WiFi: Mac-Events | cdcf2019
Program: data-curation.github.io
#CDCF2019

Canadian Data Curation Forum

Training Event - Day 1

October 16-17, 2019
Hamilton, ON

McMaster University | Library

SSHRC ≡ CRSH
Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA
portage

McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the "Dish With One Spoon" wampum agreement

(Indigenous Education Council, May 2016).

Date, Time	Room	Session	Facilitator(s)
16-Oct 0915 - 1200	OJN 111	Introduction to R for Social Scientists	John Simpson
	OJN 212	Making Ethical and Effective Data Management Work for You!	Chandra Kavanagh & James Doiron
1200 - 1300	OJN 212	**LUNCH (provided)**	
1300 - 1545	OJN 111	Data Wrangling and Versioning with OpenRefine and Github	Lee Wilson & Jay Brodeur
	OJN 212	Curating Data for Reproducibility	Sandra Sawchuk & Shahira Khair
1545 - 1600	OJN 212	**Day 1 Wrap-up**	
17-Oct 0900 - 1145	OJN 111	The DCN CURATE Model	Lisa Johnston & Cynthia Hudson Vitale
	OJN 212	Curating Data in Repositories	Lee Wilson, Reyna Jenkyns, Meghan Goodchild, Brad Covey, & Kaitlin Newson
1145 - 1200	OJN 212	**Wrap-up**	

Full program: data-curation.github.io

CDCF Program Committee

- Jay Brodeur
- Lee Wilson
- Shahira Khair
- James Doiron
- Jane Fry
- Jeff Moon
- Mark Leggott
- Sarah Wilkinson

- Meghan Goodchild
- Sandra Sawchuk
- Brad Covey
- Reyna Jenkyns
- Chandra Kavanagh
- Vivek Jadon
- Qian Zhang
- Curation Expert Group

Help us by providing feedback (if you haven't already)

bit.ly/CDCF-training-feedback

WiFi: Mac-Events | cdcf2019
Program: data-curation.github.io
#CDCF2019

Canadian Data Curation Forum

Training Event - Day 2

October 16-17, 2019
Hamilton, ON

Date, Time	Room	Session	Facilitator(s)
16-Oct 0915 - 1200	OJN 111	Introduction to R for Social Scientists	John Simpson
	OJN 212	Making Ethical and Effective Data Management Work for You!	Chandra Kavanagh & James Doiron
1200 - 1300	OJN 212	**LUNCH (provided)**	
1300 - 1545	OJN 111	Data Wrangling and Versioning with OpenRefine and Github	Lee Wilson & Jay Brodeur
	OJN 212	Curating Data for Reproducibility	Sandra Sawchuk & Shahira Khair
1545 - 1600	OJN 212	**Day 1 Wrap-up**	
17-Oct 0900 - 1145	OJN 111	The DCN CURATE Model	Lisa Johnston & Cynthia Hudson Vitale
	OJN 212	Curating Data in Repositories	Lee Wilson, Reyna Jenkyns, Meghan Goodchild, Brad Covey, & Kaitlin Newson
1145 - 1200	OJN 212	**Wrap-up**	

Full program: data-curation.github.io

Help us by providing feedback
(if you haven't already)

bit.ly/CDCF-training-feedback

WiFi: Mac-Events | cdcf2019
Program: data-curation.github.io
#CDCF2019

Canadian Data Curation Forum

Community-Building Forum - Day 1

October 17-18, 2019
Hamilton, ON

McMaster University | Library

SSHRC ═ CRSH
Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

portage

McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the "Dish With One Spoon" wampum agreement

(Indigenous Education Council, May 2016).

CDCF Program Committee = Facilitators

- Jay Brodeur
- Lee Wilson
- Shahira Khair
- James Doiron
- Jane Fry
- Jeff Moon
- Mark Leggott
- Sarah Wilkinson

- Meghan Goodchild
- Sandra Sawchuk
- Brad Covey
- Reyna Jenkyns
- Chandra Kavanagh
- Vivek Jadon
- Qian Zhang
- Curation Expert Group

Outline

- What is data curation?

