



Revitalizing At-risk Data

The Canadian Public Data Rescue Initiative



Background

Data Management

Research Data Management is a growing field dedicated to developing ways to ensure that the knowledge content of data is preserved

Computer scientists work at the bit level to ensure that files remain readable, useable and reliably uncorrupted

Data librarians work on the human side to ensure that data is documented and organized, that implicit knowledge is recorded explicitly and that data reliably represents what it purports to represent

A Crisis in Politics and Data

Data loss is usually thought of as inadvertent - a server died, the files were corrupted, someone updated something and wrote over the earlier version

Currently the U.S is deliberately creating a public data crisis - defunding data collection projects and deleting inconvenient statistics. Casualties include

- Near elimination of the National Centre for Education Statistics
- Ending international programs funded by USAid including the renowned Demographic and Health Surveys
- Newly unreliable economic data
- Vanishing environmental and disaster data
- Suppression of studies on right-wing violence



Public data

- Government-collected survey and monitoring data is a major – and difficult to ignore or discredit – source of information on social inequity, changing conditions and policy impact
- Economic, social, health, and environmental data are used to support and evaluate policy and become irrelevant when the only policy goals are to implement ideology, concentrate power and silence dissent.

The Data Rescue Project

In February 2025 several academic and nonprofit associations got together to form the [Data Rescue Project](#). Founding members included the international association of data professionals [IASSIST](#), the [Data Curation Network](#), and [Research Data Access & Preservation \(RDAP\)](#)

Since then they have worked to download data, create documentation, and prepare the data for ingest into platforms such as [ICPSR's DataLumos](#)

They have developed trackers to reduce duplication of effort and systems for coordinating the efforts of volunteers. As it is not possible to create automated processes for all these different resources, this requires immense hands on effort, and they've hosted several data rescue events to bring in volunteers

Data loss in Canada

Canada has flirted with data destruction. Harper Conservative government cut:

- 2009 onwards - several education data programs
- 2010 - Canadian Foundation for Climate and Atmospheric Sciences; ongoing eliminations of climate and environment data continue over next few years
- 2011 Census long form eliminated, disrupting the baseline for all social and health statistics collected in Canada
- 2012-2013 elimination of First Nations Statistical Institute
- Partial defunding of Statistics Canada

Census and some environmental collections reinstated by Trudeau government 2016 onwards.

Data resistance and rescue in Canada

2010 - Canadian Association of Public Data Users was one of many groups to push back on the cancellation of the Census long form among other losses. Kristi and librarian Kim Silk maintained a [blog from 2010 to 2014](#) monitoring the situation, reporting on casualties and pushback, and sharing resources

2015 - group of librarians in Ontario (including Kristi and Alexandra noticed that some of the government data being “shared” on the Open Government Portal was unusable. We formed a group to rescue this data by hunting down missing metadata and preparing the data for ingest into University portals. Included unique collections from Health Canada:

- Children’s Health and Safety Survey
- HIV Attitudes and Behaviours
- Adolescent Drug Use

Other outputs included a how-to guide on Data Rescue, an article ([Documentation as Data Rescue](#)), and conference presentations.

Larger group lapsed after a couple of years but Kristi and Alex continued to meet occasionally to discuss data rescue and hoard files in need of preservation

Do something?

After reading about the American Data Rescue Project, Kristi posted on CanLibData: “I was wondering if there were any Canadian data library type initiatives to work on this... Or any interest in starting one?”

Eugene Barsky: “Please count me in, Kristi!” and other expressions of interest in support. Oddly, no one else seemed interested in actually starting something.

Coincidentally, Alex and Kristi had one of their regular Canadian data rescue chats scheduled the following week.





