Multi-sectoral Collaborative Open Data Applications

Paulo Alencar, Donald Cowan
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Email: {palencar, dcowan}@uwaterloo.ca

Fred McGarry
Centre for Community Mapping
Waterloo, Ontario, Canada N2L 2R5
Email: mcgarry@comap.ca

R. Mark Palmer
Greenland International Consulting Ltd.
Collingwood, Ontario, Canada L9Y 1V5
Email: mpalmer@grnland.com

***Abstract*—Complex collaborative applications often involve multiple sectors and numerous heterogeneous open and private data sources. Although these applications are starting to emerge as a result of scientific and technological advances in integration and analytical data capabilities, there is a lack of understanding about the nature of collaborative multi-sectoral open data applications, their architecture and the role of open data in these applications. In this paper we describe the main features of collaborative multi-sectoral applications and illustrate our ideas using a collaborative application called "Integrated Science and Watershed Management System" (ISWMS™), which supports climate adaptation assessment and real-time predictive modeling capabilities. The collaborative efforts around building this application involve experts such as those in environment, rural and urban development, and climate change.**

I. Introduction

Collaboration among multiple sectors involving business, government, non-governmental organizations(NGOs), community groups and individuals ([1], [2], [3], [4], [5], [6], [7], [8]) has the capacity to solve systemic problems since they can draw resources from a broad range of of complementary talents. Multi-sectoral collaborative efforts are increasingly becoming an important trend in a number of areas, such as water and land resource management ([9], [10], [11], [12], [13], [2]), public services ([14]), urban development ([2]), community projects ([15], [16]), and health ([17], [18]).

Multi-sectoral collaborative approaches offer many benefits including ([19], [2]):

- increased access to resources;
- more efficient use of resources;
- enhanced accountability;
- improved innovation;
- broadened awareness;
- improved relationships among stakeholders;
- sustainable development of activities;
- broad sharing of responsibility;
- strong ownership by stakeholders;
- use of the strengths and talents of the partners;
- sharing of knowledge and technology; and
- better balanced design of projects.

One area in which collaborative multi-sectoral applications are beginning to emerge involves Integrated Water Resource Management (IWRM).

In this paper we describe progress to date in designing, building and deploying a multi-sectoral approach to IWRM. We recognize that although these applications are starting to emerge as a result of scientific and technological advances in integration and analytical data capabilities, there is a lack of understanding about the nature and architecture of collaborative multi-sectoral applications as well as their governance and long-term sustainability, especially when data is obtained from a variety of sources.

In addition we describe our current thinking on the main features of these collaborative multi-sectoral applications and illustrate our ideas using a collaborative application called the Integrated Science and Watershed Management System (ISWMS™). ISWMS™ supports climate adaptation assessment and real-time predictive modeling capabilities. The collaborative efforts around building this application is supported by experts from sectors such as the environment, urban development, climate change and computer science.

This paper is structured as follows. In Section II we briefly describe the integrated water resources management paradigm. In Section III we describe ISWMS™ in terms of its architectural structure and layers. In Section IV we analyze the main features of collaborative multi-sectoral open data applications and the role of open data in these applications. Finally, in Section V we present our conclusions and future work.

II. Integrated Water Resources Management

Integrated Water Resources Management (IWRM) has been a widely accepted approach for over two decades [20] and has been defined by the Global Water Partnership (gwp.org) as "a process which promotes the coordinated development and management of water, land and related resources, in order to maximize the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems." The founding pillars of IWRM are presented through the Dublin principles [21], which state that:

1) Freshwater is a finite and vulnerable resource, essential to sustain life, development and the environment;
2) Water development and management should be based on a participatory approach, involving users, planners and policy-makers at all levels;
3) Women play a central part in the provision, management, and safeguarding of water;
4) Water has an economic value in all its competing uses, and should be recognized as an economic good.

Although these principles are a high-level view of water management and certainly talk about collaboration they do not really tell us much about how to operationalize a complex

COLLABORATECOM 2014, October 22-25, Miami, United States

DOI 10.4108/icst.collaboratecom.2014.257375

situation such as the effective management of water resources. We need to expand and augment these principles dramatically in order to achieve their lofty goals. A number of issues are outlined in the next few paragraphs.

Although principle 1 mentions water as a resource and principle 4 mentions economic value, nowhere do these statements also recognize the destructive force of water when in full flood and not properly controlled. Recent evidence for the force of water probably owing to climate change, is epitomized by recent flooding in North America and Europe in 2013 and 2014. In other words water can have both a positive and negative economic value.

Principle 2 mentions a participatory approach and describes three stakeholder classes, namely users, planners and policy makers. Is this all? Definitely not! How do we identify the stakeholders that should participate in IWRM around a specific watershed. The group mentioned in principle 2 needs to be expanded to address both the modeling of a watershed for long-term water quality or to respond in real-time when the water is out of control such as in a major flood.

Once the stakeholders are identified, how do we assemble them into an effective team? Remembering too that watershed management is not a one-time effort, but a continuum of activity as natural or manmade changes in the landscape are occurring all the time, thus impacting the watershed capacity and the water quality.

Of course these principles do not even mention two factors of collaborative endeavours namely: governance and sustainability. Most collaborations such as IWRM are not short-term, rather they must operate and evolve forever.

Governance is critical in that all the collaborators must feel that they have a say in the operation of the joint effort. How can we ensure participation and yet not bog down in petty details? How can we be sure of making timely decisions?

Of course any long-term collaboration must be sustainable. How does the collaboration continue to draw on key personnel and re-invent itself to look at evolving situations? Any collaboration has associated expenses. For example assembling people for various tasks and in the case of IWRM assembling and maintaining open data is a significant expense. Ideally sustainability should not depend on government or similar sectors, but rather find ways to generate needed financial support from its own activities.

