Direction des bibliothèques

Université de Montréal

Ethical and legal issues of sharing and disseminating data in the social sciences:

Thoughts from the 2019 QICSS Summer School (Montréal, June 10-11, 2019)

Direction du soutien à la réussite, à la recherche et à l'enseignement
Direction des bibliothèques
Université de Montréal
Novembre 2020

Acknowledgements

Les Bibliothèques de l'Université de Montréal would like to thank the Quebec Inter-University Center for Social Statistics (QIUCSS), la Bibliothèque de l'Université Laval, professor Véronique Dupéré, professor Isabelle Archambault, both from École de psychoéducation de l'Université de Montréal, as well as everyone who participated – as an organizer, a presenter, or an attendee – in the Summer School on ethical and legal issues around the sharing and dissemination of data in the social sciences, held June 10 and 11, 2019, at the Carrefour des arts et des sciences at Université de Montréal. They would also like to thank Tierry M. Laforce who took careful notes during the event, structured and wrote this document, as well as all who reread it before its completion. Last but not least, we want to thank the Social Sciences and Humanities Research Council of Canada, whose funding made possible the event and the writing and translation of these *Thoughts.*

Table of contents

Introduction

The Quebec Inter-University Centre for Social Statistics (QICSS), in collaboration with Bibliothèques de l'Université de Montréal and Bibliothèque de l'Université Laval held a Summer School on June 10 and 11, 2019, at Université de Montréal's Carrefour des sciences. A diverse group of individuals who carry out research or provide support it attended presentations touching on some of the ethical and legal issues involved in the sharing and dissemination of data in the social sciences. The presentations, and the discussions that followed them, served as the basis of this document, while not being the sole source of its content. Several presentations can be accessed here : https://www.cigss.org/evenement/ecole-d-ete/enjeux-ethiques-et-juridiques-du-partage-et-de-la-diffusion-des-donnees-aux.

The purpose of this document is to provide an overview of the very relevant issues addressed during this QICSS Summer School to complement the presentation documents mentioned above. However, it does not claim to cover everything that has been discussed during these days, nor asserts that the perspectives presented are the only existing or possible ones on these critical and complex issues.

This document is divided into four sections. The first is an introduction to the general concept of research data and the life cycle of research data. The second discusses the context in which research is currently being conducted. In particular, we focus on the upcoming publication of the Tri-Agency Research Data Management Policy. Then, the third and fourth sections present the ethical and legal issues around the sharing and dissemination of data in the social sciences. Each subsection in sections 3 and 4 starts with an example of a research situation that could potentially raise some of the issues being discussed, and ends with some points to consider and some practical information.

1. Research data

1.1. A definition

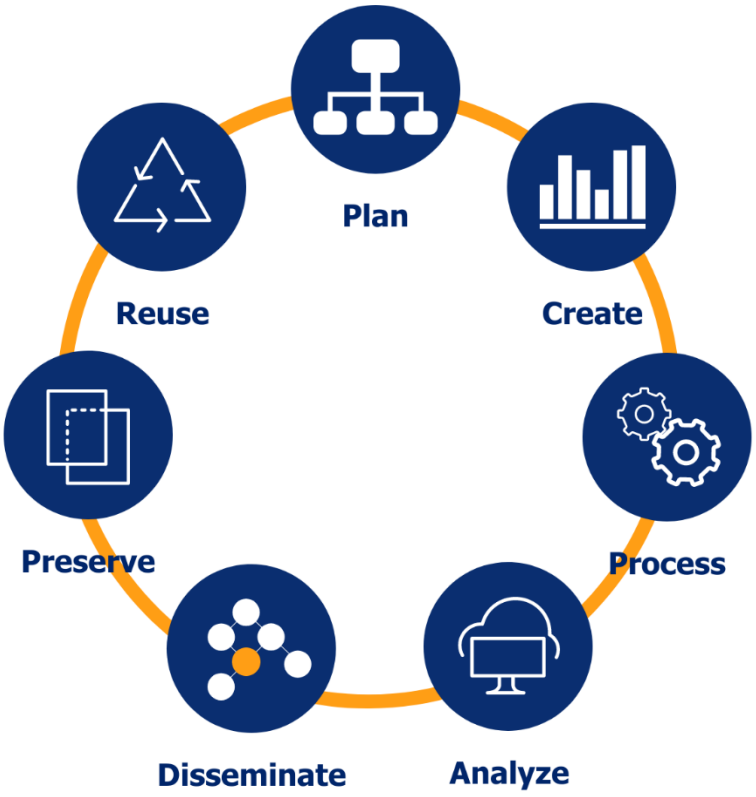
Research data comes in many forms : "Data may be in any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records" (1). One definition of research data describes them as "factual records

(numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings" (2).

Research data may or may not be in digital form; for the purposes of this document, they will essentially be considered in their digital form.

1.2. The life cycle of data

There are several models depicting the life cycle of data. This one was put forward by the Portage network for research data management, an initiative of the Canadian Association of Research Libraries (3).



2. Current context: Tri-Agency Policy Statement

Canada's three funding agencies – SSHRC, NSERC, CIHR – are on track to publish a Tri-Agency Research Data Management Policy. As a preliminary step, the organizations released to the research community a draft, based on existing

policies, in order to help the community "respond to current and future requirements as regards research data management". This draft remains relevant as a tool that helps us better understand the role of researchers in data research management.

According to the three agencies, research funded with public money should benefit Canadian society, in other words, research data should be, if possible, in the public domain, so that they may be utilized for more than one research project (4). The agencies' goal is to democratize access to research results, promote advances in knowledge, avoid duplication of research efforts, encourage the reuse of previous research results, and showcase the accomplishments of Canadian researchers (4).

In order to reach these goals, the funding agencies expect, among other elements, data management plans that are drawn up right at the start of a project, and which specify how data will be organized as the project goes along, and in what ways their reuse will be encouraged.

Data should be accompanied by metadata, that is, data, documentation or even documents, for example, the coding used by researchers to set out the findings of their research. Metadata allows for data to be more easily found and reused.

Of course, when data are reused, they should be appropriately cited. "Data are significant and legitimate products of research and must be recognized as such "(4).

3. A few issues: An ethical perspective.

The move to digital technology in research has changed perceptions towards research data. And as the context in which research is carried out evolves, so too do ethical considerations. We can see this evolution in changes to the rules that govern research that involves human beings. Through the decades, policies have been made more and more specific, in order to provide a clearer framework in which researchers can work:

- Adoption of the Nuremberg Code in 1947;
- Adoption of the Helsinki Declaration in 1964;
- Passage of the National Research Act by the U.S. Congress in 1974;
- Publication of the Code of Ethics for Research with Human Subjects by the Social Sciences and Humanities Research Council of Canada in 1977;
- Publication of Ethics Guidelines for Research with Human Subjects by the Medical Research Council of Canada in 1987;

- First publication of the Tri-Agency Policy Statement: Ethical Conduct for Research Involving Humans (1998);
- Revisions of the TCPS in 2010, 2014 and 2018.

With the advent of each of these milestones in research ethics, institutions developed policies and directives to guide researchers in their projects that involved human subjects; and the granting agencies established the relevant frameworks for the projects they fund.

In that vein, the current Canadian policy, TCPS 2, puts the responsibility for protecting data back into the hands of researchers and institutions. As well, institutions receiving funding from the three agencies must sign the Agreement on the Administration of Agency Grants and Awards by Research Institutions. Researchers must properly inform participants when they recruit them, so as to obtain a free and informed consent. TCPS 2 does provide for some exceptions to this obligation, so as to permit researchers to collect data in certain specific situations.

