

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

*Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017*

Background

The Canadian Urban Environment Health Research Consortium (CANUE) has as a mandate to develop measures of environmental factors such as pollution, green space, traffic or walkability for every neighborhood in Canada. They are collecting data from a wide range of sources, and will be developing standardized measures as well as accompanying data documentation.

The team has as an objective of maximizing the use of the data, and are therefore seeking ways and means to have them become widely available and to have as low a burden as possible for access. Also of consideration is the ongoing storage and maintenance of the data.

In May of 2017, Nancy Meagher was contracted by CANUE to engage in dialogue with major cohort studies and administrative data centres in Canada: those that are prospective clients or end users of CANUE data. Nancy Meagher collaborated with Dany Doiron of CANUE and received oversight from Eleanor Setton, the Managing Director.

Objectives

The overarching goal was to better understand the stakeholder's interests, restrictions and needs in order to inform the design and development of CANUE.

The broad models under consideration were 1) a "push" model where data is to be shared on a routine basis, pre-emptively merged with organizations' own data and to become part of their holdings to be used for research purposes, and 2) ad-hoc "pull" of data, typically on a per-project basis, as requested. In the end, the terms became modified to mean 1) pre-emptive transfer, and 2) ad-hoc transfer. In this report, "push" and pre-emptive are used interchangeably, as are "pull" and ad-hoc.

Information was sought on legal / privacy considerations, ability to connect into existing processes, barriers that should be taken into account, feasibility considerations of the two different models, and then some more operational details such as preference for receipt of documentation, experience level of staff, available physical and human resources, and transfer frequency preferences.

Interviews

A survey instrument was developed (see Appendix A), in addition to a list of prospective stakeholders from both cohorts and data platforms across Canada.

Stakeholders were invited to an interview and provided with some background information on CANUE in advance of the interview. This background information included 1) an overview slide

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

deck on CANUE (“CANUE Data Overview May 2017”), 2) a sample listing of data sources and fields to help orient stakeholders to the types of data under consideration (“CANUE_summary_table_representative_data”), and 3) in some cases, some detail on the indicator documentation to be expected (“CANUE Indicator and Documentation Examples”).

A total of 22 interviews took place between May 10 and June 22, 2017. Ten interviews were performed by Nancy Meagher alone, seven by Dany Doiron alone and five by both Nancy Meagher and Dany Doiron together.

Nine Data Platforms were interviewed, from all provinces but PEI:

Region	Organization	Person(s) Interviewed
NS	Maritime SPOR SUPPORT Unit, and Health Data Nova Scotia	Adrian Mackenzie (Maritime SPOR SUPPORT Unit, Chair of the HDNS access committee)
MB	Manitoba Centre for Health Policy	Mark Smith (Associate Director, Repository and Deliverables), Charles Burchill (Associate Director, Data Access and Use), Selena Randall (Associate Director, Planning and Development)
NB	New Brunswick Institute for Research, Data and Training	Dr. Dan Crouse (epidemiologist) and Dr. Ted McDonald (Director)
NL	Newfoundland Center for Health Informatics and Analytics	Mitch Sturge (Assistant Dir)
ON	Ontario Institute for Clinical Evaluative Sciences	Charles Victor (Senior Director, Research and Data), Sujitha Ratnasingham (Director, Data Partnerships and Development)
BC	Population Data BC	Brent Hills (Lead, Data Services)
AB	SPOR Alberta	Dr. Jeff Bakal (Lead, Health Research Methods and Analytics for the SPOR Data Platform)
QC	SPOR Quebec	Dr. Alain Vanasse (Head of the Data Platform for SPOR Quebec)
SK	SPOR SK	Kim Hill (Director, Information Governance, eHealth SK) and Tracey Sherin (SK Health Quality Council) -- co-Leads of Data Platform

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

Thirteen cohort studies were interviewed, 6 of which were birth / mother-child cohorts:

