# Toward Support for Epidemic Preparedness via Digital Twin Data + “Think Big” Proposal

Michael Wolfson, uOttawa, Banff Centre, November 16, 2023

Abstract / Summary

An essential component for pandemic preparedness is adequate data supporting ongoing analytical capacity. Since support for specialized pandemic-oriented data collection (+ analytical capacity) tends to wane between acute events, data for pandemic preparedness should, as much as possible, be designed to be useful during quiescent periods. One such kind of data is a realistic but synthetic “digital twin” that closely resembles detailed census data but is non-identifiable, hence non-confidential. Such digital twin data would provide individual-level details of Canada’s population by small area geography and a range of socio-economic and infectious disease-relevant characteristics, substantially reflecting real-world heterogeneities. In turn, these data could provide a richly textured basis for more sophisticated infectious disease modeling, especially with regard to contact patterns more readily incorporated into agent-based modeling as compared to the more usual compartment models.

Such a digital twin database could be constructed by starting first with published census cross-tabs at the census tract level using simulated annealing, and then synthetically matching (with replacement) individual and household level records from census public use files, thereby substantially preserving important kinds of correlations. In addition, data from other key microdata files like the Canadian Community Health Survey (CCHS), the Labour Force Survey (LFS), and the time use results from the General Social Survey could also be synthetically matched. As these data sources are either already in the public domain, or versions could be so constructed, the resulting digital twin would also be non-confidential, hence completely open data.

Constructing this digital twin database and regularly updating it would incur considerable costs. However, it would also have a much broader range of uses than only to support epidemic modeling. With a broad range of users, it would be more easily sustained over time as a new and important addition to Canada’s statistical system.

In order to keep the digital twin data current, “nowcasting” using a public domain version of Statistics Canada’s DEMOSIM model could be used, along with other regular monthly data sets including the LFS and CCHS.

With the advent of an epidemic and the imposition of public health measures such as lock downs, behaviors would change. These could be tracked via appropriately anonymized yet still geographically detailed real-time cell phone mobility data.

In sum, a well-conceived digital twin database could provide both a substantially improved real-time basis for infectious disease modeling and a substantial addition to Canada’s statistical system that would have a broad range of other uses, thereby assuring its longer terms sustainability.

Based on some of the preceding discussion in the conference, this presentation also offered the suggestion of “thinking big” – developing a brief for senior officials and governments outlining the idea of a digital twin along with other key improvements in linkable and coherent data flows such as infections, hospitalizations, vaccinations, and genotyping of infections. While this kind of initiative is not something fundable using the current Canadian and provincial granting council structures, there is a window of opportunity with recent reports on data from the Public Health Agency of Canada and Health Infoway, and the 2023 federal budget proposing over half a billion dollars for health data. There is also the recent Report of the Advisory Panel on the Federal Research Support System which could well open up the right kind of funding opportunities.