Higher Education Quality Council of Ontario

An agency of the Government of Ontario

CRP Blueprint: How We Built a Community Data Infrastructure

Jackie Pichette, Lorraine Valmadrid and Sally Landon

Published by:

The Higher Education Quality Council of Ontario

88 Queens Quay West, Suite 2500
Toronto, ON
Canada, M5J 0B8

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca

E-mail: info@heqco.ca

Cite this publication in the following format:

Pichette, J., Valmadrid, L. & Landon, S. (2022) *CRP Blueprint: How We Built a Community Data Infrastructure.* Toronto: Higher Education Quality Council of Ontario.

Higher Education Quality Council of Ontario

An agency of the Government of Ontario

The opinions expressed in this research document are those of the authors and do not necessarily represent the views or official policies of the Higher Education Quality Council of Ontario or other agencies or organizations that may have provided support, financial or otherwise, for this project. © Queens Printer for Ontario, 2022

Table of Contents

List of Figures

Introduction

Effective policymaking and planning in the education sector require detailed information about student experiences and outcomes, both over time and across institutions. Unfortunately, this kind of information is not readily available in Ontario (Gallagher-Mackay, 2017; Robson, 2021). Indeed, this gap in accessible information about students' educational pathways motivated the creation of the Hamilton Community Research Partnership (CRP): a coalition of six organizations committed to supporting student achievement in Hamilton through a secure data-sharing network.

This paper outlines the process and terms that governed the creation and use of a CRP pilot project. It offers a blueprint for other communities and educational institutions that are interested in learning from the CRP members, and building similar datasets for research and improvement purposes. Throughout, we detail the privacy impact assessment that informed our approach and the steps of assembling the dataset, including a two-staged de-identification process, data transfer and merging. Then we discuss the limitations inherent in our process as well as our plans to address these challenges moving forward in years to come. For the first set of findings derived from the CRP, see our companion report, *The Power of Connected Data: Charting Student Pathways to and through Postsecondary in Hamilton*.

The Hamilton CRP includes two local school boards: Hamilton-Wentworth District School Board (HWDSB) and Hamilton-Wentworth Catholic District School Board (HWCDSB); two postsecondary education (PSE) institutions: McMaster University (McMaster) and Mohawk College of Applied Arts and Technology (Mohawk); and two public organizations focused on addressing systemic issues in education: Hamilton Community Foundation (HCF) and the Higher Education Quality Council of Ontario (HEQCO). The educational partners (HWDSB, HWCDSB, McMaster and Mohawk) contribute data and their expertise as education and service providers; HEQCO and HCF serve as co-facilitators, providing strategic direction and ongoing project management support. HCF is the primary point of contact for the partnership; it works directly with partners to problem-solve and keep projects on track. HEQCO works with HCF and the partners to conduct research, mobilize knowledge and advise provincial policymakers and also provides funding for the CRP.

The CRP members joined with three goals. The first was to better understand the educational pathways of Hamilton students. Data is often constrained to individual institutions, so little is known about students after they leave one institution (e.g., high school) and before they enter another (e.g., college). This makes it nearly impossible to identify systemic factors that affect learning trajectories and contribute to student inequities. The second goal was to address this gap in available information by developing

and testing a data-sharing mechanism: we used the Ontario Education Number (OEN) to link individual-level administrative data across multiple school boards, postsecondary partners and application centres. Every student is assigned a unique OEN that follows them along their educational journeys from kindergarten to high school and, if they choose, to postsecondary. Using the OEN as the key, the CRP partners sought to link institutional administrative data to produce one dataset for research purposes. Our final goal was to build trust among CRP members and establish procedures for ongoing, collaborative research in our community — and to inspire other communities to do the same.