Region	Organization	Person(s) Interviewed
QC / National	3D study and MIREC	William Fraser (Principal Investigator)
AB	Alberta Pregnancy Outcomes and Nutrition Study	Deborah Deway (Team Lead)
AB	All Our Babies	Nikki Stevenson (Research Program Manager) and Sheila McDonald (Scientist)
NL, NB, PEI and NS	Atlantic PATH	Ellen Sweeney (Research Scientist, Atlantic PATH) and Jason Hicks (Executive Director)
BC	BC Generations Project	Trevor Dummer (Principal Investigator)
National	Canadian Alliance for Health Hearts and Minds	Sonia Anand (Principal Investigator)
National	Canadian Cohort of Obstructive Lung Disease	Jean Bourbeau (Lead Investigator)
National	Canadian Healthy Infant Longitudinal Development (CHILD) Study	Padmaja Subbarao (co-Director)
National	Canadian Longitudinal Study on Aging	Sarah Youssef (Senior Data Curator) Istvan Molnar-Szakacs (Data Access Officer)
QC	CARTaGENE Quebec	Catherine Boileau (Associée de recherche, Épidémiologie)
ON	Ontario Birth Study	Ryan Seeto (Data Manager)
ON	Ontario Health Study	Philip Awadalla (Principal Investigator)
National	The Applied Research Group for Kids	Jonathon Maguire (co-Principal Investigator)

There was only one organization with whom we sought contact with which we were unsuccessful in securing an interview: Alberta's Tomorrow Project (Dr. Paula Robson.)

Most interviews lasted an hour, some as short as 30 minutes.

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

Results and Findings

Overall **the response to the request for interviews was very strong**, as witnessed by the completion of all interviews in just over one month.

While not an interview question, enthusiasm and support for the concept of CANUE was provided, recognizing its value. A number had referenced their labour-intensive experience in pursuing one or two of the anticipated CANUE measures, or challenges with not having the relevant expertise on site to develop such measures. Many offered that environmental exposures are seen as an increasing priority, and welcomed the ability to benefit from the work of CANUE. **The CANUE mandate and offering is of high relevance.**

In nearly all cases, the requirement that the organization's sensitive data not leave the home site was made explicit by the interviewee. This quickly became part of the assumed state on the part of the interviewers. This means that the **baseline assumption is that record merging of the organizations data and CANUE data are to occur at the stakeholder organization**, not by CANUE.

Lastly, **additional information about CANUE was requested** by nearly all respondents. Questions included timing of CANUE data availability, specific data items to be available, geographical and time coverage, and size of data transfers. Providing some of this information along with summary results of this engagement process would be beneficial, as would ongoing communications and updates. This group of interviewees should be considered as part of CANUE's core data stakeholders going forward.

Specific results of the interviews are detailed in an accompanying spreadsheet titled "CANUE Consultations Summary Table."

The following are the major take-homes:

A) There is preference for the "push" or pre-emptive model of bulk data transfer on a routine basis to pre-emptively merge.

The common reasons included:

- Increase the likelihood of its use: once in house, it would be available for any data request, making it easier for researchers and thus more likely to be used. This would in turn breed increased familiarity and a virtuous cycle of increased use.
- Efficiency: A single merge vs. multiple per-project merges resulting in duplicated effort. This also relates to cost / labour implications, with the costs being borne by individual research projects in the 'pull' model.
- Eliminates a step in the data access process: related to the above, with one less step involved it brings greater efficiency for all and an increase in likely use.

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

- Quality: A single merge is likely to have more oversight and access to expertise than per-project merging.
- Feasibility assessment: In particular for cohorts, it was seen to be useful to be able to identify feasibility of certain projects, for example “how many records do we have of people with X exposure to Y pollutant?”

Concerns about such a model focused on the possible labour efforts required, which prompted a number of inquiries regarding how clean the data was going to be and how complicated the merging. A number stated that they would probably bring the data in pre-emptively, however only prepare it once a request was in place. Some indicated they would likely bring in only those data that are relevant to their own data in terms of geography, time period and content. Additionally, some questioned whether pre-emptive work could be justified and suggested that they may need to wait to see about demand.

Benefits of the “pull” or ad-hoc model identified included:

- Data would be assured to be fresh.
- No data management or storage burden on the organization.

About a third noted that some form of hybrid model would be likely beneficial. In some cases this would be having project specific demand guide holdings, and in other cases this meant that some core holdings might be identified for pre-emptive intake and merging, and other more rare ones left to ad-hoc requests.

