ASSOCIATION OF CANADIAN MAP LIBRARIES AND ARCHIVES
BULLETIN

GIS Trends

Barbara Znamirowski

DLI National Training: Accessing data and building a community

Introduction

"The Data Liberation Initiative (DLI) is a partnership between post-secondary institutions and Statistics Canada for improving access to Canadian data resources. Over the years the focus of the DLI Program has evolved from purchasing access to major Canadian datasets collected by Statistics Canada to providing training services and the continuous support required for the proper understanding and usage of an ever expanding research data collection."[i]

The above quote is taken from the Data Liberation Initiative's (DLI) web site. It's an accurate description of the DLI. However, as with many initiatives and events, to truly understand it, it helps to be there – to have experienced the program and its benefits. Many of our institutions have been members of the DLI since it began in 1996. The DLI has provided a means of accessing products otherwise unaffordable to most researchers and academic institutions – including a range of aggregate data, census geography boundary files, public use microdata files (for example, Census of Population and other surveys such as general social surveys and household income and expenditures surveys) and postal code conversion files.[ii] It has also encouraged formation of a community of experts: the DLI membership of 81 post-secondary institutions, working with a dedicated team of Statistics Canada employees.[iii] The DLI community provides mentorship and support through an active listserv, an External Advisory Committee (EAC) and Professional Development Committee, and regular training initiatives for DLI institutional contacts. I've had the privilege of watching the DLI community evolve since its inception.[iv]

Another successful national training event organized by the DLI Professional Development Committee took place from November 23 to 27, 2020.[v] Training initiatives take place annually, sometimes on a regional level and other times nationally. In response to the COVID pandemic the event was moved for the first time to an all-virtual platform.

[i] Taken from: Statistics Canada web site (https://www.statcan.gc.ca/eng/dli/dli) on 27 November 2020.
[ii] Postal Code Boundary files are obtained through the DLI through an agreement between Statistics Canada and Canada Post. Each DLI institution signs an annual End-Use Agreement with Statistics Canada for these files.
[iii] Number of members is in Data Liberation Initiative Annual Report (Spring 2020).
[iv] Further reading on the history and operation of the DLI: Boyko, Ernie and Wendy Watkins. 2011. "The Canadian Data Liberation Initiative. An Idea Worth Considering?" International Household Survey Network, IHSN Working Paper No 006. Available from: https://ihsn.org/sites/default/files/resources/IHSN-WP006.pdf
[v] Statistics Canada. The membership list is found here: https://dli-training.github.io/en/contact/

In this column I describe some of the themes covered during 2020 national training sessions in the context of trends in data and statistics and associated technologies. I've organized this discussion into six main themes. These are presented in no particular order and often cover more than one session. This summary is by no means a comprehensive account of all presentations. For further information, readers are encouraged to review the links and other follow-up materials at the end of this column. Opinions are my own.

Expansion of the Continuum of Access for Microdata

Almost two decades ago I attended a DLI training session at which Chuck Humphrey presented a graph showing the "Continuum of Access" provided by the dissemination channels used by Statistics Canada. Chuck described three characteristics of this continuum: cost (from free to expensive), restrictions or conditions (from open or no restrictions to very restricted) and type of Information (from statistics to data).[vi] Since then, Statistic Canada's initiatives have often been described in terms of this continuum – most recently to illustrate new models of dissemination of microdata.[vii] Current Statistics Canada data dissemination initiatives (some in pilot mode), and their places on this continuum of access are represented in Figure 1, adapted from the Statistics Canada website.[viii]
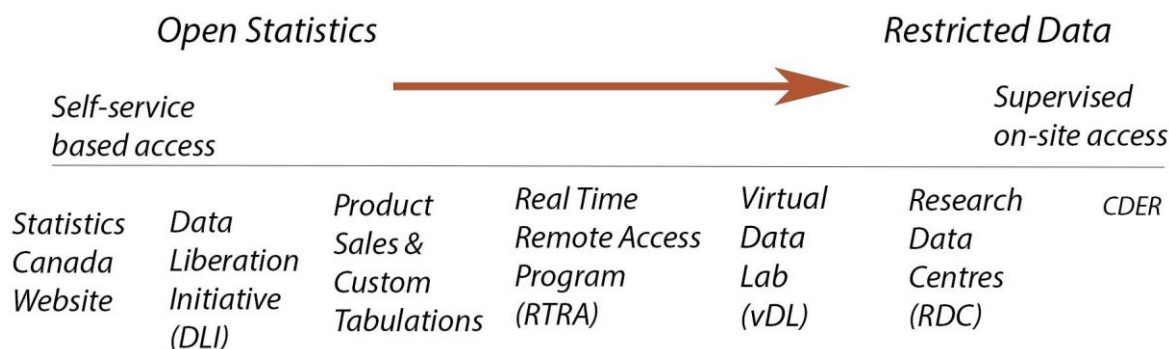