IWRM models rely on the integration of numerous aspects of watersheds including:

- catchment and stream delineation;
- digital elevation;
- soil texture;
- water holding capacity;
- erosion potential and soil drainage;
- weather station locations;
- daily precipitation;
- min/max temperature records; and
- land use.

The data sets just described are used both as input to IWRM models, and to calibrate those models to ensure that the output is credible.

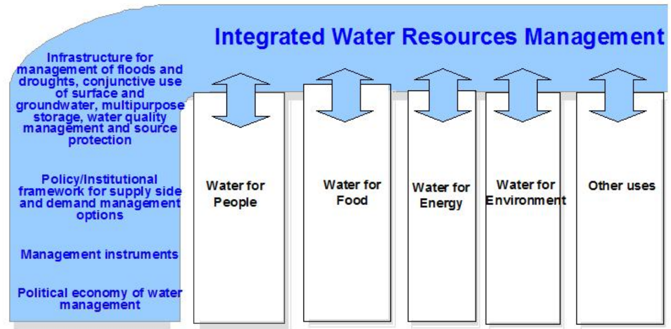

Fig. 1. Relationship between water resources and the environment (from [26]).

Of course, this data, although existing in many cases, is scattered among multiple jurisdictions but is usually available as open data [22], [23], [24], [25]. For example, open data for a watershed in any Canadian province could be held by Federal Departments, Provincial Ministries (departments), conservation authorities,[1] political regions such as municipalities, NGOs, universities or consultants.

Because the data is not accessible from a single source it becomes extremely costly to assemble, thus preventing the wider use of predictive simulation tools. The type of situation just described is exactly the kind of project that cries out for multi-sectoral collaboration.

In order to produce a multi-sectoral approach one needs to have the required data easily accessible to the tools that analyze the data and implement the models and hence to users engaged in an IWRM exercise. Probably an environmental data and software platform operating as a cloud that is accessible over a high speed network would be appropriate. Such a platform would not only contain data and software but could be connected to field personnel and sensor networks that can deliver data through both satellite and land-based communications in near real-time.

Although not all these problems have been solved for IWRM, in the next section we describe an overview of an initial attempt of developing an open data and integrated stormwater and watershed management system (ISWMS^{TM}) with accurate flood forecasting and climate change impact capabilities.

III. The Integrated Science and Watershed Management System

The Integrated Science and Watershed Management System (ISWMS^{TM}) is an example of a collaborative multi-sectoral open data application that has been developed under a partnership involving Greenland International Consulting Ltd. (Greenland), the University of Waterloo Computer Systems Group (UWCSG) and the Centre for Community Mapping

[1]In Southern Ontario, Canada conservation authorities are responsible for managing and maintaining watersheds, similar authorities exist in other regions and countries.

(COMAP) to develop new open data systems [27], [25]. The Waterloo-based partners are assisting with the design and construction of system architectures, interfaces, and other database structures which ultimately will allow for the streamlined integration of new datasets and enable users to access open data and other tools.

A. Motivation for ISWMS™

Over the last two decades, billions of dollars worth of flooding damages and incurred loss of human life has prompted the development of a flood prevention program within Canada. It has however been noted that public interest in flood prevention, and as often follows political will, does fall out of sight without major flood events (Environment Canada, 2004).

The 1997 Red River Flood in Manitoba (estimated total recovery cost in Canada: $500 million CAD) was said to have been caused by a combination of hydrometeorological factors, beginning with high antecedent soil moisture, heavy winter snowfall, and a rapid spring melt (Environment Canada, 2004). Similarly the recent and devastating 2013 Flood of Alberta (estimated total recovery cost of $6 billion CAD) has been attributed to a rapid snow pack depletion caused by climatic factors (Environment Canada, 2014). Unfortunately, the resources to have foreseen and greatly mitigate damages and loss of life from the 2013 Alberta Floods could have been operational during and before the time of the flood.

B. The ISWMS™ Decision Support System

The initial version of the ISWMS™ was developed by Greenland as a Decision Support System (DSS). Although other GIS-based software tools were developed before the millennium, there was a lack of integrated capabilities. The first ISWMS™ development phase was completed by Greenland with support from Canada's National Research Council. ISWMS™ (Version 1) was then tested successfully for subwatershed planning and groundwater studies in Simcoe, Grey and Dufferin Counties in Southern Ontario. Later, in 2001, Greenland was retained by the Nottawasaga Valley Conservation Authority to prepare a flood forecasting version of the Nottawasaga River Basin. This real-time flood forecasting system became operational in 2003.

ISWMS™ (Version 2) is a significant and timely extension of the initial system that can be used to address Canada-wide issues about accurate (real-time) flood predictions and to help prevent loss of life and reduce property damage. This system is designed to be the first ever open data and open source GIS tool with climate change impact capabilities for watershed management and stormwater management infrastructure programs, as well as reservoir dam operations for real-time flood hazard management and hydropower system operations. As one of its sub-components, ISWMS™ relies on the next version of the CANadian Watershed Evaluation Tool (CANWET™), which is an open data and open source GIS infrastructure planning and design tool with climate change modeling capabilities. In addition, ISWMS™ relies on a web-based interface designed to provide predictive capabilities about how any watershed will respond in various scenarios, such as proposed land changes with regard for climate change factors, and support the ability to model source water protection strategies from cost-benefit and management target perspectives.

