

August 30, 2018

**Response of the Canadian Historical Association
to the draft Tri-Agency Research Data Management Policy**

The Canadian Historical Association (CHA) is the professional body representing historians employed in universities, government, and private practice in Canada. The CHA supports the development of a clear and practical Research Data Management Policy [RDMP]. In a digital age, it is important for researchers to consider how they will manage their data right from the start. Moreover, the RDMP policy is very much in line with trends elsewhere in the world.

While the CHA supports the broad principle of data management we have some concerns that the attempt to create a single policy for all the disciplines covered by the tri-council will either be too vague to be useful or so specific that it will impose impractical standards on some disciplines.

The discipline of History straddles the divide between the Social Sciences and the Humanities and as a result there is a wide range of methodologies and definitions of data even within our single discipline. Each presents different issues when it comes to data management, preservation and sharing.

The draft policy FAQs defines the relationship of research materials to research data. “Research materials serve as the object of an investigation, whether scientific, scholarly, literary or artistic, and are used to create research data. Research materials are transformed into data through method or practice. Examples of research materials may include bio-samples for a geneticist, primary sources in an archival fonds for an historian.” It is our understanding that the new policy is not aimed at research material but the management and preservation of data derived

from the material. According to the draft policy “...the corresponding research data could be gene sequence data, chronological analyses of ideas and contributions, and the behaviour of the zebrafish under certain conditions, respectively.”

Many historians follow a social science methodology where there is a clear extraction of data from research material, the creation of an intermediary data set, and from that data set certain functions are performed to draw research conclusions which are presented in articles, books, video, websites and other outputs (hereafter collectively referred to as research outputs). It is to this type of research project that a Data Management Plan (DMP) fits best. Some of these data sets (census, Metis Scrip documents, homestead documents, immigration documents) have dozens of fields, tens of thousands of entries and take years to assemble and maybe funded through sources beyond the tri-council. Other datasets are so small that they are not useful to others. Consideration needs to be given to a threshold size for datasets so that repositories are not clogged with tiny, idiosyncratic datasets. For those larger projects that meet a threshold, it would be useful to promote a standardization of the data organization so, for example all geo-spatial data would contain specific fields and meta data that would allow it to be useful to other researchers. In such cases it may be more useful for certain institutions to specialize in the curation of specific types of data with a standard data structure than to have every university and government department create their own repositories with innumerable idiosyncratic data structures and codings.

Many other historians follow a humanist approach which often results in no intermediary data set between the accumulation of research material and the creation of their research output. These historians, in common with many other humanists, pore over their research materials, and commit their conclusions to publication without the creation of a dataset. In the case of these scholars, their research conclusions are supported directly by quotations from their research material, cited in their outputs. In an increasing number of cases, all the research material is already available digitally in an archive. Yet, “the draft policy proposes that grant recipients be required to deposit, in 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.” This statement in the draft policy suggests these humanists scholars should deposit their research materials which they have drawn on directly. Clarity is needed and we would strongly suggest that research materials be exempted from the expectations for deposit laid out by the RDMP. It is unlikely that historians’ intermediary drafts, research summaries or bibliographic citations, would be useful to others (except to biographers interested in a few exceptional scholars) and to preserve them would put an unnecessary burden on institutional repositories with no return to other scholars.

Some of these humanist historians use tri-council funds to travel to locations where their research material is only available in those archives. In some cases, archival holdings are covered by copyright and related policies that prohibit their reproduction online or in print. In many instances these historians create a digital copy of the research material that is pertinent to their study and use the digital copy as their research material. There may be no intermediary creation of data, but in some instances that research material itself could be useful to others. This is particularly true of photographic and video sources. In those cases, the research material would be more discoverable if the digital images were aggregated at the original archives itself, rather than scattered in small collections in university repositories around the world.

Still other historians engage in oral history which involves interviews with knowledgeable people and these are often preserved as audio or video recordings. By the definitions provided in the draft policy, these recordings would be research materials and so not subject to a DMP. These oral historians may proceed either to an intermediary data base or direct to a research output, but in both cases the research materials themselves are uniquely valuable and very useful research materials for future historians. Unlike databases, these do not require standard data structures, only basic metadata, to be discoverable and useful to other scholars. Consideration should be given to a management and preservation plan that would preserve and make available this category of research materials created with tri-council funds.

Many historians engage in research on sensitive topics with Indigenous people or marginalized groups where there may be a social science method used and an intermediary data base created, but the data may only be obtainable if a commitment is made to the research participants that the data will not be shared. This type of research is valuable and provisions need to be made for such circumstances.

Another set of historians conducts research to create digital objects such as three dimensional renderings of historic buildings, software to annotate three dimensional objects, or they create video games. Other historians study video games. Some effort will have to be done to distinguish research materials and research data in these new fields.

Finally, many historians employ a combination of research methods and data to answer their research questions. A single research project may involve interviews, text analysis and locational coding using geographic information systems, and the creation of an online educational game as an output. Requiring DMPs for each type of data may discourage innovative, interdisciplinary methods as the bureaucracy to manage the data gets too onerous.

In sum, the CHA supports the data management plan initiative so long as: 1) it is sensitive to the diverse needs of scholars; 2) that it is recognized that some scholarship does not create "data" as defined by the policy so for that scholarship no management plan is required 3) that there be a threshold for small projects that will not require data preservation and care given not to

burden researchers using a variety of data sources; 4) it is recognized that all universities and government agencies may not always be the best repository for complex data and consideration be given to supporting dedicated repositories for certain kinds of data; 5) some research material is more valuable to future scholars than the research data derived from it and those materials be identified and policies established for their preservation; 6) that some research data should be allowed to be kept confidential; and 7) that the policy acknowledge that in some cases it is hard to distinguish research material from research data and flexibility be allowed the researcher. We need always to recognize that the main goal of research is new knowledge creation and that extreme care needs to be taken so that the laudable goals behind the draft research data management policy do not have the effect of discouraging research.

Yours sincerely,

Dr. Adele Perry, president
Canadian Historical Association