

# Developing a Collaborative Cloud-Based Platform for Watershed Analysis and Management

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**—Water resources, which are a common good, are often negatively impacted through changing patterns of land use, land development, resource extraction and climate. Their sustainability through use and management are often shared responsibilities that involve multiple sectors, including federal and provincial governments and their agencies, municipalities and conservation authorities, and private organizations. In this paper we describe ongoing work on the development of a collaborative multi-sectoral cloud-based platform to support the data sharing, integration and processing requirements for watershed analysis and management. This web-based and mobile cloud-based platform is designed to be a system that integrates science and watershed management.

## I. INTRODUCTION

Environmental, geological and hydrological water resources are constantly being altered through changing patterns of land use, land development, resource extraction and climate. These fluctuating patterns can negatively impact the preservation of surface and ground water resources, the maintenance of water quality and availability, the stewardship of aquatic environments, and can require re-engineering of infrastructure for the management of stormwater and flooding ([1], [2], [3]).

To deal with these critical problems, watershed systems and platforms involve distributed, highly fragmented, data intensive collaborative efforts that depend on numerous sectors and organizations, and must follow demanding data sharing, integration and processing requirements across project lifecycles [4]. However, there are still many challenges to establishing collaborative data sharing, integration and processing in this setting, including:

- The need for an overall data management policy that can support sustainable collaboration and cope with the lifecycle of projects dealing with multi-organizational watershed-related resources (e. g., datasets);
- Project resources are often gathered by different systems and stored in different organizations, which have limited computing resources (e. g., limited computers, connectivity, persistence);
- The resources, for example, datasets, are often ad hoc and are independently managed by experts in specific sectors, and the process of managing the collaborative interactions between multiple partners can be quite challenging;

- The resources (e. g., data and process resources) are often owned by single companies and data access, in this case, must follow the set of policies of a specific company;
- The resources are associated with different projects that have different access policies;
- The users of such resources often do not have support in making changes on portable mobile devices so that the data can be processed remotely, the requested updates can be made, and the provenance information can be properly recorded for future use.

As a result, data sharing, integration and processing, as well as support for the coordination of experts in different projects and sectors become extremely difficult. For such reasons, we believe that a cloud-based platform for watershed analysis and management can provide a more efficient and robust solution for organizations in different sectors to collaborate.

In this paper we describe ongoing work on the development of a collaborative multi-sectoral cloud-based platform to support the data sharing, integration and processing requirements for watershed analysis and management by succinctly discussing some data governance requirements and presenting a cloud-based layered architecture for the system. This web-based and mobile cloud-based platform is designed to be a system that integrates science and watershed management.

As cloud computing is still an emergent technology when it comes to applications to support watershed analysis and management, many existing systems still rely on server-oriented solutions [5]. However, these solutions require infrastructure and maintenance within the organizations that are using them, and the efforts to keep the systems up to date and to maintain the applications can be very costly and time consuming. In addition, existing solutions also make use of proprietary file formats that make data integration extremely difficult. These solutions also lead to interoperability issues since the complete mapping between different formats is not always possible due to formats that have proprietary extensions.

## II. DATA SHARING, INTEGRATION AND PROCESSING GOVERNANCE

The cloud-oriented governance of the watershed resources data needs to take into account the collaboration between various stakeholders involved in the lifecycle of a watershed-related project. The governance should focus on the data

sharing, integration and processing requirements, which relate to collaborative system features, and include:

- The watershed data resources (e.g., daily precipitation, water holding capacity, catchment and stream delineation), which is scattered among multiple jurisdictions and is usually available as open data ([6], [7]), need to be provided in standardized ways and integrated. 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 data governance model needs to rely on a watershed resource information model (WIRM) that represent the lifecycle of watershed-related projects. The WIRM model needs to meet the specific data sharing, integration and processing requirements. The model can include, for example [8], watershed *resources*, *stages* and *gates*, which identify sets of mandatory or optional *requirements* that should be met at a certain stage, and the resources should be accessed (i.e., access *rights*) by *users* who are experts in specific *sectors* and can perform a specific *role*. In addition, a WIRM can be associated with project stages and a project stage can be associated with many WIRM artefacts;
- The data and process models should be accompanied with automated support for accessing and integrating multiple data sources, especially when new data sources can be dynamically introduced, some data sources may change, and there may be semantic data differences [7].
- The data sharing, integration and processing models should be able to explicitly record provenance data related to the resource changes, transformations and simulations. The provenance data should take into account roles and contextual information about the resource changes that take place in each stage [9];
- Methods should be defined to support the integration of multiple resources where some of the data sources are private. In some cases, techniques for data obfuscation should be provided [7].
- Resource access views (e.g., open data access views) should be provided may depend on sector, stakeholder, and group or individual roles [7].