The Canadian Public Data Rescue Initiative

CPDRI Committee

Project Co-Chairs

- Alexandra Cooper, Queen's University
- Kristi Thompson, Western University

Leadership Committee

- Ken Lui, University of Toronto
- Eve Paquette-Brigas, University of Montreal
- Tracy Sallaway, Trent University
- Leanne Trimble, University of Toronto

First meeting

Canadian data community met in February 2025

- Is there anything we can do collectively to help our US colleagues?
 - Sub-group worked through spring and summer contributing metadata to [ICPSR DataLumos](#)
 - Winding down this work to focus on Canadian curation work
- How can we ensure that the Canadian government data is safe?
 - Decided to start with data in the [Canada Open Government Data Portal](#)
 - All federal ministries and departments are required to publish data created within privacy and sensitivity rules
 - Data available in both languages as CSV or Excel (generally) with documentation in PDF or Excel
 - Pilot – Environment and Climate Change Canada
 - Recommended datasets from the community – [recommendation form](#)

CPDRI communication and storage

Website: <https://sites.google.com/view/cpdri/home>

Email: canadianpublicdatarescue@gmail.com

Listserv - CanDataRescue (hosted at UBC):

<https://researchdata.library.ubc.ca/learn/candata-rescue-listserv/>

Storage:

- Google Drive hosted at MacEwan University (administration, documentation, “one-off” datasets)
- Data staging: Alliance Next Cloud

Data Distribution: Borealis, The Canadian Data Repository

- Canadian Public Data Rescue Initiative - <https://borealisdata.ca/dataverse/CPDRI>

Mass Download



Challenge – 1 dataset, 3 access points

Open government portal

Maps of reporting facilities – virtual globe format

Have your say



[Rate this dataset](#)

[Provide feedback](#)

Additional Information

Creator: Environment and Climate Change Canada

The National Pollutant Release Inventory (NPRI) is Canada's public inventory of pollutant releases (to air, water and land), disposals and transfers for recycling.

Each file contains the NPRI map layers in a KMZ format that you can use with virtual globe software such as Google Earth™. Data are available for the last two reporting years.

You can filter the data by province or industry type. Select a facility to view a report that summarizes its pollutant releases, disposals and transfers.

Please consult the following resources to enhance your analysis:

- Guide on using and Interpreting NPRI Data: <https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/using-interpreting-data.html>
- Access additional data from the NPRI, including datasets and mapping products:

<https://open.canada.ca/data/dataset/004f6b03-ac84-4c7c-90cb-35949e6260bc>

ECCC data catalogue (old)

[Canada.ca](#) > [ECCC Data Catalogue](#) > [substances](#) > [plansreports](#) > [Map of reporting facilities – geodatabase](#)

Maps of reporting facilities – virtual globe format

The National Pollutant Release Inventory (NPRI) is Canada's public inventory of pollutant releases (to air, water and land), disposals and transfers for recycling.

Each file contains the NPRI map layers in a KMZ format that you can use with virtual globe software such as Google Earth™. Data are available for the last two reporting years.

You can filter the data by province or industry type. Select a facility to view a report that summarizes its pollutant releases, disposals and transfers.

Please consult the following resources to enhance your analysis:

- Guide on using and Interpreting NPRI Data: <https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/using-interpreting-data.html>

- Access additional data from the NPRI, including datasets and mapping products:

<https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/tools-resources-data/exploredata.html>

[Back](#)[View metadata](#)

Resources

Name	Last modified	Size	Description
GE.png	2023-11-10	936 KiB	File
GE_s_fre.png	2023-11-10	52 KiB	File
GE_s.png	2023-11-10	52 KiB	File

[Maps of reporting facilities – virtual globe format - ECCC Data Catalogue](#)

ECCC data catalogue - new

 ECCC Data Catalogue
ECCC Catalogue de données

Search Map

Back to search

Download Display mode

Maps of reporting facilities – virtual globe format

Map of reporting facilities – geodatabase

The National Pollutant Release Inventory (NPRI) is Canada's public inventory of pollutant releases (to air, water and land), disposals and transfers for recycling.

Each file contains the NPRI map layers in a KMZ format that you can use with virtual globe software such as Google Earth™. Data are available for the last two reporting years.

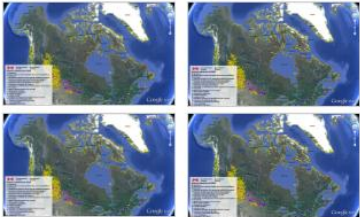
You can filter the data by province or industry type. Select a facility to view a report that summarizes its pollutant releases, disposals and transfers.