ISWMS™ has relied on the following system development objectives:

- Utilize/Update Science-based Algorithms Available from the ISWMS™ (Version 1) Source Code;
- Enhanced Flood Forecasting Module with Accurate/Advanced Notice Prediction Capabilities;
- Incorporation of Real-time and Short-term Forecast Climate Data and Snow Pack Data Collected from Open Data Sources and New Sensor Technologies;
- Use of Canadian Industry Accepted Hydrological Modeling Routines;
- Spatially Distributed Input Data;
- Incorporate New GIS Procedures from the 5th version of CANWET™ (called CANWET-5) to Accurately Calculate Basin Parameters;
- Include Yearly Snow Storage Accounting Routine and Applied to Small Catchment Conditions;
- Automatic Download, Processing, and Integration of Collected Data into Real-time Models;
- Best-Possible Calibration Routines Using Open Data Source Solver Routines.

C. Long-term Development

On February 24, 2014, the Ontario Government released a new Provincial Policy Statement, 2014 (PPS, 2014). The last Statement was published in 2005. The latest update came into effect on April 30, 2014 and contains all of the Province‘s policies concerning land use planning and future development. It will be the cornerstone of Ontario's land use planning system and all planning decisions must be consistent with the new policies. The policies provide better direction for supporting healthy active communities, strong economies and responsible resources management in a clean and healthy environment. Future decisions, however, will rely on Ontario science-based predictive tools that have life cycle costing and climate change modeling capabilities. Web-based and open data decision support systems will also be important to ensure open and transparent planning decisions too. Therefore, CANWET™ and in conjunction with the overall and ISWMS™ Decision Support System will be important to address environmental policy objectives with a climate adaptation focus, such as:

- Implement effective remedial measures for enhancing or restoring stream and river health through the integration with watershed monitoring programs and stewardship partnerships;
- Provide effective flood hazard management and resulting public safety and damage reduction;
- Implement source water protection plans that are now being implemented and which can contribute to Ontario's

business climate for ensuring sustainable community growth strategies;

- Identify life-cycle cost solutions for major water, wastewater and stormwater management infrastructure and, if feasible, reduce or defer municipal capital costs (e.g. better maintenance of stream baseflow for wastewater assimilation, stormwater management facilities, etc.); and,
- Create long-term alliances and partnerships between governments and rural and urban stakeholders in order to advance science-based solutions for all citizens of Ontario.

In the ISWMS™ system, the underlying web-based and open data system framework provides a viable/cost-effective path to consolidate it with other Greenland source code (including snowmelt, lake capacity and nutrient management tools) and into a complete watershed decision support system.

ISWMS™ will soon include a new "Biological Stream Health Predictive Module." Stream health predictions would account for measured bio-assessment data, river/stream flow and water chemistry data and other non-government program assessment techniques. Many practitioners use rapid bio-assessment techniques that rely on a variety of biotic and compositional indices to describe and assess a site. While such indices are useful because they integrate stream data into a single number, the ability to distinguish impaired from unimpaired sites, and the ability to determine what impairments are shaping the community is often compromised. Furthermore, the relationship between biotic and compositional indices and various physical catchment variables is generally not well understood.

This biological predictive module is timely: (i) to help investigate relationships between chosen biological indices with various catchment and stream variables so that we can better understand what land use changes lead to what impairments to the living stream community (and which will ultimately form the feedback loop to satisfy PPS, 2014 policies); and (ii) to improve our ability to distinguish impaired/undesirable from unimpaired/desirable sites. By recognizing that healthy is variable and that various physical parameters affect stream indicators, we need to be able to develop an expected indicator value a priori. Therefore, rather than using an arbitrary threshold for impairment, thresholds could then be calibrated to local stream conditions via the integrated system using CANWET™ and ISWMS™ data and thereby reducing the probabilities of type I and II errors (i.e. for not detecting an impaired site, or calling an unimpaired site impaired).

D. The ISWMS™Architecture

The extended Integrated Science and Watershed Management System (ISWMS™) essentially involves three layers. The first layer combines components related to open data and other sources. Components in the first layer deal with Global Climate Change Model; the Local Climate Change Impacts (which uses the Statistical Downscaling Model, SDSM); Remote Sensing, GIS, Field Monitoring and Laboratory Data; Spatially Distributed Historical Weather Data; Climate Forecasting; and Water Quality and Nutrient Source Database (which is being used in the development of The Healthy River Ecosystem AssessmenT System, (THREATS), which was developed by Dr. Monique Dubé. These components are then integrated into a regional database. At this level, ongoing work involves the construction of a Centralized Regional Database that takes advantage of the THREATS database architecture developed by Dr. Dubé to interact with and monitor the water quality and nutrient source database.

The second layer combines components related to scientific models and procedures. Components already built in the second layer include the Hydrology component, the GIS Processor component, and CANWET™. Other available components are the Flood Line Delineations and Hydraulic Modeling component and the Open Source Parameter Optimization Component. The Hydrology sub-components already built include the ISWMS™ (Version 1) DSS and a new energy-based Rainfall and Snowmelt Model.

In the second layer, ongoing work aims at producing the RealTime Flood Forecasting Tool component of ISWMS™. This component uses two new Hydrology components (the Continuous Snow Storage Accounting and the Advanced Water Storage Routing), Automated Calibration Procedures and Better Model Evaluation Techniques.

The third layer combines components related to the system front-end. Components in this layer include the Web Interface tools and the long term vision components of ISWMS™. These long-term vision components include the Biological Health Predictive Module, Statistical Evaluation and Calibration Tools, and integration components (e. g., the Integration of CANWET™ with THREATS to provide predictive modeling).

