

# **CSA-SCS**

Canadian Sociological Association / Société canadienne de sociologie

## **Canadian Sociological Association Feedback on Draft Tri-Agency Research Data Management Policy**

**Prepared by the CSA Research Advisory Sub-Committee  
September 2018**

Canadian Sociological Association  
Société canadienne de sociologie

[www.csa-scs.ca](http://www.csa-scs.ca)

416-660-4378

PO Box 98014

2126 Burnhamthorpe Rd., W

Mississauga, ON L5L 5V4

## Canadian Sociological Association Feedback on Draft Tri-Agency Research Data Management Policy

### Executive Summary

The Canadian Sociological Association (CSA) welcomes the opportunity to provide feedback on the draft *Tri-Agency Research Data Management Policy*. In preparing its feedback, the CSA consulted widely with its members and with an international expert on the matter. The CSA supports the principle of sound data management. However, members have expressed a number of important concerns about key aspects of the proposed policy and its potential impacts on research, research subjects, and researchers. These concerns are summarized below.

- 1. Model of Inquiry.** The proposed policy reflects a certain model of scientific inquiry and knowledge creation stemming from a particular positivist approach that treats data as decontextualized, and downplays other forms of inquiry rooted in different epistemological, ontological, and methodological perspectives.
- 2. Data Sharing and Protection of Privacy.** The proposed policy appears to prioritize the sharing of data over the protection of the privacy of those involved in the research process. What is more, the proposed policy is not sufficiently forceful about the need to proceed with caution when considering data sharing.
- 3. Anonymity, Consent, and Trust.** With its emphasis on a specific approach to data storage and data sharing, the proposed policy could create challenges in ensuring anonymity, confidentiality, and informed consent. It could therefore seriously undermine the ability of researchers to establish and maintain a relationship of trust with participants. In addition, making data publicly available could lead to violations of the principles of welfare and justice that are at the core of TCPS 2.
- 4. Data in Context.** The proposed policy tends to treat data, not as products of a dynamic and relational process, but as static and decontextualized objects. Without an understanding of the contexts within which data are produced, future possible users could misunderstand, misinterpret, and misuse the data.
- 5. Implications for Research Practices and Knowledge Creation.** The proposed policy could affect the very nature of research and limit the kinds of research undertaken. The pressure to share data could have negative impacts on the quality of research design, on the quality of information offered by both research participants and researchers, and on the wealth of knowledge produced. It could also silence the voices of individuals from marginalized and vulnerable communities, and stunt methodological innovation.
- 6. Indigenous Research.** The proposed policy is silent on issues pertaining to Indigenous research and methodologies. In addition, by mandating data archiving and encouraging data sharing, the policy could further institutionalize Indigenous communities' loss of control over the production of knowledge that concerns them, thereby contributing to the perpetuation of historical relations of colonialism.

**7. Research Collaborations and Partnerships.** The proposed policy's emphasis on a specific approach to data storage and data sharing could result in lost opportunities for cost-effective and highly productive research collaborations and partnerships due to partners' concerns about increased costs, loss of control over data, and potential implications of data re-analysis by unknown future users.

**8. Impacts on Researchers.** The proposed policy could significantly affect the rights and interests of researchers in various ways. It would infringe on the autonomy of researchers to determine data-management practices that are the most appropriate for specific projects. It could result in researchers having to commit a potentially significant amount of time and funds to meet additional data management expectations, which could lead to a loss of productivity, efficiency, and effectiveness. Moreover, the implications for researchers who are not able to easily comply with the policy because their work does not fall within the paradigmatic model guiding the policy are worrisome. New researchers, those belonging to traditionally disadvantaged groups, and precariously employed scholars may be especially affected by the proposed policy. Finally, the proposed policy could exacerbate current capacity-building challenges faced by researchers.

**9. Conflict with Other Government Policies.** The proposed policy could conflict with other government policies such as TCPS 2 and Statistics Canada's restrictions on the use of master data files.