B) There is a value seen in being able to self-serve CANUE data

Even with a pre-emptive approach, a number of organizations expressed a desire to be able to acquire the data on their own and not be caught up in a CANUE queue. This relates to ability to adhere to data delivery timelines that they are accustomed to, as well as avoiding the need to ‘babysit’ data transfers that are pushed out. Such an ability would also accommodate a hybrid mode where some data is pre-emptively taken in house and the more rare data is left on the CANUE servers till demand is present.

C) All are able and willing to incorporate access of CANUE data into existing processes

Access process includes the application and review steps, the release to researchers, and any legal documentation that governs release and use, including citations. Tying to the existing application and review steps provides significant advantage to CANUE as well as researchers, as there is no new steps or vetting that need to be introduced.

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

Regarding the legal documentation, these are often called “research agreements” or “data disclosure agreements.” Some organizations were open to the idea of amending their existing agreements, but in most cases it was suggested that CANUE develop its own stand-alone agreement that the organization could ensure was signed in tandem. This will be significantly influenced by what CANUE ends up developing regarding data governance, as described in recommendations.

D) Privacy and legal considerations are not likely to be barriers to uptake

Most have consents that specifically allow for linkage to external data sources. Most acknowledged further that these data do not have privacy considerations given they are not person specific.

That said, it was also noted that there are sometimes concerns, uncertainty and extra scrutiny by reviewers regarding geospatial data and potential identifiability implications. It would be beneficial for CANUE to develop some information to inform these concerns.

Organizations that may have some barriers that need to be addressed for merging to occur include the following:

- Quebec SPOR (Soutiens) and PopData would need to modify existing legal agreements
- 3D and Mirec noted that they would need to seek ethics approval to bring in the data
- APrON’s consent does not allow for linkage, however it is uncertain whether CANUE data would be considered “linked” given its geography based and not individually based.
- CARTaGENE noted some challenges with provincial legislation and use of residential information. They are seeking explicit consent for this kind of use of residential information however are only at 25% consented.

Despite the identified hurdles, it was felt that all were feasible to overcome.

E) Restrictions on CANUE data use can be managed in most cases, but need clarity.

Restrictions discussed included the possibility of requirements for co-authorship, citation of data / acknowledgement of CANUE, and limitation to academic uses. In many cases these restrictions mirrored those in place with the organization and were seen to be reasonable. Suggestion was made for CANUE to work towards standardizing

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

the requirements in order to best support their management – so for example, to have only 2 types of requirements – one for citations and one for co-authorships, not a multitude of variants for each data source.

Organizations expressed an ability and willingness to communicate the requirements, but a hesitancy to be responsible for compliance to them, which is reasonable given challenges many have experienced with compliance. This should be reflected in the ultimate data governance arrangements.

Specific clarity around the academic use restrictions as prompted by PCCF licensing were requested, as there is uncertainty of academic vs. research use, transferability of licenses, and inquiries about being able to include a software charge if non-academic uses were allowed. CANUE will need to investigate these further and incorporate into policy documentation. **PCCF licensing restrictions may affect a few organizations ability to bring in or use CANUE data.** This is only present for a few data platforms, particularly those that are based out of government or in the context of explicit government partnerships, like the Alberta and Saskatchewan SUPPORT Units and the maritime provinces.

- F) Annual transfers are likely to meet most needs.** In most cases respondents expressed a desire to update along the same timeframe as their own data is updated, which is most typically annually. Variants to the annual usually were accompanied by questions as to the labour intensity of the merging process, and the frequency of updating the CANUE data.
- G) Metadata is essential, but the location of it is not critical.** Most data platforms suggested they would make some effort to bring documentation into their own systems to allow a common look and feel and a single extract of documentation for researchers. Conversely, cohorts and in particular smaller cohorts tended to want to leave CANUE documentation at CANUE, noting that the labour of maintenance and upkeep of updates would mean that the two systems could quickly get out of synch.
- H) Other limiting factors (space, labour) are most present for the cohort studies.** For data platforms, neither space nor labour capacity were identified as rate limiting steps, although some noted that they would need to justify the use of resources. This was the same for the larger cohort studies.

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

Smaller cohorts expressed much more sensitivity to the possible space and labour implications of bringing in data, with likely labour ones being the most acute. Any efforts that CANUE is able to take to minimize labour implications for merging would have a significant effect on uptake, most notably for smaller cohorts, but even for data platforms.