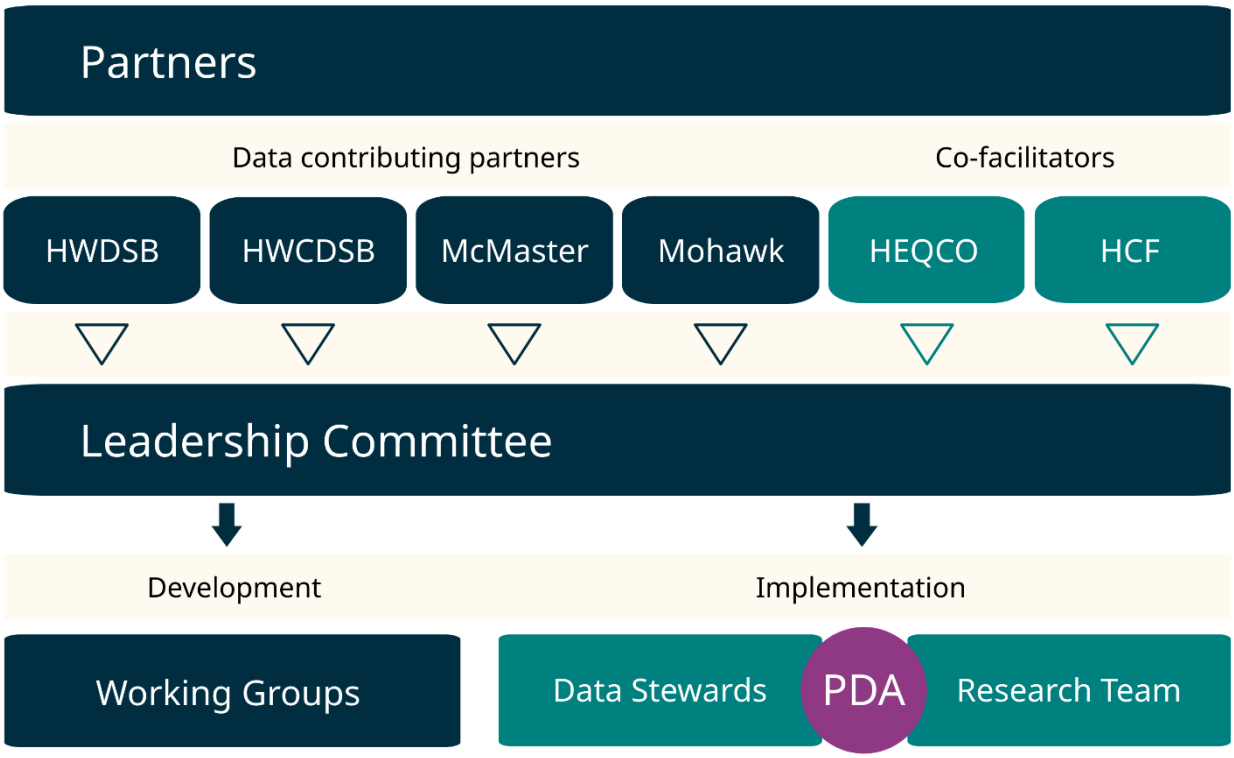
Governance Structure

Four key groups operationalize the work of the CRP: the Leadership Committee, Working Groups, Data Stewards and the Research Team.

The Leadership Committee is the primary decision-making body, setting direction and governing the overall operations of the partnership. Each partner has a representative on the Committee with decision-making authority for their organization. For example, the school board representatives are the Directors of Education; McMaster's is the University Registrar; and Mohawk's is the Vice President, Academic. The President and CEO at both HEQCO and HCF serve as non-voting members as neither organization contributes data to the CRP. The Leadership Committee meets two to three times annually at key project junctures and adheres to a set of guiding principles, described in a following section.

Figure 1

CRP Governance Structure



Note: This figure shows CRP governance structure: a Leadership Committee, made up of representatives from each partner institution, governs designated Working Groups, Data Stewards and the Research Team. The Working Groups develop policies and processes while the Data Stewards, including the Partnership Data Analyst (PDA), and Research Team implement approved policies and processes (i.e., they collect, connect and analyze data).

Working Groups are initiated by the Leadership Committee as needed to propose policies and processes that operationalize the Leadership Committee's direction. The Leadership Committee appoints Working Group members based on their affiliation with CRP partners and the skills and expertise needed at each stage of the project. For example, the Leadership Committee called on researchers at HEQCO and McMaster to help determine relevant variables to include in the dataset; information technology staff and data analysts at the educational partners to develop data transfer processes; and institutional privacy officers to assess risk and develop security measures in consultation with their legal advisors.

The Leadership Committee also designated one Data Steward at each of the four education partners as responsible for executing the assembly and transfer processes.

Data Stewards have a strong knowledge of their institution's administrative data; during the pilot, they included school board data analysts, a university statistician/programmer and a college director of corporate reporting. As a group, the Data Stewards were responsible for creating a data dictionary to ensure variables were coded consistently across institutions; ensuring privacy and security measures were upheld at their organizations; and vetting data before it was transferred. They also answered questions related to the data and advised on the feasibility of research proposals.