**10. Lack of Clarity of Proposed Policy.** A full assessment of the implications of the proposed policy is challenging because a number of issues are not addressed or require clarification. It is difficult to delineate the boundaries of the proposed policy itself because the links between the policy and related documents, especially the *Tri-Agency Statement of Principles on Digital Data Management* (2016) and *Frequently Asked Questions* (2018), are not clearly spelled out. A roadmap for implementation is also absent.

**Recommendations** to address the CSA's concerns are included at the end of the document.

## **Canadian Sociological Association Feedback on Draft Tri-Agency Research Data Management Policy**

The Canadian Sociological Association (CSA) welcomes the opportunity to provide feedback on the draft *Tri-Agency Research Data Management Policy*.

The CSA is a professional association that promotes research, publication, and teaching in sociology in Canada. Its membership comes mainly from sociology departments in Canadian universities, but there is a notable segment from other social science departments and community colleges, from a wide range of non-academic settings (government, non-governmental organization, and private sector), and from abroad. The total membership is approximately 1,000.

In preparing its feedback on the draft policy, the CSA consulted widely with its members. It also consulted with an international expert, Natasha Mauthner, who has written on issues of data management, data sharing, and data re-use for the past two decades (Mauthner, Parry, and Backett-Milburn, 1998; Mauthner and Parry, 2009, 2013; Mauthner, 2016). The following comments are based on this broad consultation.

The CSA acknowledges and welcomes Tri-Agency efforts to address an important and complex policy matter. The CSA also supports the principle of sound data management. Sociologists across Canada already practice sound data management as part of their research activities and processes, including when: designing and planning research projects; applying for research grants; obtaining research ethics approval; conducting research; storing data; analyzing data; sharing data; and publishing findings. In short, Canadian sociologists have been making and will continue to make sound data management a key focus at all stages of the research process.

The CSA is nonetheless very concerned that the proposed policy has far-reaching implications for how sociological knowledge is produced and how sociologists manage their research and knowledge-making practices. The CSA foresees major impacts of the proposed policy on research, research subjects, and researchers. In the following pages, we raise concerns about key aspects of the proposed policy and conclude by offering recommendations on ways to move forward.

### **1. Model of Inquiry**

The proposed policy reflects a certain model of scientific inquiry and knowledge creation. More specifically, the policy stems fundamentally from a particular positivist approach to research that treats data as decontextualized hard facts, and that is concerned with issues of replicability, reproducibility, and truth claims. The premise seems to be that there is one correct way to collect, analyze, and interpret data. Other forms of inquiry stemming from different epistemological and ontological perspectives that focus on interpretation, meaning construction, and reflexivity are downplayed. However, social constructionist and interpretive approaches that understand social processes and their complex relational contexts are fundamental to much

sociological inquiry. What is more, positivism in the social sciences is not monolithic, a reality that is not well reflected in the proposed policy.

The proposed policy does not speak to many qualitative methodological approaches (i.e. interpretive qualitative, post-qualitative, ethnographic, Indigenous, participatory research, community-based research, historical research, and others) in any meaningful way, thereby failing to appreciate the broad range of research carried out by sociologists. This gives the impression that the policy is driven by a strong bias in favour of quantitative methodologies and that qualitative research is perceived as less worthy and less important to the advancement of Canadian scholarship. In addition, the proposed policy does not adequately address the reality that, for the most part, both qualitative and quantitative research in sociology, and in the social sciences more generally, involves humans.

Though paradigmatic boundaries are fluid and clear lines of demarcation are therefore difficult to draw, the proposed policy is strongly influenced by particular ontological, epistemological, and methodological approaches that marginalize much sociological inquiry and, by extension, a significant number of sociologists in Canada.

## **2. Data Sharing and Protection of Privacy**

The proposed policy appears to prioritize the sharing of data over the protection of the privacy of those involved in the research process. What is more, the proposed policy is not sufficiently forceful about the need to proceed with caution when considering data sharing, an issue that is central to ethical approaches to research. These issues are especially problematic for research involving humans. The intent seems to be to give the proposed policy equal weight to, or even priority over, the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS 2), rather than subsume the former to the latter. This points to a surprising and worrisome shift in policy, especially in today's increasingly digital and globalized world where concern for protecting data and privacy has become paramount in light of frequent data breaches that have not been inconsequential. As such, the proposed policy is at odds with the trend towards increasingly stringent rules governing protection of privacy.

