REPORT TYPE
OCRE-TR-2017-017

NRC·CNRC

Ocean, Coastal and River Engineering

Online Interactive Visualization Technology

Technical Report - UNCLASSIFIED
OCRE-TR-2017-017

Document Version: 1.4

Submitted to:
International Joint Commission (IJC)
Ottawa, Ontario

Prepared by:
Ivana Vouk, Richard Burcher & Vahid Pilechi
National Research Council – Ocean, Coastal and River Engineering (NRC-OCRE)
Ottawa, Ontario

July 5, 2017

National Research Council Canada | Conseil national de recherches Canada

Canada

Report Documentation Form

NRC·CNRC

Ocean, Coastal and River Engineering – Génie océanique côtier et fluvial

Program Marine Infrastructure Energy & Water Resources	**Project Number** A1-010700	**Publication Type** Technical Report

Title (and/or other title)
Online Interactive Visualization Technology

Author(s) – Please specify if necessary, corporate author(s) and Non-NRC author(s)
Ivana Vouk, Richard Burcher, Vahid Pilechi

Client(s)
International Joint Commission (IJC)

Key Words (5 maximum) Visualization tools, online, interactive	**Pages** 45	**Confidentiality Period** NA

Security Classification UNCLASSIFIED	**Distribution** UNLIMITED	**How long will the report be Classified/Protected?** NA

Limited Distribution List (mandatory when distribution is Limited)

Date:	VER #	Description:	Prepared by:	Check by:
27/03/2017	1.0	Preliminary draft	V. Pilechi, I. Vouk, R. Burcher	
04/03/2017	1.1	Comments from IJC	W. Jenkinson	
24/05/2017	1.2	Final draft	I. Vouk	E. Murphy
30/05/2013	1.3	Comments from IJC	W. Jenkinson	
07/05/2017	1.4	Final report	I. Vouk	E. Murphy

Program Lead **Signature**

Director of R & D **Signature**

NRC – OCRE Addresses

Ottawa
1200 Montreal Road, M-32
Ottawa, ON, K1A 0R6

St. John's
P.O. Box 12093, 1 Arctic Avenue
St. John's, NL, A1B 3T5

Executive Summary

The International Joint Commission (IJC) is currently initiating the modernization of its website (www.ijc.org) and revamping its online presence. This is providing an opportunity to integrate user-friendly interactive web-based visualization tools.

With advances in online technologies, visualization tools are increasingly becoming an effective form of communication that can emphasize key messages to broad audiences. If properly designed the tools can effectively communicate the insight garnered through the work by the IJC's boards, technical working groups and the International Watersheds Initiative, as well as to clearly portray its messages to stakeholders and communities within the transboundary basins.

The National Research Council conducted an in-depth review to establish where visualization tools could be used to the benefit of the IJC, and to identify the current technologies available to create these tools. The following is a list of criteria that were applied during the review and prototype development:

- The visualization tools must have a high quality aesthetic and a visual appeal catering to a broad audience.
- The software supporting the tools must be well established and employed by other organizations to reduce risk of obsolescence.
- The overall system design must ensure components can be interchanged and reused easily in the future.
- The visualization tools must be fully automated with only minimal maintenance required by the IJC, its boards and committees.
- The majority of input data to the visualization tools must be compiled from available online resources, requiring minimal and transient data storage by the IJC.
- Simplicity must be exercised to ensure for a quick response time to the end users
- The system must embrace technology that the IJC already has access to, most notable being the ESRI tools.

Although a number of boards and committees answer to the IJC, each with distinct responsibilities, priorities and interests; five common themes emerged: (i) data informing the management and operation of control structures, (ii) water quality, (iii) streamflow, (iv) biology data and (v) precipitation, snow and climate data.

A number of technologies were reviewed and discussed and many of these were used towards the development of prototypes. Based on IJC's identified needs and objectives, appropriate technologies were shortlisted. Final recommendations were developed and are provided below.

Recommendation 1: For applications where high quality aesthetics and fully customizable interactivity are a priority (e.g. to convey crucial messages, and / or information from high impact reports to the public and wide audiences) - Use JavaScript and its associated libraries (e.g. Leaflet, D3, DataTables) to develop online visualization tools.

Recommendation 2: For applications where aesthetic quality and interactivity are important but where there is a potential need for periodic and / or rapid modification of the tools (e.g. periodic

or routine dissemination of data to the public and / or IJC boards) – Use Python and its associated libraries (e.g. Bokeh, Plotly, Pandas) to develop online visualization tools.

Recommendation 3: For applications where technical users are the target audience requiring seamless integration with other program code (e.g. dashboards integrating existing in-house statistical tools) as well as increased development flexibility and user options, or for tools that the IJC may ultimately wish to be developed / managed by boards and technical groups (e.g. rapid dashboard development) – Use R and Shiny to develop and integrate online visualization tools.

Table of Contents

Table of Figures

Table of Tables

Acknowledgments

The authors would like to thank Wayne Jenkinson for unwaning support and enthusiasm during the development of this review; to Aroscott Whiteman for timely advice, input and wholehearted support towards prototype development during a busy period; and to the working group members (Mark Gabriel, Erika Klyszeyko, Paul Johnston, Andrew Liddiard, Matt DeWolfe, Gail Faveri, Chris Sanocki, Jim Stark, Aroscott Whiteman, Peter McCarthy and Kevin Strauss) who allowed for interesting discussion and provided further advice to improvement of the prototypes. The authors would also like to thank Enda Murphy for editing the report and providing useful feedback, as well as Andrew Cornett for additional funding to complete the review.

1 Introduction

The International Joint Commission (IJC) oversees the management of US-Canada transboundary watershed boards and technical committees, each of which performs one or a number of roles including monitoring flows and water levels, regulating dams, protecting water quality and apportioning water volumes. The IJC also carries out projects through the International Watersheds Initiative (IWI), through which scientific research studies are conducted, communication is promoted and water resource management is enabled. The studies help inform decision making at all levels of governance and to abate cross-border, contentious water-related issues.

The IJC is currently initiating the modernization of its website (www.ijc.org), which provides an opportunity to consider possibilities for integration of visualization tools in parallel. In parallel the IJC intends to develop user-friendly interactive web-based visualization tools. These are used to effectively communicate the insight garnered through the work by its boards and the IWI, as well as to clearly portray its messages to stakeholders and community within the transboundary basins. Such tools are needed to support timely and efficient communication of information and to emphasize key messages or points to be communicated. A broad suite of visualization tools are envisaged, ranging from simple, interactive charts to fully customized dashboards (i.e. a single, interactive user interface presenting a diverse range of content). In addition to data held by IJC, the tools need to be capable of integrating complementary information and data provided by other government organizations and agencies, such as climatic, hydrometric, water quality, and ecological data.

The National Research Council of Canada's (NRC)'s Ocean Coastal & River Engineering (OCRE) Portfolio has been developing visualization tools for a number of years and a variety of purposes. NRC has developed a platform known as EnSim on which a number of products have been based such as: Blue Kenue and Green Kenue, tools widely used for visualization, pre-processing and post-processing of hydraulic, hydrologic, climatic and GIS-based data; specialized products such as Environment Canada Data Explorer (ECDE), Marine Kinetic Energy Explorer (MarKE), Canadian Arctic Shipping Risk Assessment Solution (CASRAS), Beaufort Sea Environmental Database (BSED) and others. ECDE alone has a large number of users throughout Canada and the US who use it to view and analyze Water Survey of Canada's hydrometric data from their water gauges. NRC has also recently developed web based interactive visualization tools for the IJC and other clients.

IJC boards have been established to perform prescribed functions within each watershed. Consequently, each board has distinct responsibilities, interests and priorities; and therefore distinct requirements for visualization tools. To allow for efficient development, easy implementation and consistency of data visualization across boards and projects, it is envisaged that a small number of tools will be developed. The tools should be flexible, such that they can be tailored to a wide range of applications and transferrable for other future needs.

To support the development of data visualization tools to meet these needs and requirements, and to provide recommendations towards appropriate technologies for consideration the IJC engaged NRC to conduct a desk-based review of data visualization technologies. NRC's review included liaison with IJC to develop an in-depth understanding of their objectives and potential needs for data visualization, as a basis for informing the selection of appropriate development technologies and approaches. The NRC also developed a number of prototypes to demonstrate the capabilities of different technologies and their potential effectiveness in meeting IJC’s needs for data visualization. Based on the review, the NRC developed recommendations for potentially suitable technologies and development approaches.

Finally, online technologies evolve at a rapid pace, which, for organizations such as the IJC that are engaged in data collection and dissemination, can be challenging to keep up with. Mitigating the risks of investment in web-based data visualization tools therefore requires a careful, strategic approach to provide interactive visualization tools that will last and provide value for IJC and stakeholders.