The Leadership Committee agreed that the data should be stored at HWDSB to protect access. As such, the Data Steward at HWDSB had additional responsibilities of merging and de-identifying the linked dataset. This was the only person who had access to the dataset once it was assembled and who performed the analyses to answer the research questions identified by the Leadership Committee. Because of this unique role, they were designated as the Partnership Data Analyst (PDA). Their role also entailed working with a Research Team, involving researchers at HEQCO and McMaster, to analyze the dataset and produce research findings.

There are two partnership agreements in place that lay the foundation for the first project (or pilot phase) and its governance structure. The first is a Memorandum of Understanding (MOU) between HEQCO and HCF that outlines the roles and responsibilities of the CRP co-facilitators. The second is a Data Sharing Agreement (DSA) between the four education partners and the co-facilitators, which outlines the governance of the partnership as well as the protocols for collecting, transferring, managing and using the data for the first round of research.[1] The partners agreed the results of the pilot phase should inform the next phase of the partnership and all subsequent agreements.

Legal Authority

The CRP members set out to do something bold — data in Ontario is rarely shared between school boards and postsecondary institutions, let alone multiple school boards and postsecondary institutions at the same time (Robson, 2021). To minimize potential concerns about sharing data, and in adherence with requirements set by the Information and Privacy Commissioner of Ontario, the Leadership Committee struck a Working Group to conduct a Privacy Impact Assessment (PIA); this was intended to identify privacy and security risks and take measures to mitigate them. The PIA also served to spell out the partners' legal authority to collect and share data and establish a process that could be shared with and used by other school boards, colleges and universities.

[1] Pilot phase research focused on the cohort of students who began Grade 9 in Hamilton during the 2010-11 academic year.

The PIA process aligned with the Ontario Information and Privacy Commissioner's *Planning for Success: Privacy Impact Assessment Guide* (2015) and began by confirming the PIA was necessary (i.e., it confirmed the project would involve the collection, use, retention, disclosure, security or disposal of personal information). The Working Group reviewed the *Freedom of Information and Protection of Privacy Act* (FIPPA), the *Municipal Freedom of Information and Protection of Privacy Act* (MFIPPA) and the *Education Act* — these are collectively referred to as the "Privacy Acts" in this section. FIPPA and MFIPPA outline public institutions' obligations in the collection, use and disclosure of personal information. *The Education Act* "is the main law under which schools and school boards operate. It governs how education is delivered to students in Ontario's publicly funded school system" (Information and Privacy Commissioner of Ontario, 2019). The Privacy Working Group found the Privacy Acts uphold the educational partners' legal authority to carry out this project and link student information across institutions (see Appendix A: Legal Authority for details about which sections uphold this authority).

Essentially, the Privacy Acts permit the disclosure of personal information for a purpose consistent with the one for which it was collected (e.g., to educate, to improve educational outcomes, to identify barriers and address them) as long as a reasonable person would expect it. The four educational partners acquired the data needed to create the CRP dataset legally, for the proper administration of a lawfully authorized activity (i.e., for the purposes of educating students). The CRP aimed to use this data consistent with the purpose for which it was collected (i.e., for educational purposes, and to improve educational outcomes for students). That said, the partners could not be confident that the students included in the dataset, beginning Grade 9 in the 2010-11 academic year, as well as their parents, would expect or be made aware that their data was to be used in this way. In fact, they may have had the opposite expectation. Ontario educational institutions have historically obtained consent for research instead of setting expectations and explaining the legal authority to collect and use student information.[2] In so doing, school boards, and to a lesser extent colleges and universities, often inadvertently insinuate that research using student data constitutes an unreasonable invasion of privacy — this is a major misconception and barrier to participation in data-sharing projects like the CRP.

Recognizing this issue, the CRP members took a conservative approach and severely restricted the contents of the dataset, reducing the risk that the data or the research it produced could be traced to any individuals. As described in the following sections, the partners developed a process to ensure the data was as 'unidentifiable' as possible, even to the detriment of our ability to uncover rich insights. In future phases beyond the pilot,

[2] This statement is based on the three authors' experience working with and at school boards (the authors are a manager of a school board research and analytics department, a research officer at a school board, and a director of partnerships at an agency (HEQCO) that frequently works with school boards).

the partners aim to expand the CRP's utility by reviewing school board notices of collection and ensuring students and parents are fully informed. Over time, and with effective communication of our initial research findings (summarized in *The Power of Connected Data*), student and parent communities will come to expect that data will be safely shared across institutions to better understand learning trajectories and barriers. This will allow the CRP to share and analyze detailed, individualized information related to academics and demographics to inform decision-making and improve educational outcomes for Hamilton students.