## **3. Anonymity, Consent, and Trust**

The proposed approach to data storage and data sharing would make it more challenging to maintain confidentiality – which is a key consideration for conducting research involving humans as per TCPS 2 – and would increase the risk of individuals being identified. Researchers take great pains to anonymize information by removing personal identifiers to maximize privacy protection, but there always remains a risk of identification. For example, when using quotes from interviews in publications, researchers are constantly concerned that the combination of details revealed in the quotes might make a person more identifiable. Under the proposed policy, researchers would be compelled to remove even more information in order to further reduce the risk of disclosure of participants' identity should the data be used by other researchers in the future. For many projects, the sample size and the nature of the sample would make privacy

protection very difficult. For example, if photo elicitation and photo voice are used as a data collection strategy, ensuring privacy and confidentiality would be practically impossible if the photographs and accompanying interviews were to be placed in a publicly available repository. Similar problems arise when carrying out studies using methodologies such as ethnography and observation. What is more, in some cases, certain types of data are very difficult to anonymize. Another issue pertaining to anonymization is that researchers may well remove so much potentially identifying information that the data, be they qualitative or quantitative in nature, become useless to future users.

Challenges pertaining to informed consent would be exacerbated under the proposed policy. Obtaining informed consent to protect participants already presents challenges under the current framework. Under the proposed policy, achieving the required level of understanding to ensure informed consent would be difficult, if not impossible, because researchers would be limited in their ability to provide information and reassurances to participants about who future users of the data might be, how the data might be used in the future, what the future potential risks might be, and how risks might be minimized. In addition, seeking informed consent for future possible uses of data by unknown users is questionable from an ethics perspective. The rights and responsibilities of both the original researcher and researchers who would re-use the data would be difficult, if not impossible, to establish and enforce. Informed consent statements would need to be changed significantly to accommodate additional and detailed information and to buttress wording to protect participants, as well as researchers in case of legal action, a change that would be practically impossible to make. And, given that anonymization is weakened where digital data linkage increases the risk of identification, the ethics of anonymizing participants' stories and, through consent agreements, informing them that this will protect their identity are questionable.

The issue of breach of trust looms large for researchers under the proposed policy. The ability to establish and maintain trust between researcher and participant is a crucial aspect of the research process and ultimately of the production of knowledge. The information that participants agree to share with researchers is a gift from participants and a privilege for the researcher to receive. While this is well recognized in Indigenous research (Kovach, 2009), other research traditions also understand data in this way. When listening to participants' stories and obtaining information, often of a very personal nature, researchers enter into a relationship of trust with participants, a relationship that implies moral responsibilities and obligations towards them. People tell researchers about their experiences in confidence because they expect that the details of those stories will not be made public. The proposed policy would make it more difficult for researchers to ensure that the data they collect are not used for purposes not intended and in different contexts, which could have harmful impacts on participants. In addition, under the proposed policy, different levels of government, private-sector entities, and other organizations, both domestic and international, could seek and obtain access to data that were never meant to be shared with, or used by, such bodies. The proposed policy thus opens up the possibility that the trust-based relationship between researcher and participant would not be easy to establish given that the researcher would not be in a position to provide certain reassurances that they are currently able to give pertaining to, for example, confidentiality, future use and users of data, and associated risks. In this context, the erosion of the perceived trustworthiness of researchers to which the proposed policy could give rise might affect future potential relationships with participants and also dissuade the public from participating in research. The above concerns are

relevant to all research involving humans regardless of whether or not they are direct and immediate participants in the process.

Concern for welfare and justice are core principles of the TCPS 2 and, ultimately, making publicly available data from research involving humans could lead to violations of these core principles and by extension of the ethics policy itself. Researchers have a responsibility to protect the welfare of participants and limit risks to which they could be exposed by making some of their information publicly available. Moreover, the principle of justice is intended to ensure that the benefits of research participation are equally distributed. Those who are marginalized and vulnerable have the most to lose from research data being made publicly available. Indeed, when marginalized or vulnerable groups cannot participate in research for fear that their information will be made public, they are not treated on equal terms with others.