The move to the digital research paradigm means that issues broader than just research participants' data may now be at stake. Researchers now study databases that include unprecedentedly large numbers of people. A larger volume of data, and new tools with which to analyze them, allow researchers to draw conclusions that could have an effect on a significant number of individuals. With the use of digital data comes the possibility that data may be amalgamated and cross-referenced. The concept of the autonomy of participants often collides with the concept of the interests of the collective. While the two ideas are not inherently opposed, the digitization of data has exacerbated a tension that does exist between the two concepts. These are complex questions to which there are no easy answers.

A critical examination is needed, and the scientific community cannot take it upon itself to answer these questions alone. The population it works with must also ask questions and take positions. Researchers must be conscious of the impact that their studies can have on individuals and on populations. They must think carefully about the issues involved in any given study from the moment they start to conceive it, right up to the publication of findings, and even beyond.

3.1. Data Sharing

An example

A researcher has been gathering data about a particular population over the course of his career. The surveys and interviews he has carried out have allowed him to build up substantial database, which he has used on a few occasions as the foundation for publishing scientific articles. Now on the verge of retiring, he sees that the database contains sufficient other information to be the basis for further articles. He knows he will not have enough time to do the work himself, so he transmits the data to a colleague, hoping that she will be able to analyze them. Do current ethical standards permit the other researcher to make use of the data?

Since it is now possible to exchange vast amounts of data rapidly and relatively easily, limitations to data sharing are less often structural. Researchers do, though, have to ask themselves which organizational, national or community rules and regulations need to be complied with. They also need to evaluate the impact that the sharing of data will have on the individuals who are the source of the information. Data sharing should always be carried out while taking into account the risks of disclosure or of inference on the individuals involved in the research. However, it can be difficult for a researcher to identify all the variables that present a risk of identification being made. Some are easy to spot, but others may be overlooked without a proper consideration of the risk that is present when they are cross-referenced with other variables. We need to develop more expertise on this point, and responsibility should not rest solely on the researchers' shoulders.

In some disciplines, a practice has been instituted of making data available upon the publication of a scientific article. That, in turn, creates issues around sharing: as mentioned above, shared data must not wind up causing harm to the individuals who have provided them. What's more, researchers don't always possess the technical know-how needed to make their data available in a form that meets the requirements of data sharing.

One first step in promoting sharing could be to establish a suitable description of the data, followed by the transmission of this description to the scientific community involved. A document accompanying publications could indicate how the raw data could be consulted, or, if they are not accessible, indicate how the researcher accessed them and what are the limitations to access.

Points to consider

Consider the interests of the participants whose data you hold before transmitting them:

- Can they be identified from your dataset?
- Would a cross-referencing of data render them identifiable?
- Have they consented to the sharing of their data with a third party?

+ Check whether the framework that governs data storage differs among the various collaborators involved in an exchange of data.

Practical notes

Open Science Framework (https://osf.io/) is a platform for storage and sharing of research data. In the spirit of open research, OSF allows researchers to access research data, and also share their own, in a secure environment.

3.2. The value of participants' consent

An example

A group of researchers start a tissue bank for use in sequencing participants' genomes. They are hoping to make advances in research on personalized medicine. In setting up their bank, they realized that the genome sequencing could lead to incidental findings, that is, information that neither they nor the participants would have expected to find. The initial consent given by the participants specified that these findings would automatically be disclosed, but discoveries sometimes come years after the signature of the consent form. Does a participant's initial decision regarding the communication of the findings still stand despite the passage of time?

Example based on Thorogood et al. (5).

The principle of free and informed consent is at the core of research carried out with human subjects. This respect for autonomy is essential in ensuring the respect of the dignity of the participants, but also in maintaining the relationship of trust that exists between populations and researchers. When data is preserved for a period of several years, we should ask ourselves whether consent obtained at the start of research is still valid after a year, or two, or three, or five, or even ten.