- Goal & objectives

- Outcomes & deliverables

- Components of a data curation network

- Data curation network models

What is data curation?

Data curation is the active management of research data as it is created, maintained, used, archived, shared, and reused.

https://dictionary.casrai.org/Data_curation

Some examples of data curation activities:

- providing QA through metadata inspection, file audit, and code review;
- creating documentation and metadata to contextualize data;
- evaluating datasets for fairness;
- implementing secure storage and version control for files and code;
- preparing datasets for deposit in a data repository or archive;
- transforming file formats to enhance reuse and long-term preservation;
- augmenting metadata and linking datasets to increase discoverability.

datacurationnetwork.org/data-curation-activities

Goal & Objectives

Develop a vision and roadmap for a national approach to data curation in Canada

Objectives

- Develop a shared understanding of the value proposition of data curation and its place in data management and stewardship.

- Define the current state of data curation capacity in Canada.

- Identify the current challenges of research data curation; determine the resources required to address these challenges and support best practices in data curation.

- Identify and describe the ways in which national coordination or support could address data curation needs.

- Identify the ways in which the Portage Curation Coordinator could support these needs.

Outcomes & Deliverables

Outcomes of the Community-Building Forum activities will be summarized in an event report, which will be drafted and reviewed by the community in the months following the event. The report will include:

- A summary of the event's discussions and significant outcomes.

- A clear statement of the value of data curation to Canadian education, scholarship, and economic development.

- A vision for a 'desired state' of curation activities and support in Canada.

Outcomes & Deliverables

Event report:

- Recommendations for strengthening a national community of practice for data curation--whether through the use of existing communities or the creation of a new one.

- A 'lightweight roadmap', which outlines short- and long-term objectives for addressing curation needs with:

    - National-level services and support (including Portage Curation Coordinator and curator positions, and within the new DRI organization, generally);

    - Training and resource development; and

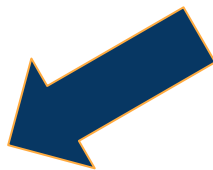   - Data sharing and preservation activities.

Forum workflow

How could / should data curation be supported at a national level?

⬇

What are the existing
needs and challenges?

⬇

What resources & services
could meet these requirements?

