Water Management and Big Data Analytics:

Examination of Opportunities and Approaches to Leverage Data Science, Analytics and AI to Support Watershed Planning and the Health of our Great Lakes' Ecosystem

JUNE 2019

POLLUTION PROBE
CLEAN AIR. CLEAN WATER.

CGLR
COUNCIL OF THE GREAT LAKES REGION
BUILDING OUR FUTURE TODAY
Discover. Connect. Influence.

Prepared with support from RBC Foundation

Table of Contents

This project is funded with support from RBC Foundation.

LIST OF ABBREVIATIONS

AI – Artificial Intelligence

AOC – Areas of Concern

CDO – Chief Data Office

COA – Canada-Ontario Agreement

ECCC – Environment and Climate Change Canada

IoT – Internet of Things

LAMPs – Lakewide Action Management Plans

NGOs – Non-government organizations

TRCA – Toronto Region Conservation Authority

EXECUTIVE SUMMARY

This report presents highlights and recommendations from the **Watershed Planning and Big Data workshop,** which took place on February 26, 2019 in Cobourg, Ontario. Stakeholders from the north shore of Lake Ontario gathered to discuss how to better leverage latest innovations in the areas of IoT, data analytics, modelling and artificial intelligence (AI) to enhance real-time monitoring, predictive capability and scenario planning to ask “what-if” questions that would assist policy-and decision-makers to better manage and protect our common water resources, including (but not limited to) our Great Lakes.

Structure

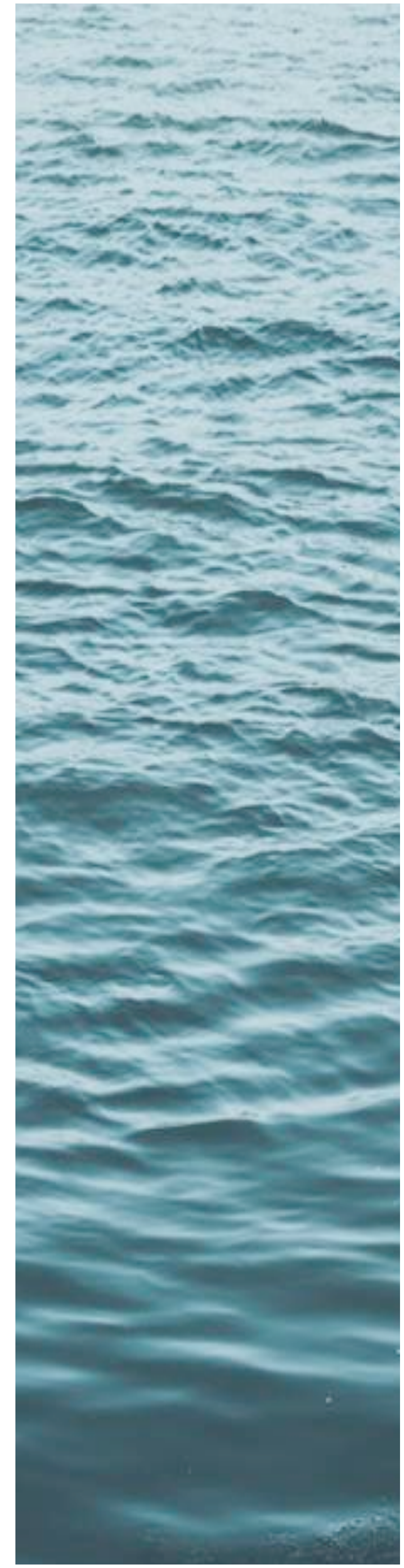
The workshop was hosted by Pollution Probe and the Council of the Great Lakes Region and was funded by RBC Foundation.

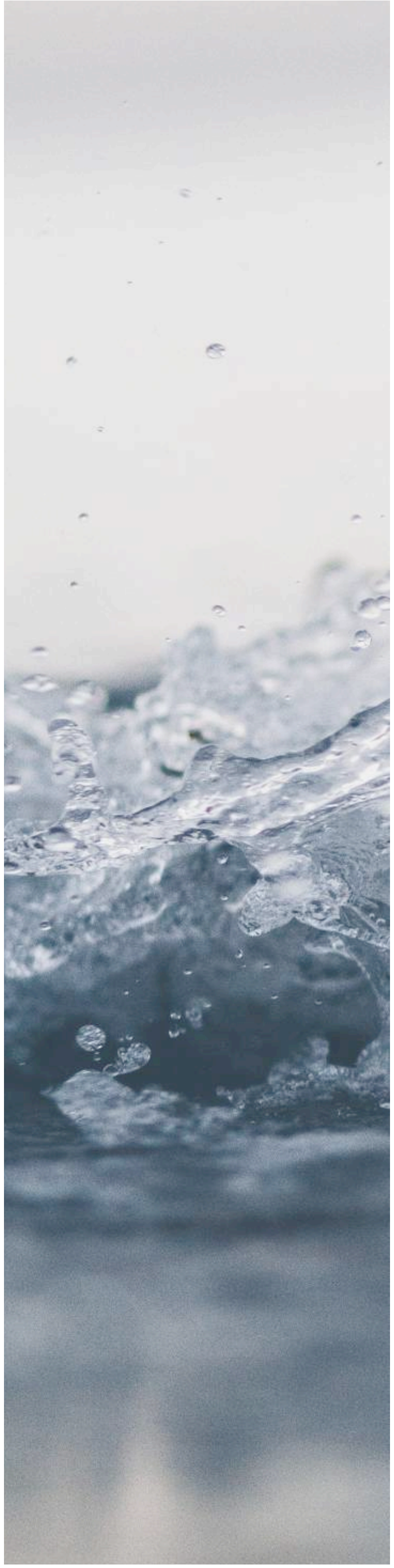
The workshop convened over 50 leading watershed planning experts and data specialists from different levels of government, conservation authorities, not-for-profit organizations, industry and academia. The format of the day involved a series of speaker presentations in the morning and open dialogue breakout and forum sessions in the afternoon to facilitate collaboration and idea sharing between the data and watershed planning experts.

Discussion

Some key themes that were explored were data gaps and quality concerns, applicability of data, variability of data science capability, monitoring techniques and technologies, access and ownership of data, and data context.

While a broad range of water asset and resource management issues that could be improved by using big data tools and approaches were discussed, 3 topics surfaced as being of priority interest and relevance to the room.