Please consult the following resources to enhance your analysis:

- Guide on using and Interpreting NPRI Data: <https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/using-interpreting-data.html>
- Access additional data from the NPRI, including datasets and mapping products: <https://www.canada.ca/en/environment-climate-change/services/national-pollutant-release-inventory/tools-resources-data/exploredata.html>

On going

Overview



Resource events

✓

Publication date
2025-04-23

➤

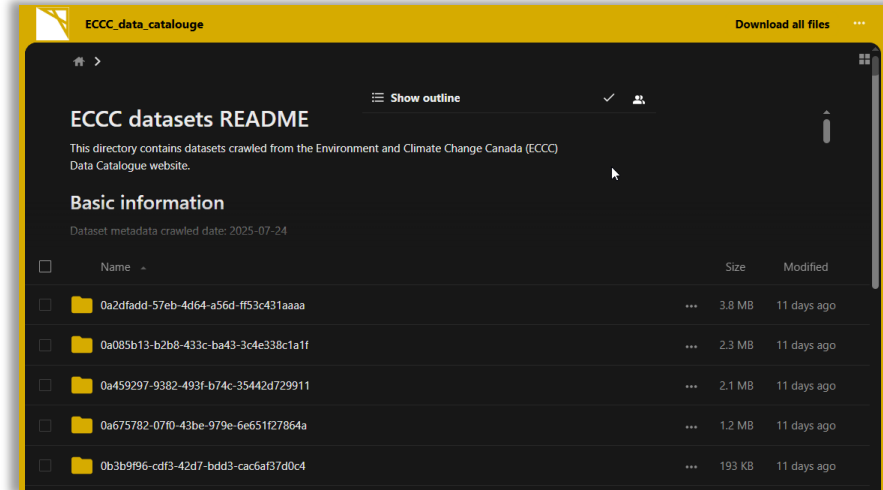
Creation date
2025-04-23

Temporal extent

[ECCC Data Catalogue / ECCC Catalogue de données](#)