The system provides predictive watershed modeling with cumulative effects. CANWET™ is a watershed evaluation tool that deals with climate change, water balance, nutrition and contaminant reduction (BMPs), water stress (takings), assimilative capacity, and land use changes and wastewater infrastructure planning. ISWMS™ deals with storm water management and flood forecasting. THREATS is tool for healthy river ecosystem monitoring and assessment.

Further, the system supports adaptive watershed management. The system is managed by watershed and basin authorities. Modeling is done by public and private sectors: watershed authorities and municipalities; insurance and financial; resource development; and land development industries.

The system also supports shared knowledge services: secure cloud service; multi-party open data; state-of-the art proprietary algorithms, analytics; multi-party governance with role-based access administration; multi-party data modeling with mediated model accumulation; and Community of Practice (CoP) mediated social network.

E. Predictive Models

Sophisticated model calibration techniques will be included in ISWMS™ to reduce or eliminate any model deviation from monitored data. In addition, the following elements will be integrated:

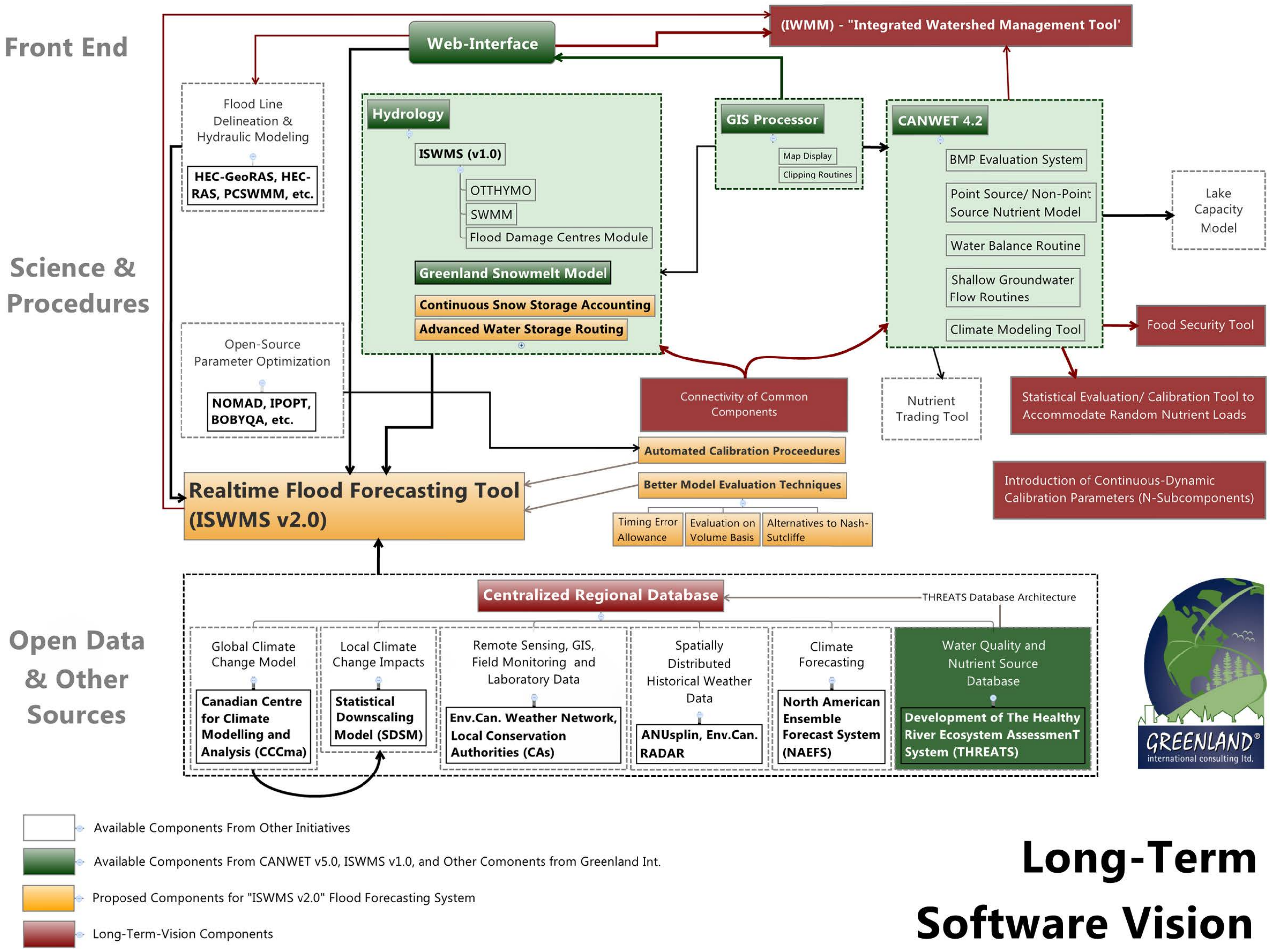

Fig. 2. ISWMS™ architectural layers and components.

- ISWMS™ (Version 1) simulation and routing routines combine the best elements of the Canadian industry accepted HYMO and SWMM hydrologic and stormwater management models;
- Introduction of advanced storage routing techniques applied to natural topography and constructed storage features (natural land attenuation , constructed storage units); and
- Customized application of the Generalized Snow Melt Equation (GMSE) for major land storage coupling current atmospheric models with hydrological phenomenon.

A major challenge of flood forecasting for medium and larger watersheds in Canada is a lack of reliable snowpack data. Through the use of a calibrated and spatially distributed water and energy balance model and a thorough understanding of snowmelt energy dynamics, snow storage can be reliably determined and further validated against either field measurements or satellite imagery. The best possible estimates of inflows will be determined through the use of spatially distributed precipitation data. To ensure accurate and precise outflows, a set of empirical calibration parameters will be applied to the hydrological routing methods. For the best calibration, a variety of open source solver-routines will be used.