These areas were consistently raised and discussed in context of pinpointing the greatest opportunities (and need) for improved data and data science support. The following is a summary of the recommendations that that were formulated from the workshop:

1. The application of road salt and its impact on nearshore chloride levels in the surrounding ecosystems;
2. the impacts of storm water on water quality, based on data such as incidence of extreme weather events;
3. the impacts of land-based contaminants on native species in the lakes.
4. Nutrient runoff (with emphasis on phosphorus)

These 4 topics were consistently raised and discussed in context of pinpointing the greatest opportunities (and need) for improved data and data science support.

Recommendations

The following is a summary of the recommendations that that were formulated from the workshop:

1. **Demonstration Project**: A collaborative multi-stakeholder demonstration project should be established to create a living, real world example of how best to apply modern tools and techniques from data science, based on a manageable scope, common interests/needs and a shared desire to build a tool – or set of tools – relevant to a common concern.

2. **Build on Existing Collaborations**: The demonstration project should build 'on the shoulders' of existing collaborative relationships that complement the proposed demonstration project's objectives instead of attempting to establish a completely new collaboration.

3. **Data Science Steering Group**: A data science/technology-literate steering group should be formed to conduct an in-depth review and audit of any water quality data-related projects and platforms already in development and use these to evaluate how to serve and/or feed into the development of a platform for better future collaboration.

4. **Focused Dataset Audit**: The dataset audit should initially be focused on the proposed pilot project content. Rather than asking stakeholders to provide 'all of their data', the proposed collaboration would be to secure data sharing centred on the specific areas of focus.

5. **Topic of Demonstration Project**: The demonstration project could focus on one of the relevant and somewhat interrelated topics that are top-of-mind for water quality and resource decision makers that could be explored from a number of levels.

6. **Phased Approach**: The demonstration project should be laid out as a phased, actionable, clearly outlined approach for how best to engineer this collaboration toward the eventual development of a predictive capability to generate insights and explore "what-if" scenarios based on machine learning combined with IoT enabled sensor data, physical model forecasting, remote sensing, analytics and visualization techniques that – combined with analytical and visualization techniques – can directly serve water resource decisions makers seeking better, more valuable input into regulation, policy and specific water- and land-use permissions impacting upon Ontario's water resources and water infrastructure.

INTRODUCTION

Historic monitoring and on-going data collection touching all facets of water quality, flows, aquatic life, as well as water asset and infrastructure performance, amass to comprise a vast store of information and data relevant to the lakewide management and the specific tasks of managing watersheds and water resources. However, many of these data assets are unutilized, underutilized or being used in isolation of other relevant datasets.

As part of this emerging understanding and capability to deploy and manage data ecosystems, new challenges and opportunities are emerging. These include questions around data resolution (frequency, density and geography), data ownership vs. access, tools and means to evaluate performance/veracity between data and metadata assets and approaches, perpetual, real-time data quality assurance and quality control, and computing capacity to drive high resolution physical modeling forecasts and to effectively process and analyze datastreams and data assets within timeframes to maximize their value to science, planning and decision-making. Furthermore, modern capabilities of 'internet of things' (IoT) networks combined with sophisticated sensing technologies, advanced machine learning, physical modelling, and related innovations offer an unprecedented ability to understand, visualize and predict how water ecosystems are responding to various factors and stimuli.

"We're all going to this open data portal world – we pretty much have to."

– Pam Lancaster, Ganaraska Region Conservation Authority

The opportunity before us is two-fold:

- **Capitalizing upon the data assets we already have** (by way of current day data analysis and management approaches and techniques) to assess, understand and visualize historic patterns and complex relationships among different factors affecting water resources; and

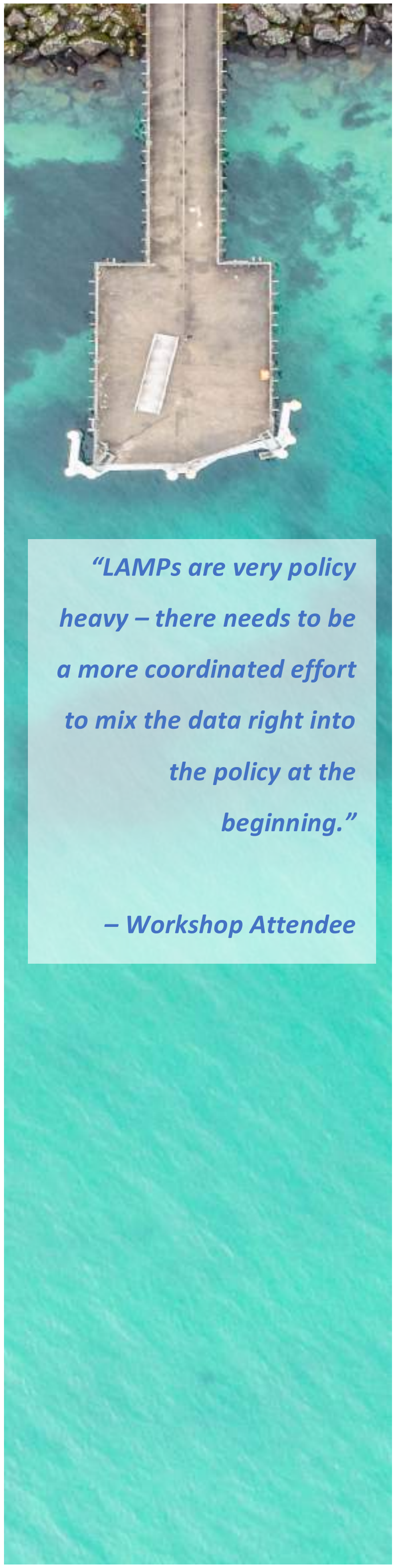
- **Leveraging this historic data** – hand in hand with strategically selecting new monitoring and data collection priorities – using IoT, physical modelling, analytics and machine learning to begin to formulate predictive capabilities, by developing complex models that will ultimately be able to accommodate a combination of environmental, climactic and socio-economic factors.

The primary benefit of this work would be fairly evident: providing new capabilities to support planning decision-making within communities and across regions, by regulators, and by water asset/infrastructure managers, often within budget-constrained conditions.

Secondary and tertiary benefits undoubtedly also exist, many of which may be identified from more in-depth collaborations and discussions around the needs and opportunities in/across water and community planning-related sectors.

Overall, the outcome of these efforts should support the optimization of water resource protection and management; improved capability to manage water infrastructure in context of both environmental sensitivities (and economic needs/constraints); and a more resilient and predictive understanding of how future challenges – such as climactic shifts, economic growth, and varying socioeconomic factors – interrelate and will impact on our sensitive water ecosystems.