2 Background

2.1 Objectives

The primary objective of the desk-based review is to identify a web-based interactive visualization technology or technologies that:

- effectively communicates information and issues related to IJC and board activities across the Canada-US international border;
- leverages existing IJC technology and expertise;
- is flexible enough to incorporate and display the wide variety of data types of interest to the IJC, including binational data and data from both the U.S. and Canada; and
- is highly automated, or could easily be automated, to facilitate efficient maintenance, stewardship and dissemination of data.

2.2 Overview of Methodology

NRC conducted an in-depth review of existing data visualization technologies. This report, which documents the review, is intended to be a comprehensive reference document for the IJC. It provides an overview of the interactive visualization technology domain and how it relates to the IJC's user requirements for data visualization tools.

A key focus of the review was the development of two prototypes to demonstrate current technological capabilities. The first focused on providing near real-time hydrometric and meteorological information to non-technical stakeholders or the general public. The second prototype was developed to help communicate transboundary water apportionment conditions to technical experts and stakeholders. The prototypes were developed taking into account the overall objectives of the study (Section 2.1) and key considerations presented in Section 2.3. The development process for the prototypes involved testing a number of alternative technologies and options.

2.3 Key Considerations

The following points were key considerations during the review of technologies and development of prototypes.

HIGH QUALITY OF AESTHETIC
IJC is striving for a high quality, interactive data visualization experience. The appearance of the data visualization interface (or interfaces) is important to the IJC to facilitate clear communication of messages and information.

TECHNOLOGY OF PREFERENCE
IJC prefers to use technologies that have been successfully adopted by other scientific organizations. This demonstrates that the technologies are accepted, tested and are capable of meeting IJC's requirements. Furthermore, it mitigates investment risks to IJC in several ways: (i) the technology will continue to be developed and supported into the future; (ii) a community of resources exists to solve known development issues, and (iii) IJC can access the expertise of knowledgeable developers, who are familiar with the technology.

LOW DEVELOPMENT EFFORT, REUSABILITY
IJC is emphasizing the design of reusable software components that will assist in reducing development effort for future components. Beyond the initial development phase, the additional development effort to implement interactive components should be low. For instance, many of the interactive components will be generic (e.g. water level time series, or maps displaying gauging station locations) with only background data sources and textual descriptions changing. Initial development and aesthetic refinements of these generic components should be easily applied to similar future components.

FULLY AUTOMATED
IJC is inclined towards automating processes that will generate and serve interactive components of the data visualization tools. This will serve to mitigate demands on IJC's internal IT resources for maintenance. It will also facilitate efficiency of development by allowing the system to scale to accommodate additional interactive components with minimal human input.

MINIMAL DATA STORAGE
IJC wishes to minimize the need for storage and retention of scientific data in-house, recognizing potential concerns with regard to the demands and resources needed for data storage and archiving as it supports a large number of projects. Requirements to present historical datasets or to conduct complex analysis may necessitate some limited storage (likely with a finite storage period) to serve the interactive components of data visualization tools. This should only be where warranted on a case-by-case basis, and provided that the storage mechanism does not require substantial internal IT resources.

SHORT RESPONSE TIME
IJC is striving for simple presentations of data that facilitate short response times. The desired web user experience for requesting and loading a webpage resource is less than 2 seconds, based on IJC's current web site renewal requirements. The majority of interactive components will simply present data (with minimal processing / analysis) and it should therefore be feasible to meet the desired response time.

LEVERAGE ESTABLISHED ESRI TECHNOLOGY AT THE IJC
The IJC has invested considerable resources into ESRI technology. ESRI software provides comprehensive Geographic Information Systems (GIS) and spatial analysis capabilities, upon which the IJC depends. Where possible and appropriate, ESRI software should therefore be embraced and integrated with the data visualization tools.

3 Review of IJC's user base and technical capabilities

IJC has expressed an interest in creating user-friendly interactive web based visualization tools that will help the IJC raise stakeholder awareness of the IJC's and its boards efforts and contributions, increase public understanding and knowledge of the transboundary watersheds and help build consensus and support for IJC's endeavours. The tools are meant to be interactive in order to engage the user with provided information, but to do this the tools must effectively communicate the information provided. The NRC worked with the IJC to better understand IJC's user base and in identifying appropriate information that the visualization tools are to present.

The IJC operates its own website and owns some IT infrastructure and therefore already has in-house capabilities with respect to managing online technologies. The NRC with IJC's help has completed a preliminary assessment of IJC's IT capabilities, the results of which will be used to recommend an appropriate framework for the integration of visualization tools to the IJC's website.

3.1 IJC's user base and the need for visualizations

The availability of accurate data, organized and presented in a timely and easily understood manner is critical to support IJC's decision-making and communication with stakeholders within the potentially affected watersheds. This information can be accessed online through simple to use interactive tools that are easily understood by the end users (i.e. users who are accessing the tool online and who are not necessarily interested in the underlying data sources). Interactive tools have shown to be educational and engaging and have shown to improve retention of the provided material. However, the content and information should be simple, easy to understand and should never overwhelm the end user. In other words, the intention of the tool must be clearly identified. Thus, the following questions are of important consideration:

- Who is the user of the tool?
- What does the user require?
- Why does the user need such a tool? What problem is being solved?
- Are there examples or use cases that demonstrate the requirement and how it might be met?
- What can be built to address that need?

To help answer the above questions, and therefore identify the needs and interests of the IJC (as well as its stakeholders and website users), the NRC actively communicated and met with the IJC advisors, the boards and IJC's partners (e.g. USGS, ECCC). Additionally, IJC advisors reviewed a number of reports and products produced by the IJC and its boards and consulted additional members and liaisons of the boards to evaluate the needs for visualization tools across the organization. The advisors identified the data that are currently collected and required to fulfill the needs of boards' directives, identified the sources of data currently used by the boards and provided suggestions as to how the data might be presented. The results of this evaluation are available in Appendix A. Although the information needs vary between various boards and committees, five common themes emerged:

1. **Control structures**
   The IJC has a responsibility for managing and operation a number of dams and hydraulic structures. For these control structures there are often rules of operation, allocation,

water level management, etc. which directly concern stakeholders in the watersheds. Up to date information on water levels, discharges, inflows and key elevation or flow targets such as rule curves are of great interest to the boards and stakeholders.

2. **Water quality**
   Several watershed boards as well as the Great Lakes boards have a mandate to consider water quality issues in their basins. Although not every board has this mandate, several do, and it typically involves cooperating with agencies to review multiple measurement parameters at key locations (transboundary locations), collected by grab sampling multiple times per year every year, with key alert levels or objective targets. The presentation of these data along with trends year over year would be valuable to stakeholders.

3. **Streamflow**
   In general, the presentation of streamflow data at boundary or transboundary locations is important for all boards, and directly impacts issues like apportionment, flooding and drought management. These data involve hydrometric data (levels and flows) from stream gauges managed by federal, provincial or state agencies.

4. **Biology data**
   Although less common among the boards, certain boards provide regular data collected relating to fish and benthic invertebrates.

5. **Precipitation, Snow and Climate Data**
   Although not generally collected by the boards at this time, the importance of historical and forecasted precipitation was identified to be of great importance to the boards and their stakeholders. Those boards with control structure management obligations require these data to assist with decision making, and to help explain to stakeholders why decisions regarding control structures are being made. With recent progress in data reanalysis tools (i.e. CaPA, MPE and SNODAS) new opportunities are provided for the collection and presentation of precipitation, snow accumulation and forecasted water supply information.

Most of the data identified through discussions and by the above list have spatial and/or temporal aspects. Spatial information includes data with geographical properties which can be presented on maps (e.g. basin outlines, gauge station locations or maps showing how much precipitation has fallen). Temporal data generally denotes a change in a system over time. In fact, the majority of the parameters of interest in this project are temporal. The frequencies of time vary from hourly to seasonal (e.g. hourly water level measurements from a hydrometric gauge, or seasonal fish samples). Thus, the majority of the visualization tools will involve a representation of temporal and spatial data.

A number of datasets that the IJC and its boards use are served by third-party data providers. These generally originate from federal, state or provincial agencies that collect or process field data that are necessary to operations of organizations such as the IJC and its boards (e.g. streamflow, water levels, climate, water quality, etc.). Thus, the visualization tools will have to be served by the same data providers and will require the capability to access such data.

In some cases the IJC and its boards produce special calculations using available data, such as calculations of natural flows, development of water quality indicators, identification of hydraulic conditions, development of runoff maps, calculations of transboundary water allocations, etc. This aspect is an important consideration towards the development of visualization tools, because it requires processing, intermediate work and engineering knowledge to accomplish. Thus, some of the visualization tools will necessitate the processing of existing data to produce value-added data to the IJC and its boards.