As a part of computerized solving mechanisms, more advanced model evaluation techniques must be applied (unlike traditional calibration, computers cannot eyeball a good calibration result, a variety of objectives, relative weights, and constraints must be clearly defined). Therefore, the proposed model evaluation techniques for ISWMS™ will include:

- Alternatives to the traditional Nash-Sutcliffe analytical approach;
- Volumebased accounting;
- Acceptable limits for errors in flood timing; and
- Unique performance evaluations specific to low flow and less frequent return period events.

ISWMS™ will include catchment clipping routines, a web-based interface, and hydrological routines which could be utilized to develop the initial version of ISWMS™. Thereafter, initial pilot testing and verification would be completed on selected watersheds in Southern Ontario, including large catchments in the Nottawasaga River Watershed, Muskoka Region and Trent-Seven River Basin. Verification is to include real-time predictive analysis as well as historical flooding scenarios utilizing data available only before the period of the flood.

F. Applications of the ISWMS™

ISWMS™ and related niche software tools for water resources engineering and watershed management applications have been used successfully by Greenland and its clients for many projects in Canada, including:

- To develop legislated watershed management plans (e.g. Lake Simcoe Protection Plan);
- To evaluate water balance and nutrient reduction options (e.g. Lake Winnipeg Basin);
- To develop municipal drainage standards with regard for climate change impact factors;
- To investigate water stress (taking) impacts (e.g. Trent-Severn Waterway Basin);
- To investigate assimilative capacity and municipal servicing gaps, as well as water reclamation opportunities, for Public-Private implementation strategies;
- To assess nutrient target setting and Best Management Practices (BMPs) usage in order to achieve sustainable nutrient loads and compliance with regulated water quality objectives; and
- To examine land use changes associated with new legislation and develop mitigative stormwater management plans for community growth projects.

IV. Features of Collaborative Multi-Sectoral Open Data Applications

ISWMS™ is an example of a collaborative multi-sectoral open data application. It involves multiple stakeholders from different sectors such as government, NGOs and experts from information technology, water resources, urban development and climate change. Based on our experience in developing numerous multi-sectoral open data applications, we present in this section some of the general features and specific open data-related features of these applications.

A. General Multi-sectoral Collaborative Features

Being a multi-sectoral collaborative application, its development has been impacted by the numerous factors that affect multi-sectoral collaboration, which include ([2], [19]):

- Context and motivation: This factor refers to the reasons that different stakeholders engage in collaborative initiatives;
- Leadership: This factor refers to the approach and process of social influence that is used in collaborative initiatives both in the organizations involved and the project team;
- Structure: This factor refers t the formal and informal frameworks and settings that support a collaborative initiative;
- Resources: This factor relates to the funding and staffing available to support the collaborative initiative;
- Governance: This factor relates to the structure and practices in place to deal with issues of power and control in a collaborative initiative;
- Trust: This factors relates to the confidence that other organizations and individuals will take decisions that are in the best interests of the collaborative initiative;
- Risk: This factor refers to refers to the chance or possibility of loss or negative consequences emerging from loss of autonomy and capacity for unilateral action in a collaborative initiative;
- Collaborative capacity: This factor can be defined as the experience and knowledge of collaboration possessed by individual organizations involved in a collaborative initiative;
- Empowerment: This factor refers to the authority and power to affect change, which can be a pre-condition to successful collaboration;
- Information and evidence: This factor relates to the body of knowledge and (credible) data upon which the collaborative initiative bases its efforts (e. g., discussions, planning and decisions);
- Monitoring: This factor relates to the framework and process put into place to track progress on achieving the goals and objectives of the collaborative initiative;
- Organizational capacity: This factor relates to the ability of organizations to participate in the collaborative initiative (e. g., in terms of resources, expertise, leadership and experience);
- Common understanding: This factor refers to shared meaning and understanding developed among the members of the collaborative initiative;
- Sustainability: Many collaborative open data projects such as those related to environmental analysis and modelling continue for a long period of time with multiple public and private partners. In this case sustainability refers to the ability to find an operational model that ensures the project's long-term viability particularly with respect to governance and financial support;
- Evolution: As noted previously many collaborative open data projects continue for a long period of time with multiple changing public and private partners. Governance and operation of such projects requires a different approach for both the management and the staff recognizing the need to evolve.

Although each instance of a multi-sectoral collaborative initiative is unique and involves different circumstances and challenges, some steps can be followed as general guidelines to establish a multi-sectoral collaborative initiative and define the resulting integrated system ([2], [19]):

- Identify the stakeholders;
- Make a commitment to collaborate;
- Establish procedural agreements;
- Make sure participants have the necessary process skills;
- Build trust;
- Develop and instill a culture of sustainability;

- Develop and instill a culture of evolution;
- Identify problems;
- Create a vision for the collaboration including governance, sustainability and evolution;
- Create options for solving problems;
- Formulate goals, objectives and an action plan;
- Implement the action plan;
- Evaluate the results;
- Define the next steps of the collaboration.

From the standpoint of software development, multi-sectoral collaborative applications have the following features:

- These applications require collaborative efforts from stakeholders from multiple sectors;
- The integration of components is based on the requirements associated with different stakeholders in different sectors;
- The non-trival combination of multiple open data sources has to be supported by the underlying software infrastructure;
- The coordination of decision-making is based on multi-sectoral components and multiple open data sources;
- The applications require the integration of multi-sectoral standards and policies.

