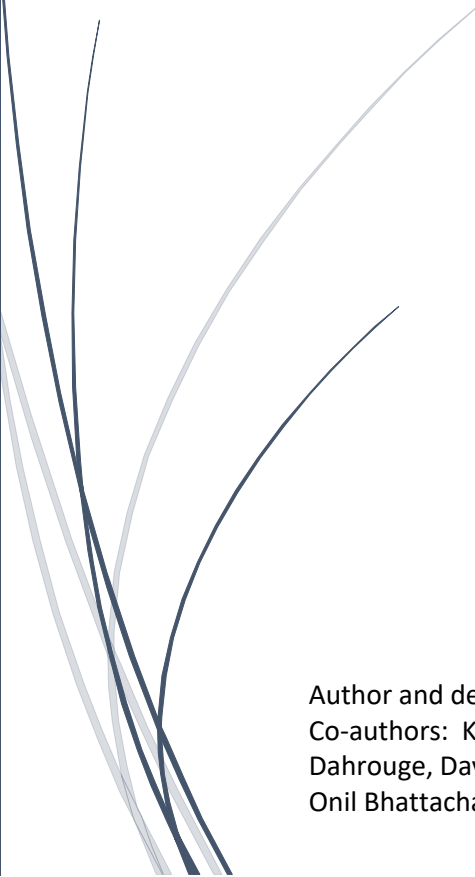# Ontario Primary Care Learning Network: ORACLE

EMR data to support research, quality improvement and innovation

A white paper submitted to the New Digital Research Infrastructure Organization (NDRIO)

Participating Practice Based Learning Networks:
UTOPIAN, University of Toronto
OPEN, University of Ottawa
DELPHI, Western University
EON, Queen's University
NORTHH, Northern Ontario School of Medicine
MUSIC, McMaster University
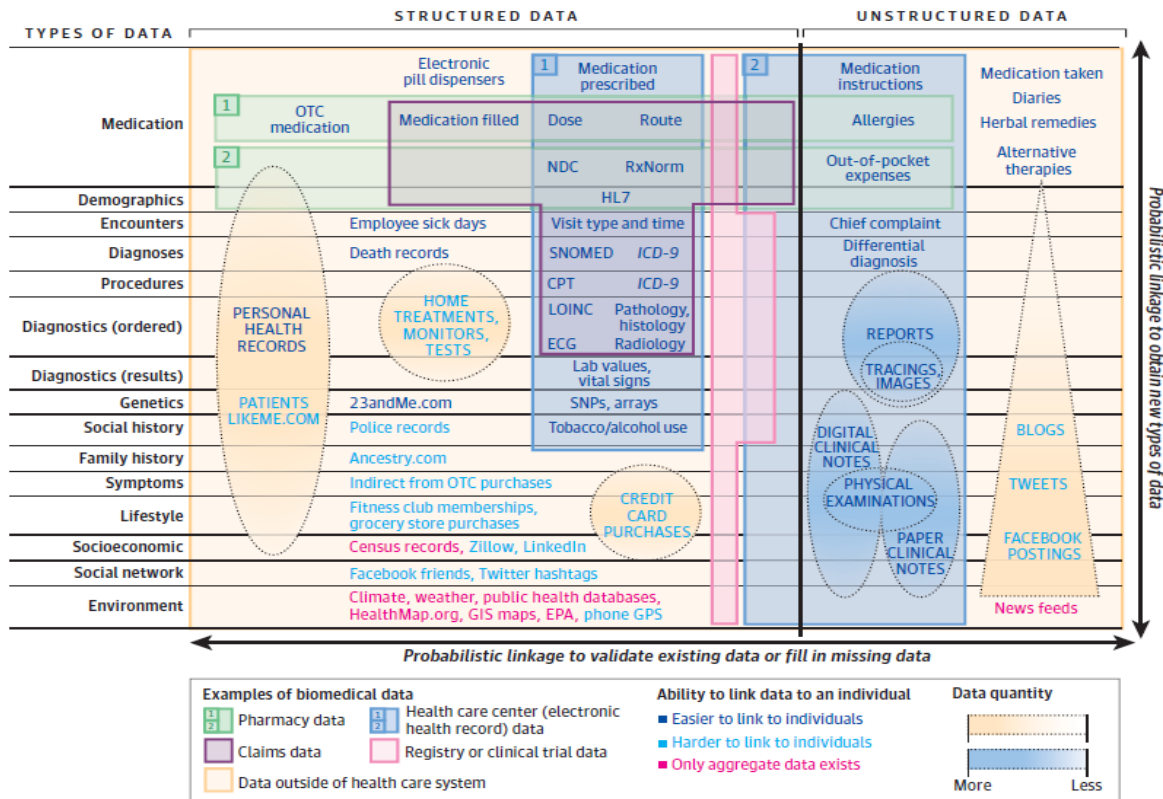EPIC, the Alliance for Healthier Communities