### III. THE CLOUD-BASED LAYERED ARCHITECTURE

The collaborative platform for watershed analysis and management has been developed under a partnership involving Greenland International Consulting Ltd (Greenland), the Computer Systems Group at the University of Waterloo (UWCSG) and the Centre for Community Mapping (COMAP), which are cooperating to design, implement and deploy an initial cloud-based version of the system. This web-based and mobile cloud-based platform is designed to be an Integrated Science and Watershed Management System (ISWMS<sup>TM</sup>) [10]. A version of the system is shown as three layers in Figure 1. The three layers are: (1) Open Data and Other Data Sources; (2) Science and Engineering; and (3) Human Interface.

A possible design involves using the CometCloud platform [11], in which a virtual space for storing data can be implemented. This platform supports both data sharing and process execution (e.g., simulation). CometCloud enables on-the-fly integration of local computational capabilities (e.g., data centers, grids) and public on-demand cloud services such as Amazon, Eucalyptus and EC2. The latter feature is one of the benefits of CometCloud, which leads to the possibility of dynamically expanding the system based on user demand. The provision of advanced methods to support the aforementioned integration and dynamic expansion are relevant challenges for existing cloud platforms.

#### A. Open Data and Other Data Sources Layer

The Open Data layer indicates the source of data, much of which will come from governments and NGOs such as Conservation Authorities (CAs)<sup>1</sup>. Current weather data comes from the Federal government, while a significant amount of water quality data comes from the provincial government while gathered by local CAs through such systems as the Flowing Waters Information System. Present and future land use data is typically provided by municipalities, while drainage characteristics such as catchment and stream delineation are derived from digital elevation maps. Remote sensing and field data may come from CAs, university researchers, professional consultants or government. An Open Data layer will provide multiple advantages:

- Data will not have to be gathered every time there is a requirement to analyze or model aspects of a watershed, it will be accessible through the Open Data layer;
- Projects will be able to build upon the results of previous analysis and modeling runs, thus able to anticipate cumulative effects of watershed developments; and
- Costs of understanding and managing the watershed should be significantly lower.

The issue with the Open Data layer is to provide easy access and to ensure that the data is current. This is a significant but manageable problem that can be addressed.

#### B. Science and Engineering Layer

The Science and Engineering layer in Figure 1 shows the tools that have been developed or are under development. Many of these tools have become mainstream in the last decade and so are relatively new to the watershed management community.

Box (1) summarizes the functionality of the CANWET<sup>TM</sup> (CANadian Watershed Evaluation Tool). Since 2004, CANWET<sup>TM</sup> (Versions 1 through 4) has been used to develop a Lake Protection Plan for a large lake; identify new infrastructure solutions and sustainable community planning policies; prepare regional Official Plan directives; and complete project related climate change impact assessments.

---

<sup>1</sup>In Southern Ontario, Canada conservation authorities are responsible for managing and maintaining watersheds that are in one province.

