CAREB-ACCER Response to the Draft Tri-Agency Research Data Management Policy

September 28, 2018

The main focus of this CAREB-ACCER response relates to how the Policy may affect research involving humans. This response reflects the position of the CAREB-ACCER Board of Directors and members of our association. It was also influenced by feedback received at several institutions following a consultation on the draft policy.

1. Preamble, Policy Objectives and Scope

CAREB-ACCER agrees with the overall intent and objectives of the policy, most notably to "support Canadian research excellence" (section 2) and that research should be conducted ethically and with the highest standards.

Scope

The Policy is inconsistent with respect to whom it applies. For example, section 3 indicates that the "policy applies to grant recipients and to institutions administering tri-agency funds" and in the "Institutional Strategy" section, that institutions should ensure "that their researchers have data management plans in place." It is unclear why the Policy "does not apply to scholarship, fellowship or Chair holders" (section 3). CAREB-ACCER supports an approach that is consistent with other Tri-Agency policies, e.g., the Tri-Council Policy Statement (TCPS 2) and Tri-Agency Framework on the Responsible Conduct of Research (RCR Framework). While we understand there may be a need for specific requirements that apply to Agency-funded research, we hope the Agencies use this policy to identify and uphold standards of data management for all research. The minimum would be that all researchers have a data management plan.

Terminology

We recommend that the terminology of this policy be consistent with other Tri-Agency documents, including the TCPS 2 and the RCR Framework. Many comments we received raised issues around defining various terms. It would also be helpful to include definitions of words used in the policy within the document itself, rather than in an FAQ.

- "research data:" the CASRAI definition is used, rather than the TCPS 2 (data set).
- "research" is not defined in the document nor in the FAQ: the TCPS 2 definition of research, i.e. "an undertaking intended to extend knowledge through a disciplined inquiry and/or systematic investigation" (article 2.1), could be used.
- The term "Institution" should be used throughout as the policy is not just for Universities and Colleges. Use of "Campus" implies a University or College.

Concerns were raised that the policy and its implications were focussed on quantitative types of research and would be less applicable to qualitative methodologies (e.g., mixed-methods, participatory research). An example is a reference to the possibility of replicating findings, which is not possible nor desirable in certain fields of research.

2. Institutional strategy

This section mentions that institutions should "develop their own data management policies and standards" for data management plans (DMPs). This would likely lead not only to a needless duplication of work, but also potentially to inconsistent standards across institutions. For research involving humans, this would have a major impact on REB review, particularly multi-jurisdictional REB review. It would seem preferable to develop federal policies and standards which integrate a multi-field perspective and are flexible enough to apply to all institutions in order to ensure that research conducted in multiple institutions does not have to meet varying standards. REBs/REB administrators should be actively engaged in the development of these strategies given that much of the data affected by this policy is human participant data.

There were also concerns surrounding the gradual implementation of the policy and the statement that it would be reviewed « as appropriate ». It is not clear how the former would proceed nor how the above issues would be avoided. A set timeline for review of the policy would be preferable, as this would facilitate planning and implementation.

3. Data Management Plans

In CAREB-ACCER's consultations, most agreed that DMPs are an important component of good data management practices. In order to review and properly assess the risks and benefits of research involving humans, REBs already require information regarding how data will be collected and safeguarded throughout the life cycle of the research project. We believe that proper data management, which includes a data management plan, should be in place for all research conducted within an institution, not merely for research that is Tri-Agency funded.

We also support the requirement to create a DMP at the application stage (recognizing that it will evolve as the project progresses), as well as common tools and set of standards for developing and evaluating DMPs.

4. Data Deposit

The issue of Data Deposit is a complex one, as it also touches upon data ownership (does the data belong to the researcher, the participants, the funding body, or the institution), data use, curation of data and intellectual property.

The policy states that grant recipients would be "required to deposit into a recognized digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications, pre-prints, and other research outputs that arise from agency-supported research."