Author and designated contact person: Michelle Greiver (michelle.greiver@nygh.on.ca)
Co-authors: Karen Tu, Andrew Pinto, Amanda Terry, Bridget Ryan, Dee Mangin, Simone Dahrouge, David Barber, Barb Zelek, Brianne Wood, Jennifer Rayner, Michael Green, Onil Bhattacharyya, Joanna King, Eliot Frymire, Ayat Salman

Current Issues

Research has shown that the strongest evidence for improving health outcomes in the population is through strengthening primary care.

Clinical data in electronic health records represent one of the largest sources of Big Biomedical Data. Linked with other data sources, these structured and unstructured data have the potential to provide a nearly complete picture of health and health care processes.

TYPES OF DATA	STRUCTURED DATA				UNSTRUCTURED DATA	
Medication		Electronic pill dispensers	[1] Medication prescribed	[2]	Medication instructions	Medication taken / Diaries
	[1] OTC medication	Medication filled	Dose — Route		Allergies	Herbal remedies
	[2]		NDC — RxNorm		Out-of-pocket expenses	Alternative therapies
Demographics			HL7			
Encounters		Employee sick days	Visit type and time		Chief complaint	
Diagnoses	PERSONAL HEALTH RECORDS	Death records	SNOMED — ICD-9		Differential diagnosis	
Procedures			CPT — ICD-9			
Diagnostics (ordered)		HOME TREATMENTS, MONITORS, TESTS	LOINC — Pathology, histology / ECG — Radiology		REPORTS	
Diagnostics (results)			Lab values, vital signs		TRACINGS, IMAGES	
Genetics	PATIENTS LIKEME.COM	23andMe.com	SNPs, arrays			
Social history		Police records	Tobacco/alcohol use		DIGITAL CLINICAL NOTES	BLOGS
Family history		Ancestry.com			PHYSICAL EXAMINATIONS	
Symptoms		Indirect from OTC purchases				TWEETS
Lifestyle		Fitness club memberships, grocery store purchases	CREDIT CARD PURCHASES		PAPER CLINICAL NOTES	FACEBOOK POSTINGS
Socioeconomic		Census records, Zillow, LinkedIn				
Social network		Facebook friends, Twitter hashtags				
Environment		Climate, weather, public health databases, HealthMap.org, GIS maps, EPA, phone GPS				News feeds

Probabilistic linkage to obtain new types of data (vertical axis, right side)

Probabilistic linkage to validate existing data or fill in missing data (horizontal axis, bottom)

Examples of biomedical data
- [1][2] Pharmacy data
- Claims data
- Data outside of health care system
- [1][2] Health care center (electronic health record) data
- Registry or clinical trial data

Ability to link data to an individual
- ■ Easier to link to individuals
- ■ Harder to link to individuals
- ■ Only aggregate data exists

Data quantity
More — Less

Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA. 2014 Jun 25;311(24):2479-80.