Using the knowledge gained from above, the NRC defined four broad user categories that engage with the IJC based on their unique functional and data needs. The NRC developed Table 1 to help assist in the future design and development of visualization tools. This was a necessary step in order to i) identify the user groups that will be served by the visualization tools; ii) facilitate user needs with respect to available data types and interactions and iii) define how a technology may support i) and ii) above. Although the list of Table 1 is not exhaustive, it ensures that the NRC used a people-first approach to determine needs before reviewing and recommending potential technologies.

Table 1: Identification of IJC user groups, their requirements and potential solutions.

User group	Requirements	Purpose	Example	Potential Solution
Technical individuals who understand complex data (e.g. hydrologists, scientists, etc.).	Provision of historic, near real-time and pre-calculated data in an automated fashion.	To present regulatory requirements. To indicate collaborative work. To increase visibility of available information. To share tools internally amongst members of a board or a technical committee and other researchers.	An Accredited Officer wants to easily communicate and query water apportionments between Canada and the U.S. to a stakeholder. Dam operators want to compare forecast and near real-time data with historic data to better manage dam gates.	Interactive charts that include current data, historical analysis, ability to select between various languages, units of measure, etc. Maps. A summary table.
Public who wants to understand complex data in a simplified fashion.	Provide historic, present, and forecast data both spatially and temporally	Transparency of IJC's activities Understanding of current and/or future conditions. Understanding of reasoning behind regulation decisions. Comparison of current and historical conditions. Comparison of conditions between basins.	A lakeside property owner wants to understand why a control structure is being operated in a particular way that may affect boat access, or flood risk. A resident wants to know why lake levels are low in a controlled lake when there was no recent rain storm. A farmer wants to know the water level in the lake next to their property in order to understand possible flood risk.	Dynamic maps that have a number of spatial layers that can be turned on or off. Interactive charts. Linkage between a map and charts (e.g. filtering the chart will highlight points on a map). Processing of available data to communicate a particular message (e.g. what is the cumulative precipitation over the last seven days in a basin).
Engaged public (e.g. researcher, reporter, or politician), IJC staff and stakeholders.	Overview of key basin metrics and data	To provide a summary of basin conditions (e.g. health of the environment, inter-jurisdictional water demands or cross-border flows). Current and/or planned activities. Currently available research. A summary of available data.	Public would like to understand how fish counts have been changing over the years for a particular lake and be able to access the associated research reports. A politician wants to know how the average of total precipitation has changed during last ten years in a specific basin during the summer month of August.	Static maps with overlays and filtering ability (e.g. selecting multiple basins). Charts with filtering of (e.g. grouping, filtering out or aggregating) data for the selected basins. Processing of available data to communicate a particular message (e.g. summary statistics of historical water flow at several hydrometric gauges).

User group	Requirements	Purpose	Example	Potential Solution
IJC staff (e.g. communi-cations department).	Provide insight about the IJC and stakeholders activities that are related to public interest.	Public education and outreach. Transparency of activities.	The IJC wants to provide a social media post (e.g. blog) about recent SPARROW model results related to Lake Erie phosphorus loads. The IJC wants to provide a summary overview of the transboundary watersheds water status (e.g. drought status).	Maps that are linked to the social media post (e.g. blog) Interactive charts that are linked to the social media post and map.

The diversity of broad user groups in this project makes customizability of the selected technologies a key requirement. The information provided to the end user must communicate the meaning of the data in an efficient and effective manner. The visual appeal of the potential visualization tools, especially those available publically through the IJC website, is very important to this project and thus to the selection of technologies used to deliver the final tools. Aesthetics of the visualization tools can be less of a deciding factor for very technical audiences, and more specifically if the tools are used internally amongst the IJC and its boards' members to share technical information.

3.2 IJC's Technical Capabilities

The NRC conducted a review of the IJC's current technical capability in order to better understand the internal resources, technical inclinations and infrastructure. The gathered information summarized in Table 2 assisted the NRC in determining an appropriate technical solution to integrate interactive web components.

Table 2: Assessment of IJC's technical capabilities and constraints

Area assessed	Assessment
Review of IJC's internal technical staff and resources	• Small internal team. • Detailed skill sets of the team are unknown but familiarity with Windows administration and ESRI products is present. • One staff member responsible for the internal ESRI GIS infrastructure. • Unlikely that staffing levels will increase.
Clarification of technologies the IJC currently uses, licenses and has preferences to.	• Preference for software that is predominantly of commercial nature supplied by commercial companies (e.g. Microsoft, ESRI, etc.) where support for technical issues is available. • The operating system is Microsoft Windows based for both desktop and server needs. Third-party hosted website server is Linux based. • GIS is complemented by ESRI, where IJC has invested in desktop licenses, the ArcGIS Server, and ArcGIS Online. • Open Source technologies are supported where it provides business value and has a strong community support. • Preference to reduce cost of IT ownership and internal IT resources by utilizing third-party server hosting. This is based on the IJC's current web site renewal tender requirements that explicitly state third-party hosting.

Configuration of IJC's current IT infrastructure as it pertains to hardware, software and available facilities.	• Internal servers:
	  • Multiple servers are present. IJC uses Microsoft virtual servers. It also owns an FTP server.
	  • www.gis.ijc.org is running the Microsoft webserver IIS (version 7.5).
	• Third-party servers:
	  • www.ijc.org is externally hosted using Open Source technologies (e.g. Linux operating system, Apache webserver).

Area assessed	Assessment
Identification of internal and external requirements that may impact technology considerations.	• IJC is a bi-national organization and as such has a degree of autonomy over how it's IT security is managed (i.e. it does not fall under the Treasury Board of Canada requirements).
Review of NRC's capability to access IJC's infrastructure remotely in order to determine and allocate NRC and IJC staff resources for potential future production of tools.	• During this phase, NRC provided prototypes to the IJC through a software versioning and revision control system on NRC's CHyMS environment, which the IJC staff uploaded to IJC's web server space. • IJC's FTP server is actively used by project partners for the IWI SPARROW project where it is used as a collaborative space for sharing information.

Review of IJC's previous experience with various web based technologies to understand IJC's comfort in the area	• The main IJC website contains embedded ESRI web maps served from IJC's internal ArcGIS Server.
	• Evidence of internal experimentation with interactive technologies.
	• IJC hosts a wide variety of web based technologies that serve a wide range of public and internal interests:
	  ◦ IJC website and internet application servers
	  ◦ Interactive tool for the International Rainy River Lake of the Woods Watershed Board and the Lower Pembina River Flooding Task Team (currently under development)
	  ◦ Online maps and story-maps that are linked to websites of the boards
	  ◦ Collaborative democracy tool at: http://www.participateijc.org/public-meetings
	  ◦ Great Lakes-St. Lawrence research inventory at: http://ri.ijc.org/
	  ◦ Great Lakes association of science ships at: http://www.canamglass.org
	  ◦ Online videos explaining issues in the binational basins (currently under development)

4 Data Providers

The IJC and its boards are served by a diversity of data sources and data providers and use a variety of parameters for their studies. A large portion of these data originate from external or third-party data providers, which are generally collected and published by federal, state or provincial departments and agencies that collect data to fulfill their own mandate. The remainder of data are collected or compiled by the IJC and its boards internally.

Data by third-party data providers are nowadays provided online disseminated using an assortment of web based approaches. Many of the providers offer data using *web services*, a form of communication between a data requester and a data provider through some program code. This makes data retrieval automatic and can be embedded within the automation of a visualization tool. The NRC has taken time to understand the protocols associated with web services of some of the more common data providers (discussed further below) the IJC is served by, as discussed further below. More specific details for some of the third-party data providers discussed below can be found in Appendix B.

4.1 NOAA

National Oceanic and Atmospheric Administration (NOAA) provides a variety of datasets related to conditions of the oceans and the atmosphere. NOAA's National Center for Environmental Information (NCEI) – Climate Data Online (CDO) offers several free web services to access data that are available on their website. The IJC and its boards employ NOAA's historical and forecast data for precipitation, temperature and wind within their projects, which are offered through two different web services named *API v2* and *API Legacy*.

4.2 CHS

Canadian Hydrographic Service (CHS) at Canada's Department of Fisheries and Oceans (DFO) provides a wide range of hydrographic data including water depths, tides, currents and water levels, and sea bottom characteristics. IJC's Lake Superior and Lake Ontario – St. Lawrence River watershed boards use CHS's water level data, for which predictions, observations and forecasting are available. CHS provides three web-based water level services (Fisheries and Oceans Canada, 2016).

4.3 USGS

U.S. Geological Survey (USGS) provides a broad range of data related to surface water (water flow and levels in streams, lakes and springs), water quality and groundwater in the U.S. The *USGS Daily Values Site Web Service* provides daily statistical data about hydrologic sites served by the USGS (U.S. Geological Services, 2016). A number of IJC's watershed boards are served by this data.