Guiding Principles

The CRP is governed by a set of guiding principles, laid out in a research protocol and reflected in the DSA. The first of these principles is to conduct research under the common goal of supporting students through their various pathways. Access to longitudinal data is critical to understanding how policymakers and practitioners can equalize opportunities for students — especially those who have historically been marginalized by and underrepresented in postsecondary, including first-generation students (those whose parents didn't complete postsecondary), low-income students, racialized students and students with disabilities (Chatoor et al., 2019, 2022; Ford et al., 2019; Robson et al., 2019). The CRP Leadership Committee members recognize the important role the partnership can play in building the evidence needed to equalize educational opportunities. Other guiding principles related to research include collaboratively agreeing on all research projects, balancing knowledge generation and action, and working collaboratively to protect the reputations of all partners.

In drafting the guiding principles and terms of a DSA, the Leadership Committee also acknowledged that each educational partner has its own Research Ethics Board (REB) with rules and procedures that must be followed. REBs are independent, multidisciplinary committees that ensure research is conducted ethically, protecting the rights and well-being of research participants. Their advice, as well as the advice of institutional legal teams, add value and security to the partnership.

To build trust among stakeholders (such as parents and students) and create a culture where data sharing is seen as a powerful tool for understanding systemic gaps and barriers, the Leadership Committee agreed upon data privacy and protection guidelines informed by the PIA described above to be compliant with relevant legislation. These guidelines resulted in risk-mitigation measures, described in the data assembly section, including applying de-identification methods and limiting data access. The Leadership Committee also developed the CRP using a phased approach. The first phase was designed to serve as a proof of concept to confirm that the partners' data could be

connected successfully and safely, relying on historical data for one cohort of students (i.e., students who entered Grade 9 in the 2010-11 academic year).

The partners were also concerned about reputational risks. To ensure that partner reputations were not harmed because of the CRP, the DSA stipulates that research cannot be used to rank organizations or make nefarious comparisons between them. As a check and balance, the governance structure states that all research and operational decisions must be made by the Leadership Committee to consensus. Research publications must also be vetted by the Leadership Committee.

The full list of guiding principles is included in Figure 2 (and in standard text in Appendix B: Guiding Principles). The partners developed the principles collectively during Leadership Committee and Working Group meetings in the early stages of the partnership.

Figure 2

CRP Guiding Principles

Note: This figure shows the four CRP guiding principles: research; data governance; knowledge mobilization; and collaboration. A standard text (i.e., non-visual) description can be found in Appendix B: Guiding Principles).

Data Assembly and Storage

Assembling the CRP dataset for the pilot phase was a multi-step process.[3] As a first step, Data Stewards at the two school boards pulled together their institutional data on students who started Grade 9 in the 2010-11 academic year. The PDA provided a list of hashed OENs to McMaster and Mohawk so that the PSE Data Stewards could mine their student information systems for data corresponding to the same students for up to eight years (through to fall 2018). Each Data Steward ensured the dataset they created contained the variables agreed upon by the relevant Working Groups. The Data Stewards conducted a first stage of de-identification, removing all direct identifiers, such as student names, and generalizing[4] sensitive quasi-identifiers, such as full date of birth and full six-digit postal code. To minimize the risk that students could be identified by their OEN, the PDA provided each Data Steward with a hashing formula[5] to obscure the OEN in their datasets. The resulting datasets containing the hashed OEN were stored separately from institutional student databases (which include the OEN) to further reduce the risk of re-identification.

The Data Stewards transferred their datasets to the PDA using a Secure File Transfer Protocol (SFTP). SFTP enables secure transfer of data over an encrypted network for a specified period. To test the data transfer process, the Data Stewards first transferred a sample of their datasets to the PDA. When that proved successful, they used the same process to transfer the remaining data.