Another limiting factor in the case of 3D and Mirec is that only 3 digit postcode has been collected.

- I) Organizations would like advisory support and expertise by data source.** In the absence of having people in-house with knowledge and expertise in specific areas such as noise pollution and air quality, many noted that they would like to be able to have access to experts in these areas. This might be to field researcher questions about feasibility or design, or support in development of a cohort for a research project that uses CANUE data. There was further suggestion made that a knowledge or methods bank would be valuable to expanding the uptake of the data usage. Related to this would be support in interpreting the data.

A few organizations requested guidance on merging, noting that there could be many choices in terms of which postcode and which timing to employ, and that this may even vary by research question. Scenarios in merging was put forth as a suggestion to help with this, for example 'merge using the last known postcode in December of each year.'

- J) Desire for support with promotion and marketing.** Most organizations have little to no experience with use of environmental and exposure data, though most saw its value. They would like to see support from CANUE in terms of generating interest in the data, providing ideas of the data's potential, and raising awareness of the data. In some way, it would be to prime the pump for use, or as one described it, to demonstrate the "art of the possible." Another described it as help 'getting over the inertia' of (not) using the data. A few recommended that CANUE present to executive or steering committees in order to gather strategic support for the data.

K) Miscellaneous additional input or commentary includes:

- a. Nova Scotia has created a new more accurate geographic unit, "community cluster"; others looking at latitude and longitude merging.
- b. New Brunswick: Their address history is only good from 2000 forwards, as previous addresses were overwritten.

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

- c. There is a possible need for materials to be translated into French, in particular for Quebec and New Brunswick.
- d. Look at the possibility of incorporating access to CANUE data via existing online health atlases which are present in many jurisdictions.
- e. Suggestion to create sample datasets for researchers to gain an understanding of the available data.
- f. Some may wish to charge researchers for their services in accessing the CANUE data. Will this be a problem at all with agreements with CIHR or elsewhere?
- g. Strong interest was identified by one respondent in trying to add person level exposure information.
- h. One mother-child cohort respondent requested built environment information on parks and playgrounds.

Recommendations

The following are recommendations for action on the part of CANUE in order to keep moving forward this component of the initiative.

1. Develop the governance framework for CANUE data.

A governance framework relates to the characterizations and authorities for collection, use and disclosure of data. For CANUE, this clarity is needed for a) the contributors of CANUE data to CANUE, b) CANUE itself, c) the organizations that will use and / or house CANUE data, and d) the end-users. This needn't be overly complex, however it will be essential to have this laid out to inform the needed agreements between each of those four players. Likely there will be 3 types of agreements: 1) information sharing between the original data providers and CANUE, 2) information sharing between CANUE and the organizations where the data to which it will be merged resides, and 3) (if needed) research agreement with the end-user and either the organization provisioning data or CANUE (TBD based on the governance framework.)

Some key questions that will be addressed by such a framework include:

- What is the chain of accountability for the data, or who is responsible?
- Can that responsibility be delegated, and how / what are the conditions on that?
E.g. is CANUE able to be structured such that they have all decision making authority over uses of the data they collect / collate? And can they transfer this to any of the organizations through agreement? Does this in turn get transferred on to the researcher?

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

- What are the specified uses of the data? Restrictions on use?
- What is CANUE's legal structure? Is it technically at a specific university, a not for profit, or other?
- Is there any legislation that applies to collection, use and disclosure?
- Are there any considerations for inter-provincial movement of data?

Ideally CANUE will secure sole decision making control and legal authority over the data it houses. This transfer of authority will allow for lower burden of approval and agreement on the stakeholder organizations. In the absence of this, CANUE may need to structure agreements by data sources with each stakeholder organization, rather than a single agreement with CANUE.

2. Develop standardized data transfer / information sharing agreements, and seek to maintain a single standardized agreement.

With the information and clarity achieved from a data governance framework, the core contents of an information sharing agreement (ISA) between CANUE and the stakeholder organizations can be clarified.

While the burden is likely much less than what most are used to with record level privacy-sensitive data, it would be highly beneficial for organizations to see the core CANUE requirements in the form of a draft ISA, so that they have a better understanding of what it would take to maintain those requirements.