This comprises the foundational context for efforts to better assess opportunity and improve how we leverage the data-related assets and resources that underpin best practices of water resources and assets.



"LAMPs are very policy heavy – there needs to be a more coordinated effort to mix the data right into the policy at the beginning."

– Workshop Attendee

OBJECTIVES

Protecting our Great Lakes and its watersheds/subwatersheds is a priority for governments of all levels, agencies, a legion of stakeholders including First Nations, businesses, NGOs, foundations, and – ultimately – individual citizens. This interest and commitment is clearly evidenced in the wealth of policies, regulation and framework agreements that shape and govern bilateral, multiparty efforts to restore and maintain the physical, chemical and biological integrity of the waters of the Great Lakes.

This framework of policies, agreements and stakeholder relationships is – in its own right – a complex ecosystem that requires support, with both traditional funding, partnerships with private sector and resources as well as fresh insights; supported by latest developments and technologies in data science, project management, information sharing and scientific monitoring.

Central to this effort are the cooperative mechanisms set out under the Canada-Ontario Agreement (COA), which asks parties centred around each of the individual Great Lakes to input, support, and deliver individual actions collected in the form of a 5-year jointly agreed collaborative framework called "Lakewide Action and Management Plans" (LAMPs). These LAMPS have been in operation since 1987 and have enjoyed a measure of success in remediating a number of Areas of Concern (AOCs) as well as instituting measures that have protected individual lakes from a number of on-going threats, such as wetland protection and invasive species. COA is a key mechanism for cooperation between the Ontario Provincial and Federal governments to jointly work toward LAMP objectives.

Contextually, LAMPs have provided a useful model for collaboration that can be examined, both from a continuous improvement of governance and processes, and also for opportunities to provide better and more useful inputs. Two obvious and vital elements of this input are the focus of this report:

1. How watershed and subwatershed management practices, microscosms of larger lake systems, can be improved and integrated to more effectively understand and manage important land-lake connections and collaborations to protect the health of these systems; and

2. How modern practices of data science, including analytics, updated monitoring technologies, physical modelling, and artificial intelligence (AI), can provide enhanced predictive capability and insight to assist policy- and decision-makers to better manage and protect our common water resources, including (but not limited to) our Great Lakes.

From the foundational ‘starting point’ of Ontario watersheds, this project sought to assess the current data environment and context surrounding management of water resources in Ontario, for the purpose of developing practical recommendations that can lead to concrete tools and practices.

> *"For the larger collaboration there's a stepped process, let each grouping determine what that is, to feed into the larger data world which is totally organic – not easily controlled"*
>
> *– Pam Lancaster, Ganaraska Region Conservation Authority*

WATER-RELATED DATA ASSETS, DATASTREAMS, AND METADATA

To be able to find improved data-enabled tools and approaches for water resource management, first a more complete and holistic understanding of existing data assets, on-going monitoring and data collection activities, and related metadata must be determined. Efforts to audit and clarify what is already established, collected and underway will yield insights into both the possible value that can be mined or extrapolated and data gaps as a first result.

Additional work to evaluate the quality and veracity of these datasets/datastreams (including temporal and geographic data resolution, reliability, 'ownership' and accessibility) will also lead to a more sophisticated understanding of how these resources can be effectively leveraged.

Finally, this work will make it possible to develop our data resources to the next level of sophistication: How can we make best use of emerging opportunities such as citizen science, remote sensing and satellite imagery to supplement and/or replace traditional methods of data collection? How do we ensure we have properly validated collected data from both traditional (including indigenous knowledge) and modern sources? What further data needs to be collected, analyzed and leveraged to maximize our ability to fully understand, visualize, model and predict the ways in which our water systems and ecosystems can best be protected and managed in context of on-going and future stressors?

BIG DATA, ANALYTICS AND MACHINE LEARNING: GUIDING QUESTIONS

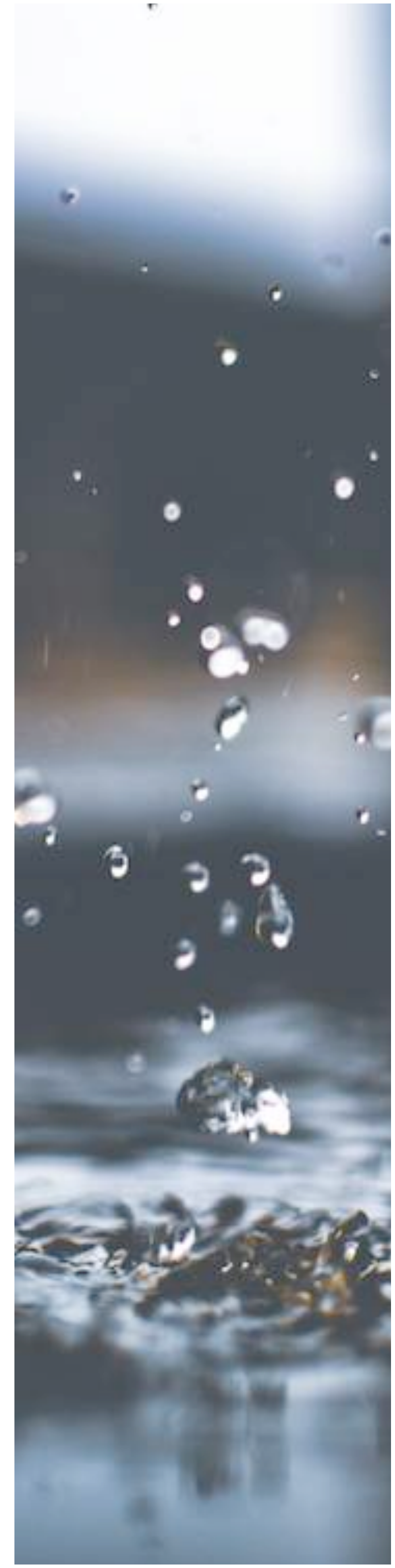
As the workshop was developed, a number of core guiding questions were identified to use in framing the presentations and discussion. It was acknowledged that not all of these questions would be directly answered, but that that they should anchor our discussion and present medium- and longer-term goals as the work evolves:

1. How can we determine what relevant data assets are already amassed, being collected, and analyzed by different levels of government, agencies and organizations surrounding our watersheds, water infrastructure, and larger lakes and lake ecosystems?

2. What additional data can and should be collected to most effectively broaden and extend our ability to achieve desired outcomes?

3. What tools and approaches need to be brought to bear in the exercise of measuring performance of varying models and approaches as effectively and simply as possible, against a backdrop of an exponentially increasing data ecosystem?