Once the datasets from each educational partner were received, the PDA connected them using the hashed OENs. As an extra step, the PDA conducted probabilistic matching[6] using the remaining quasi-identifying variables (e.g., gender, calculated age). They then linked postsecondary application data from Ontario's college application service (OCAS), and the Ontario University Application Centre (OUAC) using the same process. The resulting dataset had a 99.6% match rate,[7] meaning we were able to connect school board and

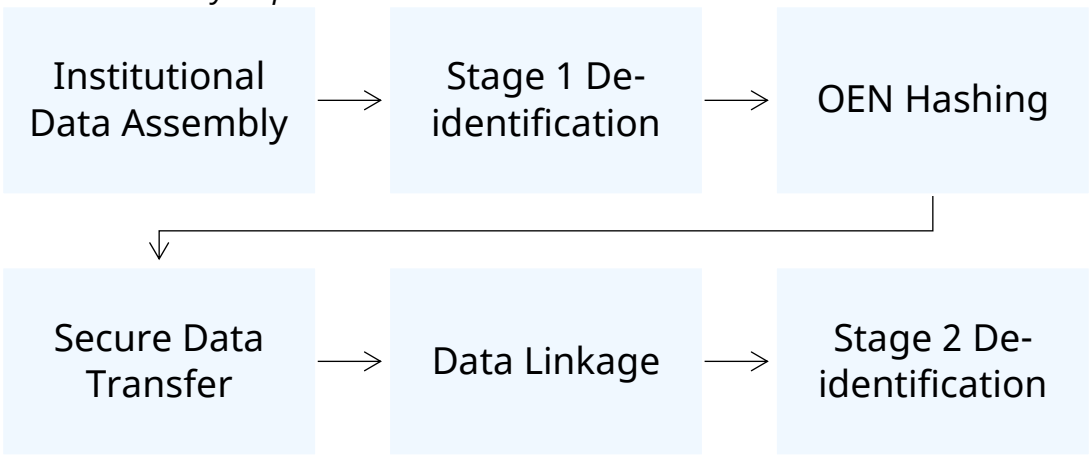
[3] Our process was informed by other models, including British Columbia's Student Transitions Project: a joint initiative between B.C.'s education ministries and postsecondary institutions that links K–12 data to public postsecondary data to better understand student success (British Columbia, n.d.).
[4] Generalization removes the granularity of data; in this case, Data Stewards only included the month and year for students' date of birth and the first three letters of students' postal codes.
[5] The formula was delivered in person by the PDA to individual Data Stewards on a USB. Data Stewards were the only individuals at their respective institutions with access to the hashing formula. As an added layer of security, the Data Stewards applied a pepper (i.e., a randomized value that is added to the hashing formula and kept separately from the formula).
[6] Probabilistic matching is a statistical approach used to measure the probability that two records in a dataset represent the same individual.
[7] Without identifiers such as dates of birth and names, we could not match students who may have an incorrect OEN assigned to their records.

postsecondary data successfully for nearly all students who entered Grade 9 in 2010 in Hamilton and continued their postsecondary education at Mohawk College or McMaster University. The PDA thereafter conducted a second stage of de-identification: first, by removing the hashed OEN and replacing it with a unique identifier to be used for research purposes. They then removed or generalized the remaining quasi-identifying variables. Once the datasets were linked, the hashed OENs and the hashing formula were destroyed.[8]

Since being assembled, the dataset has been stored separately from HWDSB's student information system, behind multiple layers of security, each requiring a key. Per the DSA, the dataset will be maintained by HWDSB for 20 years, or the length of the partnership (whichever is longer). The PDA has sole access to the dataset. If members of the Research Team wish to collaborate on analyses, they can only do so on-site with the PDA, or by providing instruction to the PDA remotely. No other researchers or organizations outside the partnership are allowed access to the dataset during the pilot phase.

Figure 3

CRP Data Assembly Steps



Note: This figure is a visual, flow-chart summary of the CRP's data assembly steps.

[8] The de-identification of the final dataset did not need to be bulletproof because it is kept under lock and key, inaccessible to the internet and accessed by only one individual (the data steward from HWDSB). For anyone familiar with the de-identification process, there are degrees of risk that can be tolerated based on contextual factors such as whether or not the dataset will be made public, the number of people using it and the access controls. We used these De-identification Guidelines for Structured Data released by the Information and Privacy Commissioner of Ontario for the CRP project.

Limitations

Despite a clear legal authority to conduct research for educational purposes, the CRP Leadership Committee decided to take a conservative approach and build trust among each other and stakeholders. They took the steps outlined above to minimize privacy and security risks while maximizing the potential to inspire improvements for Hamilton students. Striking this balance meant that the dataset we produced had significant limitations from a research perspective.