Primary care is the cornerstone of any health care system. As 95% of Ontarians have a regular source of primary care, information in primary care electronic medical records (EMRs) is one of the single largest sources of population-level data and represents a rich source of data in our communities. EMRs contain data on acute symptoms as patients reach out to their providers, information on underlying chronic conditions, mental health, vital signs, social determinants of health and risk factors. These data are updated regularly and can be mined for clinical, quality improvement, planning, and research purposes. Importantly, these data are embedded within provider networks that can use the results of these big data to enhance care.

Practice-Based Learning Networks (PBLNs): Managing, cleaning, and analyzing primary care EMR data is complex. This requires deep clinical knowledge about the context in which the information is recorded and used, coupled with knowledge and expertise in statistical methods and computer science. For more than ten years, EMR data have been managed by PBLNs, which have nurtured relationships between researchers and primary care providers serving their communities. PBLNs have also developed partnerships and connections with experts in Data Science, resulting in productive combinations of clinical knowledge, data science and

relationships with front line clinicians in our communities. PBLNs connect innovators, quality improvement leaders, and researchers experienced in community engagement and participatory action studies. This provides a platform for epidemiological surveillance, clinical research and for data-informed quality improvement relevant to the context of practices and communities.

The data are extracted from the EMRs of consenting primary care providers; we currently extract data from the EMR software of four different Vendors, representing between 85% and 90% of the EMR market share in Ontario. The data are managed within a secure Compute Ontario facility, the Centre for Advanced Computing (CAC) in Kingston, Ontario. We clean, standardize, and merge the data into research ready datasets. After review and approvals, we provide data cuts for analysis in a secure virtual analytic environment at the CAC, along with analytical expertise and knowledge about the data and its context.

A Provincial Effort: Ontario currently has seven PBLNs that extract data; six are affiliated with a University Department/Section of Family Medicine (U of Toronto, Western U, Queen's U, NOSM, McMaster U, U of Ottawa), and one is part of the Alliance for Healthier Communities, which includes Community Health Centres, Nurse Practitioner Led Clinics and Aboriginal Health Access Centres from across the province. Our networks are collaborating to form a single province-wide database as part of **ORACLE**, the Ontario pRimAry Care LEarning Network. ORACLE will **combine and harmonize the processing of primary care EMR data across the province** to allow for standardized and efficient province-wide reporting of indicators. ORACLE will also coordinate efforts to monitor performance across the networks and implement quality improvement initiatives. Currently over 1,000 Ontario family physicians using four different EMR vendors contribute their EMR data on the almost **1.8 million** patients they serve to their respective PBLN. With the support of the Ontario Medical Association and the Ontario College of Family Physicians, we have launched a recruitment strategy to increase our membership base and the representativeness of the data available to ORACLE.

Potential for ORACLE: ORACLE can rapidly scale up province-wide initiatives which have been tried and tested in a particular PBLN.1) EMR data is cleaned, de-identified and standardized to produce reliable provincial and regional data for health system planning. 2) Processed data can be returned and re-identified at the sites of care to identify patients eligible for targeted quality improvement efforts. 3) That data can also be supplemented with census data aggregated at the dissemination area level (a population of 400-700 individuals). This data can help clinicians to understand the social context of their practice and inform clinical decision making. It can also identify patients at higher risk and direct them to appropriate health and social resources. 4) EMR data can be linked to hospital clinical records to study local care trajectories as well as to administrative data at ICES to study disease patterns and evaluate quality improvement efforts. 5) ORACLE can also help coordinate clinical research across Ontario, connecting interested participants to trials and be a platform for health services research.

Within ORACLE, UTOPIAN has recently received approval to extract the whole chart, opening opportunities to use the rich free text found in clinical records through NLP, ML and AI projects. This requires expanded DRI for storage and analytical computing capacity, as well as increased safe and vetted access for data scientists to participate in collaborative projects.

