

Lessons in Establishing and Operating a Research Data Centre Network: The Canadian experience

Chuck Humphrey, University of Alberta

Beginning in 1996, a fixed-cost agreement between Statistics Canada and the post-secondary sector in Canada lowered the access barrier to anonymised microdata produced by Canada's national statistical agency for many of its household surveys and population censuses, dating back to 1971. This programme, known as the Data Liberation Initiative (DLI), opened access to standard data products, i.e., anonymised microdata, aggregate databases and spatial data but excluded data from surveys for which anonymised microdata were not produced. In particular, several valuable longitudinal surveys started in the 1990's were not available to researchers. A joint-study between Statistics Canada and the Social Sciences and Humanities Research Council in 1998 recommended establishing a handful of data enclaves with strict access rules to provide access to the confidential data from these longitudinal surveys. Implementing this proposal resulted in the Canadian Research Data Centre Network, which started in 2000. This paper discusses key lessons learned over the past decade in funding and operating a data enclave network. Unlike the DLI, which has low costs for data access, a Research Data Centre (RDC) requires a substantial annual investment by participating institutions. To confront these inhibitive costs, a national financing package was established to lessen the costs to universities. Several lessons about financing RDCs are discussed from the Canadian experience. Initially, Statistics Canada required each RDC to operate within a closed local area network, prohibiting the establishment of a national infrastructure connecting these services. More recently, funding has been secured to link all RDCs into a network through secure lightpath technology running over Canada's optical high-speed research network. Lessons about the operation of individual RDCs and the emergence of a true network are also addressed. Finally, lessons about the management of the data within the RDC network are reviewed, including the need for newer metadata technology.