To start with, the pilot phase of the CRP relied on historic data. Our dataset included information about students who entered Grade 9 for the first time during the 2010-11 academic year. This focus limited the partners to an examination of pre-existing and common variables between research partners. It also meant our findings could only be retrospective.

The data is also limited to a 10-year timeframe, which is restrictive; high school graduation rates are conventionally calculated within five years, and postsecondary graduation rates are conventionally calculated within seven years for university and three to seven years for college, depending on the program. Some students who graduated high school after seven years and then entered a four-year postsecondary program would not be shown as having graduated from their college or university programs, even though they may have done so in the years following.

A third limitation involves geography. The CRP dataset allows us to explore postsecondary access only for students who used a centralized application service (OCAS or OUAC) to apply to and confirm an acceptance at a publicly assisted, Ontario-based college or university. We could not account for those students who may have applied to and confirmed offers of acceptance at institutions outside of Ontario (in other provinces or countries). Beyond confirmations, we could only examine the postsecondary outcomes for the students who remained in Hamilton, at either McMaster or Mohawk, for their postsecondary studies.

The fourth important limitation relates to the de-identified dataset, relying on generalization and suppression. Both de-identification techniques safeguard individual privacy, but from a research perspective, they limit the number of variables we can include in our analysis and the level of nuance we can explore. For example, due to the small number of students in our sample identifying as Indigenous, that information had to be suppressed; moreover, the language variable had to be generalized to “English” or “Other” categories. As a result of these privacy protections, the pilot phase of research could not shed light on the pathways of equity-deserving students who identify as

Indigenous or who speak first languages other than English, among other important characteristics.

Lastly, we encountered a limitation from working with application centre data, specifically with the classification of program codes. Statistics Canada offers a classification structure that includes a total of 10 categories, split into Science, Technology, Engineering and Math (or STEM) and Business, Humanities, Health, Arts, Social Science and Education (or BHASE) (Statistics Canada, 2018). This would be an excellent structure for sorting and analyzing data about program choice were it not for one problem: postsecondary program codes in Ontario are not standardized across institutions. Each Ontario postsecondary institution has the discretion to categorize their program codes as they see fit, and there is no key to map these codes to Statistics Canada's Classification of Instructional Program (CIP) codes. HEQCO researchers recoded each program code that appeared in the postsecondary application data files into one of two broad CIP categories: STEM or BHASE. Though labour intensive, this was the best available option. Of course, the binary classification of postsecondary application program codes led to significant data loss, meaning we could explore very little about program choice and the factors relating to it.

Lessons Learned and Next Steps

The Hamilton CRP members joined and built a unique dataset with the potential to better understand the factors that influence access to, persistence through and completion of PSE at a community level. In the process, the partners developed a blueprint that we hope other communities will draw inspiration and direction from. We also gleaned two important overarching lessons for others endeavouring to do this work beyond the lessons for government and institutions articulated in *The Power of Connected Data*.

1. Trust-building is essential

The Hamilton CRP sought to build trusting relationships — with each other and their stakeholders — and establish procedures that would enable ongoing, collaborative educational research in their community. In developing and implementing the principles and processes above, the CRP members developed trusting relationships across their organizations. These will serve as a foundation for ongoing research and improvements to educational policy and practice in Hamilton.

The partners also had to work to build trust amongst stakeholders, including students and their parents — an ongoing pursuit that starts with being open and transparent and demonstrating accountability. In response to parents' concerns about use of personal information, school boards tend to err on the side of caution when interpreting privacy legislation and Information and Privacy Commissioner rulings. This conservative approach

to dealing with student data can inadvertently create barriers to doing robust educational research. Being transparent about the benefits of educational research and the security protocols in place would help to remove these barriers.

The Privacy Acts described above spell out a path for cross-institutional research collaborations designed to enrich our understanding of learning trajectories. When school boards clearly communicate intentions to utilize data for improvement purposes, they can help foster a culture of research within their stakeholder communities that moves with those stakeholders along educational pathways — from K-12 through to college or university (should they choose one of those paths). As a result of their participation in this project, the educational partners in the CRP provided notices of collection that set the stage for a public school to postsecondary continuum of academic research. This represents a positive departure from historic practices. We encourage other educational institutions to make similar strides.