The proposed policy could exacerbate power differentials between researcher and participant, a structural inequality that sociologists have long recognized as problematic. For example, in order to comply with the policy's expectations and to secure funding, researchers could pressure participants to provide blanket informed consent to the use of their data. In this context, there is a risk that data sharing would further marginalize vulnerable communities and populations, both locally and globally, thereby bolstering power inequalities as well as relations of colonial exploitation.

#### **4. Data in Context**

As noted earlier, the proposed policy is strongly influenced by a particular model of science, and this is reflected in the way in which data seem to be understood. This model treats data as static and decontextualized objects that carry inherent meanings. Data are therefore assumed to be separate from the contexts in which they are produced and in which knowledge is created (Mauthner and Parry, 1998). Sociologists and other social scientists – as well as researchers from the natural and health sciences – have put into question this assumption and have instead argued that data are the product of a dynamic process, and are inextricably tied to the personal, social, emotional, intellectual, moral, political, historical, and material conditions of their production.

At the core of an understanding of the production of data as relational and socially constructed is the recognition of the specificities of the encounter between researcher and participant as well as between researcher and research materials, which the proposed policy does not reflect. What is more, researchers are not mere technical data collectors, but rather are central to the construction, interpretation, and theorization of data. Data are therefore imbued with the personal, social, emotional, intellectual, theoretical, moral, and political lenses or filters of researchers. The proposed policy assumes that data are decontextualized objects that any future user can selectively retrieve and reconstitute for new purposes. While the re-use of certain types of data may be less problematic, these data nonetheless need to be understood in the context of their production. Data taken out of context can be misunderstood, misinterpreted, and misused. A policy that adheres to notions of context-free data and data-user interchangeability runs the risk of fostering poor research practices that may not only cause harm to participants and researchers, but may also result in bad science that has no value to policy or to society more generally.

Sociologists appreciate the distinctive nature of particular kinds of inquiry, such as realist/positivist, interpretive/constructionist, critical/transformational, post-structuralist, Indigenous, post-qualitative, ethnographic, participatory, and community-based. These epistemological and ontological orientations stem fundamentally from researchers and their standpoint, and shape the entire research process, including the nature of the data collected. It is not clear how the epistemological and ontological orientations of the original research project could be communicated to those wishing to use the data for other purposes. It is perhaps unethical and even bad research practice to use data collected by a researcher working under one paradigm to answer questions posed by a researcher working under another paradigm.

It would seem that testability of findings is a central goal of the proposed policy, which again reflects a particular ontological, epistemological, and methodological stance. However, the statements and conclusions of researchers in all disciplines are constantly evaluated, confirmed, and challenged by new research. In most cases, the new data that are produced to evaluate, confirm, or challenge existing research are not the same data as the original data. Also, as mentioned earlier, much social science data are produced through methods such as field research, surveys, and interviews that are confined to a particular time, place, and group of people.

The stipulation in the proposed policy that metadata be included as part of the archiving of research data, especially for use by other researchers, is problematic. The assumption that the inclusion of such information will make the anonymized data whole for the purpose of generating new questions, findings, or theories fails to understand contextual information as constituting rather than adding meaning (Mauthner and Parry, 2009, 2013; Mauthner, 2016). It is also critical to keep in mind the fundamental ethical concerns about confidentiality when background information is supplied on how, why, by and for whom the research was carried out, and when research notes and other documents are made available.

## **5. Implications for Research Practices and Knowledge Creation**

Many of the concerns raised above highlight that the proposed policy could affect the very nature of research and limit the kinds of research undertaken. Indeed, the policy could result in a shift in the type of research methods used, the type of data collected, the questions asked, the relationship between researcher and participant, the level of public participation in research and, more broadly, the knowledge that is produced. Both qualitative and quantitative research could be affected. A few examples will help illustrate this point.

The pressure to share data could stunt methodological innovation as researchers become concerned about a variety of issues, such as the identification of participants, as in the case of photo elicitation and photo voice mentioned earlier.