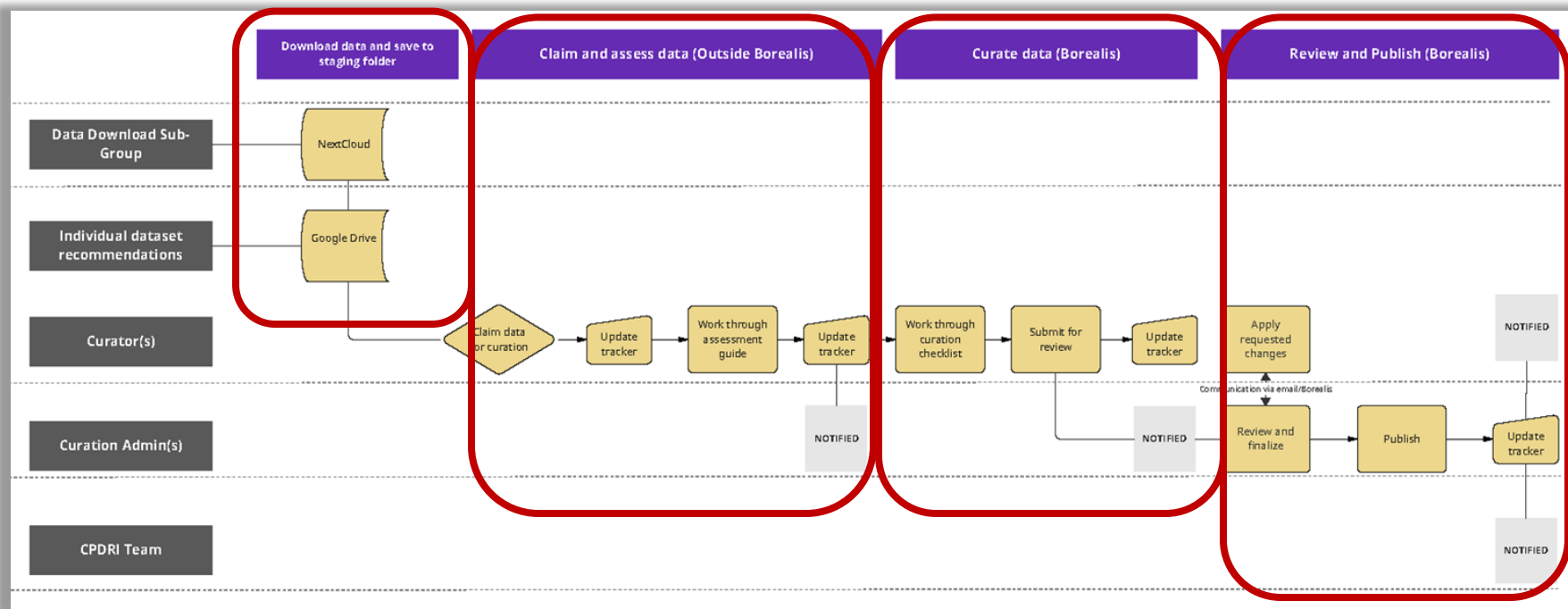
Workflow

- Webscraped open data portal to get datasets and metadata
 - Data files from ECCC Data Catalogue - Old - then;
 - Metadata of the dataset from ECCC Data Catalogue - New
- Compiled spreadsheet of the files downloaded
- Used Alliance's Machine (Fir) to download the files and put in folder