4. What existing projects, models, approaches and technologies are already being effectively deployed related to our desired outcomes? Which of these should be studied in order to accelerate our desired advancements in water resource protection and management?



5. What are the most pressing needs and priorities toward which we should be building our capability for more sophisticated uses of data, analytics and machine learning (including but not limited to LAMPs, water asset management and community growth planning and decision-making)?

6. Who are the stakeholders that need to be consulted and incorporated into the future discussions and work as part of these efforts?

7. Which are the most promising areas of academic review and experimentation underway that might be most likely to have direct relevance to current and future priorities?

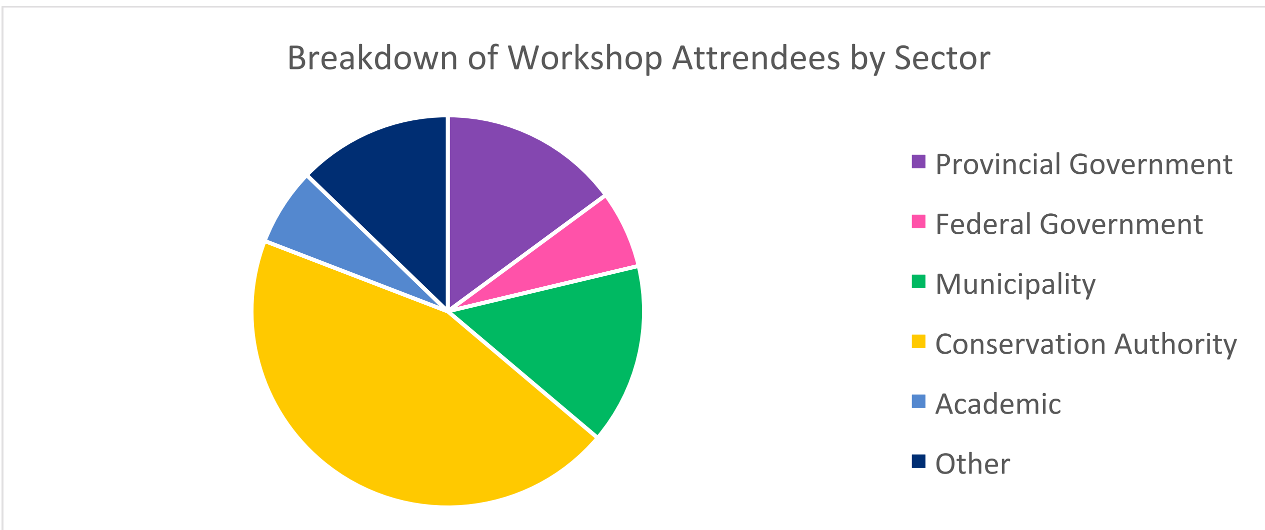
8. How can/should this work interrelate and establish common frameworks with other, similar and related efforts to leverage insights from Big Data in context of our watersheds, communities, and the Great Lakes?

WORKSHOP STRUCTURE

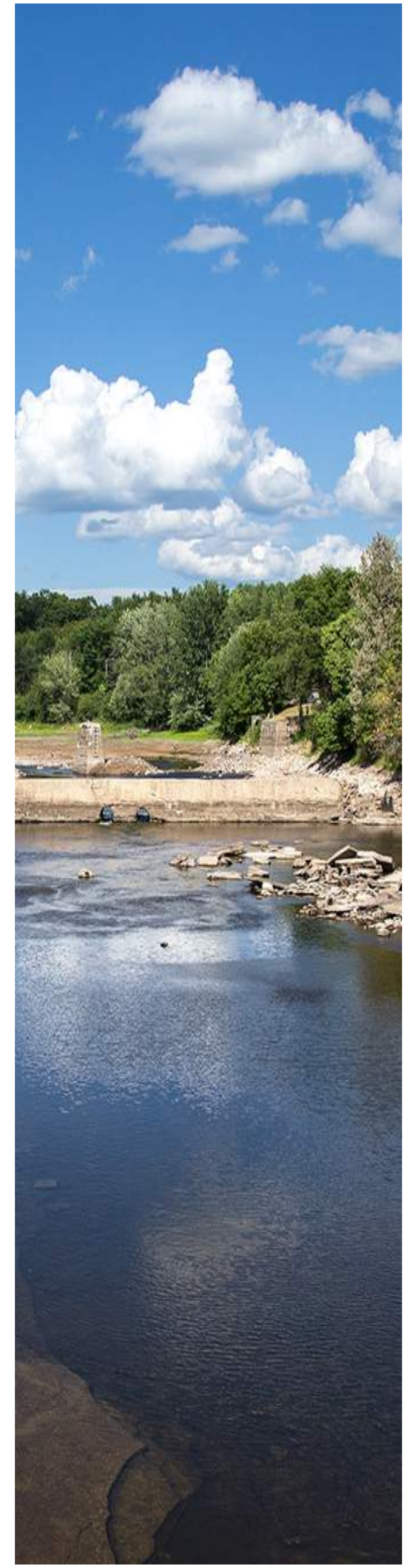
The workshop was held on February 26, 2019, at the Best Western Plus in Cobourg, Ontario. Cobourg was chosen as the location due to its proximity to the two watershed case studies – Carruthers Creek in Toronto Region Conservation Authority (TRCA) and the Moira River in Quinte Conservation Authority.

The breakdown of attendees by sector is as follows:

Figure 1: Breakdown of workshop attendees by sector



Source: Watershed Planning and Big Data workshop



The workshop was ultimately focused on eliciting valuable outputs: priming attendees to think creatively and collaboratively during ideation sessions that were held end-of-day.

The morning and lunchtime sessions were essentially briefing sessions; geared to provide important background, context, and an inspiring example (the Jefferson Project at Lake George, NY) supporting a 'level set' to align perspectives among the participants drawn from varied backgrounds and water- and data-related roles.

To lead in to the ideation phase of the day, the facilitator led a general discussion session focused on modelling and visualizing change in the Great Lakes basin watersheds. The session encouraged all participants to discuss existing data assets that are or are not being leveraged on water and non-water sides to support watershed planning. This set the context and oriented the group towards the questions and issues around the use of big data in watershed planning.

The breakouts (populated by representatives from different backgrounds and roles) were overseen by leading data science experts in the room (Kaveh from ECCC Data, Trevor Boston from Greenland Consulting).

WORKSHOP BREAKDOWN

The following summaries of the day's presentations and sessions:

- **Protecting the waters of the Great Lakes:** An introduction to the Great Lake Water Quality Agreement and the development of Lakewide Action and Management Plans (LAMPs) with a focus on Lake Ontario.