4.4 ECCC

Environment and Climate Change Canada (ECCC) provides climatic data for current and forecasted weather conditions across Canada and into the U.S., stream flows and water levels for Water Survey of Canada gauges located on Canadian lakes and rivers, and water quality monitoring data of national and international interest. The data is available freely to the public. However, for a fee, ECCC also offers limited services for the delivery and customization of specific information. All of the IJC's watershed boards utilize ECCC data for water levels, streamflow, water quality and climate data.

ECCC also offers *GeoMet Geopspatial Web Services* which provide access to weather prediction data layers and the weather radar mosaic. These layers are presented as graphic images to the requesting user. The service is free; however, the data are not appropriate for meteorological data processing.

4.5 IJC's internal sources

In addition to the above mentioned third-party data providers, the IJC and in particular its boards collect and assemble unique and valuable datasets that would serve the visualization tools. Some examples of these sources include precipitation data from IJC International Watersheds Initiative (IWI), water quality data collected by agencies throughout the basin, data related to hydraulic control conditions and control structures owned by different boards, and data obtained from various coordinating committees within the IJC (please refer to Appendix A).

5 Review and discussion of available technologies and development

User-friendly interactive web-based visualization tools are used to draw the audience to key messages and story lines. For the IJC, the intent is to raise stakeholder awareness of the IJC's efforts and contributions, increase public understanding and knowledge of their watershed, and help build consensus and support for IJC's efforts. The knowledge and information gained by the boards and from the IWI studies are of considerable importance to many stakeholders. Therefore, provision of visualization tools is an important component to IJC's communication efforts.

Visualizing the data can be provided in different methods, but are primarily in the form of maps, charts and tables. If developed properly, these components deliver features and functions that can support in an organization's decision-making. In simple terms, construction of the components require a particular system setup which contains software, a place to house the necessary building blocks of the components and a place to view the components by the end user. This section provides an introduction and a discussion to the technical background and development required to achieve the intended visualizations.

5.1 Introduction to system setup

Online data visualization tools are generally facilitated by a *browser* and a *server*. In simple terms, a browser is a software-based viewing platform which allows the user to view and interact with a visualization tool, while a server is a container (hardware and software) where most of the background processing occurs, the program code resides and where data may be located. The browser and the server communicate back and forth when the user is accessing the visualization, thereby enabling the viewing and interactivity of the tool. The choice of browsers and servers depends on the types of visualization tools and desired functionality, as well as the end users' devices and operating systems. Browser and server are further discussed in Sections 5.2 and 5.3, respectively.

Communication between the browser and the server is managed by a program code that generally resides on the server. This program code also manages and coordinates a number of other necessary functions that are further discussed in Section 5.3.3.

Potentially the most significant portion for this review is the technology that serves the visualization tool and builds the visualization components (i.e. maps, charts and tables). This is discussed in detail in Section 5.4. Although NRC reviewed a large number of different technologies, only those deemed significant are discussed in this section.

The NRC also provided a brief review in Section 5.5 of some essential portions to the system setup required for the functioning of IJC's visualization tools which include data processing, data storage and additional programming code developed by the IJC and its partners.

5.2 Browser

More specifically, a browser is a piece of software which retrieves information from the server to present to the end user. A number of different browsers exist (e.g. Google Chrome, Microsoft Windows Explorer, etc.) each with its own features and specifications, which can affect the performance of a visualization tool. The end users' devices (e.g. desktop computer, laptop, notepad, mobile device, etc.) and computing specifications also affect the operation of the browser and therefore the performance of the visualization tool. For example, the processing speed of a browser depends on the processing capability of the device. Moreover, different browsers use different approaches for interpreting information and displaying graphical features. As shown in Figure 1 location of the horizontal axis font and the drop-down menus are different between the two examples even though the exact same program code is used to generate the visualization. The IJC has a very diverse user base (as illustrated in Section 3.1) and the number of browsers utilized by these end users can become large. For example, as of April 30, 2017 six different versions of Chrome made the top 10 list of web browsers by browser market share (W3Counter, 2017). Additionally, some browsers do not support some available visualization technologies. For example, *D3 JavaScript* (further explored in Section 5.4) is a popular online visualization technology which cannot be used in older versions of Microsoft browsers. All of these challenges require serious consideration when developing visualization tools. Therefore, minimizing computing requirements by the browser ensures that the interactive components load quicker for a better user experience. This can be done by minimizing the options of browsers a visualization tool is served by, by limiting file sizes that are downloaded by the browser and by limiting the processes performed on the browser. It is also important to notify the end users immediately when the requested components are not compatible with the version of their browser, which is often accomplished using warning messages, pop-up dialogues to upgrade a browser or other such fallback methods. During review and testing of various visualization technologies NRC determined that the following browsers provide a high level of support and have the highest user base: Firefox, Chrome, Safari, Opera, IE9 and above, Android and iOS.

The browser views information through webpages. Each webpage is backed by a language known as *Hypertext Markup Language* (HTML) which determines the structure and content of the information. HTML also embeds different types of resources such as style information (i.e. CSS), media (i.e. images, sounds and videos) and program code snippets (i.e. scripts) to add interactivity to the webpage. In summary, HTML files hold the information for the appearance and interactivity of a web page viewed through a browser.

Within the browser visualization tools are generally embedded into a webpage and identified by their own unique web accessible address. Simple tools are generally surrounded by some other content within the webpage, while larger tools (such as dashboards discussed in Section 5.4.4) become its own full page.

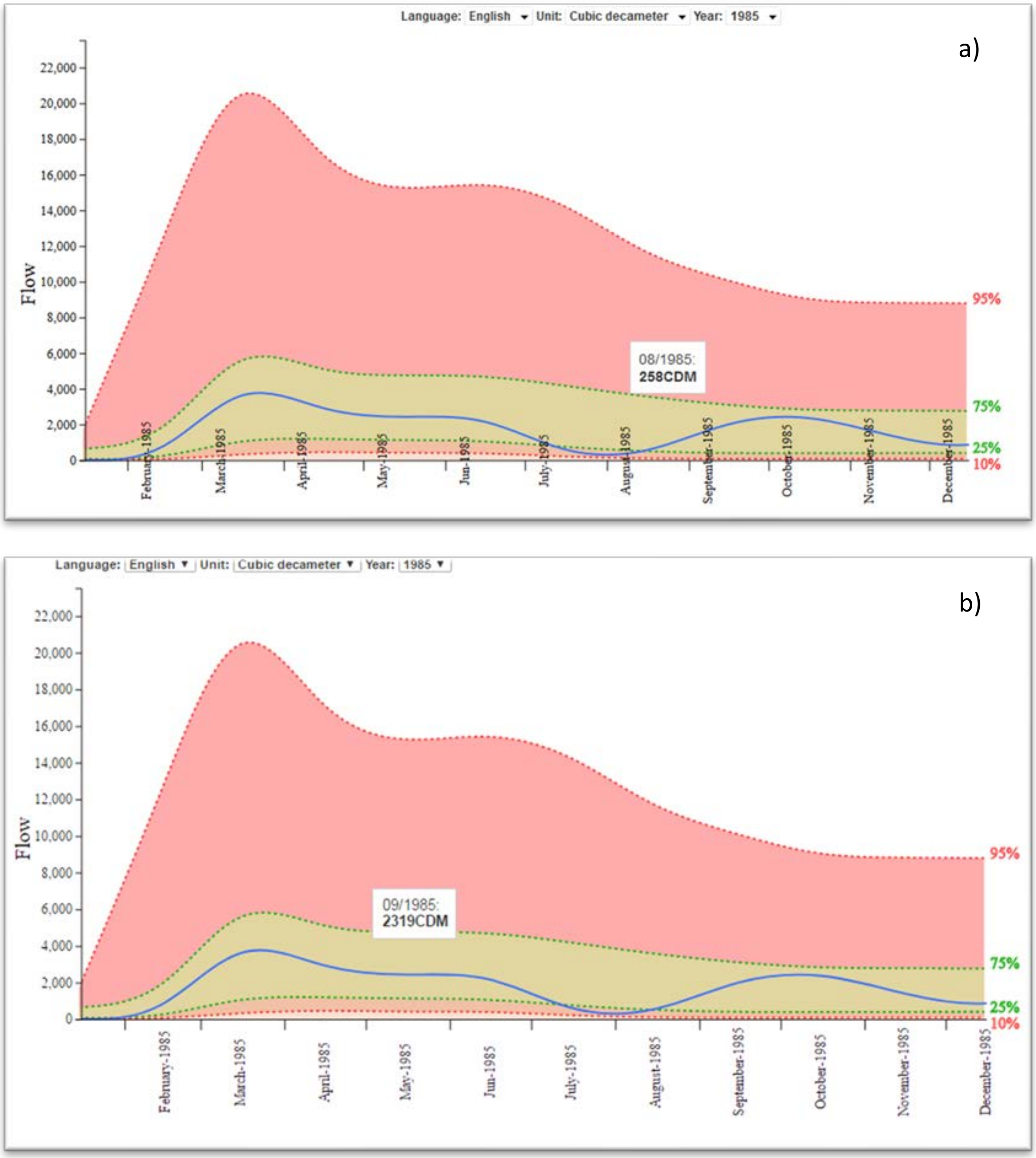


Figure 1: Example of slight differences seen in a visualization tool between two different browsers: a) Firefox and b) Google Chrome.