Curation workflow overview



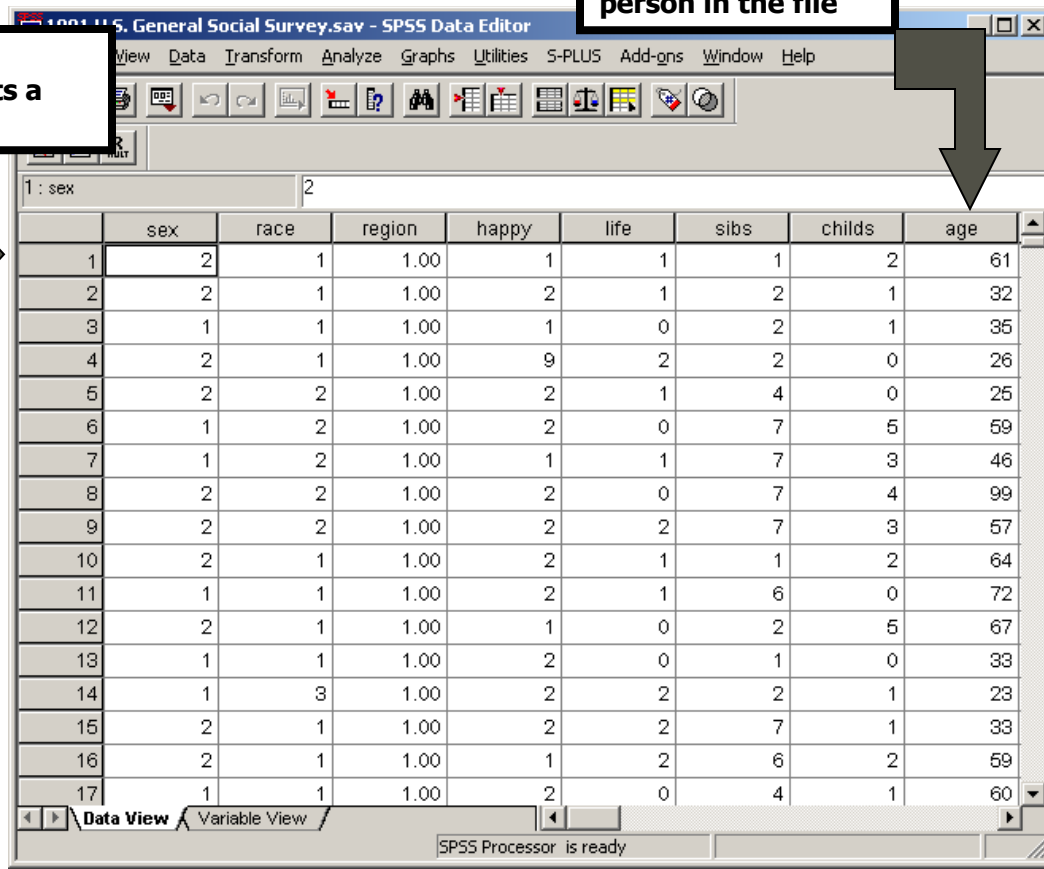
Metadata and Data Repair



Viewing a data file in SPSS

This row
represents a
person

This variable gives
the age of each
person in the file



1991 U.S. General Social Survey.sav - SPSS Data Editor

View Data Transform Analyze Graphs Utilities S-PLUS Add-ons Window Help

1 : sex 2

	sex	race	region	happy	life	sibs	childs	age
1	2	1	1.00	1	1	1	2	61
2	2	1	1.00	2	1	2	1	32
3	1	1	1.00	1	0	2	1	35
4	2	1	1.00	9	2	2	0	26
5	2	2	1.00	2	1	4	0	25
6	1	2	1.00	2	0	7	5	59
7	1	2	1.00	1	1	7	3	46
8	2	2	1.00	2	0	7	4	99
9	2	2	1.00	2	2	7	3	57
10	2	1	1.00	2	1	1	2	64
11	1	1	1.00	2	1	6	0	72
12	2	1	1.00	1	0	2	5	67
13	1	1	1.00	2	0	1	0	33
14	1	3	1.00	2	2	2	1	23
15	2	1	1.00	2	2	7	1	33
16	2	1	1.00	1	2	6	2	59
17	1	1	1.00	2	0	4	1	60

Data View Variable View

SPSS Processor is ready

SPSS variable view

The screenshot shows the SPSS Data Editor window titled "1991 U.S. General Social Survey.sav - SPSS Data Editor". The "Variable View" tab is active, displaying a list of variables with their properties. The variables listed are: sex, race, region, happy, life, sibs, child, age, educ, paeduc, maeduc, speduc, prestg80, occcat80, tax, usint, obey, and popular. The "occ80" variable is highlighted, and its "Values" column shows a list of values: 1, 2, 3, 4, 5. A "Value Labels" dialog box is open, showing the mapping of these numeric values to their corresponding occupational categories. The dialog box has a "Value" field, a "Value Label" field, and a list of labels. The labels are: 1.00 = "Managerial and Professional", 2.00 = "Technical, Sales, and Administrative", 3.00 = "Service", 4.00 = "Farming, Forestry, and Fishing", and 5.00 = "Precision Production, Craft, and Repair". An arrow points from the "occ80" variable in the Variable View to the dialog box.

Value Labels

Value:

Value Label:

Add Change Remove

1.00 = "Managerial and Professional"
2.00 = "Technical, Sales, and Administrative"
3.00 = "Service"
4.00 = "Farming, Forestry, and Fishing"
5.00 = "Precision Production, Craft, and Repair"

This box of value labels tells us what each value in the "occ80" variable represents. Without some way of knowing what the numbers represent, the data is useless.

Canadians' Understanding of Issues Related to the Aging Population - Employment and Social Development Canada

							Metadata			
id	argid	q1	q1a	quota	q2	q3				
20002	102391	2	6		Question C	Question	Answer Code	Answer		
20004	106422	2	4		id	ID				
20005	103034	2	6		argid	ARGID				
20006	102201	2	1		q1	1. To begin, in which offici	1	English		
20007	104639	2	3		q1	1. To begin, in which offici	2	French		
20008	100386	2	2		q1	1. To begin, in which offici	3	Don't know/Refused		
20009	103010	2	6		q1a	1A. What age group do you	1	18 to 24		
20010	105108	2	5		q1a	1A. What age group do you	2	25 to 34		
20011	104370	2	4		q1a	1A. What age group do you	3	35 to 44		
20012	106018	2	3		q1a	1A. What age group do you	4	45 to 54		
20013	103392	2	4		q1a	1A. What age group do you	5	55 to 64		
					q1a	1A. What age group do you	6	65+		
					q1a	1A. What age group do you	7	Refused		

Data

Beat it half to death in Excel

Encoding: UTF-8.