These questions are usually brought up in the initial discussion that leads to consent. The researcher and the participant assume that the consent decision will remain valid over time, until the researcher destroys the data. The latter thus has a responsibility to respect the commitments made to the participant until long after the data has been gathered. The need for the researcher to think about the issue of the ultimate fate of the data therefore does not disappear once approval is received from the research ethics committee. Obviously, the extent of the

reflection on the issue should be proportionate to the sensitivity of data collected, and to the risk that the participants are exposed to.

Points to consider

- A participant's free and enlightened consent must be ongoing:
  - Identify the key moments in time at which consent must be reaffirmed;
  - Do not make any assumptions about a participant's consent: when in doubt, contact them;
  - Try to anticipate broader forms of consent for secondary usages of data, as the case may be, bearing in mind that "meta consent" is accepted in Canada, but not American-style "blanket consent".
- Regular communication allows the researcher to keep the participant informed about the project they are involved in. Consider utilizing methods for keeping participants informed of the purposes for which you are using their data.

Practical notes

Scientific literature in ethics abounds with ideas for improving procedures for obtaining free and informed consent from participants. Before you plunge into extensive research, though, you can start with TPCS 2's online tutorial (https://ethique.gc.ca/eng/education_tutorial-didacticiel.html) or that of the Ministère de la Santé et des Services sociaux (https://ethique.msss.gouv.qc.ca/didacticiel/?lang=en). They cover quite a bit more than the consent process, but they will give you a solid base from which to pursue your research subsequently.

3.3. Big Data

An example

Since 1955, the New South Wales State Emergency Service (NSW SES) has used various sources of information (archived and real-time data from weather agencies, satellite images, social media feeds, photos and videos available online, etc.) to plan and organize rescue responses to natural disasters in Australia. In 2009, its IT framework was overhauled and modernized, so as to facilitate the transfer of information among the actors involved in rescue operations. The new infrastructure allows for the integration of information from multiple sources in order to identify the potential risk of natural disaster that different regions may be exposed to. In combining information from several sources, NSW SES can now

anticipate the impact of a catastrophe, and devise intervention plans, based on the Big Data it analyzes.

Drawn from Wamba et al., 2015 (6).

The term Big Data refers to the large-scale and rapid accumulation of data regarding a multitude of facts, and their storage in equally vast databases. They are unlike data traditionally collected in research contexts. Apart from their aforementioned attributes, they are difficult to analyze using the normal statistical tools. (7). Big data are defined in terms of the 5 "Vs":

- Volume: The quantity of data collected;
- Variety: Their different types, formats and structures;
- Velocity: The speed with which they are gathered and stored;
- Value : The potential worth of the data being collected;
- Veracity: The quality, accuracy and reliability of the data (8).

In research, the use of Big Data can raise critical ethical issues with regards to the identification of individuals and of aspects of their private life, but also when it comes to the ownership of the data. In terms of identification, some differentiate between anonymous data (presuming that maintaining anonymity is still possible); data that is rendered anonymous; and data that is de-identified. The first type are collected without any direct identifier having been associated with them. The second type is data from which identifying markers have been removed so as to render them anonymous. The third type have had their identifying marker replaced by a code. What we would call sensitive data can be defined by the moral value it has for the individual or institution who shares it with the researcher. The security measures put in place by the researcher to protect the data should be proportional to its sensitivity.

How, then, do we protect the individuals who are the source of data? There are technologies to help protect privacy – Privacy Enhancing Technologies or PETs – that offer users and researchers more control over shared data. For instance, a text messaging system can be encrypted so that even the application's developer cannot access the content of the messages. The principles underlying the technology are 1) minimizing the amount of data collected 2) the user is master of the data they share. In this way, the data that are stored and collected are limited to those that the user agrees to share.

Points to consider

+ Create a plan for data protection even before starting to collect the data;

- Are the data to be analyzed anonymous, sensitive or personal? The level of protection will differ in each case.