2. Project management and coordination are essential

The CRP was a collaborative effort based on principles of shared decision-making. While all partners were essential to its success, it is worth highlighting the project management function taken on by the co-facilitators of the project. Embedding the CRP in the portfolios of HCF and HEQCO staff was critical to moving the partnership forward. It meant staff at both organizations had dedicated time to support the education partners who were primarily focused on their core mission of educating students (which was particularly challenging during COVID-19). For example, HEQCO and HCF staff completed research ethics board applications; organized meetings of the Leadership Committee and Working Groups; drafted documents to action decisions; and addressed questions and concerns expressed by each partner. We encourage other communities to ensure at least one partner play a similar project management and facilitation role.

We intend that the partnership and data infrastructure established through the CRP pilot can be a foundation for long-term, self-sustaining and collaborative research in the Hamilton community. In the years to come, as CRP members and their stakeholders become increasingly comfortable using connected data, we anticipate future phases of the CRP will be broader in scope. We expect the next iterations of the CRP will include multiple cohorts of students (including the most recent group of students for whom we can examine pathways from Grade 9 through high school and their transition to postsecondary) and include additional and more nuanced variables for analysis. Aligned with Ontario's privacy legislation, open communication with stakeholders about research intentions and outcomes will allow the partners to continue improving our dataset. This

will in turn allow for a more complete understanding of the factors — including barriers and supports — that affect student outcomes.

References

British Columbia, (n.d.). “Student Transitions Project” https://www2.gov.bc.ca/gov/content/education-training/post-secondary-education/data-research/student-transitions-project#:~:text=The%20Student%20Transitions%20Project%20uses,post%2Dsecondary%20education%20and%20graduate

Chatoor, K., Mackay, E., & Hudak, L. (2019). *Parental Education and Postsecondary Attainment: Does the Apple Fall Far From the Tree?* Toronto: Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2020/02/Formatted-Parental-Ed3-1.pdf

Chatoor, K., Courts, R., Han, J., Barclay, V., & Colyar, J. (2022) *Access Programs in Ontario: OPAIP and Pathways to Education*. Toronto: Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2022/01/Access-Programs-in-Ontario_FINAL1.pdf

Education Act. R.S.O. (1990). https://www.ontario.ca/laws/statute/90e02

Freedom of Information and Privacy Act. R.S.O. c.F.31. (1990). https://www.ontario.ca/laws/statute/90f31

Ford, R., Hui, T. S., & Nguyen, C. (2019). *Postsecondary Participation and Household Income*. Toronto: Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2020/02/Formatted-SRDC-PSE-Access-and-Income-1.pdf

Gallagher-Mackay, K. (2017). *Data Infrastructure for Studying Equity of Access to Postsecondary Education in Ontario*. Toronto: Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2020/03/FINAL-Data-Infrastructure.pdf

Information and Privacy Commissioner of Ontario. (2015). *Planning for Success: Privacy Impact Assessment Guide*. Toronto: Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/2015/05/planning-for-success-pia-guide.pdf

Information and Privacy Commissioner of Ontario. (2016). *De-identification Guidelines for Structured Data*. Toronto: Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

Information and Privacy Commissioner of Ontario. (2019). *A Guide to Privacy and Access to Information in Ontario Schools*. Toronto: Information and Privacy Commissioner of

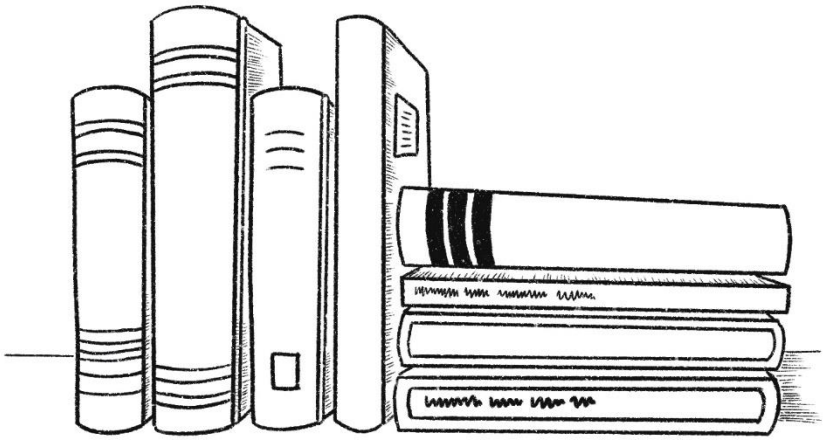
Ontario. https://www.ipc.on.ca/wp-content/uploads/2019/01/guide-to-privacy-access-in-ont-schools.pdf