Even long-standing methods, such as in-depth interviewing, could be affected negatively. More specifically, researchers might structure the interview to limit participants' use of personal identifiers when telling their story, a strategy that could impoverish the data collected and stifle methodological innovation. Showing interest in the personal details of participants' lives and using this information to generate additional questions and to gain a deeper understanding of the

story are key practices of successful interview-based research, in terms of both developing a relationship of trust and gathering rich and textured data. When participants know that their stories could be used by other researchers in the future, the details they choose to share and the nature of their stories will likely be different. For researchers, knowing that the interview could be accessed by other users may temper their willingness to talk about personal experiences as part of building trust with participants, because doing so could have unintended negative consequences. With the possibility that data would be publicly available, certain questions could no longer be asked and certain answers would no longer be given. This would be particularly the case in interviews with individuals from vulnerable and marginalized groups. We do not want to silence these voices, but instead provide opportunities for such voices to be heard. Also, because potential participants would be informed that their data could be used by other researchers in the future, people would most likely be more reticent to take part in studies.

The case of historical research, which many sociologists pursue, offers another example of how the proposed policy could limit the kinds of research undertaken because of the undue burden it would impose on researchers. Historical research often requires spending years painstakingly compiling and analyzing data from a variety of sources (which are often publicly available), and researchers typically create various types of files for their own use. Should they be required or feel pressured to make all their documents publicly available, putting these records in a format that would be accessible to others could take a considerable amount of time. It seems unfair to expect these researchers to make public their personal research notes when other researchers have equal access to the original public records. It may well be that researchers will decide that, given the expectations of the proposed policy, it is not worth their time to invest years in historical data analysis, when others can immediately enjoy the fruits of their labour with no effort. The issues raised with this example of historical research are applicable to many other types of research in the social sciences. What is more, a token gesture such as an acknowledgement from those who would re-use data would fall short of fully recognizing the true value of the work and products of the initial researcher. As importantly, acknowledgement, or any form of recognition might, not be welcome or desirable if future researchers were to misinterpret or misuse the data.

These different examples illustrate how the proposed policy could have a huge impact on the quality and wealth of knowledge that is produced, thereby limiting what can be learned about the social world. These limitations would have serious implications for the development of sound evidence-based policy. They would also stifle innovation that is highlighted as a benefit of the proposed policy.

## **6. Indigenous Research**

The proposed policy is silent on issues pertaining to Indigenous research and methodologies, a consideration that should have been an integral part of the policy from the outset. This absence is quite perplexing given the emphasis on this matter by the Truth and Reconciliation Commission and by the three federal research funding agencies in their grant application criteria. As SSHRC has recognized well in Chapter 9 of the TCPS 2 and in Guidelines for the Merit Review of Indigenous Research, the history of Indigenous peoples' and communities' experiences of

Western research has been marked by exploitation, oppression, paternalism, subordination, and colonialism. The proposed policy could nonetheless perpetuate these destructive relations for a variety of reasons, some of which are highlighted here. Mandated data archiving and the pressure to share data would further institutionalize a process whereby Indigenous communities lose control over the production of knowledge that concerns them, a situation that would be contrary to decolonization and indigenization commitments made by government and various other institutions. In addition, Indigenous methodologies that flow from tribal epistemologies are not reflected in the proposed policy. Rather, the methodological approach and assumptions that imbue the proposed policy are centred in Western epistemologies that dominate academic knowledge production (Kovach 2009).

## **7. Research Collaborations and Partnerships**

The proposed policy could have a significant impact on researchers' ability to develop cost-effective and highly productive collaborations or partnerships with other entities, such as research bodies, community organizations, not-for-profit or philanthropic organizations, governmental departments and agencies at different jurisdictional levels, unions, and private-sector companies and associations. The requirement to deposit data as prescribed by the proposed policy and the pressure to make data publicly available could engender additional costs for such organizations. In addition, these organizations may not be interested in sharing data with anyone other than their collaborators or partners. They may also be deeply concerned about how data would be re-used and who would re-use the data, especially in situations where information is of a sensitive nature or when research focuses on marginalized groups. Knowing that data would need to be deposited in prescribed locations and within specific time limits, and that data could be made publicly available might result in organizations being less willing to partner for research out of concern that future users would publish findings that could be threatening to the group. Therefore, for a variety of reasons, the proposed policy would add a layer of complexity that could severely limit collaboration opportunities and exacerbate the difficulty for researchers to gain access to some types of data. Consequently, socially beneficial research that would otherwise result from such collaborations would not take place. Methodological innovation could also suffer.