Linking ORACLE to existing structures

ORACLE offers an unparalleled opportunity to use primary care EMR data to proactively monitor health and healthcare as provided across Ontario's communities. **Similar surveillance through primary care clinics has already been deployed across the United Kingdom.**[1, 2] Copies of approved elements from our EMR data are forwarded to Canada's national EMR database, the Canadian Primary Care Sentinel Surveillance Network (CPCSSN), where they are merged with data from other provinces and territories to generate a national-level research-ready dataset. These data, following approvals, are also forwarded to Diabetes Action Canada's

National Diabetes Repository. EMR data from our networks have been combined with national data at CIHI,[3] provincial and regional datasets, including clinical hospital data to support local/regional initiatives such as Ontario Health Teams[4] and administrative data at ICES.[5] They could be combined with larger hospital initiatives, including GEMINI and LKS CHART, genomic data, biobank data, environmental data or data collected through Statistics Canada.

In addition, because ORACLE is a network of seven learning systems, it can link the monitoring system directly to clinical care provision, allowing patients to receive timely care in the community. We have used this data to identify some of the most vulnerable patients in our communities, including those living with multiple health conditions, and have been able to reach out to them for care interventions.[6] The structure can support service planning for integrated health data initiatives such as Ontario Health Teams and future pandemic management.

We are now taking the necessary steps to strengthen the data collection system so that it is **more standardized, extracted more often**, **expanded to more practices, and reported more quickly,** providing provincial monitoring data that is broader and more timely than currently available datasets.

Significant challenges

Challenges include the lack of infrastructure support for the lifecycle of EMR data: its extraction, curation, merging, storage, provision for analysis in high performance computing centres, and linkage with other data sources.

There is a lack of Canadian legislation mandating interoperability and forbidding commercial EMR Vendor activities that block authorized forwarding of data. As a result, some Vendors have imposed financial and administrative barriers to data provision, limiting the availability of EMR data for research, quality improvement and innovations.[7] There are significant limitations on trusted third parties being allowed to provide data extraction services, resulting in monopolistic extraction practices by some of the largest vendors. By comparison, in the US, the 21st Century Cures Act and the Final Rule from the ONC has mandated interoperability, enacted requirements for standards-based APIs, forbidden data blocking activities and processes and has imposed significant financial penalties for lack of compliance.[8] This has resulted in innovations and expansion of computational health activities that currently outpace those in Canada.

Processes and procedures to set up data linkage with other datasets have been time consuming and resource intensive, often taking years to achieve. Nimbler processes and procedures for linkages and Data Sharing Agreements are needed; one example being Data Trusts mentioned in several other submissions. Additionally, understanding and cleaning these complex primary care EMR data into research-ready datasets is a non-trivial activity; managing, cleaning, and standardizing EMR data so that they are transformed into a common internationally relevant model such as OHDSI's OMOP[9] should be done once by a team combining primary care clinical expertise with data science expertise. ORACLE provides such a team for Ontario. To reflect this and ensure efficiencies and standardization of data transformation, we propose extending COUMT to **C-COUMT** (Collect **and Clean** Once, Use Many Times). Agreements on C-COUMT are required.

Finally, sufficient IT resources are required for the storage and analysis of data, given the significant expansion to full chart, additional practices, and planned more frequent extractions. Collaborations with High Performance Computing Centres should include the management of security and provision of safe access to vetted scientists following the Five Safes framework (Safe Projects, Safe People, Safe Settings, Safe Data, Safe Outputs).[10]

These challenges can all be addressed. Similar solutions, extending to other sectors requiring health related DRI, are presented in NDRIO white papers led by Sabrina Wong (CPCSSN), Alison Paprica (HDRN Canada), Gary Lewis (Diabetes Action Canada), Francine Lemire (CFPC) and Laura Rosella (Data Science Platform).