An additional piece of guidance is that CANUE seek to maintain a single version of their ISA with outside organizations, and that any suggestions or requirements to include information with the ISA by any organization are incorporated into the single version. This becomes important the more organizations that are dealt with, because any variation means keeping track of that variation, introducing more oversight intensity and room for error. To this end, once a draft ISA is developed, consider sending it out to all stakeholders for input and feedback.

Note that significant work has been put into this whole area of ISA development, and there is a lot of opportunity to take advantage of existing agreements to inform the development of CANUE's agreements, including use of Creative Commons licensing for publicly accessible data: <https://creativecommons.org> .

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

- 3. Develop guidance on restrictions based on PCCF software licensing.** This is likely to be rolled into the draft ISA contents, but is highlighted specifically because of the need for completely understanding the licensing restrictions since this was so prevalent in the interviews.

CANUE is seeking simplicity based on academic use licenses. Many respondents asked for more detail and clarity. Because effort is needed to seek additional information to flesh this out, it is recommended that slightly more effort be put into negotiating with postal code file owners (Canada Post, DMTI Spatial) for use scenarios.

The following questions need clarity:

- How is academic defined?
- Is it academic or research use? How do you distinguish between the two?
- What if the organization has PCCF software already, does it transfer?
- What if they organization has the alternative to PCCF – can that be used interchangeably?
- Can an agreement be developed that would allow for per-use software payments to PCCF?

- 4. Develop guidance on data merging.** A number of organizations asked for support on this, including organizations that had experience in this type of merging. An assessment can be made in terms of guidance that may be possible, and could include:

- Merging instructions per data source,
- Sample merging code
- Case scenarios for merging
- Challenges and solutions
- Advisory contacts for merging

- 5. Determine if cost recoveries for use of CANUE data are allowed.** A number of data stakeholder organizations suggested they may need or want to charge for use of the CANUE data – not for the data itself but for the provided services in supporting access to it. There may be restrictions on the part of CANUE funders or other data organizations which would be useful to sort out if allowed, and relevant restrictions at this early stage.

- 6. Develop guidance on privacy considerations in merging geospatial data to record-level data, and use of geocoded data.** While many acknowledged that CANUE data itself is not privacy sensitive, it was noted by many that merging spatial data with record level

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

data often triggers uncertainty in the review and approval process. In the spirit of CANUE providing centralized initiatives to avoid duplication of effort, policy guidance on known hurdles such as this would be welcome. It would also provide some data steward comfort if such a policy document were to be vetted externally and used in a uniform way across jurisdictions, demonstrating some standardization.

- 7. Formalize the group consulted into a data stakeholder group, expanding membership as needed.** There is a high level of engagement and interest on the part of those interviewed, and it would be helpful to CANUE to formalize their role with the organization. It would continue to build on the enthusiasm and allow these organizations to see themselves as having a stake in CANUE. This will benefit CANUE in the longer term, not necessarily in the short term. CANUE leadership will want to assess what kind of relationship this might be: information, guidance, decision making? Such a group could even result in cross-connecting between groups and provide crowd-sourced input in e.g. merging routines and issues. Additionally, CANUE may wish to formalize into a technical sub-group and a scientific sub-group.

- 8. Expand the group by the currently known missing organizations.** Based on interviews, the following are additional groups that should be considered for interview and participation in a stakeholder group:
 - a. Research Data Centres / Stats Can
 - b. <https://policywise.com/initiatives/sage/>
 - c. Newfoundland Centre for Health Information – a better contact than the one we had interviewed.
 - d. PEI data platform under SPOR SUPPORT Unit
 - e. START Study and Aboriginal Birth Cohort in Hamilton (5,500 mother / baby pairs)

- 9. Develop regular communications with the data stakeholder group.** Initially, this would involve summary results of the stakeholder engagement, and also address some of the burning questions that came from the interviews:
 - How big will the data be?
 - What is the timing?
 - How “clean” will it be?
 - What are the specific restrictions on use?
 - What kind of work is being done at CANUE to support the data stakeholder uptake, e.g. privacy guidance, data merging guidance, expert advisory support

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

10. Clarify anticipated support CANUE will provide to data stakeholder groups and researchers. In the absence of this, assumptions may range from nothing to extensive call desk support. This includes also addressing requests for support in marketing and promotions: how will this be designed, and what level of commitment is CANUE able to make?