- **The Water's Edge:** A spotlight on 'real life' examples of watershed management planning in the Carruthers Creek and Moira River Watersheds; and how integrated, multidisciplinary planning supported by current and desired capability to assess and manage the respective monitoring datasets can improve the health of these watersheds and the effectiveness of the Lake Ontario LAMP.

- **Lessons from the Jefferson Project at Lake George in New York State:** An introduction to high-level opportunities created by coupling IoT, high resolution physical modeling, predictive analytics/AI, and adaptive decision-making, and the opportunity potential of bringing this kind of tool into water resourcing planning and protection.

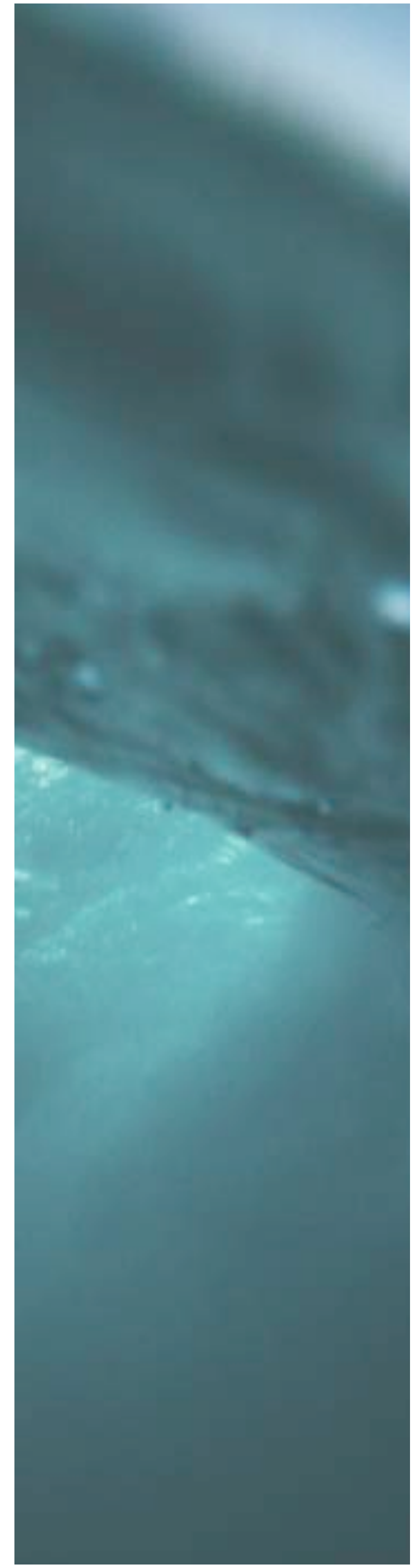
- **Introduction to Data Science for Public Policy Making:** An introduction to the work of ECCC's Chief Data Office (CDO) and how they enhance and develop data science and analytics at ECCC.

- **Discovery Session 1 – Watershed Planning and the Lake Ontario LAMP:** Attendees discussed the following:

    - Process Mapping – Why? What? How? When?

    - Gap Analysis – Strengths? Weaknesses? Lessons Learned?

    - Opportunities – Best Practices? Process Modifications?



- **Discovery Session 2 – Modelling and Visualizing Change**: Attendees discussed the following:
    - What watershed and lake data is collected now?
    - What data (e.g. socio-economic) is collected but not analyzed in watershed planning? Why? Who owns/manages the data?
    - What data is not being collected but should be?
    - How is available data analyzed? Integrated? And by whom?
- **Ideation breakouts**: Breakout Groups mixing both watershed planning, data analytics and municipal leads into five groups. Ideas and concepts were further developed and were then brought to an ideation plenary.
- **Design Thinking Plenary Session**: A free-flow facilitated session to fully explore stakeholder perspectives on what steps could be done to:
    - Improve the development of LAMPS, such as incorporating socio-economic factors into the planning and scientific process, using the Canadian side of Lake Ontario as the test case.
    - Leverage data, data science, and machine learning to deepen insights about watershed trends, visualize the information and changes, and enhance decision-making and future planning.

IDEATION PROCESS

For the breakout sessions, the planning team designed a bottom-up process that facilitated self-selection of problems, open collaboration and participative generation of ideas.

Five breakout groups were convened mixing watershed planning, data analytics and municipal leads in each group. Each group included a range of representatives from various roles (federal, provincial, municipal, conservation authorities) and multi-disciplinary backgrounds, bringing a broad range of knowledge of available data sets into the discussions.

The groups were tasked with selecting a priority water quality or resource management issue that could be more effectively addressed by using Big Data tools and approaches. Group discussions focused on identifying the nature of the problem and specific questions that could be explored, relevant data assets that already exist, and additional data, models and approaches that could be leveraged to achieve the desired outcome.

Following one hour of discussion and ideation, the groups had a few minutes to share their ideas, concepts and highlights of the conversation at the ideation plenary.

A number of common themes have emerged from the breakout sessions, which were identified through pattern searching.

THEMES AND INSIGHTS

A number of themes and topics were highlighted in the discussions held, stemming from the ideation break-out groups and then fleshed out in the plenary discussion:

Priority Topics for Possible Focus: While a broad range of water asset and resource management issues that could be improved by using big data tools and approaches emerged, four broad scope topics surfaced as of interest and relevance to the room:

1. The application of road salt and its impact on nearshore chloride levels in the surrounding ecosystems;
2. the impacts of storm water on water quality, based on data such as incidence of extreme weather events;
3. the impacts of land-based contaminants on native species in the lakes.
4. Nutrient runoff (with emphasis on phosphorus)

These 4 areas were consistently raised and discussed in context of pinpointing the greatest opportunities (and need) for improved data and data science support. Further detail on this discussion is summarized in objective four (4) on page 24.

Data gaps and quality concerns: It was generally agreed that a great range of relevant data sets and sources/platforms already exist, but these are poorly catalogued, disconnected, uncurated and piecemeal. Specific quality issues and concerns around existing data were identified:

"It's definitely a situation of herding cats, some data sets are very difficult to corral together because everyone is collecting them in their own way for their own purpose."

– Trevor Boston, Greenland Consulting

- ***Incompatible or non-existent datasets***: In a number of areas, datasets simply do not exist – or are not known to exist – to help address priority questions or concerns. In some cases, data MAY have been collected, but be in formats (print, microfiche) that are unwieldy and largely inaccessible to be analyzed, given resourcing levels.