3.3.1 "Data lakes"

An example

The CHUM (Centre hospitalier de l'Université de Montréal) has implemented a Center for the Integration and Analysis of Medical Data (Original title: Centre d'intégration et d'analyse des données médicales or CITADEL) so as to create a structure for the totality of data produced and stored by the hospital centre. CITADEL retrieves and organizes patient care data in order to make them usable by researchers to whom they may be relevant. Once they have obtained the necessary authorizations from the Office of Professional Services and the Office of Research Ethics, researchers can communicate directly with the CITADEL team to obtain a dataset. That team is responsible for ensuring the confidentiality of the data transmitted to researchers and drawn from the "data lake" in which it operates.

For more information: https://www.chumontreal.qc.ca/en/crchum/facilities-and-services

"Data lakes" are databases that store large quantities of data that may be raw, structured or semi-structured. They are used by institutions that collect the data, and by institutions that offer storage space to researchers who are analyzing the data. A researcher who uses them to store or analyze data needs to think about the impact this might have on the individuals involved. Could the information in question compromise a person's dignity? Could it lead to their being identified? Does it only touch on a certain segment of the population? Did the participant's consent envisage this type of storage?

The confidentiality of individuals' information is without a doubt at the core of the issues relating to data lakes. By warehousing a very large quantity of information in one single space, and giving access to that space to researchers, an institution makes itself vulnerable to a breach of confidentiality. The more variables a lake contains regarding an individual, the easier it is to identify them by cross-referencing variables. How, then, can we ensure that confidentiality is protected? One way is to withdraw variables that might lead to an identification prior to sharing the data. Thus, instead of providing the researchers with all the variables pertaining to a population, the ones that could, when cross-referenced, lead to a re-identification of individuals would be withdrawn. In this way, confidentiality is protected.

Can an individual, be they a participant in research or the user of a service, decide what happens to their data? In the digital age, it is difficult to withhold one's consent to data gathering when the use of technological tools is predicated on the collection of data about users. On the other hand, that is not the case with research. A researcher can decide what will happen to the data they collect, and is able to inform the participants who will be involved in their study. In fact, just informing them is not enough; one must also assess the various ways in which data can be gathered and stored. Such decisions require thought for the duration of the project, and even after.

Points to consider

+ Check whether the variables you cross-reference from a data lake might lead to individuals being identified;
+ Think about the sources of the data that you obtain from a data lake;
    - Who are the individuals behind the data? Are they participants in a research project? Patients? Users of a public service, or a privately provided service?
    - Are you using the data in a manner that corresponds to the interests of these individuals?
+ Prepare your data protection measures before you receive the data;
+ Think about what will eventually happen to the data prior to starting your project.

Practical notes

The Center for Internet Security specializes in data protection for organizations. It is a vendor of cybersecurity products, but nonetheless also provides information that can be helpful to researchers. One of its resources explains how to create a data protection plan. It's available here: https://www.cisecurity.org/blog/how-to-create-a-data-protection-plan/.

3.3.2 The role of artificial intelligence

An example

A researcher develops an algorithm that gathers data on the users of an application created for research purposes. The algorithm uses machine learning so as to retain only the data this will be especially relevant to the researcher. Once collected, the data are then organized by the same algorithm so that they can be presented to the researcher in a form that is usable for a research project.

Based on Determann, 2018, p. 19 (9).

After data gathering and storage comes the analysis of Big Data. Traditional research tools are not capable of procuring us the maximum amount of benefit from the collection of Big Data; these data are so numerous and so different in scope and format that a researcher can not delve into them without being equipped with modern tools. One tool offered by artificial intelligence (AI) is machine learning. The researchers who use these tools will be responsible for them. Such tools are expected to be:

- Loyal to their users (they behave in the manner stipulated by their developers)
- Fair and not discriminatory;
- Transparent;
- In compliance with existing laws.