11. Technical design considerations:

- a. **Seek to develop ability for users to self-serve for the CANUE data.** While the pre-emptive merge model was supported by most, a third liked the hybrid model, and some of those in support of pre-emptive merging stated they would prefer to grab the data themselves. Self-serve capacity would allow multiple models to be supported, and further it would mean lower maintenance requirements in the longer run.
- b. **Explore opportunities to automate and even synchronize data with remote sites.** To the extent that extractions can be created as stored procedures that can automatically transfer, or even better, synchronize with other data environments, this would reduce the burden on CANUE as well as the data stakeholder organizations.
- c. **Explore opportunities to support and automate merging.** The effort required to merge was a question amongst nearly all those interviewed, and any work to simplify or support this would be welcome and increase chances of uptake. This will be particularly relevant for smaller cohorts that might otherwise not prioritize CANUE data. Further, guidance would ensure better standardization of merging methodologies, important for pan-jurisdictional research.

12. Additional notes:

- a. Population Data BC has excellent guidance on data citations, and have offered it to CANUE.
- b. CANUE Metadata should be exportable in case organizations want to try to import.

Conclusion

The stakeholder engagement process that was undertaken was highly beneficial for CANUE as well as the ultimate stakeholders. It provided an important initiation of the relationship, as well as content that will be useful for not only the design and development of CANUE but also for feedback to CANUE funders.

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

Appendix A: Interview Guide

1. COHORT Q: Do you anticipate any ethical / legal / privacy considerations to merge environmental data with your cohort and support access to it? Describe.

ADMIN Q: For research projects that intend to use merged health and environmental data, what implications might there be for securing approval? (ethics, data steward) Do data stewards need to be aware of the inclusion? Are there any 'grey areas' that you can think of that might affect the ability for these data to be approved for use?

2. Some of the environmental data has conditions of use. How might the addition of CANUE data to a project be able to fold into your current process for authorizing (i.e. research agreements), and monitoring or managing conditions of use? (i.e. vetting research outputs, other monitoring)?
3. We'd like to get some understanding of the opportunities and implications of possible approaches for CANUE data to be incorporated with your data for research uses. To this end, we'll be asking you questions about two opposing models.

One model is to have CANUE data "pushed" or transferred to you pre-emptively on a routine basis, and to become part of your holdings. The other end of the spectrum is an ad-hoc "pull", for project specific data requests to be handled and merged as they come in.

Let's start with the "push", or where you receive and hold CANUE data. <<would you anticipate pulling in only a subset of the data?>>

- a. Is there any interest in such a model? Would you anticipate pulling in only a subset of the data, or all of them?
- b. Would you have the capacity to house such data? If no, what might be missing?
- c. Would this affect any of your existing legal agreements (e.g. information sharing, consent)?
- d. Do you have experience linking environmental data?
- e. Would you have the capacity to merge such data? What questions do you have around data merging? Do you anticipate it could be done pre-emptively, or would it be done on a case by case basis? Are you able to send out postal codes?

CANUE Stakeholder Report

Understanding technical, procedural and functional needs of major data sources

Prepared by: Nancy Meagher (Lumine Research Inc.) with input from Dany Doiron (Maelstrom Research/CANUE)
August 2017

- f. Would you anticipate including CANUE documentation in your documentation, or pointing to CANUE's documentation? *(If the former, ask for example for their kind of documentation to consider.)*
 - g. Transfer frequency preferred: monthly, quarterly, annual, as it comes, other?
 - h. How would you characterize the cost and resource implications for including these data? Would you need additional resources to be able to partner on CANUE data?
4. Now to move to the other end of the possible service delivery spectrum: Where data is "pulled" on an ad-hoc basis. The data may be provisioned by CANUE, or we could also consider a possible model where you pull from CANUE as needed.
 - a. Where would you see data merging take place? Could you do it? Would it need to be at your centre? For what reasons? (Can you give out 6-digit postcode?)
 - b. What are factors that come to mind when considering CANUE pull vs. you directly pulling? Do you have a preference?
 - c. How would you characterize the cost implications on you or the researcher for this model?
5. Do you have a preference among the options?
6. Are there other factors that come into consideration that we have not yet covered?

Appendix B: Consultations Summary Table

(see excel spreadsheet of same title)