## **8. Impacts on Researchers**

The proposed policy could significantly affect the rights and interests of researchers in various ways as discussed below.

Researchers have a long history of storing and sharing data, but the way this is achieved has been a matter of individual choice. The proposed policy would remove the autonomy of researchers to determine data-sharing practices that are the most appropriate for specific projects. By explicitly encouraging researchers to share their data for re-use, the proposed policy puts the onus on researchers to justify why data should not be shared. In addition, Section 3.3 of the proposed policy is a major source of concern given that it would require researchers to deposit research data, metadata, and code into a recognized digital repository. Over and above technical issues

pertaining to digital storage, this requirement would erode academic freedom and autonomy by removing researchers' right to determine the best strategy for managing their data from individual projects. By extension, researchers' control over their research processes and products and over their intellectual labour would be significantly diminished.

The collection and analysis of data is often time-consuming, challenging, and expensive. Data collection and current data management practices necessitate huge personal and professional investments. Additional data management requirements could exacerbate these already demanding conditions. What is more, data are part of researchers' intellectual capital and are key to securing career advancement, rewards, and recognition. This career imperative should not be dismissed given the current context of high and ever-growing research and publication expectations. New scholars, who experience considerable pressure to publish early, could potentially lose significant research time to data management imperatives, depending on the type of research they conduct. In addition, the ability to comply with the proposed policy would be compromised for those without research funding, for those conducting particular types of research, and for those who, for a variety of reasons, are unable to publish relatively quickly from their research.

The proposed policy would result in researchers having to commit a potentially significant amount of time and funds to meet additional data management expectations. Indeed, in most instances, preparing documentation, such as data files, metadata, code books, and other files for storage and for potential access and use by others would be time-consuming and would require specific skills and therefore investment in training. Researchers are not archivists and should not be expected to prepare archives for use by others. It should also be noted that data gathering agencies such as Statistics Canada have specific departments that prepare data and supporting documentation for public use, and it often takes them one to two years after the initial gathering of the data to prepare them for public use. Sociologists already carry out research on very limited budgets, and new expectations flowing from the proposed policy would exacerbate an already difficult situation. For those with no or limited research funds, the implications would be more serious because their ability to conduct research could be greatly affected, which would have negative consequences for their career. What is more, the time required to prepare documentation for research ethics boards and secure ethics approval would likely increase given the added pressure to make data publicly available. These many demands on researchers could result in a loss of productivity, efficiency, and effectiveness. They would also be a source of additional stress. These various issues are compounded for precariously employed scholars.

The proposed policy raises a number of concerns about compliance. The model of inquiry on which the policy is based does not sufficiently reflect the kinds of research and knowledge-making activities that many sociologists pursue. This means that the data-archiving and data-sharing expectations were developed without adequately considering forms of inquiry stemming from different epistemological, ontological, and methodological perspectives. Sociologists who are guided in their work by different models should therefore not be expected to comply with requirements that were not designed with their types of research and knowledge-making activities in mind. The implications for researchers who are not able to easily comply because their work does not fall within the paradigmatic model guiding the policy are worrisome. The proposed policy seems to imply that, in order for research to be considered worthy, researchers

need to work within the dominant model of inquiry. Some researchers could feel compelled to abandon their worldview and change their approach to research for fear of losing access to funding support. Not only would this be inequitable, but it would impoverish research and stunt innovation, which in turn would have detrimental consequences for knowledge and policy-making. Such an outcome would be contrary to the objective of research excellence pursued by the proposed policy.

The proposed policy could have particular ramifications for researchers, such as women, members of the LGBTQ+ community, racialized persons, Indigenous peoples, and persons with disabilities, who have traditionally faced differential and unequal treatment in institutions of higher learning. The proposed policy could exacerbate inequities that individuals from these groups experience when managing their research and knowledge-making practices, which would have significant implications for the production of sociological knowledge. The proposed policy does not indicate whether care was taken to avoid negative impacts on these groups.

