Marcel Fortin is the Head of the Map and Data Library at the University of Toronto (UofT) and has been the GIS and Map Librarian since 1999. He supports geospatial teaching and research on campus through data collection development, consultation services and teaching. Marcel is a sessional instructor in the UofT’s Department of Geography and Planning where he teaches a course called “The Power of Map and Geographic Information.” He is the co-editor of the book “Historical GIS Research in Canada,” published in 2014 by the University of Calgary Press. In 2015, Marcel was awarded a three year federally funded Social Sciences and Humanities Research Council (SSHRC) Partnership Development Grant for a project called the Canadian Historical GIS Partnership (http://geohist.ca). The HGIS Partnership team included Historians, Geographers and Librarians from both Canada and the United States. Marcel is also a co-investigator on a CANARIE funded Research Data Management (RDM) project. CANARIE is a non-profit corporation that supports the digital research infrastructure for Canada’s research, education and innovation communities.

Perspective

As a librarian who supports research, and as a researcher myself, I have always been concerned with the preservation of original data, and I have long-struggled to find appropriate technologies and workflows that allow for the description, safe preservation, and the discoverability of original data. Especially worrisome has been the lack of appreciation by most Research Data Management Programs in academic institutions for the complex and unique nature of the curation and discovery needs of original geospatial data.

In the last few years, most academic institutions across North America have created Research Data Repositories. The most popular tool used to build these repositories is the Dataverse platform. Dataverse is the defacto platform used across Canada for RDM repositories. While Dataverse does allow for the ingestion of many types of data, including geospatial data, the challenges of exposing and querying geospatial data or visualizing geographically within this platform are well-known. Out of the box, Dataverse is currently only searchable via text-based queries and results are returned textually as well, which seriously limits the discovery of most geospatial datasets.

The Canadian Historical GIS Partnership project’s main aim in 2015 was largely to investigate the needs of HGIS researchers across Canada and in the United States. We found that issues surrounding the storage, sharing, preservation and re-usability of geospatial data were also some of the main concerns of faculty and librarians involved in HGIS research projects.
In the fall of 2018, Geohist.ca teamed up with a CANARIE-funded research project team based out of the University of British Columbia to investigate the use of geospatial technology (Geoblacklight and Geoserver), as an add-on to Dataverse in order to allow for geographic and visual querying of Dataverse. The completed product will consist of a federated map-based user interface for data querying. The tool is meant to help discover geospatial data within repositories. But the tool is also envisioned as enabling the exposure of other Research Data with geographic components attached. Examples of these types of components include geographic units captured in Survey Data such as City, Province/State, Postal/Zip codes, etc.

The ultimate goal of this project is to develop a tool to work in the same fashion with Canada’s Federated Research Data Repository (FRDR, http://frdr.ca), which supports and is the one-stop RDM search and discovery tool for approximately 30 data repositories in Canada. FRDR harvests metadata using various protocols, including the Open Archives Initiative (OAI) protocol amongst other feeds to populate its service.

It is hoped that this CANARIE project will not only fulfil the need for querying spatially RDM catalogues of data, but that it will also help raise awareness amongst RDM communities of the necessity of special care and treatment of original geospatial data for long-term preservation and access.