Figure 1. Continuum of Access

This year's DLI Training hosted a panel entitled "From Open to Restricted – Statistics Canada's Continuum of Data Access," with six speakers presenting on the following topics: Public Use Microdata Files at Concordia (Alex Guidon); Advantages and Disadvantages of Public use Microdata (PUMF) (Charles Fleury); ODESI: Reflecting on 10 years of Collaboration in Library Data Services (Amber Leahy); Real Time Remote Access (Paul McDonald); Virtual Data Lab (Sara Tumpane) and Research Data Centres & Centre for Data Development and Economic Research (Grant Gibson)

[vi] Humphrey, Chuck. DLI Orientation: Concepts. A Framework for Thinking about Statistical Information, April 2004.
[vii] For illustration and further information on the Continuum of Access for Microdata Access see: https://www.statcan.gc.ca/eng/help/microdata
[viii] See Statistics Canada, https://www.statcan.gc.ca/eng/help/microdata (accessed 27 November 2020)

This session explored how PUMFs (Public Use Microdata Files) are used within universities, as well as new models of dissemination for other forms of microdata. In considering these models it is important to understand that the level of sensitivity of microdata can vary significantly depending on how a file is processed to minimize participant identification. PUMFs available through the DLI are very different from microdata master files available in RDCs. We look forward to learning more about Real Time Remote Access (RTRA) and Virtual Data Lab (vDL) options in context of their accessibility to our researchers, and to defining our roles in support of these microdata access initiatives.[ix]

Adapting to the pandemic through remote instruction – a panel

In this engaging panel speakers Kelly Schultz, Andrew Nicolson, Marcel Fortin, and Leanne Trimble (moderator), reviewed their experiences with delivering remote workshops to their university communities during the pandemic. There was also considerable audience engagement through the chat function, demonstrating the relevance of this topic and need for our community to share experiences and methodologies.[x] There was general consensus that diverse methods are required, including both asynchronous and synchronous workshops, and sometimes a hybrid of both. The University of Toronto Map and Data Library has also tried (the wonderfully named) "Snacking on Bits and Bytes" as a substitute for some of their open workshops, offering abbreviated sessions meant to introduce a topic (for example, on "R" or Programming with Python in ArcGIS Pro) through approximately 40 minute online live demos, with time for questions.[xi]

There was less consensus on experiences relating to the volume of requests for teaching and support: the chat revealed that some institutions saw a decline, some saw no change, some have experienced more or different requests since the start of the pandemic. Speakers and chat participants noted that virtual workshops require significant preparation time. Some felt synchronous involved less time to prepare than asynchronous, but also noted that synchronous could be more stressful given the uncertain stability of remote connections and / or the ability to provide support or feedback to students during workshops. Access to and knowledge of technologies to support creation of virtual training material was a common thread in panelist presentations and the chat.

[ix] In addition to reviewing the slides from the presentations (all posted at: https://cudo.carleton.ca/dli-training/4360) readers may wish to review the following references for more information: Presentation by the Chief Statistician of Canada to the Canadian Research Data Centre Network (CRDCN) (October 2018) available at: https://crdcn.org/sites/default/files/arora_-_presentation_by_the_chief_statistician_of_canada.pdf. See also: Kelly Cranswick, Virtual data labs - A more flexible approach to access Statistics Canada microdata, presented at UNECE Conference of European Statisticians, Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality Session 1.1 16 September 2019, available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2019/mtg1/SDC2019_S1.1_Canada_Cranswick_AD.pdf

[x] Chat from training sessions is being transcribed and will be available at: https://cudo.carleton.ca/dli-training/4360

[xi] See: https://mdl.library.utoronto.ca/bits-and-bytes

Kelly Schultz of the University of Toronto Map and Data Library shared an excellent slide on some of the tools she found useful to prepare virtual training (Figure 2).[xii] Chat contributors supplemented some of their own experiences including use of Mentimeter (for keeping students engaged and actively responsive during workshops), pros and cons of students sharing screens during workshops and diverse approaches to doing this, built in metrics for evaluating workshop participants engagement (for example in Quercus), and the introduction of Virtual Computing Commons / Labs on campus to ensure broader availability of software.

Figure 2: Tools for Virtual Training

Kelly also gave a fantastic asynchronous workshop, open to all DLI training participants, on Data Visualization: Theory and Critique (Part 1) and Practice with Tableau (Part 2). I think we all aspire to produce workshops as good as this! Among other things, it has pushed me to explore Tableau mapping capabilities – but that's for another time.

Data Management Planning

James Doiron from the University of Alberta delivered a detailed session on the creation of Data Management Plans (DMP). He reminded participants of the importance of DMPs in the context of the Tri-Agency's draft recommendations.[xiii] He discussed the DMP in terms of key research lifecycle categories, taking participants through the creation of a DMP using Portage's DMP Assistant.[xiv] Portage's DMP Exemplars were noted as good reference points for researchers writing their own DMPs. He also reminded participants of the importance of choosing depositories (such as Dataverse) that assign permanent identifiers (DOI) to datasets.