➡

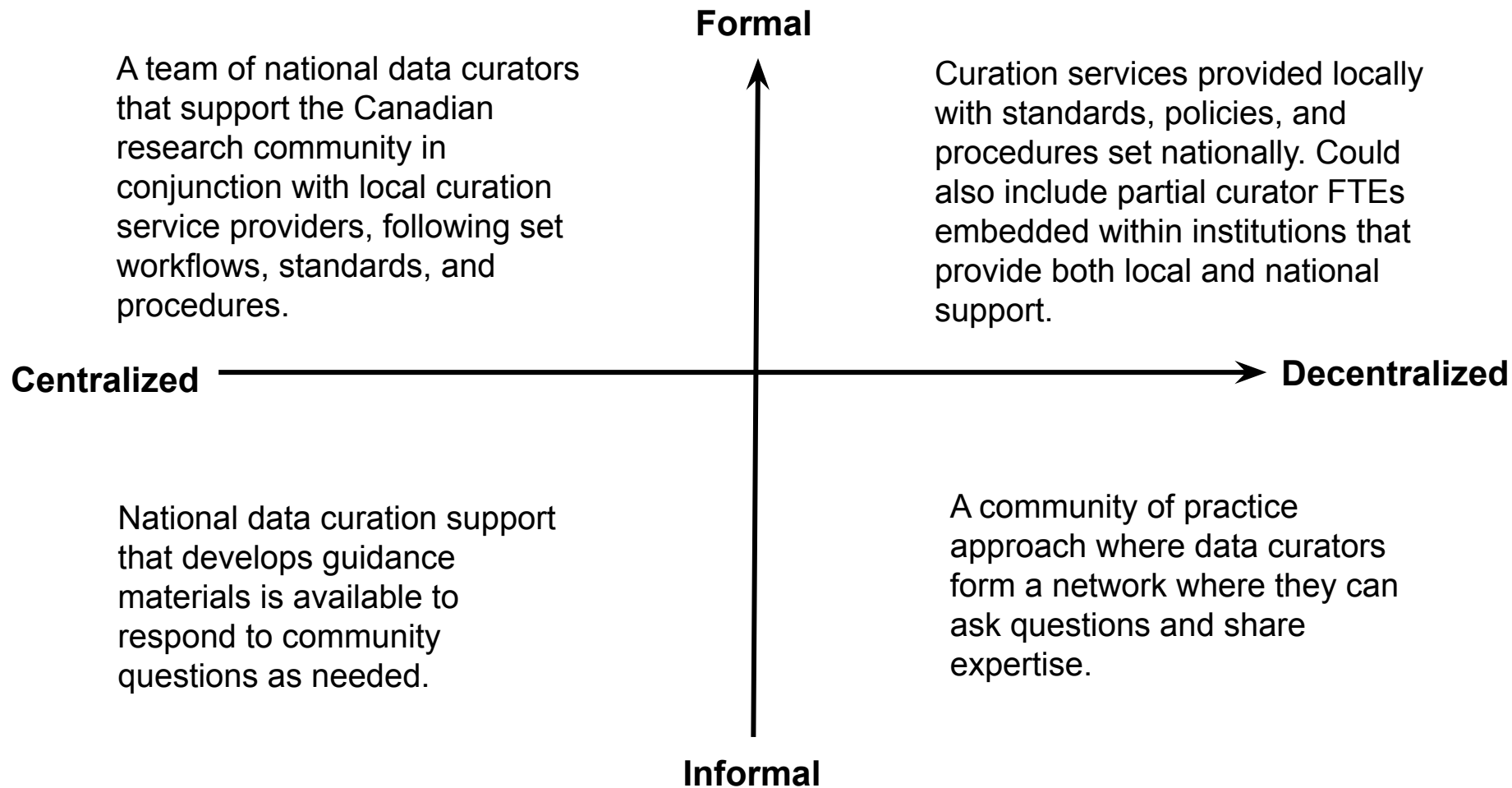What are the strengths,
weaknesses, and constraints?

⬆

How might these be provided and
coordinated (i.e. as a model)?

⬅

Possible Components of a Data Curation Network

- **Formal communication channels for data curators**, e.g., Slack or Google Group, where curators could ask questions and share experiences and expertise.

- **A knowledge base of guidance and training materials** and email templates of typical curator recommendations, including best practice handouts and guidelines.

- **A central support role for data curation** across Canada that would support the needs and queries of local data curators working in institutional, domain-specific, and other kinds of repositories.

- **Increased support locally** for curating research data at your institution.

- **A shared curation model** to promote standardization and quality of data publications.

- **Coordinated training opportunities** to develop curation skills and capacity.

Formal

A team of national data curators that support the Canadian research community in conjunction with local curation service providers, following set workflows, standards, and procedures.

Curation services provided locally with standards, policies, and procedures set nationally. Could also include partial curator FTEs embedded within institutions that provide both local and national support.

Centralized ——————————→ **Decentralized**

National data curation support that develops guidance materials is available to respond to community questions as needed.

A community of practice approach where data curators form a network where they can ask questions and share expertise.

Informal

Day 1 Schedule

Time	Session
1300 - 1320	Opening remarks
1320 - 1400	Opening Plenary: The Canadian DRI Landscape
1400 - 1500	Panel & Audience Discussion: DRI, RDM, & The Role of Data Curation
1500 - 1530	**Break**
1530 - 1630	The Data Curation Network
1630 - 1700	Wrap-up discussion

Full program: data-curation.github.io

Let's begin!