5.3 Server

A server is both a physical device and a piece of software that is required for functionality, processing as well as storage of HTML files, data files, and program code that are all necessary to support the operation of visualization tools on a browser.

5.3.1 Server location

Two types of servers worth differentiating are in-house and third-party servers. An in-house server is located on local premises and is fully administered and managed by local IT staff. A third-party server is located off-site, and can either be managed by the local IT team or a third-party provider. Managing a third-party server locally is similar to managing an in-house server.

Both methods require IT resources and skills to set up, configure, administer and manage the server. For that reason, many third-party server providers offer *managed servers* that remove the aforementioned requirements leaving only content management to local IT staff. Thus, managed third-party servers can reduce the effort required to setup and maintain servers as well as providing such services as: monitoring, administration, on-going maintenance, phone support, security, server back-ups, specialized and preconfigured software if required, appropriate amount of processing capability, data storage, and access to multiple servers if required. Using services by managed third-party server providers allows more time to be allocated on the development and maintenance of key software such as the visualization components IJC is planning on integrating to its website. In fact, IJC's current website is hosted by a third-party server provider.

IJC's current GIS server is hosted in-house. The NRC configured and tested an in-house server similar to IJC's. NRC found that using an in-house server provided full control over computer configuration, software installation and data storage. It also provided an ideal development platform for visualization technologies. Installation and basic configuration of the server were relatively easy. Configuration aspects addressing security and other such requirements were not carried out as this is currently not a testing requirement.

The NRC also tested a free version of a managed third-party server (i.e. shinyapps.io by RStudio) during the development of a prototype discussed in Section 6.2. This only required that the NRC's program code be uploaded to the server simplifying the testing process. Although used for NRC's testing, the free version limits a number of features and capabilities which are otherwise recommended.

Some clients choose to use multiple servers to divide operational needs between different servers. Most common factors determining the use of multiple servers are the amount of traffic (i.e. the number of end users) accessing a website (or in this case visualizations) and the size of information that is exchanged between the browser and server. If both are very large the use of multiple servers is explored. As it currently stands, neither IJC's traffic nor the amount of information provided to the end user warrants the need for multiple servers.

5.3.2 Server operating systems

An operating system (OS) is the software platform which supports all of the functionalities required to operate a device, or in this case, a server. Two different computer operating systems (OS) were tested in terms of server operation: Microsoft Windows (hereafter referred to as Windows) and Linux. Windows requires a license while Linux is available for free.

The Windows based OS was utilized almost exclusively during the review and testing of various technologies and development of prototypes. Overall there were no major issues in working with Windows. Using Windows OS in conjunction with Microsoft Internet Information Service (IIS) server several visualization outputs were tested (which are further discussed in Section 6.3.4). A Windows based development environment employing the Open Source Apache server (XAMPP) was configured for local (i.e. not online) use only, as well. The server was utilized in the development of a prototype discussed in Section 6.1. Both the Windows and Apache servers provided nearly the same functionality. Overall, the Windows based OS yielded excellent results. NRC's internal understanding and familiarity with Windows ensured that only minor issues were encountered.

IJC's current GIS server lies on a Windows based OS known as ArcGIS Server to which the IJC staff have expertise. The ArcGIS Server, by ESRI, was developed to provide sharing of mapping and GIS capabilities online and can be accessed by any devices connected to the web (i.e. smartphone, laptops, desktop computers, etc.). Information is requested by the end user through a browser and processed by a server.

A Linux based OS was installed to test the set up and configuration of a prototype discussed in Section 6.2. The development environment was deployed on a server named Shiny Server, which is not available on a Windows OS. For individuals lacking familiarity with Linux, the setup can be challenging and configuration of various OS and server requirements such as directories and associated permissions is not straight forward.

Given the above experience and challenges the NRC strongly suggests the use of Windows OS to comply with the IJC's IT staff expertise.

5.3.3 Server software

As well as needing an OS to function, a server (or a group of servers) requires software that can manage and coordinate a range of activities, such as: downloading third-party data at varying intervals, processing or transforming large datasets (generally greater than 10MB), analyzing incoming data, calling existing scripts to process data, storing data for short periods of time, writing data to storage, etc. All of these processes are accomplished through the use of a programming language. Some of the primary factors which affect choosing an appropriate programming language include the following:

- internal developer proficiency in the programming language;
- availability of software libraries which can increase development productivity;
- available support in terms of documentation and learning resources; and
- utilization of available infrastructure and resources (human/technical).

Python is one example of such a language that is widely used and flexible. It functions on both Windows and Linux based OS. It is a popular programming language used by numerous businesses, some of the more prominent being NASA, NOAA, Dropbox and Google (Github, n.d.; High Scalability, 2011; Python, 2017; Quora, 2017). Given such popularity, a large pool of developers exists to provide new features, maintenance and support into the future. Thus, Python is complemented with a strong development community; has extensive modules to assist with common programming tasks; and is well understood by software developers. Compared to other programming languages for a similar purpose, Python can be easily learned. Python is often referred to as a *glue* language, meaning it is used to connect a variety of software components together. Moreover, Python possesses well developed scientific software packages – or simply *packages* – for analysis some of which are already used by the IJC. For example, ArcPy is one such package developed by ESRI that is used to automatically perform many of the functions enveloped within ESRI's popular ArcGIS software. ArcPy was recently used in the IWI's SPARROW project by the NRC for a variety of spatial processing. Another example package is pyEnSim that provides advanced hydrological data analysis routines through Python. It was developed by the NRC to envelop functionalities within its EnSim program, or the background program code that enables NRC's Green Kenue and BlueKenue advanced pre- and post-processing hydrologic and hydraulic software. IJC's stakeholder, the Lake of the Woods Control Board, currently utilizes pyEnSim for precipitation analysis.

5.4 Technologies that serve the visualization tools

Interactive solutions for the broad groups of IJC users (as identified in Table 1) indicate that interactive maps, interactive charts and summary tables are key components required to render the required visualizations. The layout, style and behaviour of content viewed by the user are determined by the program code used on a browser. The server (amongst many other functions) determines the background content and any complex processes in response to a request made by the end user through the browser. Program code can be written in a number of different programming languages, but those most frequently used to create interactive components (e.g. interactive maps, interactive charts and summary tables) are described below.

Javascript is designed for creating dynamic, interactive components such as visualization tools. It processes the requests from the end user in the browser and makes corresponding changes to a webpage or a visualization component, including embedded maps, charts and tables. JavaScript is also the foundation of a lot of commonly used software libraries. An advantage of using JavaScript and its libraries is that the local development environment is simple, it does not take much computing space, and software libraries are referenced online meaning that they do not need to be installed on a server. As well, an abundance of interactive examples exist online which are free to use.

R is an environment where data manipulation, calculation and graphical display are easily enveloped in a program code (The R Foundation, n.d.). A wide variety of statistical computing (e.g. linear and nonlinear modeling, classical statistical tests, time series analyses, classification, clustering, etc.) and related graphical techniques (e.g. capability to develop publishable plots) are possible. The language is highly extensible by offering a number of packages useful to the scientific community such as mapping, spatial analysis, modelling, plotting and web based interactivity of generated plots and tables. R utilizes a web application framework, known as Shiny, for development of online visualization tools. The tools (once developed) are placed on a specific server (i.e. Shiny Server) created to web-enable the tools.

Python is described as a general-purpose programming language as well as a language to use for analytical and quantitative analysis (Continuum Analytics, n.d.). It offers a number of packages useful towards the development of visualization tools, but it is not as fully featured in interactivity as JavaScript. Note that Python can be used to program the necessary functions of the server (as discussed in Section 5.3.3) as well as a development language used to create visualization tools.

All of the above programming languages are well established and employed by a number of academic, government and commercial organizations worldwide. All three can support the design of reusable software components and full automation of tools. Instead, the onus is on the designer to develop most proficient program code.

Considering available resources and infrastructure at the IJC, ESRI was considered as the first option to creating interactive components. ESRI has unique capabilities in analysis and management of spatial data. It provides interactive maps with different spatial layers, static charts and tables all in one application. However, ESRI is limited by customizability options; only interactivity is available through maps while charts and tables are static.

The NRC reviewed and tested JavaScript, Python and R programming languages, their available libraries and packages to test the breadth of applicability towards development of interactive visualization tools. However, the NRC did not invest much time to reviewing ESRI technology given that the IJC has in-house expertise in this product. Instead, ESRI products were considered when viable.