VARIABLE LABELS

id	"ID"
argid	"ARGID"
q1	"1. To begin, in which official language would you prefer to be interviewed?"
q1a	"1A. What age group do you fall into?"
quota	"Quota"
q2	"2. Thinking of the issues facing Canada to
q3	"3. How would you rate the current state of
q4	"4. How would you rate the current state of
q5	"5. Thinking again about issues facing Can
q6	"6. Are you currently retired?"
q7	"7. Are you fully retired or still working part-t

Excel



ADD VALUE LABELS	q1	1	"English" .
ADD VALUE LABELS	q1	2	"French" .
ADD VALUE LABELS	q1	3	"Don't know/Refused"
ADD VALUE LABELS	q1a	1	"18 to 24" .
ADD VALUE LABELS	q1a	2	"25 to 34" .
ADD VALUE LABELS	q1a	3	"35 to 44" .
ADD VALUE LABELS	q1a	4	"45 to 54" .
ADD VALUE LABELS	q1a	5	"55 to 64" .
ADD VALUE LABELS	q1a	6	"65+" .



**Actual SPSS
Syntax**

Ottawa Police Services data

https://www.ottawapolice.ca/en/news-and-community/race-archive.aspx Go AUG SEP 19 2020 MAR 2021 About this capture

9 captures 19 Sep 2020 - 3 Jul 2024

HOME REPORT A CRIME TRAFFIC COMPLAINTS FAQs ACCESSIBILITY f Twitter Instagram LinkedIn YouTube FRANÇAIS CALL US

 OTTAWA POLICE SERVICE
SERVICE DE POLICE D'OTTAWA
A Trusted Partner in Community Safety
Un partenaire fiable de la sécurité communautaire

Search Online Services

NEWS AND COMMUNITY CRIME SAFETY AND CRIME PREVENTION ABOUT US CAREERS AND OPPORTUNITIES

CONTACT US

Race Data Background Info

A A A | Print | Share

The Ottawa Police Service (OPS) is committed to developing and maintaining the trust of the community it serves by providing responsive, effective and equitable policing to all

<https://borealisdata.ca/dataverse/crime>

Citizen Science – Health Canada

Data

A	G	H	J	K	L	M	N	O	P
ID	region/rég	gender/gen	birth_year	age/âge	q1_1	q1_2	q1_3	q1_4	q1_5
1	Manitoba	Female	1983	35 to 44	#NULL!	#NULL!	#NULL!	4	#NULL!
2	Quebec	Female	1951	65 or older	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!
3	Ontario	Male	1963	55 to 64	Participate	#NULL!	3	4	#NULL!
4	Atlantic	Female	1978	45 to 54	#NULL!	#NULL!	#NULL!	4	#NULL!
5	Ontario	Male	1980	35 to 44	#NULL!	#NULL!	#NULL!	4	#NULL!
6	Ontario	Female	1999	18 to 24	#NULL!	2	#NULL!	#NULL!	
7	Quebec	Male	1962	55 to 64	#NULL!	#NULL!	#NULL!	4	#NULL!

Columns: q24 to q44= respondents answered on a scale from 1-4; 1=Not at all; 4=Very much

Columns: q46_1 to q46_20 and q47_1 to q47_7

1= Emails (active notification)

2= Smart phone notifications and/or text messages (active notification)

3= Social media feeds

4= Community meetings

5= Newsletter (active notification)

6= Updates on webpage/Participatory Research/Citizen Science platform when I

Metadata



Future steps

Where from here?

- Need to create a way for this project to sustain itself long term
- Build documentation and infrastructure
- Look into rescuing older public data on CD

The Canadian Public Data Rescue Initiative

Co-chairs

- Kristi Thompson, Western University, kthom67@uwo.ca
- Alex Cooper, Queen's University, coopera@queensu.ca

Leadership Committee

- Leanne Trimble, University of Toronto, leanne.trimble@utoronto.ca
- Ken Lui, University of Toronto, kenlh.lui@utoronto.ca
- Alex Cooper, Queen's University, coopera@queensu.ca
- Tracy Sallaway, Trent University, tracysallaway@trentu.ca
- Ève Paquette-Bigras, Université de Montréal, eve.paquette-bigras@umontreal.ca

[Web site](#)

[Borealis archive](#)