2 outlines steps for handling electronic data, from first loading data onto a computer to secure deletion. First, the evaluator must scan the computer or device for malware (short for malicious software) and ensure that the operating system is up-to-date. Next, evaluators should password-protect confidential data, and encrypt the data when possible. Finally, evaluators should securely delete confidential data using a “file shredding” software. We strongly recommend
<table>
<thead>
<tr>
<th>Level of risk</th>
<th>Steps for securing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Low-risk confidential information</td>
<td><strong>Storage:</strong> Data must be stored on a password-protected computer or drive. <strong>Sharing:</strong> It is recommended that files sent via email be password-protected. Password must be sent through a different medium.</td>
</tr>
<tr>
<td>2 Sensitive confidential information</td>
<td><strong>Field collection:</strong> Data should be collected on a password-protected device. <strong>Storage:</strong> Data must be password-protected; encryption is recommended. <strong>Sharing:</strong> Files sent via email should be password-protected and encrypted. Password must be sent through a different medium. Use of an organization’s shared drive, however, is preferred to email.</td>
</tr>
<tr>
<td>3 Information that would likely cause harm if disclosed</td>
<td><strong>Field collection:</strong> Data should be collected on an encrypted and password-protected device. Use of paper material is discouraged, but if used should be handled with extreme care and not left unattended unless in a locked and secure environment. <strong>Storage:</strong> Data must be encrypted and password-protected. <strong>Sharing:</strong> Data should not be shared by email. Files must be encrypted when using an organization’s shared drive. <strong>Access:</strong> Should be controlled by lead evaluator, who should keep a list of individuals who have been granted access to data.</td>
</tr>
<tr>
<td>4 Information that would cause severe harm if disclosed</td>
<td><strong>Field collection:</strong> Data should be collected on an encrypted and password-protected device. Use of paper material is discouraged, but if used should be handled with extreme care and not left unattended unless in a locked and secure environment. <strong>Storage:</strong> Data must be stored in a physically locked room (preferably secured by an alarm) on a password-protected and encrypted hard drive or computer with limited or no connection to a data network. <strong>Sharing:</strong> Sharing at this level should be limited; data should be accessed only in a secure location. <strong>Access:</strong> Should be strictly controlled (e.g., by keeping a list of individuals who have been granted access to data.)</td>
</tr>
</tbody>
</table>

*Note. Levels of risk adapted from Harvard Information Security (n.d.)*

that all of these steps be taken when handling any confidential data. Of course, it is up to the evaluator to decide on how cautious they should be with their data, depending on data type and level of risk. We acknowledge that our examples outlined in the two tables (e.g., use of email for data sharing) will not be applicable to all, as there are many other alternative ways to handle data (e.g., use of USB rather than email) that inevitably present their own challenges. We do, however, highlight the fact that the principles remain the same, regardless of medium used.
Table 2. Steps for securing electronic data

<table>
<thead>
<tr>
<th>Steps</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1: Find a clean computer / create a secure environment</strong></td>
<td>First and foremost, it is essential that computers be up-to-date. System providers release operating system updates on an ongoing basis to protect users against vulnerabilities that hackers have identified and exploited. Next, it is important to scan computers for malware. To be considered secure, a computer should first be scanned for malware, and if any threats are detected, they should promptly be cleaned or removed.</td>
</tr>
<tr>
<td><strong>Step 2: Password protect, encrypt</strong></td>
<td>After a computer is scanned and threats are removed, an evaluator may load their sensitive data onto the computer (e.g., interview recordings, non-anonymized data, etc.). At this stage, it is important to de-identify or anonymize data to the greatest extent possible and password-protect documents. It is recommended that the computer’s entire drive be encrypted and strongly recommended that the individual files be encrypted.</td>
</tr>
<tr>
<td><strong>Step 3: Permanent file removal</strong></td>
<td>Files with sensitive information should never be deleted using standard “trash bins” installed on computers to “remove” files. This only removes the file directory, yet the file and its information are still stored physically on the computer’s hard drive and remains accessible. “File shredding” applications can be used to permanently delete files. These programs overwrite the file with random letters and/or numbers, often multiple times, before removing their directory. It is important to note that this applies to files that contain sensitive information and are no longer needed (e.g., an audio recording that has been transcribed.) A file that has been anonymized and will later be used for research does not need to be destroyed; it can simply be password protected, and preferably encrypted.</td>
</tr>
</tbody>
</table>

to manage data. An evaluator ought to be aware of the multiple ways along the spectrum of security to handle their data and choose the options most suited to their situations.