Given the storage and analysis capacity that is now available, the relationship between research and the individual is necessarily evolving. In this new relationship, respect for human dignity remains crucial. There are more and more possibilities for making inferences, and progress in AI allows for numerous types of analyses. In this environment, we must find a balance between the economic and social benefits AI promises and the respect for the privacy of individuals. Efforts to find such a balance include the Asilomar principles and the Montréal declaration for a responsible development of artificial intelligence.

AI allows for the analysis of large quantities of data. As well, machine learning holds out the possibility for a researcher to guide initial analyses but then have them continue with a minimum of human oversight. One issue raised by the use of such tools is the opacity of machine learning, that is to say, between the time the analysis is launched and the results are produced, a sort of "black box" phenomenon results.

This black box can limit researchers' capacity to detect discriminatory aspects of programming. Technical and organizational measures must be implemented in order to avoid such discrimination. For example, data can be sampled prior to the analysis so as to evaluate them as they are about to be entered. Or, the analyzing algorithm can be altered so as to factor for discrimination and correct it in the course of the analysis.

The study of these questions is still in its infancy, and it is difficult at this point to evaluate their impact on data analysis in research. While data can be analysed from multiple angles, we can choose to limit the options out of respect for the

individuals who have provided the data. Just because a researcher is in possession of data, and the tools with which to analyze them, that does not necessarily mean they must do so. These questions are important and must be examined; but it is difficult for a lone researcher to come up with all the answers.

We must be conscious of the impact on individuals. Cross-referencing variables, reusing data for new purposes, and the risk of large-scale leaking of data can all be harmful to those who contributed to the data gathering operation.

In concluding this section on the ethical issues involved in data sharing, we should remind ourselves that not everyone has equal access to the digital world. This phenomenon has been called the "digital divide" and refers to inequality in terms of access to, and contribution to, information, knowledge and networks (10). Research projects anchored by information and communications technology exclude certain individuals from the outset. TCPS 2 requires that the advantages and disadvantages of research be equitably distributed among individuals and groups in a community. The scientific community therefore must consider the digital divide when it conceives research projects. Even though it is certainly not the researchers' responsibility to equip these individuals with knowledge, or initiate them to the digital universe, they should nonetheless be conscious of this reality and the limitations it imposes on research endeavours.

Points to consider

AI allows for analyses that cannot be performed by humans, but humans can decide what analyses will be carried out;

+ Identify the biases of the program and the team behind an AI project. Machine learning is first and foremost conceived by a human being; that individual must minimize their own biases in order to avoid AI perpetuating them;
+ Use AI that will allow you to explain the results and analyses you obtain;
+ Consider the potential impact of the analysis of your particular dataset on the targeted community;
+ The tools you use in gathering your data may exclude certain individuals from your research. Identify them and see if you can include them in other ways.

Practical notes

Since 2017, the Quebec Artificial Intelligence Institute has been bringing together renowned researchers to promote collaboration and knowledge sharing in the development of artificial intelligence. You can learn more about their activities here: https://mila.quebec/.

4. The issues: A legal perspective.

Digital data are now well integrated into the research universe. Up to a certain point, reuse of data is accepted. But this added advantage can also be a source of difficulty. The range of stakeholders goes from individuals seeking to protect their personal information to businesses seeking to use the data they collect for commercial purposes. (11).

The granting agencies' declaration of principle sets out responsibilities that are incumbent on the following actors: researchers, research communities, research institutions and research funders (4).

Researchers must develop data management plans, comply with the requirements of the granting agency and professional standards, acknowledge and cite datasets that contribute to their research and stay abreast of standards and expectations in their discipline.

Research communities must develop data management standards that will apply to their community, and identify the repositories and platforms that could be used by their researchers.

Research institutions must educate and support researchers in data management. The establishment of data management practices that are consistent with ethical, legal and commercial obligations on both provincial and national levels is crucial. Any institution that administers tri-agency funds must establish an institutional data management strategy. They must also recognize data as an important research output.