If the requirement to deposit data in recognized repositories and the bias in favour of data sharing are in any way reflected in criteria for adjudication and release of research funds, the proposed policy risks marginalizing certain types of research and even large segments of disciplines, a situation that would have a significant impact on researchers.

The proposed policy runs the risk of increasing institutional and Tri-Agency control over yet another component of the research process that has traditionally fallen under the purview of individual professional academics. The assumption seems to be that current peer-review processes are inadequate, and that academics have neither the capacity nor professional integrity to assess the validity of their own research, are not being honest about the data they are collecting and how they are using them, tend to waste the grant monies they receive, and therefore require yet more administration and evaluation mechanisms, which ultimately undermines the autonomy of academics, interferes with their research processes, and has at its core a fundamental disrespect for the profession. This is problematic because the existing academic review process already provides a strong and effective overview mechanism for research.

Given the model of inquiry that infuses the proposed policy and also in light of different kinds and levels of resources available across educational institutions, there is a concern that the proposed policy would be applied unevenly across Canada. This not only has implications for researchers within disciplines, but also among disciplines within the same institution. Unevenness in application would compound an already unfair situation where certain institutions, disciplines, and models of inquiry are privileged, and others are disadvantaged. In this context, the challenges related to capacity building are tremendous and require urgent attention. Data archiving, preservation, protection, sharing, and accessing raise capacity issues pertaining to, for example, infrastructure, technology, training, skills and knowledge of research-support staff, costs, and funding support. Capacity building challenges flowing from the proposed policy are compounded by current challenges faced by researchers and institutions as they continue to adapt to major policy changes governing research, for example, policies on ethics.

## 9. Conflict with Other Government Policies

The proposed policy seems to be in conflict with other federal government policies. As mentioned earlier, making publicly available data from research involving humans could lead to violations of core principles of the TCPS 2, especially those pertaining to concern for welfare and concern for justice.

Another example relates to Statistics Canada survey data. More specifically, over the past two decades, Canadian sociologists who conduct quantitative research have been increasingly relying on the use of Statistics Canada's master files of survey data (as opposed to public use versions). These data contain sensitive information about respondents and access to these data has been restricted to their use in Research Data Centers. Such data are not, and will not become, available to researchers, let alone to the public, outside the confines of Research Data Centres. Violation of these restrictions would result in researchers being denied access to Research Data Centres. Canada's national data agency thus currently makes access to its data more difficult than what the proposed policy suggests researchers do for other users. What is more, Statistics Canada data that researchers may use are often diluted to the point of being of little value.

## 10. Lack of Clarity of Proposed Policy

There are many issues that are not addressed or that require clarification, which make it difficult to properly assess the full range of implications of the proposed policy. For example, key terms are not adequately defined in the policy itself and the policy is not sufficiently detailed. The proposed policy includes links to documents such as the *Tri-Agency Statement of Principles on Digital Data Management* (2016) and *Frequently Asked Questions* (2018) that provide useful and important information. However, it is not clear how these documents relate to the draft policy itself. These documents include information that seems to be core to the policy, yet this is not clearly spelled out. The proposed policy does not include a clear roadmap, which is problematic in light of the significant and fundamental shift that the policy entails for research and knowledge-making processes and practices. Also, the problem that the proposed policy aims to address is not clear, and information is lacking to support the policy direction being proposed.

## **Recommendations**

As indicated at the beginning of this submission, the CSA agrees with the broad principle of sound data management and appreciates that the three agencies have committed time and resources to engage with this complex matter. However, the CSA is concerned about the potential impacts of the proposed policy on data and knowledge production, especially when considering the many issues to which the policy gives rise. In the spirit of contributing to the formulation of policy that is respectful, prudent, flexible, fair, effective, and transparent, the CSA offers the following recommendations.