Data Curation vs. Data Deposit

The terms data "stewardship" and "deposit" are used throughout the Policy. It is unclear how merely depositing data will meet the Agencies' goal of making data accessible for reuse, verification, and replicability (where applicable). Many researchers and research/REB administrators raised issues around the usability of the data, file formatting, etc. Depositing data may help researchers meet their responsibilities regarding record keeping and ensuring that data is available for verification but if the goal is for other researchers to use the data, or reanalyze it, curation is needed.

Potential Challenges

It is unclear if or how the requirements related to DMPs and Data Deposit will be reflected in the TCPS2 and whether these requirements affect the end point for ethics review. The potential risks to participants however, do not necessarily end at this point, particularly if the data will be made available more widely. This has many ethical and practical consideration for ethics review, which need to be discussed before applying this requirement to all research involving humans.

Additional information / clarifications would be helpful on the following:

- What is a "**recognized** digital repository"? On what basis is this determined (e.g., is there a list of criteria, who makes the determination)?
- Does the location of data storage matter? If so, does this apply for all types of research or only for some types (e.g., identifiable human data on Canadian vs. U.S. servers (also see TCPS 2, Article 12.2e))?
- Will a list of such repositories be created? If so, by whom; where; how will redundancy and a waste of time and resources be avoided? Will the Tri-Agencies facilitate this work?
- When should the data be stored, e.g., after publication, after collection is finished, once the grant is over (if applicable)?
- For how long must data be stored? When is it acceptable to destroy it?

It may be prudent to limit mandatory deposits to data that is low risk, for example non-identifying data with a low probability of re-identification or where risks are low and participants have agreed to waive anonymity (see TCPS 2, Chapter 5, section A "Ethical concerns regarding privacy decrease as it becomes more difficult (or impossible) to associate information with a particular individual. These concerns also vary with the sensitivity of the information and the extent to which access, use or disclosure may harm an individual or group."). This would minimize potential risks to participants and would provide some time to discover and manage operational issues that could affect the safety and wellbeing of participants.

Other potential ethical and operational issues, recognizing that some of these could be addressed by defining standards associated with acceptable repositories:

- If someone uses a researcher's data in an irresponsible way, which may cause harm to participants, who is responsible for the transgression?
- Withdrawal of data – can data be withdrawn if the data are in a repository? If so, how? Until when? Who is responsible for this?

- What effect will this have on how participants consent to research and the information regarding data storage that is provided to them?
- Even if there are no direct identifiers in the data, it is difficult to predict how access to very large data sets could affect the confidentiality of data.
- IP: how do you ensure that recognition is given to the researchers who collected the initial data?
- Quality of data and reuse – the premise seems to be that all collected data is of good quality, which is not necessarily the case (e.g., research methods were flawed, data is not robust). Should an assessment be made on the quality of the data before it is made available for use by others?
- Many data can only be analyzed within context. There are potential risks to participants should this data be made available widely.
- Secondary use of data – would these deposits automatically mean that data is "public", and therefore that it is not necessary for a researcher to submit to a REB for secondary use of the data (see TCPS2 articles 2.2a and 2.2b)?
- Data security: how do you protect data from a potential breach?
- Security of researchers: in some cases, the researchers themselves could be in danger should their data be made public.

Conclusion

CAREB-ACCER fully supports the use of DMPs in research. Regarding data deposits, it is too early to tell what the risks are for human participants, be it on an individual or a community level. Human rights and the protection of personal data are not directly addressed in the document. There is no mention of research with or by Indigenous peoples, nor of consultations with these groups in order to obtain their opinion on how, for example, application of OCAP (Ownership, Control Access and Possession) will be respected. Considering the impact that this could have on human participants, we feel that these issues need to be widely discussed.

It may be preferable to proceed incrementally when it comes to the mandatory deposit of research, and perhaps begin with data that is not highly sensitive and does not involve humans at the outset, followed by low-risk, medium-risk data and so on.

We thank the Tri-Agency for the opportunity to comment on this draft. We have seen through the CAREB-ACCER consultation, and in the high response rates and very relevant feedback received during institutional consultations that participant safety and research excellence are important to all those in the research community. Dialogue is therefore crucial in order to ensure that policies and guidelines regarding data management reflect the commitment of this community to conducting ethical research of the highest quality and appropriately safeguarding their data.