5.4.1 Maps

Maps provide the spatial context for items with geographic location. They can provide locations of control structures, streamflow gauging stations, spatial flooding conditions within a locale, etc. Online maps are useful as they are usually interactive and provide the end user with immediate and useful details. For example, Figure 2 provides a commonly used example of a Google interactive map indicating a live traffic feed that can potentially allow the user to avoid highly congested areas.

Most prominent technologies for creating interactive web maps reviewed by NRC were OpenLayers and Leaflet both of which are open source and based on JavaScript. Leaflet is considered to be light-weight, meaning that it only comes with a small number of functionalities. Therefore, it is quick to load onto a browser and simple to program with, manage and maintain. Leaflet relies on third-party functionalities (i.e. plugins) for additional capabilities which developers use only when needed. Amongst many capabilities provided it accommodates the inclusion of map tiles for rapid interactivity. Leaflet can be integrated with ESRI to leverage some useful features of the ArcGIS platform. Leaflet has also been packaged to work with R and Python. The Python version of Leaflet is not as fully featured as the JavaScript and R versions. OpenLayers is considered much heavier in terms of development and functionalities, and most likely too excessive for IJC’s needs. OpenLayers was found to be most useful for complex mapping functionalities, while Leaflet sufficed for common mapping tasks the IJC will require in terms of providing visualization tools online. During development of prototypes and testing of visualization technologies NRC found that excellent documentation and numerous online examples existed for Leaflet allowing for rapid developer onboarding.

Figure 2: Local traffic conditions provided in an interactive Google map to help the user select a route to another destination, where light traffic is indicated by green lines and heavy traffic by dark red lines.

5.4.2 Charts

Charts or plots provide graphical representation of data, most commonly shown as line and bar plots, but numerous possibilities are available. For example, Figure 3 provides an example of a stock chart commonly used by the public to track stock earnings and losses. The chart allows the user to see clearly whether the stock prices are rising or falling.

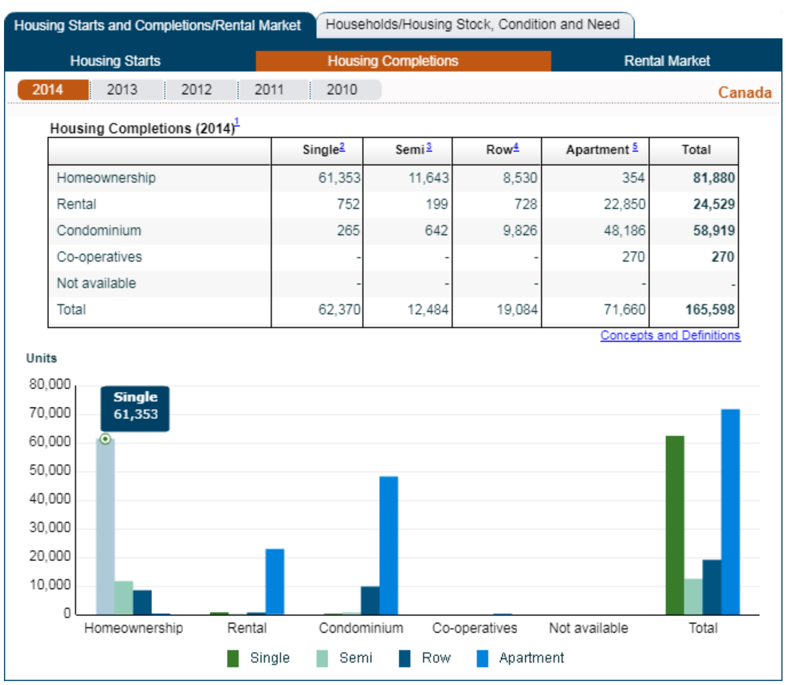
Figure 3: Historic and up-to-date prices of a stock on a free interactive stock chart made available by Yahoo Finance

Some of the more commonly used technologies for development of interactive charts were explored, and included alternatives based on JavaScript (i.e. D3, C3, NVD3 and DC), Python and R (i.e. Bokeh and Plotly). In summary, programming based on D3 was generally found to be more complex and time-consuming than some of the other alternatives, but it facilitated more customization and resulted in more aesthetically pleasing charts. Numerous programming code examples are available online, which helped to accelerate the development process in D3. This technology was used for the development of a prototype described in Section 6.1 (where an in-depth discussion on its use is available).

5.4.3 Tables

Some data types are best suited to tabular forms of presentation (e.g. historic water levels, climate normals, statistical summaries, etc.). NRC reviewed a number of different methods for displaying tabular data using historic water levels as a sample test dataset. Creating a simple table in a HTML file proved to be easy, but time-consuming and susceptible to human error when a larger dataset was employed (i.e. where number of rows was greater than ten). DataTable is a common technology based on JavaScript used to develop interactive summary tables online. Generating a table using text files is relatively trivial and best suited for the IJC's needs. R and Python versions of the DataTable JavaScript library are also available if required. Python Pandas is a package that provides useful capabilities where calculations can be applied before outputting data to the visualization. D3 can be utilized to generate tables but NRC did not experiment with this technology. A review of the documentation indicated that this was possible but the development process was significantly more complicated than the technologies described above.

An online, interactive tabular-based visualization tool is provided as an example in Figure 4. The table provides detailed values which can be further manipulated by the user. The tool is an example that Canada Mortgage and Housing Corporation (CMHC) uses to share their data.


Housing Completions (2014)[1]

	Single[2]	Semi[3]	Row[4]	Apartment[5]	Total
Homeownership	61,353	11,643	8,530	354	**81,880**
Rental	752	199	728	22,850	**24,529**
Condominium	265	642	9,826	48,186	**58,919**
Co-operatives	-	-	-	270	**270**
Not available	-	-	-	-	-
Total	62,370	12,484	19,084	71,660	**165,598**

Figure 4: Example of detailed data provided in an interactive online table by CMHC.

5.4.4 Amalgamation of Components

Mapping, charting and table components can be combined together in a multitude of ways on a webpage to help provide additional context. For example, data associated with a chart is also displayed in a table for a quick reference; or a map is introduced to provide spatial context for a group of given time series charts.

One common approach to communicate several key pieces of information related to a specific topic is a *dashboard* by providing several pieces of information in both textual and simple visualization components (i.e. maps, charts and tables). To address the information that needs to be conveyed, selected components must be well thought out in advance (i.e. why include a component and what purpose does it serve) of development so as not to waste time writing and debugging program code that may not be used.

Two technologies for developing complex component scenarios were reviewed and tested. One prototype included a two panel dashboard showing US-share and Canada-share data made available by the St. Mary and Milk River Technical Working Group (see Section 6.3.4). Python's Bokeh and Pandas libraries were employed here. The other prototype provided a dashboard for the general public with information on lake and precipitation levels with the International Rainy River Lake of the Woods Watershed Board (IRRLWWB) as described in detail of Section 6.2. The R programming language and the interactive Shiny library were used, with the following simple component libraries discussed above: Plotly, Leaflet and DataTable. Unlike the dashboard developed using Bokeh and Pandas libraries containing static data, the Shiny library allowed for dynamic data updates, as they become available on the server (i.e. in real-time) or are requested by the end user. The level of development effort was rapid for both prototypes. Note that the use of the Shiny library allows for greater flexibility in terms of development and user options. However, it requires additional server infrastructure (i.e. Shiny Server) which incurs greater costs through: i) in-house technical resources required to setup and maintain a server; or, ii) a monthly paid subscription to utilize a third party managed Shiny Server service. A JavaScript version of a dashboard was not created or tested at this time, however based on NRC's experience the development time required is greater than that of both R and Python, but provides the greatest control for rich interactions and highly customized aesthetics.

5.5 Data storage and IJC's program code

Additional considerations were investigated to ensure that all of the details required to enable the inclusion of the visualization tools to IJC's website have been considered. The inclusion of data accessibility to the visualization tools as well as program code that has been developed in the past by the IJC and its boards are briefly discussed below.

5.5.1 Data processing and storage

Data used by the IJC and its boards are obtained from various sources as discussed in Section 4. They are available in a variety of formats and update frequency. Depending on the need, some data are available for integration into the visualization tool as is, while other data require further processing to create additional information for the visualization tool. The processing can be complex (e.g. averaging precipitation values provided by ECCC or NOAA by basin) or simple (e.g. calculating means, maxima, minima and percentile values from historic time series data). Complex processes typically require greater computational time and resources, and additional tools such as spatial analysis, advanced statistical computation and analysis, etc. Such complex processes are best performed on the server where the computational and processing resources are known and are much more powerful. On the other hand, simple processes can easily be performed on the browser. To handle the integration and processing of data, a mechanism that retrieves required data on a frequent basis, adjusts the data to required file formats and stores them with a simple directory structure is a straightforward option. Storage of data can be as simple as using self-contained databases such as SQLite or simple text based files on a server. Considering IJC's reluctance to store large datasets, storing data in such a manner provides a reasonable option. Some, albeit minimal, data storage is necessary as it ensures that the end users' browsers do not take the strain of computationally heavy data processing and that the visualization tool loads quickly. Also, these data can be reused across multiple visualization tools without further processing if stored.