  - An example of this is the absence of municipal/lake-wide data on salt levels, including data relating to how much salt used in a specific municipality contributed to the overall salt concentrations in the lake. This inhibits ability to understand the potential impacts on ecosystems and species, including cumulative impacts.

- ***Insufficient temporal and spatial resolution*** (data not sampled frequently enough or with insufficient geographic density, positioning and distribution) inhibit the utility of existing data sets. Specifically mentioned in this discussion:

  - Monitoring of fish populations is inconsistent and inadequate for the purpose of detecting effects of water quality stressors on fish that can be anticipated during sensitive times such as spawning.

  - Water quality information is often not measured with sufficient frequency given that storm water discharges are very episodic. Higher resolution/frequency measurement is required to assess what happens in mixing areas known to be most likely affected by these storm water discharges.

- ***Insufficient historical data*** such as the datasets required to understand the causes of Atlantic salmon loss. While some relevant data exist (such as tree loss and overfishing) there are gaps in the data needed to understand the history and biology of the species and other historical causes of species loss. (It was noted, however, that understanding these causes would require data that would have been collected for over 100 years.) Notwithstanding this, it was noted that traditional indigenous knowledge and oral histories could be a valuable source of supplementary information about past conditions in the watershed.

"For example, with Lake Ontario, there's a way of centrally predicting how it's being affected by what's coming out of our watersheds, based on this kind of information and monitoring, then doing a mass balance."

– Gary Bowen, Toronto and Region Conservation Authority

> *"In the age of Big Data, information managers are moving away from the idea of centralized data portals. Conservation Ontario's focus is to enable our member conservation authorities to make their own data discoverable, accessible, and open."*
>
> *– Bonnie Fox, Conservation Ontario*

Applicability of data: It was noted that in a number of instances, questions have been posed of certain datasets that are disconnected to the original purpose for which the data was collected. This misapplication can lead to poor results, including wasted investment and application of resources. Consequently, it is vital that users of data be informed to the extent possible of limitations, and data providers must be able to provide this context as data is made available. Data quality and validation can then be applied to ensure that data is used in a manner that does not distort or misrepresent outputs.

Variability of Data Science Capability: Discussions also highlighted a significant variation in degree of existing data- and data science capability. Larger and/or more urban municipalities tend to have better data availability and quality than more rural regions. This means that relevant data may be spotty depending on the municipality or CA in question. However, it was agreed that opportunity also exists to explore how relevant data and analyses (e.g. storm water) could be accessed from neighboring municipalities and CA's for the purpose of comparative and inferential analysis.

- For example, the data and learnings with regards to the impacts of storm water on water quality in storm ponds in Toronto could be useful for other municipalities (such as Cobourg) because storm ponds of similar dimension and characteristics may result in common impacts. Further work is needed to explore and validate these kinds of 'two birds, one stone' opportunities.

Monitoring Techniques and Technologies: There are significant gaps in resourcing and capability to conduct on-going monitoring in a number of core priorities and specific to local conditions. There was consensus that budget and resource issues will continue to constrain this foundational activity. Solutions will be needed to 'even out' this capability to support lakewide collaborations, including enhanced collaborative efforts to find efficiencies by pooling resource and capability to foster more sophisticated monitoring practices. Ecosystem physical modeling can also be used to inform where sensors can best be placed to locate sensors/sensor systems.

Access and Ownership of Data: Concerns were raised that it is sometimes difficult to access data collected or owned by private parties.

- For example, several groups identified a lack of data relating to the use of salt by private companies, contractors, home owners and big box parking lots. It was stated that collecting this data would be particularly challenging.
- In some cases data has been collected and analyzed on behalf of a third party. The primary collector does not have authority to release the data without consent; but also, there is no incentive to even make public that the data exists.
- Many studies may only collect data for a short duration or a limited scope. Resources required to process data and make it openly available are often not available.

Data context: Foundational work is also needed to contextualize socio-economic factors and barriers, hand-in-hand with the data science and analytical capabilities. For example, in the discussion about how to improve road salt application practices, a clearer understanding of policy implications around insurance and legality would be invaluable to better frame contextually the structural barriers and potential solutions to the issue.

"You don't have to do a lot – just telling us what to do would help."

– Municipal Lead, Workshop Attendee

RECOMMENDATIONS

Based on inputs and comments gathered, a number of opportunities and suggested next steps were identified:

1. A collaborative multi-stakeholder demonstration project should be established to create a living, real world example of how best to apply modern tools and techniques from data science, based on a manageable scope, common interests/needs and a shared desire to build a tool – or set of tools – relevant to a common concern. The pilot should serve as a platform for enhanced collaboration, establishing a foundation for expansion as the project demonstrates value and attracts further interest and participation. The guiding philosophy of this project should follow the advice of the ECCC CDO's representative: "Crawl, walk, run." At its starting point, the project should start with a clear articulation of the specific questions that decision-makers and similar stakeholders would like to see answered, to avoid developing solutions that do not directly connect to desired value.

2. Rather than trying to establish a completely new collaboration, it was recommended that this project build 'on the shoulders' of existing collaborative relationships that complement the proposed demonstration project's objectives. Based on such a foundation, the intent would be to build a common view and purpose amongst disparate stakeholders with varying levels of resourcing and capability, as well as differing priorities and perspectives about what (in a period of shrinking budgets) deserves investment of both internal capacity and incremental capital. Two examples of existing collaborations from which to draw were identified as the Lake Ontario Partnership (the multi-stakeholder collaboration focused on the development of the Lake Ontario LAMP) and the Quinte AOC collaboration. These should both be approached and evaluated for interest and applicability.

3. Similarly, rather than 'reinvent wheels', an early input into this proposal would be to conduct an in-depth review and audit any water quality data-related projects and platforms already in development and use, to evaluate how they could potentially serve and/or feed into the development of a platform for better future collaboration.

A number of such existing data management tools and platforms have been developed through partnerships led by SOSCIP in partnership with participating academic institutions. These include but are not limited to:

a) CANWET is a watershed modelling platform developed in conjuction with the University of Guelph that allows scenario consideration for planning purposes once calibrated models have been developed.
b) THREATS is a tool for Cumulative Effects Assessment and exploring data and indicators around a variety of stressor and environmental response themes
c) ISWMS is a real-time flood prediction tool that integrates weather prediction data, hydrology/hydraulic modeling and a web-based platform. The platform is currently operating on 2 pilot watersheds.

It was recommended that a data science/technology literate steering group be formed to examine these – and other – platforms as they conduct this feasibility evaluation, based on agreed criteria for both usability and on-going open source access for future uses.