[xii] Figure presented by Kelly Schultz in her talk given as part of panel "Adapting to the pandemic through remote instruction – a panel" (panel members: Marcel Fortin, Andrew Nicholson, Kelly Schultz, Leanne Trimble) at DLI National Training, 26 November 2020. Figure reproduced with permission.
[xiii] See: http://www.science.gc.ca/eic/site/063.nsf/eng/h_547652FB.html
[xiv] For further information about Portage's DMP Assistant: https://assistant.portagenetwork.ca/

Product Reviews

An important part of DLI Training is the opportunity to engage with data producers directly, hearing from subject divisions within Statistics Canada about survey and other product content, dissemination formats, technologies and deliverables. These sessions are invaluable and serve not only to inform the DLI community, but often function as an opportunity for dialogue where the DLI Community can ask questions and provide feedback to subject divisions on patterns of use and requirements of faculty and students.

This year's sessions included:

- Overview of the content of the 2021 Census, presented by Sarah Franklin and Andrea Levett from the Census Subject Matter Secretariat
- Statistics Canada COVID-19 Data Projects, presented by Kathleen Fowler and Melanie Kowalski from the Centre for Social Data Integration and Development
- The Gender, Diversity and Inclusion Hub, presented by Sylvie Guilmette from the Centre for Gender, Diversity, and Inclusion Statistics
- Key Justice and Victimization Data Sets presented by Kathy AuCoin from the Canadian Centre for Justice and Community Safety Statistics
- Update on DAD/DLI EAC/PDC presented by Gilbert Bede, Alex Cooper, Arden Kayzak

Dataverse, DDI and ODESI

As part of the pre-recorded Lightning Talks, Chantal Ripp from the University of Ottawa presented "Projet visant à faire découvrir les collections de données secondaires dans Scholars Portal Dataverse". She reported on a project designed to evaluate the feasibility of using the University of Ottawa's Scholars Portal Dataverse for discovery of secondary data collections. These would include one-time purchases, including collections that are locally managed and not available on other platforms. The study looked at existing practices used by institutions using Scholars Portal Dataverse, including their pros and cons, such as challenges for geospatial dataset metadata descriptions.

Jane Fry from Carleton University gave a lightning talk on DDI, a standard for documenting data files. In introducing DDI she showed some nice visuals on how she introduces the importance of metadata to students. She also reviewed the elements of the DDI lifecycle and DDI Codebook, providing links for future reference and training. Nesstar or Colectica are the two main tools used to interpret DDI (xml files).

Amber Leahy from Scholar's Portal presented a talk on "ODESI: Reflecting on 10 years of collaboration in library data services" as part of the panel "From Open to Restricted - Statistic Canada's Continuum of Data Access". She reviewed the history of ODESI, including aspects of community development and collaboration such as the OCUL Markit! Program. Slides presenting the evolution of the graphic interface reminded us that we have come a long way in content and design. Future challenges include the evolving research data landscape and data discovery needs, as well as the future of Nesstar.

Postal Codes Conversion files and Geocoding

Pre-recorded lightning talks included a session by Daniel Brendle-Moczuk from the University of Victoria on geocoding options for matching postal codes or addresses to locations (for example to census boundaries). He highlighted key factors to consider when evaluating options, such as whether you have addresses or just postal codes, the number of locations, and cost. Options covered include free and priced solutions such as: the PCCF, PCCF+, BC Address Geocoder (for BC addresses), QGIS with plug-in MMQGIS, US Census Bureau (for US addresses), Esri, DMTI as well as some options for geocoding with browser based online options.

Alex Cooper from Queen's University gave a lightning talk introducing Postal Code Conversion Files: what it is, the differences between the PCCF and PCCF+, licensing and access. She also generously shared Queen's guide to the PCCF, which I recommend to anyone needing assistance with processing the PCCF or PCCF+.[xv]

Conclusion

This year's DLI training had 155 registrants, with an average of 54 to 89 participants at each session. These numbers illustrate the interest in and need for such events. Although many indicated that they looked forward to a return to in-person meetings, initial feedback from participants demonstrated that the event was a resounding success. All participants were given an opportunity to complete surveys on individual sessions which will no doubt inform future training activities. Thanks to everyone involved for such an excellent event!

Further Information

DLI 2020 Training:

Sessions slides and Lightning Talk Videos are available in English and French in the CUDO repository (https://cudo.carleton.ca/dli-training/4360). Responses to questions and chat transcripts will also be provided once ready in CUDO.

DLI Citation Guide: https://www.statcan.gc.ca/eng/dli/guide/section7

DLI EAC: https://www.statcan.gc.ca/eng/dli/eac

DLI Listserv: DLILIST@IDD-DLI.STATCAN.GC.CA

DLI Survival Guide: https://www.statcan.gc.ca/eng/dli/guide/index

DLI Training Repository: https://cudo.carleton.ca/collection/dli

Statistics Canada Real Time Remote Access: https://www.statcan.gc.ca/eng/rtra/rtra

---

[xv] See https://guides.library.queensu.ca/pccf

GIS Trends: Note from the Editor

Submissions and Feedback

GIS Trends is a place to share ideas, observations and discoveries in the area of data visualization, GIS and other spatial technologies. If you have something you would like to share please write to me. We also welcome feedback on GIS Trends articles. Proposals for articles and feedback should be sent to: bznamirowski@trentu.ca Thanks for reading and contributing! Barbara Znamirowski, Editor, GIS Trends