… with a quiz

What are you hoping to take away from this event?

- Insight for response to triagency data policy
- A better understanding of how domain specific repositories should fit into the canadian data librarian world. Where are the points of collaboration, how do we stay in touch, etc.
- Ways to address data curation needs in Canada as a consortium
- The scope of data services and needs across the country.
- connections
- what services and practices I can implement at my organization
- general guidelines about data curation
- Community connections
- gauge the state of curation activities in institutions across the country, see how we can cooperate
- National RDM team building.
- Ethical problems related to data sharing
- international collaboration to support data curation
- To get a sense of how recently fits into the rid landscape
- Learn about gaps and challenges in the landscape
- Scaleable workflows that work for institutions of various sizes
- Better understanding of national picture

- Better understanding of national picture
- A template to take home
- How to get my institution on board with resourcing data curation and participating in network.
- Clear definition of data curation vs data management and how both work together
- Start brainstorming what we can do on a national level around curation
- Understanding the issues that individual institutions are facing around RDM and data curation
- Collaborate to make our services better
- A vision for how we are going to support curation within research institutions, especially in light of the new DRI org
- Learn about how small and mid-sized institutions can contribute and benefit from a national curation network
- Direction from the community on the future of data curation support in Canada.
- collaboration between data/GIS/RDM/liaison librarian communities
- Vision of Canadian Data Curation Landscape
- Learn about what others are doing across the country

What are you hoping to take away from this event?

- Informed idea of what's coming nationally
- Resources
- networking
- Insight into others' thoughts on collaboration
- Understand the broader landscape and the role my centre might play
- share First Nations perspectives and learn
- "Consensus on a national curation strategy."
- An understanding of where my organization fits within the larger landscape of data curation and RDM
- To create a pathway for curators to discuss best practices and to advance our services to users
- Services and resources for data curation
- Have more background and contextual information to help design our local data curation workflow.
- An update on academic data curation activities in Canada, which I will apply to a plan to set up data management/curation/sharing among public libraries in Ontario, nationally, etc.

- First steps on what to ask my institution to commit to and ways to get this written into my job description
- How do we best leverage our collective resources to support data curation (and can we do this work with some degree of coherence)?
- "Identify major gaps to be addressed.
- Better understand the role of the institution beyond the library.
- Better understand the role of disciplinary communities."
- Who pays for this?
- Connecting, Community, and discussing all things Data Curation!

Wrap-up discussion (30-minute lightning round!)

Step 1: Find your groups (hint: it's on the back of your nametag)

Step 2: Find a space for your discussion:
- Groups 1, 2, 3, 4 in room 212
- Groups 5 & 6 in room 111

Step 3: In groups, respond to the questions presented in our shared Google Doc:
bit.ly/CDCF-day-1-discussion

1. What are the main barriers/challenges to data curation activities within your organization and in your work?

2. What are the key elements of a national approach that addresses these barriers/challenges?

WiFi: Mac-Events | cdcf2019
Program: data-curation.github.io
#CDCF2019

Canadian Data Curation Forum
Community-Building Forum - Day 2

October 17-18, 2019
Hamilton, ON

Day 2 Schedule

Full program: data-curation.github.io

Time	Session
0845 - 0900	Welcome & Day 2 Organization
0900 - 1015	National Repository Updates
1015 - 1030	**Break**
1030 - 1200	Breakout Session 1: Conceptualizing a National Approach to Data Curation
1200 - 1300	**Lunch**
1300 - 1430	Breakout Session 2: Model Evaluation & Discussion
1430 - 1445	**Break**
1445 - 1530	Final Discussions and Closing Remarks

Day 1 Summary

What are the main barriers/challenges to data curation activities within your organization and in your work?

Common themes:

- Capacity
  - Human resource
    - Staffing
    - Expertise
    - Time
  - Service scaling
  - Financial
- Standards & workflows
  - Roles and responsibilities
  - Curation models/expectations

- Coordination
  - Institutional & national
- Culture
  - Incentives for researchers
  - Lack of knowledge about RDM & best practices
- Access to tools & software
- Training & Guidance
  - Training opportunities
  - Toolkits

What are the key elements of a national approach that addresses these barriers/challenges?