Research funders also recognize data as an important research output. They must develop policies and requirements that enable responsible data management, and include data management considerations in the process that assesses applications for funding.

4.1. Who are the authors, and the owners, of data?

An example

An organization collates data regarding housing in a North American city. On its website, it provides information such as house sizes and land area, numbers of rooms and floors, property values as assessed by the city, year of construction and so on. The organization stipulates in its terms of use that the data presented is the property of the collaborators who gathered them. A researcher uses the site as a source of data for the purpose of tracking the evolution of housing in the city over the decades. Is the researcher violating copyright law?

Case presented by Dominique Lapierre at QICSS Summer School

When it comes to data, Canadian and Quebec laws stipulate interests, access rights and restrictions to access (9). For a more comprehensive response, one must turn to intellectual property rules, which are divided into two branches: industrial property and copyright. That second branch includes copyright over literary, dramatic, musical and artistic works. While the Copyright Act, RSC (1985), c. C-42, does not protect raw data, it does set out rules for certain compilations of data.

Under copyright law, there is a balance to be struck between the protection of the rights of the author and the rights of users. A digital protection mechanism like a password can, however, be a way for an author to limit the use of their work. (11).

Apart from the Copyright Act, a research institution may have its own rules regarding the ownership of research data. For example, at Université Laval, the regulation on intellectual property stipulates that:

> *The University shall be the owner of a set of holdings constituted by a member of the University or by a group of members of the University when the member or group of members has used the name or the time or the services or the premises of the University, or has benefited from a grant from a sponsor requiring that the grant be ratified by the University.*

Such legal provisions can take the form of university regulations, but they may also be enshrined in collective agreements with researchers or in agreements reached by the research institutions with granting agencies. A researcher should therefore contact the appropriate resource at their research institution to find out exactly what framework governs the ownership of research data at that institution.

In another legal setting, in Europe, a general regulation on data protection came into force in May of 2018 (12).

Points to consider

+ Data may be subject to copyright law: identify the owner of the data you are using prior to commencing your research project;
+ Use a digital form of protection such as a password to limit the use by others of a dataset that you are not yet ready to share;
+ Check the rules and regulations of your institution!

Practical notes

It's quite likely that your research institution has implemented a policy regarding the attribution of intellectual property rights. The policy may have been adopted by a University Board, an office of ethics in research, or another body responsible for supervising research within the institution. Contact the person in charge of that body in order to find out about the current policy.

4.2. Who can authorize access to research data?

An example

For the purposes of her Master's, a student has collected and stored data in her research supervisor's research lab. She is currently the sole user of the data, while she actively works on her Master's thesis. But following a conflict with her research supervisor, she loses her access to the lab. She can no longer consult the data she was analysing; the access code has changed. In this situation, who is the owner of the data?

Case presented by Dominique Lapierre at QICSS Summer School.

Apart from the issue of ownership of the data, it may be appropriate to look at the issue of access to data, too. Who can authorize access to data? Every data storage platform has its own terms of use. Some platforms are located right where the researcher works, while others live in an entirely different legal jurisdiction. How, then, to determine who authorizes access? Using a digital lock, for example a password, is one way of protecting access to data. The person with responsibility for the lock is certainly in a position of authority vis-à-vis any users who wish to have access. This approach recognizes that data, in order to be accessible, must be in some physical medium. That medium then becomes an object which is the central point in the determination of who owns the data or at the very least who has responsibility over them (12). In general, when one has authority over access,

one also is responsible for the costs associated with the storage of the data, their transfer and their security. Rules that will govern the access to data must be established, as well as mechanisms that will guarantee compliance with these rules. As well, there must be maintenance of the data, and of the access to the data, so that they remain available. Many researchers get their access through an organization that is responsible for such access. In that case, institutional rules will be utilized to resolve any conflicts that may arise.