1. That the three agencies proceed slowly and with caution in formulating and implementing the policy on research data management.
2. That the policy explicitly acknowledge the existence of different and equally valid models of scientific inquiry, and that it advocate in favour of a flexible approach to research data management that recognizes and accounts for such differences.
3. That the policy explicitly recognize the paramount importance of privacy protection and make such protection a guiding principle.
4. That the Social Sciences and Humanities Research Council develop a stand-alone policy on research data management geared to its particular stakeholders.
5. That, before the policy is finalized, a comprehensive analysis be carried out to determine whether the policy is consistent with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS 2) and other relevant government policies; and that, in the event conflicts are identified, action be taken to resolve these discrepancies.
6. That the policy be consistent with federal government commitments to reconciliation with Indigenous peoples by: acknowledging the specific circumstances surrounding Indigenous research broadly speaking; recognizing Indigenous information governance as a priority; supporting Indigenous communities' control over whether and how data about them are gathered, stored, and shared; and committing the three agencies to abide by the principles of *Ownership, Control, Access and Possession* (First Nations Information Governance Centre 2014).
7. That the policy include a requirement to involve Indigenous communities in the development of data management plans for Indigenous-related research.
8. That, in consultation with relevant Indigenous organizations, a mechanism be established for ongoing and meaningful engagement with Indigenous communities on the implementation of the policy.
9. That, consistent with the federal government's commitment to Gender-based Analysis Plus, an analysis of the implications of the proposed policy for researchers from traditionally

disadvantaged groups be undertaken before the policy is finalized, and the results of this analysis be shared with all stakeholders.

10. That, for research involving collaboration with partners, the policy include provisions for the involvement of these partners in the development of data management plans.
11. That an analysis of costs and benefits of the policy be undertaken with a view to understanding the full range of implications of the policy for institutions, researchers, research participants, and Canadians more generally, and that the results of this analysis be shared with all stakeholders.
12. That the policy explicitly allow researchers to exercise discretion with regards to the depositing and sharing of data.
13. That, should the policy pursue a more directive approach to data management planning, social science research involving humans be exempt from the provisions of the policy dealing with the depositing and sharing of data.
14. That the depositing and sharing of data not be a consideration in the adjudication of grants and release of funds for social science research involving humans.
15. That, consistent with an incremental approach to implementation proposed in the policy, pilot projects first be undertaken to test the soundness of the policy, identify problems, and propose changes to the policy accordingly.
16. That a national committee comprising funding agency, institutional, faculty, Indigenous, and archiving community representatives be established to oversee the implementation of the policy, including the above-recommended pilot projects.
17. That the proposed policy include a commitment to secure new funding for the establishment of a Tri-Agency program to support – technically, financially and otherwise – the development of individual researcher and institutional capacity to meet the requirements of the policy.
18. That the links among the different Tri-Agency documents associated with the policy (especially the draft *Tri-Agency Research Data Management Policy*, the *Tri-Agency Statement of Principles on Digital Data Management*, and the *Frequently Asked Questions*) be more clearly established in order to delineate the boundaries of the policy itself.
19. That the three agencies develop an effective strategy to consult, communicate, and engage with various stakeholders as the initiative evolves.

## References

- First Nations Information Governance Centre (2014). *Ownership, Control, Access and Possession (OCAP™): The Path to First Nations Information Governance*. Ottawa: The First Nations Information Governance Centre.
- Kovach, M. (2009). *Indigenous Methodologies: Characteristics, Conversations, and Contexts*. Toronto: University of Toronto Press.
- Mauthner, N. S. (2016). Should Data Sharing Be Regulated? In: Will C. van den Hoonaard and Ann Hamilton (eds.), *The Ethics Rupture: Exploring Alternatives to Formal Research Ethics Review*. Toronto: University of Toronto Press, 206-229.
- Mauthner, N. S. and Parry, O. (2009). Qualitative Data Preservation and Sharing in the Social Sciences: On Whose Philosophical Terms? *Australian Journal of Social Issues*, 44(3), 291-307.
- Mauthner, N. S. and Parry, O. (2013). Open Access Digital Data Sharing: Principles, Policies and Practices. *Social Epistemology*, 27(1), 47-67.
- Mauthner, N. S., Parry, O. and Backett-Milburn, K. (1998). The Data Are Out There, or Are They? Implications for Archiving and Revisiting Qualitative Data. *Sociology*, 32(4), 733-745.