4. The content of this collaboration should cross-cut both geographically and at varying scales, with relevance from watershed/subwatershed levels up to lakewide scope and beyond.

During our breakout groups, a number of relevant area of focus were raised fairly consistently. These somewhat interrelated themes have in common that they are top-of-mind for water quality and resource decision makers and could be explored from a number of levels applying data science:

a) **Chloride**

i. The application of road salt and its impact on chloride levels in the surrounding ecosystems (from nearshore areas ultimately extending to lakewide chloride levels);
ii. informed by episodic stormwater discharge events caused by precipitation spikes,

iii. considering source protection in context of sensitive aquifers and/or drinking water management (with human health impacts/considerations)
iv. better understanding how chloride use on land impacts on both native and introduced species as well as species dynamics and interactions in the lakes.

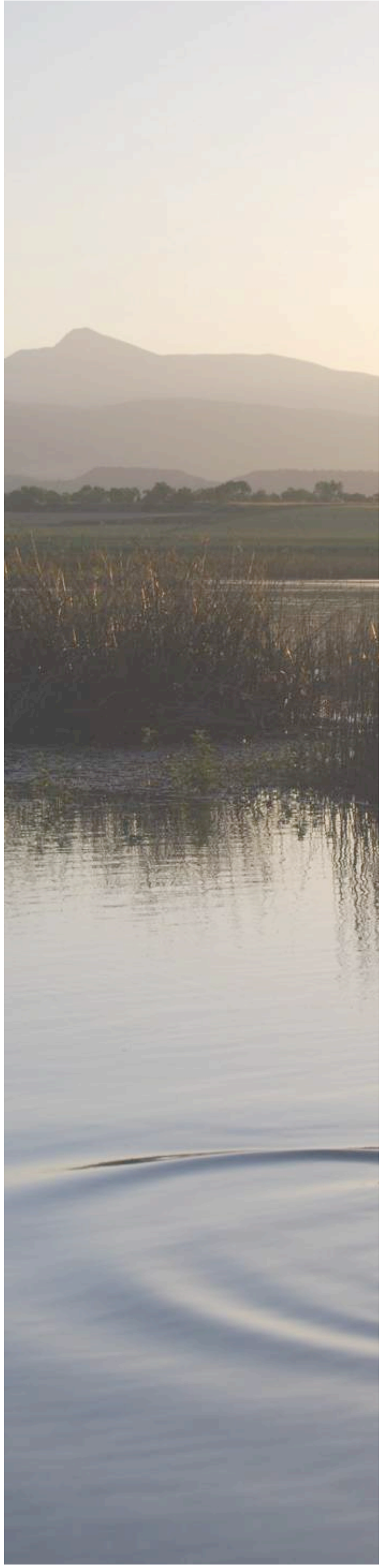
b) **Nutrient run-off** (with emphasis on phosphorus levels)

i. Agricultural land-use and related activities' impact on watersheds and larger water bodies;
ii. exploration of how nutrient runoff in combination with chloride contamination affects nutrient dispersal in water bodies;
iii. implications for land-use planning and fertilizer applications practices.

c) **Stormwater management**

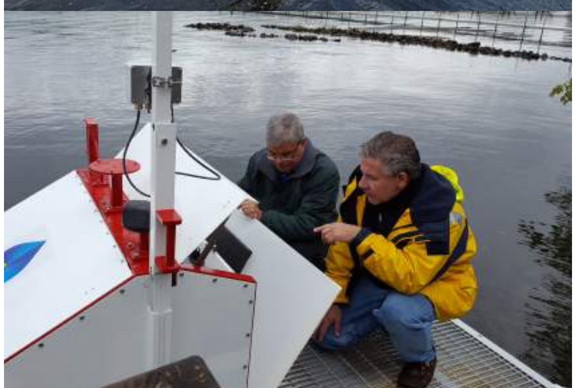
i. Enhancing current understanding of the relationship between weather and precipitation datasets in conjunction with stormwater release events;
ii. developing better understanding of how current stormwater management practices are directly and indirectly affecting – in particular – waterbodies and near-shore ecosystems;
iii. providing visualizations of how stormwater flows and disperses in different geographic profiles;
iv. seeking opportunities for joint modelling of (for example) how stormwater collected in ponds of consistent size/structure and materials may be created and shared.

5. Hand-in-hand with this, greater effort is needed to audit and understand the scope of existing datasets that have been gathered by different Ontario water-focused stakeholders. This should include their basic characteristics (temporal, geographical, historical) current usage and possible/likely compatibility with other datasets. As suggested, this dataset audit would be focused on the proposed pilot project content. Rather than asking stakeholders to provide 'all of their data', the proposed collaboration would be to secure data sharing centred around the specific areas of focus.



A core finding from our discussions was that a number of exhaustive attempts at broadscale data audits have floundered, as the resources required to collect and codify existing and historic datasets is significant, and does not immediately yield value. Consequently, the expert advice was to begin with more focused data collection and cross-collaboration exercises to elicit datasets, concentrated on a specific project or issue (such as chloride).

6. The above should be laid out as a phased, actionable, clearly outlined approach for how best to engineer this collaboration toward the eventual development of a predictive capability, that – combined with enhanced IoT, physical model forecasting, remote sensing and analytics / AI and visualization techniques – can directly inform water resource decision makers seeking better, more valuable insights for scenario planning and input into funding regulation, policy and specific water- and land-use permissions impacting upon Ontario’s water resources and water infrastructure.