5.5.2 Additional program code available at IJC

The IJC, its boards and stakeholders have developed a number of useful scripts and libraries for processing and analysis of collected and available transboundary data. For example, Visual Basic for Applications (VBA) scripts were developed for MS Excel by members of the St. Mary and Milk Rivers Technical Working Group to perform natural flow calculations in the St. Mary and Milk Rivers systems. Many of these scripts and libraries could be reused by other IJC related projects and easily integrated to potential visualization tools as needed. The future IJC server which hosts the visualization tools and all of the processing required to enable them is a place where these technologies can reside and from which they could be called by Python for execution.

6 Prototypes

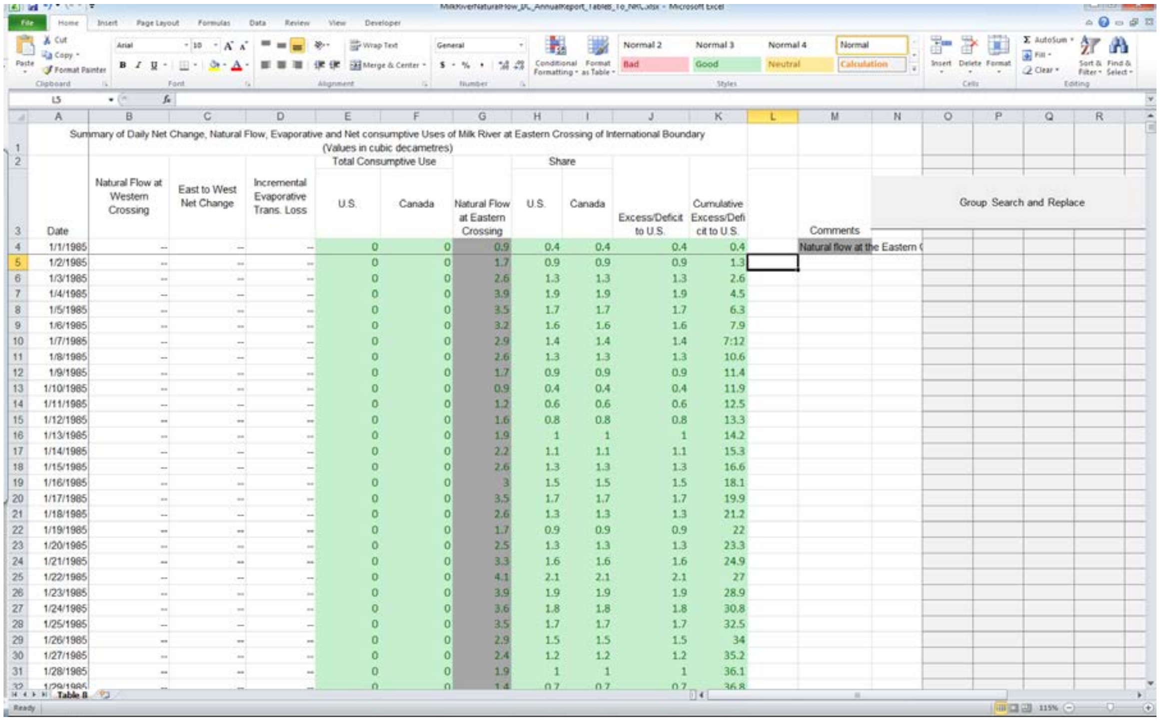
The NRC developed a number of prototypes based on discussions with the IJC and IJC's stakeholders. The prototypes act as preliminary and demonstration tools that can be showcased as future potential. The development process also helped: i) shape recommendations (of Section 7); ii) further understand IJC requirements; iii) enable interaction with stakeholders where valuable feedback was received; iv) inform the review undertaken for this project; v) determine future approaches to technology; and vi) provide future development potential.

Two main prototypes were produced highlighting significance and uses of interactive time-series charts and mapping products. These are described in detail in Sections 6.1 and 6.2. A number of other prototypes were developed as discussed in Section 6.3.

6.1 Graphing Tool Prototype

The St. Mary and Milk Rivers are international rivers that flow between the US and Canada. The flows of each river are shared between the two countries, where the accessible portion of the flow for each country is determined in accordance with the *1921 Order of the IJC respecting the St. Mary and Milk Rivers* (IJC, 2014). These flow apportionments vary based on the amount of flow in the river, and the time of year (i.e. whether within irrigation or non-irrigation season).

Communicating the natural flow and flow apportionment is necessary in a transboundary basin where the waters are shared between two countries such as the St. Mary and Milk Rivers. Communities on both sides of the border rely on the information for a variety of water uses. The St. Mary and Milk Rivers Technical Working Group (SMRTWG) provide support in the apportionment agreement (IJC, 2014) and thus are interested in becoming more transparent and responsive to the communities served by the waters. The SMRTWG currently employ a spreadsheet, as illustrated in Figure 5, in which flows are hand entered and calculation is performed through a series of software scripts. The information is currently shared with the stakeholders through the spreadsheets. However, the SMRTWG would like to have an online tool that provides necessary data to their stakeholders directly.



Summary of Daily Net Change, Natural Flow, Evaporative and Net consumptive Uses of Milk River at Eastern Crossing of International Boundary
(Values in cubic decametres)

Date	Natural Flow at Western Crossing	East to West Net Change	Incremental Evaporative Trans. Loss	Total Consumptive Use U.S.	Total Consumptive Use Canada	Natural Flow at Eastern Crossing	Share U.S.	Share Canada	Excess/Deficit to U.S.	Cumulative Excess/Deficit to U.S.	Comments
1/1/1985	--	--	--	0	0	0.9	0.4	0.4	0.4	0.4	Natural flow at the Eastern (
1/2/1985	--	--	--	0	0	1.7	0.9	0.9	0.9	1.3	
1/3/1985	--	--	--	0	0	2.6	1.3	1.3	1.3	2.6	
1/4/1985	--	--	--	0	0	3.9	1.9	1.9	1.9	4.5	
1/5/1985	--	--	--	0	0	3.5	1.7	1.7	1.7	6.3	
1/6/1985	--	--	--	0	0	3.2	1.6	1.6	1.6	7.9	
1/7/1985	--	--	--	0	0	2.9	1.4	1.4	1.4	7:12	
1/8/1985	--	--	--	0	0	2.6	1.3	1.3	1.3	10.6	
1/9/1985	--	--	--	0	0	1.7	0.9	0.9	0.9	11.4	
1/10/1985	--	--	--	0	0	0.9	0.4	0.4	0.4	11.9	
1/11/1985	--	--	--	0	0	1.2	0.6	0.6	0.6	12.5	
1/12/1985	--	--	--	0	0	1.6	0.8	0.8	0.8	13.3	
1/13/1985	--	--	--	0	0	1.9	1	1	1	14.2	
1/14/1985	--	--	--	0	0	2.2	1.1	1.1	1.1	15.3	
1/15/1985	--	--	--	0	0	2.6	1.3	1.3	1.3	16.6	
1/16/1985	--	--	--	0	0	3	1.5	1.5	1.5	18.1	
1/17/1985	--	--	--	0	0	3.5	1.7	1.7	1.7	19.9	
1/18/1985	--	--	--	0	0	2.6	1.3	1.3	1.3	21.2	
1/19/1985	--	--	--	0	0	1.7	0.9	0.9	0.9	22	
1/20/1985	--	--	--	0	0	2.5	1.3	1.3	1.3	23.3	
1/21/1985	--	--	--	0	0	3.3	1.6	1.6	1.6	24.9	
1/22/1985	--	--	--	0	0	4.1	2.1	2.1	2.1	27	
1/23/1985	--	--	--	0	0	3.9	1.9	1.9	1.9	28.9	
1/24/1985	--	--	--	0	0	3.6	1.8	1.8	1.8	30.8	
1/25/1985	--	--	--	0	0	3.5	1.7	1.7	1.7	32.5	
1/26/1985	--	--	--	0	0	2.9	1.5	1.5	1.5	34	
1/27/1985	--	--	--	0	0	2.4	1.2	1.2	1.2	35.2	
1/28/1985	--	--	--	0	0	1.9	1	1	1	36.1	
1/29/1985	--	--	--	0	0	1.4	0.7	0.7	0.7	36.8	

Figure 5: Screen capture of the SMRTWG spreadsheet.

NRC worked closely with the SMRTWG and the IJC to develop a prototype of an interactive visualization tool. The prototype was intended for a technical audience that is familiar with hydrologic data and is comfortable with technical interpretations of the data. The final delivery needed to provide natural flows and historic division of waters between the Canadian and the US shares of the St. Mary and Milk Rivers.