Municipal Freedom of Information and Protection of Privacy Act. R.S.O. c. M. 56. (1990). https://www.ontario.ca/laws/statute/90m56#BK20

Robson, K., Maier, R., Anisef, P., & Brown, R. S. (2019). *High School Success and Access to Postsecondary Education*. Toronto: Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2020/02/Formatted-CRP-report_FINAL.pdf

Robson, K. (2021). An Essay on the Challenges of Doing Education Research in Canada. *Journal of Applied Social Science, 15*(2),183–196. doi:10.1177/19367244211003471 https://journals.sagepub.com/doi/full/10.1177/19367244211003471

Statistics Canada. (2018). *Variant of CIP 2016 – STEM and BHASE groupings.* Statistical classifications. https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=401856



CRP Blueprint: How We Built a Community Data Infrastructure

Appendices

Appendix A: Legal Authority

The Privacy Working Group found that the legal authority to carry out this project and link student information across institutions is upheld by the following:

1. Section 266.3(3) of the *Education Act* states that "[t]he Minister and a prescribed person or entity may collect, use or disclose or require the production of Ontario education numbers for purposes related to education administration, funding, planning or research." This supports our use of OENs as the primary key to conduct education research in the manner we described.

2. Section 14(1)(e) of MFIPPA and section 21(1)(e) of FIPPA state that an institution can disclose personal information:

> "(e) for a research purpose if,
>
> (i) the disclosure is consistent with the conditions or reasonable expectations of disclosure under which the personal information was provided, collected or obtained,
> (ii) the research purpose for which the disclosure is to be made cannot be reasonably accomplished unless the information is provided in individually identifiable form, and
> (iii) the person who is to receive the record has agreed to comply with the conditions relating to security and confidentiality prescribed by the regulations"

3. Section 10 of Regulation 460, a regulation under FIPPA, sets the requirements for security and confidentiality that must be agreed upon before an institution may disclose personal information for a research purpose. These provisions ensure compliance with clause (iii), above:

> 1. The person shall use the information only for a research purpose set out in the agreement or for which the person has written authorization from the institution.
> 2. The person shall name in the agreement any other persons who will be given access to personal information in a form in which the individual to whom it relates can be identified.
> 3. Before disclosing personal information to other persons under paragraph 2, the person shall enter into an agreement with those persons to ensure that they will not disclose it to any other person.

4. The person shall keep the information in a physically secure location to which access is given only to the person and to the persons given access under paragraph 2.
5. The person shall destroy all individual identifiers in the information by the date specified in the agreement.
6. The person shall not contact any individual to whom personal information relates, directly or indirectly, without the prior written authority of the institution.
7. The person shall ensure that no personal information will be used or disclosed in a form in which the individual to whom it relates can be identified without the written authority of the institution.
8. The person shall notify the institution in writing immediately if the person becomes aware that any of the conditions set out in this section have been breached.

Appendix B: Guiding Principles

Research

1. Conduct research under the common goal of supporting students through their various pathways.
2. Agree collaboratively on all research projects.
3. Conduct research in a way that both integrates and creates a balance between knowledge generation and action for the mutual benefit of all partners.
4. Avoid research questions or projects that focus on specific institutions. Research will not identify individual students or institutions; research will not deliberately seek to shame, embarrass or otherwise damage any organization in the partnership.
5. Acknowledge and be mindful that each partner has its own Research Ethics Board, each of which has its own rules and procedures that must be followed and respected.

Data Governance

6. Agree on and abide by data privacy and protection guidelines that respect the requirements of each partner.
7. Elect a Data Steward for each partner who will be responsible for the data rights and responsibilities of their organization.

Knowledge Mobilization

8. Share findings from our collective research with the broader community, allowing other organizations to learn about how we work and what we now know.
9. Agree upon how and when research results are communicated collectively.
10. Use our research findings to inspire policy change that will improve opportunities and outcomes for our students.

Collaboration

11. Provide an equal role and voice for each partner. Decisions will be made collectively; when context does not allow for collective agreements, decisions will be negotiated around our common goal of understanding student pathways.
12. Use this project as an opportunity to improve how data is collected across the education partners and work together, when possible, to identify and create streamlined solutions.
13. Identify opportunities for students to be involved in the research.