Develop Capacity
- Education and Training
- Sustainable funding
- Toolkits, FAQs

Develop Culture
- Engaging early-career researchers
- Demonstrating impact
- General advocacy
- Embedding curation in research

Develop Standards
- Standards & Principles
- Common vocabularies

Expand & sustain a network
- Supporting multilingual
- Join the DCN?

Develop Services
- Institutionally & nationally

Governance
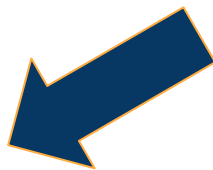- Policies & guidelines

Forum workflow

How could / should data curation be supported at a national level?

What are the existing needs and challenges?

What are the strengths, weaknesses, and constraints?

What resources & services could meet these requirements?

How might these be provided and coordinated (i.e. as a model)?

Breakout discussion 1 - Questions

GOAL: To produce potential models that can be evaluated during the next breakout discussion.

1. What are some approaches/models of a Canadian 'data curation network' that meets the identified needs and requirements?

2. Consider the challenges of, or barriers to, the identified models?

3. What are the linkages or relationships between your curation model and other RDM services and systems? E.g., how does the plurality of Canadian data repositories fit into a curation network?

4. Provide a short summary (3-4 sentences) and a sketch of your preferred model, including a timeline for rollout (diagrams encouraged!).

Breakout discussion 1

Step 1: Find your groups (again), and a space for discussion:
- Groups 1, 2, 3, 4 in room 212
- Groups 5 & 6 in room 111

Step 2 [55 mins]: In groups, respond to the questions presented in our shared Google Doc: bit.ly/CDCF-breakout1

Step 3 [11:45]: Regroup and discussion

Reviewing our models

Breakout discussion 2 - Questions

Considering the preceding discussion, for your model, address:

1. What are the strengths and weaknesses of your model?

2. What resources and support would be required to implement this model?

3. What parts should be coordinated at a national level? What should be addressed at a local or regional level?

4. Sketch out a proposed timeline for the roll out of your model. Consider:

   a. What needs to be implemented first? Last? (and everything in between)

   b. How long will to take to get to the 'idealized' state for your model?

Breakout discussion 2

Step 1: Find your groups (again), and a space for discussion:
- Groups 1, 2, 3, 4 in room 212
- Groups 5 & 6 in room 111

Step 2 [50 mins]: In groups, respond to the questions presented in our shared Google Doc: bit.ly/CDCF-breakout2

Step 3 [30 mins]: Regroup and discussion

Final wrap-up

Final Summary

Summarize the conversation within your group.

What changed? What was resolved? What wasn't?

What are the commonalities of our models?

- no more pillars
- Need for excellent communication
- Tiered approach
- decentralized
- Emphasis on local
- Collaboration.
- national coordination
- Funding needed
- network
- Community of practice
- working together
- incorporating standards and policies
- Decentralized
- Build on existing networks
- inclusive
- Various stakeholders in each model
- Time as an additional axis
- need to build capacity at local level
-

- Standards
- national coordination of resources and capacities, layers of services, local and domain expertise.
- Easy to morph existing org's and systems to new environment if/when needed.
- Networked, use expertise of Portage and successes of DLI.
- existing national organizations
- Sustainable (funding)
- Collaboration
- Aligned with existing library service models/networks - less so with realities of researcher economies?
- Goal: achieve coherence at scale
- Training and mentoring

What hasn't been considered in our models? What's missing?