**Encryption and password protection**

Not all passwords are equally secure, and encryption is useless if someone can easily guess the password. Although there may be some debate as to what kind of password is most secure, it is often accepted that longer passwords (at least eight characters) and with varied characters (upper case, lower case, numbers, and special characters) are more secure (more difficult to guess by an individual or software designed to crack passwords). We also note that a password should
be sent through a different medium than data (e.g., if a password-protected file is sent by email, the password should not be sent in that same email, but instead should be sent by text message, mail, or phone). Passwords should be sufficiently complex so that they cannot be easily guessed. What is “sufficiently complex” ought to fit prevailing standards, and the complexity of a password should be proportionately greater than the level of risk associated with a potential confidentiality breach.

Encryption can be very simple, and many individuals may use devices encrypted by default and not even be aware. Full-drive encryption is available (and often automatically enabled) on Windows computers (Bitlocker) and Mac computers (FileVault). However, this may give some a false sense of security, since once one is logged on, the system decrypts the data, thus rendering a user’s data exposed. Therefore, we recommend file-level encryption for files containing sensitive data; a range of encryption software is available, from free, open-access software to paid professional applications.

**Data-management plans**

A data management plan (DMP) refers to a document that outlines how evaluators should handle data before, during, and after an evaluation project. A DMP can help address many concerns raised in this article and act as a checklist to ensure that an evaluator has considered how they will handle the data at every stage of the project. A DMP can also be useful in reassuring concerned program stakeholders that the evaluator(s) will handle data with care. On projects with multiple evaluators and stakeholders, teams can also use the document to ensure that everyone is handling data in a consistent manner (e.g., a DMP can outline how often data should be backed up, in what format data should be saved, and how long it is kept before it is securely destroyed).

**CONCLUSION**

In a constantly evolving environment, having knowledge related to technology, especially the secure storage of electronic data, is imperative. We argue that an evaluator’s abilities to appropriately secure data is a core evaluation competence, and as such it is essential that evaluators be proficient in this area. Akin to the reality that an evaluator may spend hours learning a new statistical method required for an upcoming project, we argue that an evaluator should spend the time necessary to learn how to handle the data securely.

The responsibility should not rest solely at the level of an individual evaluator but extend to an organizational level. Therefore, it is important that associations and organizations provide their members with adequate guidelines, instructions, and resources to facilitate their capacity to handle data securely. It is essential that organizations and associations provide their members with specific guidelines that consider the type of data, level of risk associated with a possible breach, and detailed instructions related to how they should handle data.
such practices would increase the confidence of concerned program stakeholders and add to the integrity of the evaluative process.

REFERENCES


AUTHOR INFORMATION

Konrad Czechowski is a doctoral student studying clinical psychology at the University of Ottawa and a researcher at the Centre for Research on Educational and Community Services. His research and evaluation interests include mental health services, approaches to methodology, and human sexuality.

John Sylvestre is a professor in the School of Psychology, and a senior researcher at the Centre for Research on Educational and Community Services, at the University of Ottawa. His interests lie in the study and evaluation of community mental health programs, with a focus on the issues of housing and homelessness.
Katherine Moreau is an assistant professor in the Faculty of Education at the University of Ottawa, an affiliate investigator at the Children’s Hospital of Eastern Ontario Research Institute, and a senior researcher at the University of Ottawa’s Centre for Research on Educational and Community Services. Her research interests include participatory program evaluation and patient engagement in medical education.