Points to consider

Designate someone who will have responsibility for controlling access to data.

- It will not necessarily be the owner of the data;
- In some cases, the organization housing your data will have this responsibility.

4.3. Some thoughts about personal information

And what about personal information? This is indeed a form of data that can be collected and used in research. But here is a key nuance. A piece of data exists whether or not it is used in research. Even if they are not the owner of this data, an individual is nonetheless capable of consenting to their use by a researcher. The researcher does not become the owner of the data, even they are obligated to control access to the data (R. c. Stewart [1988] 1 RCS 963). The researcher thus becomes responsible for the protection of the personal information that they will have gathered in the course of their research. In Quebec, the Commission d'accès à l'information (CAI) is by law the moral owner of personal information and oversees its protection. When a researcher wishes to use data gathered without the express consent of the individuals involved (e.g. by cross-referencing databases), the CAI has the power to speak on their behalf. It can thus authorize a researcher to receive the personal information for research purposes.

It must also be realized that confidential data must be protected by the researcher who gathered them and wishes to use them in their research. The researcher must therefore implement reasonable measures to avoid them being used or consulted by a third party without authorization. Such measures may include protection by password, storage of the data in a physical enclave accessible only to certain authorized persons, etc.

Conclusion

This document is first and foremost meant to be an informational resource for researchers wanting to explore selected ethical and legal issues regarding the sharing and dissemination of data in the social sciences. It is far from exhaustive, but it does examine some of the issues discussed during the QICSS Summer School held on the 10th and 11th of June 2019. The explorations of data sharing, the value of consent and Big Data were only the tip of the iceberg, an iceberg that keeps getting bigger as new technologies and new ways to collect and keep data are developed. The governing ethical and legal frameworks that determine rights and responsibilities regarding data are still evolving. In the near future, the three Canadian funding agencies will bring forward a data management policy. Researchers and research institutions will be able to refer to it in order to better frame their work. But until then, all researchers must remain alert to the issues that may be raised by their research.

References

1. CASRAI Dictionary [Online] Consortia Advancing Standards in Research Administration Information. Data [cited June 25 2019]. Available at: https://dictionary.casrai.org/Data

2. OECD Secretary General. OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: Organisation de coopération et de développement économiques; 2007.

3. Réseau Portage. Untitled; unpublished, 2019.

4. Tri-agency statement of principles on digital data management. Canada: Canadian Institutes of Health Research, Natural Sciences and Engineering Council of Canada, Social Sciences and Humanities Research Council of Canada; 2016.

5. Thorogood A, Joly Y, Knoppers BM, Nilsson T, Metrakos P, Lazaris A, et al. An implementation framework for the feedback of individual research results and incidental findings in research. BMC Medical Ethics. 2014;15(1):88.

6. Wamba SF, Akter S, Edwards A, Chopin G, Gnanzou D. How big data can make big impact: findings from a systematic review and a longitudinal case study. INT J PROD ECON. 2015;165:234-46.

7. Mégadonnées [Online]. Québec: Gouvernement du Québec; 2017 [cited July 10, 2019]. Available: http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=26507313

8. Bourany T. Les 5V du big data. Regards croisés sur l'économie. 2018;23(2):27-31.

9. Determann L. No one owns data. Hastings LJ. 2018;70:1.

10. Cliche D. La ville intelligente au service du bien commun – Lignes directrices pour allier l'éthique au numérique dans les municipalités au Québec. Québec: Commission de l'éthique en science et en technologie; 2017.

11. Scassa, T. Who Can Own Data in a Data Economy? CIGI [Online]. 2018 [cited June 25 2019]. Available at: www.cigionline.org/articles/who-can-own-data-data-economy

12. Boerding A, Culik N, Doepke C, Hoeren T, Juelicher T, Roettgen C, et al. Data ownership – a property rights approach from a European perspective. J Civ L Stud. 2018;11:323.