June 19, 2014

Research Data Canada welcomes the chance to comment on the CFI proposed initiative in cyber-infrastructure. It is gratifying to see the recognition of the importance of data and the means to make optimal use of data, share data, and preserve data.

Data as both product and foundation - We believe that a focus on data must recognize that data is no longer simply the product of research activity; it is increasingly the foundation of research activity, often in fields significantly different from those in which the data were created. Data as the foundation of research is true in all areas of discovery and especially true in the social sciences and humanities. Building data structures that enable appropriately open access should be an explicit goal of any project that addresses the second challenge outlined on page 2 of the consultation document. The policy initiatives by the Federal Government to require open access to government funded data including the data from federally funded research make such a goal imperative.

Inter-operability in an international context - Projects funded by this CFI initiative should demonstrate detailed knowledge of disciplinary and cross-disciplinary international efforts in data standards development. Canadian researchers must be in a position to easily access international data, integrate that data with Canadian data, and make their own data available to international collaborators and other researchers. The day is fast approaching when data-citation will be important to career advancement in many disciplines. Attracting and retaining research talent will become dependent on data environments that follow global standards and provide appropriate open access. Research data interoperability should be a key goal and evidence of integrating international standards should be a criterion for success.

Strategic or domain specific? - The first paragraph on the second page speaks to "areas of strategic importance to Canadian research institutions", the section on Timelines says, "Concurrently, the CFI will launch a competition for the domain-specific data projects and the Review and Decision Making-Process section says "…are broadly generalizable across an area or domain of research". "Domain-specific" is distinctly more restricted than "areas of strategic importance". The criteria be open enough to permit, for example, a consortium of researchers and university research libraries to build a data repository for researchers not currently served by big science repositories as well as to put in place the appropriate software to facilitate the data deposit and re-access as well as analytical tools. There is significant international data expertise within Canadian universities and research libraries to drive such projects. It is important to note that data intensive research is not necessarily "Big Data" intensive. There are many researchers working from and creating comparatively small data-sets who are doing ground-breaking work. They need data infrastructure for work that is truly data dependent. And their data must be appropriately accessible to others. In many disciplines particularly in the humanities, the coterie of "digital" scholars is still small and may not yet have the capacity to mount a domain specific project, but those scholars would be eager participants in a collaborative effort with the research library community to build solutions for data management that could have broad application. We would argue that such efforts would be decidedly strategic in advancing digital scholarship in Canada.

The role of Compute Canada – Compute Canada is an active member of Research Data Canada. We welcome the recognition of the important role Compute Canada plays in Canadian infrastructure for data. We also welcome Compute Canada's commitment to build capacity to provide the physical storage for data and its commitment to collaborate with those with the curatorial skills to manage data in those storage environments.

The competition – We are somewhat concerned that the proposed procedure for the competition does not adequately reinforce collaboration with Compute Canada and the benefits of inter-institutional, inter-disciplinary, inter-sectoral, as well as international collaboration. We do not believe that appropriate infrastructure is likely to be built within single institutions. As we have already made clear, anything we build in Canada must be compatible with international efforts. There is no way better to ensure international compatibility than to collaborate in the building of the data infrastructures.

In conclusion – CFI has been active in the work of Research Data Canada and its predecessor organization. RDC, as the gathered community of all with a responsibility for data, is committed to work with CFI and the other federal funders as the physical resources, skills, and policies are built to ensure effective research data management in Canada.