SAMPLE CASE: JEFFERSON PROJECT AT LAKE GEORGE, NY

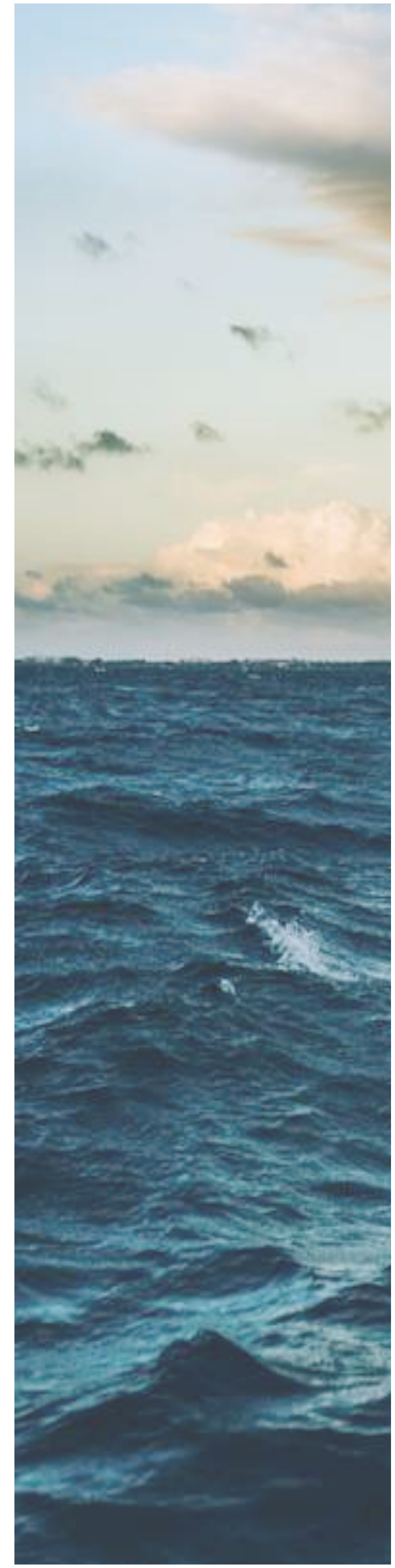
Around the world there are a variety of water-related projects that highlight the power of technology and analytical science to enhance environmental monitoring, develop innovative solutions and promote global freshwater security. One such project is the Jefferson Project at Lake George, NY where IBM Research, Rensselaer Polytechnic Institute and the FUND for Lake George are working in partnership to advance real-time water monitoring and data collection. The lake is a source of drinking water for New York State so it is important to understand and mitigate the impacts of human activity on freshwater systems. The project has been operational since 2013 and builds on decades of monitoring work by Rensselaer's Darrin Fresh Water Institute scientists.

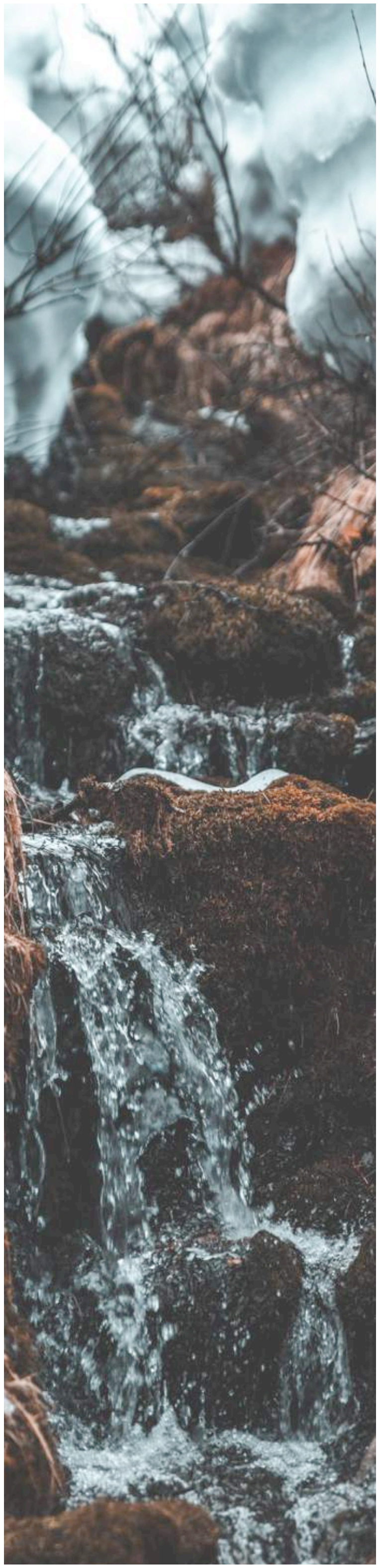
Currently, there are over 50 multi-sensory platforms with solar panels located on the lake, which are connected wirelessly using 3G and 4G networks. The advanced sensing network monitors various water quality indicators at a range of depths and provides scientists with a real-time data stream in a "living lab" setting rather than having to collect and transport samples to laboratories. All of the platforms are capable of adaptive sensing, which means that they are contextually aware of the other sensors and can detect environmental changes. For example, if scientists want to track a specific event in greater detail or a storm is coming and power needs to be conserved, sensors can alter sampling frequencies. This data stream allows the complex interrelationships between stressors to be analyzed including chloride levels, nutrient loading, invasive species and other threats to the water supply. With this data, scientists can compare what is expected, based on traditional scientific knowledge, with current lake conditions. Beyond this, scientists can utilize advanced modelling to create a sophisticated, high resolution picture of the lake's ecosystem and predict the future impacts of weather events, water runoff from surrounding mountains, road salt and water circulation. The technology has already helped measure the impacts of reduced road salt usage further reinforcing local programs and commitments. The project is relevant to a broader audience as it can be scaled and extrapolated to help advance understandings of other lakes and water bodies, including the Great Lakes basin.

BIG DATA, ANALYTICS AND MACHINE LEARNING: GUIDING QUESTIONS

As discussions are initiated related to the opportunity to improve water asset and resource management using Big Data tools and approaches, a number of core questions are helpful as a starting point:

- What relevant data assets are already amassed, being collected, and analyzed by different levels of government, agencies and organizations surrounding our watersheds, water infrastructure, and larger lakes and lake ecosystems?
- What additional data can be collected to most effectively broaden and extend our ability to achieve desired outcomes?
- What tools and approaches need to be brought to bear in the exercise of measuring performance of varying models and approaches as effectively and simply as possible, against a backdrop of an exponentially increasing data ecosystem?
- What existing projects, models, approaches and technologies are already being effectively deployed related to our desired outcomes? Which of these could accelerate our desired advancements in water resource protection and management?
- What are the most pressing needs and priorities toward which we should be building our capability for more sophisticated uses of data, analytics and machine learning (including but not limited to LAMPs, water asset management and community growth planning and decision-making)?
- Which stakeholders need to be consulted and incorporated into the future discussions and work as part of these efforts?
- Which are the most promising areas of academic review and experimentation underway which will be most likely to have direct relevance to current and future priorities?
- How can/should this work interrelate and establish common frameworks with other, similar and related efforts to leverage insights from Big Data in context of our watersheds, communities, and the Great Lakes?





LEAD REVIEWERS

Pollution Probe gratefully acknowledges the invaluable contributions of the following Lead Reviewers in the preparation of this report.

- Gary Bowen, TRCA
- Patricia Moleirinho, Conservation Ontario
- Kaveh Afshar-Zanjani, ECCC
- Amanjot Singh, Credit Valley Conservation
- Trevor Boston, Greenland Consulting
- Michael Passow, IBM