- Money
- (more) funding!
- government input/participation
- Mechanisms for local capacity building
- Researchers
- Needs analysis
- Researchers
- Global/international geography of research teams (data flows are not national)
- Researchers
- Sustainability
- Contextualize data curation in the open science ecosystem
- Tools for data curation
- What do we stop doing so we can take this on?
- Non institutional entry points

- alignment with government data repositories
- impact plan
- Again, time. What we do at the start will (and should) morph over time.
- governance
- Tri agency policy
- Efficiency of access for back end users
- Links with researchers
- Individual capacity/risk of burnout
- Implementation plan (i.e., phases)
- non-researcher data providers (e.g., non-profits, govt, First Nations)
- A mechanism for capturing direct user feedback to the networks to ensure fit-for-purpose. A way of consulting with the community directly

- How much will be an off the side of the desk workload increase
- Sensitive data - how can we utilize a national curation network to curate this data if it cannot be shared outside of a given institution?
- The amount of datasets could grow at a faster rate than we can handle. The Curation we are discussing is a boutique service. Scalability?
- cost models given the different sizes of our institutions and we're thinking about this at a national level.
- support for NON funded research
- Build FN/I/M (Indigenous) within the national Canadian approach but also look to align with FN, Inuit, and Metis strategies, respectively
- level of curation

Final Thoughts

… with one more quiz

Final Thoughts

- Let's not lose the momentum here.
- united goals
- How do we continue this dialogue?
- I would write something silly but you can only submit once
- Thanks for your work!
- More national conversations need to happen, more often
- Bring the NGO's in.
- We are not all at the same curatorial place
- Building and demonstrating end user (public) value.
- Let's continue the collaboration! So much more to do!
- Great to convene such a rich array of talent and perspectives! Thanks.
- Thanks for organizing and bringing everyone together
- Clear, actionable goals are needed
- please explain your next steps.
- Sustainability is hard.
- Start the conversation around Data Duration!!
- great group of people doing good, important work - inspired by the energy and sustained level of engagement

- Excellent content, networking opportunities and people - looking forward to CDCF 2020!
- Optimal time to be doing this work with DRI on the horizon, DCN in the works, and supportive policies on the table. Much to be done, conducive landscape for making waves!
- only the beginning
- I learned a ton. Thanks everyone
- First Nations are developing their own national Data Governance Strategy. Canadian (national) initiatives need to plan to align with FN strategies, not assume the opposite - or subsume FN initiatives (on purpose or by accident).
- Great training and talks. Look forward to next steps.
- Of immediate use is list of people who can help curate.
- Don't forgot the govt research scientists and their data. Federal provincial / territorial, and municipal.

Next Steps

Draft event report for community review
- Draft recommendations + Lightweight 'roadmap(s)' for short and long term

Archived forum materials
- Presentations, slides, products
- Questions —> event FAQ

Join the Canadian Data Curation Network Google Group: bit.ly/join-can-dcn

Add 'Data Curation' to your ORCiD profile keywords

✏ **Keywords**

Biometeorology, Geographic Information
Systems, Data Curation

All sorts of thanks!

CDCF Program Committee

Workshop facilitators

Panelists and presenters

The Data Curation Network

McMaster CCE

McMaster University Library

CARL Portage

SSHRC-CRSH

You all!

Shahira Khair
James Doiron
Jane Fry
Jeff Moon
Mark Leggott
Sarah Wilkinson
Meghan Goodchild
Sandra Sawchuk
Brad Covey
Reyna Jenkyns

Chandra Kavanagh
Vivek Jadon
Qian Zhang
Curation Expert Group
Lee Wilson
Jay Brodeur
Jennifer Rutkowski
Chris McAllister
Kyle Fletcher
Lisa Johnston
Cynthia Hudson Vitale

How to get out of here

To Pearson Airport:

GO Bus 40

- To Pearson Terminal 1 every 30 minutes.
- Pickup at GO sign at King @ Caroline (Close to Staybridge)

To Union Station

- GO Bus 16 (King @ Caroline)
- GO Bus 18 (King @ Caroline) to Aldershot —> Lakeshore West GO Train

Buy tickets

- Online: https://tickets.gotransit.com/en-us/
- Station: 36 Hunter Street