Future DRI State

Our vision for a cohesive Canadian DRI ecosystem rests in collaborations and partnerships across sectors. ORACLE's Networks already collaborate with Diabetes Action Canada (data from our Networks and sister PBLNs in other provinces have formed the National Diabetes Repository), CPCSSN (Ontario constitutes over 40% of national data), ICES and others. We envision partnership in the nascent Canada Health Data Research Alliance to enable broader provision of information on EMR data holdings to the research community, in partnership with CPCSSN. Beyond data, we also advocate addressing the very large gap in Canada's clinical research productivity and inefficient trial infrastructure;[11] the UK's highly effective Clinical Research Network infrastructure has recently demonstrated that a better system can be achieved. Our PBLNs are ideally positioned for this new model, leveraging data and integrating research into primary care practices through ongoing connections between researchers and front-line clinicians.

ORACLE's members are embedded in their communities and have extensive involvement with organizations representing Family Physicians and Primary Health Care, including the Ontario Medical Association's Section of General and Family Practice, the Ontario College of Family Physicians, the Alliance for Healthier Communities, and the College of Family Physicians of Canada. In addition, we are embedded in Academic Departments / Sections of Family Medicine in every Ontario School of Medicine.   Several ORACLE leads are members of the Temerty Centre for Artificial Intelligence in Research and Education in Medicine (T-CAIREM). We have collaborations with the Vector Institute for Artificial Intelligence. ORACLE members lead CIHR funded national and provincial Primary and Integrated Care Innovations Networks.

ORACLE is part of the Ontario SPOR Support Network funded INSPIRE-PHC program, which already includes a provincial advisory committee. A stakeholder group meets twice yearly, and includes patients/citizens on the management team, provincial professional organizations (Ontario College of Family Physicians, Ontario Medical Association), PHC service provider organizations (Association of Ontario Family Health Teams of Ontario, Association of Ontario Health Centres), Ministry of Health representatives, and Health Systems Leaders.

Patient Engagement is a key aspect of PBLNs; ORACLE's members already have extensive experience and expertise in this area. Patient representatives from every network will be on ORACLE's provincial advisory council.

While we represent the largest provincial primary care EMR data system in Canada, our vision for a DRI ecosystem extends beyond primary care, to collaborations with data providers and experts in other sectors, including Genomics, environmental health, public health, administrative data centres, and acute care. The potential collaborations extend beyond health care, to other sources of data including social media and private sector data held by large technology companies, provided there is appropriate oversight and engagement of people whose data these are.

How to Bridge the Gap

Significant value can be gained through investment in areas that have traditionally lacked DRI funding in Canada, and where the gap with international comparators is particularly large. Investment and activities should focus on new DRI with international evidence of value and productivity, broad community-based value, potential for

population-based coverage and a high likelihood of success in making significant contributions to Canada's overall DRI ecosystem. Infrastructure support for Canada's primary care EMR data system represents remarkably high DRI value;[12] this can be a game changer for our patients and populations.

Our immediate actions for an enhanced DRI to support Ontario primary care EMR data in 2021 include:

1. **Establishing an Ontario EMR based dataset that is linked to existing data holdings.**

We will link EMR data to administrative data at ICES, to enhance existing administrative data. Four ORACLE networks linked EMR data at ICES in 2015 as a project. The Alliance has a linkage with ICES as a controlled use dataset; UTOPIAN signed a Data Sharing Agreement with ICES in April 2020, based on the Alliance document. The technical aspects of linkage are now underway and will be expanded to ORACLE.

2. **Enhancing and standardizing data management provincially.**

We will centralize the data management and enhance its efficiency to produce a robust data hub that will maximize synergies and improve economies of scale. This will improve the timeliness of reporting and enhance the ability to generate policy relevant reports using EMR data. We will work with privacy experts and legal consultants to ensure the data hub is aligned with all relevant privacy legislation and the highest standards of data protection.

3. **Expanding the number of participating practices.**

We have already partnered with the Ontario College of Family Physicians and the Section of General and Family Practice of the Ontario Medical Association to support us in the recruitment of community practices that do not have an affiliation with universities. We will be able to include more than 2 million patients in Ontario within a year.