B. Open Data-Related Features

Regarding open data, a multi-sectoral collaborative application such as ISWMS™ involves multiple sources of open data. The open data sources for this application include federal departments, provincial ministries (departments), municipalities (land use), conservation authorities, political regions (municipalities), NGOs, universities, businesses, and consultants. Among the well-accepted definitions of open data [28], [29], [30], the Open Knowledge Foundation has suggested that from a technical standpoint, open data is [28]: "A piece of content or data is open if anyone is free to use, reuse, and redistribute it subject only, at most, to the requirement to attribute and/or share-alike." Open is becoming an increasingly important direction in information technology as governments are releasing more and more data to become more transparent and claiming that open data has substantial economic value [31], [32], [33]).

The multi-sectoral available datasets interact in complex ways. In general, climate data is available to populate the proposed spatially distributed precipitation and energy balance routines for the rainfall simulation and snowmelt input routines. Interactions between various data sets create a space for an even wider set of data through interpolation. Open data sources used in the development ISWMS™ include statistics data, weather-related information, historical data, and land-related data. The open data sources and constraints related to this application are provided in Table 1.

More specifically, from the standpoint of software development, the features of multi-sectoral collaborative open data applications include ([25]; See Table 2):

The data is scattered through each and multiple jurisdictions: ISWMS™ deals with multiple sources of open data, which come from federal departments, provincial ministries (departments), municipalities (land use), conservation authorities, political regions (municipalities), NGOs, universities, businesses, and consultants. The multiple open data sources are often difficult to find. In addition, there is a need to ensure that the multipe open data sets are appropriately maintained since some of the data, such as weather data, must be available in real-time, but, in contrast, other data, such as soil composition, is fairly constant and may not need to be updated for years.

There are open data identification problems: In these applications it is often not straightforward to identify which data needs to be or can be open. This problem is exacerbated when there are multiple stakeholders and each of them has numerous distributed open data sources. Sometimes decisions about what data should be open are delayed by bureaucratic problems and by a lack of understanding concerning the nature of open data. Other times these decisions can become very difficult given their legal implications such as liabilities.

Open data standards should be adopted: These applications also present issues related to the choice of appropriate open data standards. The issues also involve choice of which meta data standards should be adopted in the case of specific open data applications (e. g., XML, DDI, RDF). Standards facilitate the integration of the multiple data sources as well as the integration of the components that use these sources. Are standards for data enough? Most open data is derived from databases that contain relationships among data elements that may be valuable.

There is a need for open data evaluation by experts: In some cases, open data has to be evaluated by experts, who assess them based on some chosen criteria in order to improve their quality. Open datasets provided by the public or non-experts that are to be used in predictive analysis procedures, for example, have to be assessed and approved. In this way, depending on the quality level of the open data used in a specific application, more effort is required to make sure the data is properly evaluated.

There is a need for stakeholder communication: The applications need to support the communication of stakeholders from multiple sectors. This communication support can be achieved though community of practice mediated social networks. These social networks allows individuals and groups to present their views and work in a collaborative fashion while providing mechanisms to help them manage the group and inter-group communication process. Such mediated social networks benefit multi-sectoral collaboration in many ways by providing, for example: support for timely knowledge sharing, context for effective knowledge exchange, mechanisms to identify people with

ISWMS™ Open Data	Description
Open Data Sources	- Federal departments - Provincial ministries (departments) - Conservation authorities - Political regions (municipalities) - NGOs - Universities - Businesses - Consultants
Open Data Constraints	- Accessible to Communities of Practice (CoP) - Editable by contributing agencies - Location obfuscation is used to hide exact location of endangered species (to prevent developers and eco-tourists to go to a location creating more of a problem; e. g., some developers may remove the species on their own)

TABLE I
OPEN DATA SOURCES AND CONSTRAINTS.

Multi-sectoral Colaborative Open Data Applications	Description
Application features	- The data is scattered through each and multiple jurisdictions - There are open data identification problems - Open data standards should be adopted - There is a need for open data evaluation by experts - There is a need for stakeholder communication - The applications require support for collaborative mapping - The applications require support for negotiation - The applications require support for proper access control - The applications lack support for using open data combined with secure data

TABLE II
FEATURES OF MULTI-SECTORAL COLLABORATIVE OPEN DATA APPLICATIONS.

specific skills, encouragement of social cohesion, creation of a shared space for geographically dispersed people, create a memory for group brainstorming and deliberation, and support for collective thinking.

The applications require support for collaborative mapping: The applications require support for stakeholder collaboration. Since many of these applications involve a geographical component, stakeholders in different sectors need to collaborate using maps as a central component. Collaborative mapping supports this collaboration by allowing various stakeholders to share maps with information about their decision-making choices (e. g., the borders of an urban development effort).

The applications require support for negotiation: Sometimes these applications require support for negotiation. Negotiation can be supported by collaborative elements (e. g., documents, maps) that can be updated in real-time by stakeholders in different sectors. In this way, not only the updates of each stakeholder are recorded on the elements, but the sequence of changes made by each of the stakeholders or in a particular session in which they were negotiating can be recorded for future use.

The applications require support for proper access control to open data: The application should allow the definition of which stakeholders (groups or individuals) can have access to which data. In this way a proper level of access control is provided to each member and unauthorized access to private or confidential information is disallowed.

The applications lack support for using open data combined with secure data: In some cases, the applications need to support the integration of multiple sources, where some are open and others are private. In this case, the private information that is used, for example, for the purpose of analysis, can not be released to the public or to other unintended groups. As a result, the data is sometimes obfuscated or aggregated in order to hide private details.

V. DISCUSSION

The paper has focused on one specific project and used that project to derive several observations about collaboration and collaborative projects. However, much more can be said based on the over 80 person-years thta the authors have had in working in formal and informal public-private collaborative partnerships.