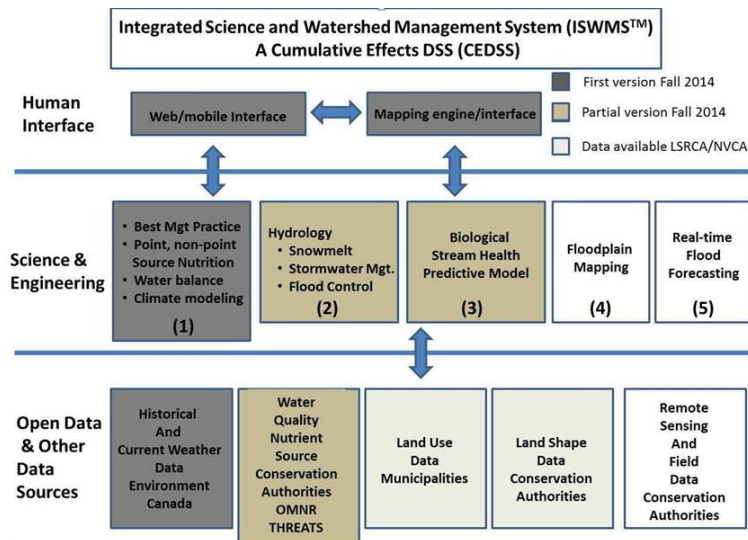


Fig. 1. ISWMS<sup>TM</sup> - Cloud-Based architectural layers and components.

This part of the diagram - Box (1) in Figure 1 represents CANWET<sup>TM</sup> (Version 5), which is a web-based cloud application based on CANWET<sup>TM</sup> (Version 4.2) and being deployed in the Fall of 2014.

Box (2) in Figure 1 shows the first version of the ISWMSFF<sup>TM</sup> (Integrated Stormwater Management and Flood Forecasting tool) that is now being upgraded to a web-based cloud application with added functionality.

Box (3) in Figure 1 shows the possibility of using Predictive Models for Biological Stream Health through measurements of water course shape, fish and plant populations, and benthic data. The Flowing Water Information System (FWIS), a joint project of the Ontario Ministry of Natural Resources, Ontarios CAs and the Centre for Community Mapping contains a significant amount of this data for Ontario.

Boxes (4) and (5) Figure 1 identify significant problems that we have encountered with climate change. We do not have accurate floodplain maps considering the extreme weather events that are occurring more frequently. These floodplain maps could be developed using digital elevation mapping acquired by Lidar<sup>2</sup>-equipped drones at a reasonable cost. The system could use available in-stream/river bathymetry data. Real-time flood forecasting is another area where progress could be made by analyzing snowmelt and weather data.

### C. Human Interface Layer

The Human Interface layer Figure 1 allows the user of ISWMS<sup>TM</sup> to view the results of modeling and analysis. The mapping engine/interface provides the ability to outline the area to be studied and to visualize the results related to a map. The Web/mobile interface incorporates the maps as well as reports and graphical presentation of the results. The mobile version could also be used for data collection in the field. The web-based cloud system contains the interface just described when deployed in the Fall 2014.

<sup>2</sup>Lidar is a remote sensing technology that measures distance by illuminating a target with a laser and analyzing the reflected light. Lidar is popularly used as a technology to make high-resolution digital elevation maps.

## IV. CONCLUSIONS

In this paper we have briefly described ongoing work towards the development of a collaborative cloud-based platform for watershed analysis and management. The system relies on a governance model based on the notion of watershed resource information model, which meets the strong data sharing, integration and processing requirements of these applications.

## REFERENCES

- [1] GWP, "Global Water Partnership: Integrated Water Resources Management," Background paper 4, Stockholm, Sweden, 2000.
- [2] R. Hirji and R. Davis, "Strategic Environmental Assessment: Improving Water Resources Governance and Decision Making," The World Bank, April, 2009.
- [3] 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.
- [4] 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.
- [5] P. Mel and T. Grance, "Nist Definition of Cloud Computing," National Institute of Standards and Technology, 2009.
- [6] "G8 Open Data Charter," [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/207772/Open\\_Data\\_Charter.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf), 2013.
- [7] 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, pp. 24–33.
- [8] T. Beach, O. Rana, Y. Rezgui, and M. Parashar, "Cloud Computing for Architecture, Engineering and Construction Sector: Requirements, Prototype and Experience," *Journal of Cloud Computing: Advances, Systems and Applications*, no. 2(8), pp. 1–16, 2013.
- [9] J. Freire, D. Koop, E. Santos, and C. Silva, "Provenance for Computational Tasks: A Survey," *Computing in Science and Engineering*, no. 10(3), pp. 11–21, 2008.
- [10] "The Integrated Science and Watershed Management System (ISWMS<sup>TM</sup>)," <http://www.grnland.com>, 2014.
- [11] H. Kim and M. Parashar, "CometCloud: An Autonomic Cloud Engine," in *Cloud Computing: Principles and Paradims*. John Wiley, 2011, pp. 275–297.