4. **Negotiating cheaper and faster access to EMR data with vendors.**

A major limitation to frequent extraction is the high cost imposed by EMR vendors. Issues of access and data ownership should be addressed in collaboration with governments and other organizations with a mandate to improve Canada's DRI.

Suggested strategic directions for NDRIO

The pandemic has focused attention on important gaps in the availability of comprehensive health data, on health inequities resulting from gaps in our ability to measure, report and address disparities, and on the high societal value of timely and safe access to data and analysis. Canada currently lags behind other nations; ensure that health, social determinants of health, and health data are strategic priorities for DRI investment and resource allocation.

Promote and support DRI activities that:[13] design-in interoperability; design-out approaches that restrict sharing, cooperation, flexibility and mutuality; establish policies and procedures that adequately protect the interests of participants; ensure the longevity of the system through carefully managed access policies and stewardship of resources; and ensure that governance policies and mechanisms remain fit for purpose over time.

ORACLE White Paper Submission to NDRIO

References

1.	RCGP RSC Workload Observatory. Accessed August 6, 2020. https://clininf.eu/index.php/rcgprscworkloadobservatory/

2.	Mannan F, Chaudhry Z, Gibson-White A, et al. Outputs and growth of primary care databases in the United Kingdom: bibliometric analysis. *BMJ Health & Care Informatics*. 2017;24(3):284. doi:10.14236/jhi.v24i3.942

3.	Rayner J, Khan T, Chan C, Wu C. Illustrating the patient journey through the care continuum: Leveraging structured primary care electronic medical record (EMR) data in Ontario, Canada using chronic obstructive pulmonary disease as a case study. *Int J Med Inform*. Aug 2020;140:104159. doi:10.1016/j.ijmedinf.2020.104159

4.	An J, Keshavjee K, Mirza K, Vassanji K, Greiver M. Bridging a divide: architecture for a joint hospital-primary care data warehouse. *Stud Health Technol Inform*. 2015;208:45-9.

5.	Birtwhistle R, Green ME, Frymire E, et al. Hospital admission rates and emergency department use in relation to glycated hemoglobin in people with diabetes mellitus: a linkage study using electronic medical record and administrative data in Ontario. *CMAJ Open*. Jul 11 2017;5(3):E557-e564. doi:10.9778/cmajo.20170017

6.	Greiver M, Dahrouge S, O'Brien P, et al. Improving care for elderly patients living with polypharmacy: protocol for a pragmatic cluster randomized trial in community-based primary care practices in Canada. *Implement Sci*. Jun 6 2019;14(1):55. doi:10.1186/s13012-019-0904-4

7.	College of Family Physicians of Canada Position Statement: supporting access to data in electronic medical records for quality improvement and research. The College of Family Physicians of Canada. Accessed December 13, 2020. https://www.cfpc.ca/en/policy-innovation/health-policy-goverment-relations/cfpc-policy-papers-position-statements/position-statement-supporting-access-to-data-in-el

8.	SMART Health IT. Computational Health Informatics Program, Boston Children's Hospital. Accessed December 13, 2020. https://smarthealthit.org/

9.	Observational Health Data Sciences and Informatics (OHDSI), OMOP Common Data Model. Accessed December 13, 2020. https://www.ohdsi.org/data-standardization/the-common-data-model/

10.	The Five Safes. Accessed December 13, 2020. http://www.fivesafes.org/

11.	Lamontagne F, Rowan KM, Guyatt G. Integrating research into clinical practice: challenges and solutions for Canada. *CMAJ*. Dec 2020;doi:10.1503/cmaj.202397

12.	Katz A. Opportunity beckons for electronic medical record data. *Can Fam Physician*. Aug 2020;66(8):559-560.

13.	Laurie G. Reflexive governance in biobanking: on the value of policy led approaches and the need to recognise the limits of law. *Hum Genet*. Sep 2011;130(3):347-56. doi:10.1007/s00439-011-1066-x