Collaborative partnerships rely very much on shared leadership. Each organization involved in the multi-sectoral project must have a leader who will take the responsibility for ensuring that the participating organization delivers on their commitments in a timely manner. Of course the leadership of the partnership although multi-faceted must operate in an effective manner. For example, the leader-group could select their leader. This person must know when to defer to the other “leaders” on the team as no one has complete knowledge of the project. In other words leadership is not hierarchical but a shared responsibility.

Sustainability is a key issue in many collaborative projects. All the multi-sectoral collaborative projects that the authors have experienced, particularly related to the environment, do not have a finite life, but certainly appear as if they will go on forever. How do we ensure that:

- the partners live up to their original obligations;
- the governance model is long-lasted; and
- there is an adequate funding model, preferably one involving income generation to keep the project going.

From the latter point we should recognize that governments are rarely long-term funders of most projects as their priorities change over time. Of course there are exceptions such as NASA, but even there funding priorities have changed.

Evolution is another key concept that is closely related to sustainability. In fact evolution and sustainability go hand-in-hand. Projects that have a long life undergo changes over time. For example, modifications can occur in that:

- more or different partners can become involved;
- organizational leaders move on and thus the leadership group changes;
- the governance model can progress;
- the science and engineering practices can progress; and
- solutions as embedded in software and hardware differ over time as new approaches appear.

An evolutionary mindset has to be built into any collaborative project that has a long life as change is inevitable. New people that join the project must adopt the culture. They must see change as a positive driver and not as a thereat to their livelihood.

Collaborative projects that rely on open data have interesting additional problems. Interoperability presents some new challenges when addressing open data. To use the phrase found on may toys and other consumer goods, "some assembly is required." First the data will have to be located. One can envision an open data registry approach that is likely to evolve over time. Such approach is described in [25].

Open data as applied by governments is likely to be supplied on a level 3 format on the five-point scale originally conceived by Tim Berners-Lee [29]. This approach means making data avaialable in a non-proprietary format such as character or comma separated variables (CSV) or XML. Current practices indicate that this is and will continue to be the case. Thus, only the data will be available, not the relationships provided by a database, which means the database must be constructed separately from the data and the be structured based on the type of applications that use the data.

How will the database synchronize with the open data as published by a government? Pre-loading the data into the database and "watching" for updates to the open data and coordinating the two sources would seem to be a practical approach. As we start to acquire near real-time data such as weather and sensors, so-called big data, such a method will be problematic and another approach will need to be devised.

Open data is not always open. What about data related to endangered species or other sensitive habitats? For example, access to open data about habitat of an endangered species could be used to pinpoint and eliminate the species thereby opening up the related land to development.

Collecting of environmental open data will become more intense as we try to monitor and understand our surroundings. However, we will have to enable citizen scientists with appropriate data collection protocols as the ability to monitor will exceed the capacity of government scientists. Thus, we will establish and extremely large multi-sectoral collaborative open data application. This form of crowd-sourcing of environmental data will require some new form of mediation structure, automated or manual, that will be able to vet this open data before it is "published." This is a problem that still needs to be addressed, although it has been handled for data capture for invasive species [25].

VI. Conclusions and Future Work

Collaborative approaches and systems relying on wider systems thinking are essential to organizing and sustaining the efforts to improve multi-sectoral collaborative efforts, especially when the numerous involved stakeholders need to combine multiple distributed open data sources. In this paper we have described key features of multi-sectoral collaborative open data applications and presented an instance of such an application in the area of integrated water resources management.

ISWMS^{TM} involves a combination of a robust physically based model with computerized calibration techniques, and includes a real time flood forecasting system with the designed capability of accurately predicting major flooding events. The complex multi-sectoral collaborative open data application can not only allow for the reduction of infrastructure damage, and loss of life incurred during floods, but can also aid in the prediction and understanding of yearly spring-freshet events and allowing more succinct timing operations of reservoirs and dam structures. Informed decisions on a real-time basis will also allow watershed managers, policy makers and scientists to seek optimal use of water resources and to balance a river basin's ecological functions flood prevention and hydropower generation potential.

In terms of future work, from a software development perspective, there are many open research questions related to collaborative multi-sectoral open data applications, including [25]:

- What methods can be used to identify the open data needs of the enterprises and communities of practice involved in a multi-sectoral collaborative open data application?
- What methods and criteria can be used to identify appropriate data stewards and to systematically (and maybe automatically) allocate data to them?
- How to coordinate open data governance activities in the case of complex multi-sectoral collaborative open data applications that involve multiple partners, multiple data sources and multiple sub-systems?
- What methods can be used to resolve data issues and inconsistencies when data comes from multiple source and in different formats?
- What methods can be used to make sure that open data

can be located easily and to ensure that it is current in a meaningful way?
- What tools can be defined to support the access and integration of multiple data sources where new data sources can be dynamically introduced, some data sources may change, and there may be semantic data differences?
- How to define the requirements of open data applications that involve changing and dynamic datasets (i. e., datasets in which data changes or new datasets that can be introduced into an application dynamically)?
- How to define the requirements for the data, the extension and the integration of open data applications that involve multiple datasets in which some data has different semantics?
- What open data change processes can be adopted when open data applications involves multiple data sources provided by different stakeholders?
- How to design open data applications that involve dynamic event-based notifications and contextual data changes?
- How to define methods that can support the integration of multiple open sources where some of the data sources are private?
- How to define and manage open data access views that depend on sector, stakeholder, and group or individual roles?
- How do we define governance and operational policies for sustainability and evolution of a multi-sectoral collaborative project?

Overall, although multi-sectoral collaborative efforts in general face many complex problems, we believe these efforts are certainly laying the groundwork for tackling "big" societal problems (e. g., water resources management, urban development) and for establishing a new paradigm in which the diverse views of stakeholders in different sectors are brought together to share information and lead to improved decision making. In this sense, automated approaches and processes (e. g., novel software systems and applications) are certainly playing a key role in advancing the state of the art of multi-sectoral collaborative initiatives.

Acknowledgment

The authors would like to thank the Ontario Research Fund, Natural Sciences and Engineering Research Council and the Government of Canada FedDev program for supporing portions of this research.

References

[1] V. de Ven, A. H., and G. Walker, "The Dynamics of Inter-organizational Coordination," *Administrative Science Quarterly*, vol. 29(4), pp. 598–621, 1984.
[2] M. Morris, "Getting it Right: Understanding Multi-Sectoral Collaborative Change Initiatives A Study of Regional Development Agencies in England," Master's Thesis, University of Northern British Columbia, 2010.
[3] C. Huxham and C. Eden, "The Negotiation of Purpose in Multi-organizational Collaborative Groups," *Journal of Management Studies*, vol. 38(3), pp. 373–391, 2001.
[4] C. Huxham and S. Vangen, "Working Together: Key Themes in the Management of Relationships between Public and Non-profit Organizations," *International Journal of Public Sector Management*, vol. 9(7), pp. 5–17, 1996.
[5] C. Huxham and D. Macdonald, "Introducing Collaborative Advantage: Achieving Interorganizational Effectiveness Through Meta-Strategy," *Management Decision*, vol. 30(3), p. 50, 1992.
[6] M. Guffey, "Empowering Collaborations inthe Hollow State," *International Journal of Public Administration*, vol. 29(8), pp. 561–575, 2006.
[7] P. Mattessich, M. Murray-Close, and B. Monsey, *Collaboration: What Makes it Work*. Wilder Foundation, Saint Paul, Minnesota, 2004.
[8] J. Bryson and B. Crosby, *Leadership for the Common Good: Tackling Public Problems in a Shared-Power World*. Jossey-Bass, San Francisco, California, 1992.
[9] GWP, "Global Water Partnership: Integrated Water Resources Management," Background paper 4, Stockholm, Sweden, 2000.
[10] ——, "Global Water Partnership: Informal Stakeholder Baseline Survey: Current Status of National Efforts to Move Towards Sustainable Water Management using an Iwrm Approach," Stockholm, Sweden, 2004.
[11] UNESCO, "Water: A shared Responsibility. The United Nations World Development Report 2," UNESCO, Paris, France, 2006.
[12] R. Hirji and R. Davis, "Strategic Environmental Assessment: Improving Water Resources Governance and Decision Making," The World Bank, April, 2009.
[13] P. Molinga, A. Dixit, and K. A. (Eds.), "Integrated Water Resources Management: Global Theory, Emerging Practice and Local Needs," SAGE Publications, New Delhi, Thousand Oaks, London, 2006.
[14] K. Cheever, "Collaborations in Public Service: Memphis Experience," *International Journal of Public Administration*, vol. 29(7), pp. 533–555, 2006.
[15] A. Himmelman, "Communities Working Collaboratively for a Change," The Himmelman Consulting Group, Minneapolis, 1992.
[16] S. Morse, *Building Collaborative Communities*. Pew Partnership for Civic Change, Charlottesville, Virginia, 1996.
[17] J. Woulfe, T. Oliver, S. Zahners, and K. Siemering, "Multisectoral Partnerships in Population Health Improvement," *Preventing Chronic Disease*, vol. 7(6), 2010.
[18] Institute of Medicine, "The Future of Public Health in the 21st Century," National Academy Press, 2003.
[19] A. Costello and A. Zumla, "Moving to Research Partnerships in Developing Countries," *British Medical Journal*, vol. 321, pp. 827–829, 2002.
[20] WB, "Water Resources Sector Strategy," World Bank, Washington, DC, 2003.
[21] M. Solanes and F. Gonzalez-Villarreal, "The Dublin Principles for Water as Reflected in a Comparative Assessment of Institutional and Legal arrangements for Integrated Water Resources Management," Global Water Partnership, TAC Background Paper No. 3, 1999.
[22] "Open Data Institute Business Plan," http://theodi.org/about-us, 2012.
[23] "Open Data White Paper Unleashing the Potential," https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf, 2012.
[24] "US Open Data Institute," http://flowingdata.com/2013/10/29/u-s-open-data-institute/, 2013.
[25] D. Cowan, P. Alencar, and F. McGarry, "Perspective on open data: Issues and opportunities," in *Software Summit 2014, International Conference on Software Science, Technology and Engineering (SWSTE)*. IEEE Computer Society Press, 2014.
[26] WB, "Strategic Environmental Assessment and Integrated Water Resources Management and Development," World Bank, Washington, DC, 2007.
[27] "The Integrated Science and Watershed Management System (ISWMS^{TM})," http://www.grnland.com, 2014.
[28] "Open Knowledge Foundation," opendefinition.org, 2014.
[29] T. Berners-Lee, "Linked Data Design Issues," W3C, 2009.
[30] EC, *Data an Engine for Innovation, Growth and Transparent Governance*. European Union, 2011, ch. Open Data an Engine for Innovation, Growth and Transparent Governance.
[31] J. Gurin, *Open Data Now*, 1st ed. McGraw Hill, 2014.
[32] A. Koller, "Opening Open Data," Opening Open Data http://www.w3.org/2013/04/odw/papers, 2013.
[33] "G8 Open Data Charter," https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf, 2013.