During the development of the prototype NRC, SMRTWG and IJC worked through a number of iterations, initially to establish the exact requirements and subsequently to clarify the appearance and aesthetics. The following features were identified as the requirements for the visualization tool:

- Display of natural flow for each river;
- Display of daily apportionment for each country;
- Ability to select a targeted year to display data;
- Display of available historical statistics along with the target year (i.e. maximum, minimum, 25th, 75th and median flows);
- Ability to choose user-selected units (i.e. metric or imperial);
- Ability to choose user-selected language (i.e. French or English);
- Ability to choose user-selected vertical axis scale (i.e. linear or log);
- Ability to display actual data when hovering over the graph; and
- Ability to read from a standard time series format.

Development required iterations from the conceptual design to the final development stage which resulted in an interactive, online, visualization prototype illustrated in Figure 6.

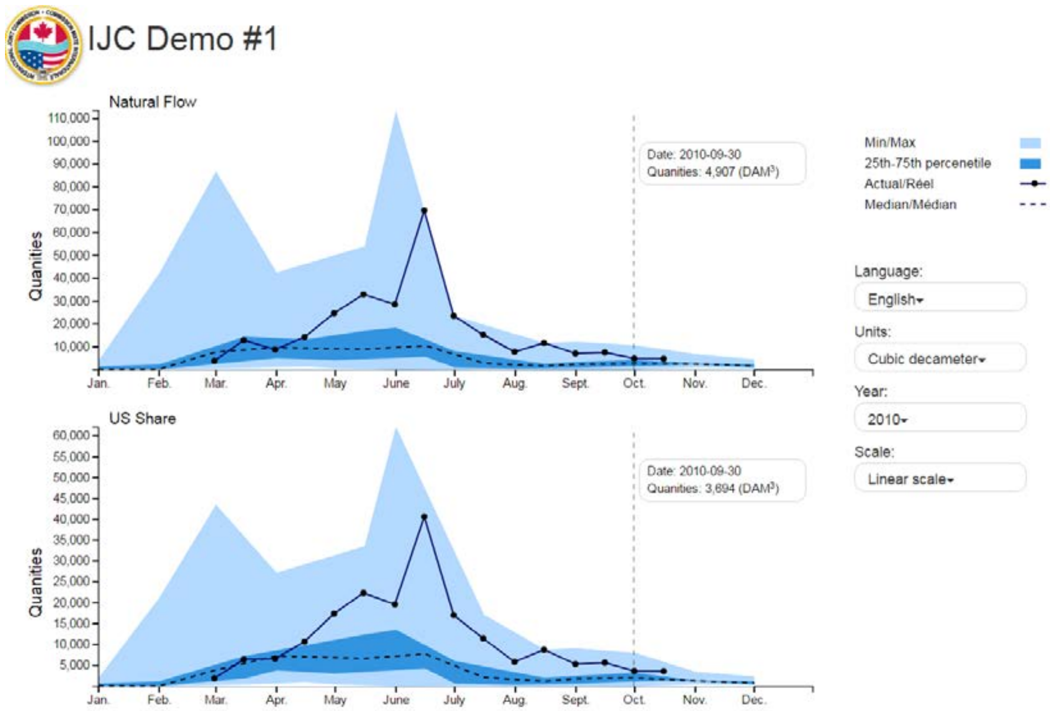


Figure 6: Screen capture of the graphing tool prototype.

A number of chart types (bar, line, multiline, pie) were investigated. A line chart was chosen because it provides the simplest means to communicate the necessary data, and is more popular amongst the targeted user groups. NRC explored, reviewed and tested several JavaScript based technologies during the early stages of the prototype development: DC, D3 and C3.

Initially, the prototype was developed using DC technology. DC is built on D3 technology. It is designed to facilitate exploration of datasets containing multiple attributes. The user is able to filter displayed data using multiple criteria, a functionality termed cross-filtering. However, enhancing interactivity beyond DC's scope was found to be very tedious. Thus, DC technology was found inappropriate for development of this prototype. As well, SMRTWG found no need for cross-filtering of their data. See Section 6.3 for further details to the use of DC technology.

C3 was explored, but quickly abandoned because of its limitations with respect to customizability.

D3 was chosen as the technology to develop the prototype tool. D3 required more programming than C3, but was preferred due to the high level of customizability needed to develop the prototype. D3 provides full control over style and interactivity of a chart.

The type of information provided and the aesthetics are clearly important to the success of the tool. The appearance of the prototype as shown in Figure 6 was completely altered several times. Initially, the tool displayed all of the parameters of interest on the same chart (as is further described in Section 6.3), however it was found to be too busy and confusing. Moreover, comparison of historical statistics and the target year was not possible within this version.

Subsequently the prototype consisted of appropriate information as illustrated in Figure 7 but was not as aesthetically pleasing and as modern looking as Figure 6.

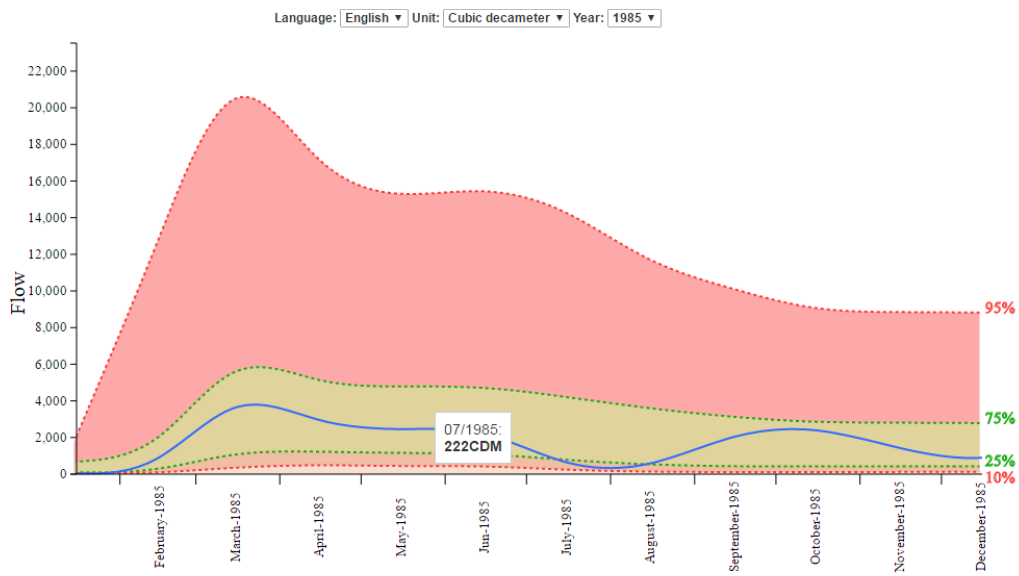


Figure 7: Example of the graphing tool prototype before it was visually enhanced.

6.2 Dashboard Prototype

NRC worked with the IJC, the International Rainy River Lake of the Woods Watershed Board (IRRLWWB), the Lake of the Woods Control Board (LWCB) and ECCC to develop an interactive tool that examined sub-watershed data related to precipitation and lake water levels. An issue that the NRC repeatedly heard from the group related to the public's concerns with respect to lake regulations. Many of the lakes (e.g. Lake of the Woods, Rainy Lake, Namakan Lake, etc.) in this region are controlled. For example, water levels may be lowered prior to a forecasted precipitation event that may cause water levels to rise. The lakes are connected, where upstream precipitation events can impact downstream lakes, which is not always obvious to a local community located downstream of the event. Providing a tool that illustrates historic, current and future precipitation and lake water levels would assist in raising public awareness.

During early phases of the project, the NRC shared examples of various technologies that might be of interest to the IJC, IRRLWWB, LWCB and ECCC. A dashboard based tool emerged as the most promising tool, where the group was interested in seeing how it would look to suit their needs. Dashboards are a simple way to present a few items of importance that have been determined ahead of time. Each piece of information provided in the dashboard should be descriptive, easily understood and contribute to assisting the audiences' understanding of the topic at hand.

The scope of work for this prototype aligned with the capabilities of some well used libraries available in R that aim to provide interactive web based tools. In consultations with the IJC, several mockups were developed to demonstrate that the proposed technology would satisfy the desired output. The prototype was developed using Shiny.

Prior to designing the dashboard, the NRC identified the type of user that the prototype should target. This assisted in identifying what questions the public would have and how the dashboard would assist in answering these questions. For this prototype, the user type was determined to be a general public individual who owns or rents property in the watershed, uses the nearby lake for recreational boating activities and is concerned about potential flood risk to the property. Using this information the NRC identified several requirements to establish the user's interaction with the tool discussed below.

The overall layout was divided into five information panels (as illustrated in Figure 8). The top most panel is the informational dialog that informs the user how much precipitation has fallen in the basins. The user can select the time period of this calculation from a drop down menu available to the left. Four informational panels consist of a map, a table and two charts.

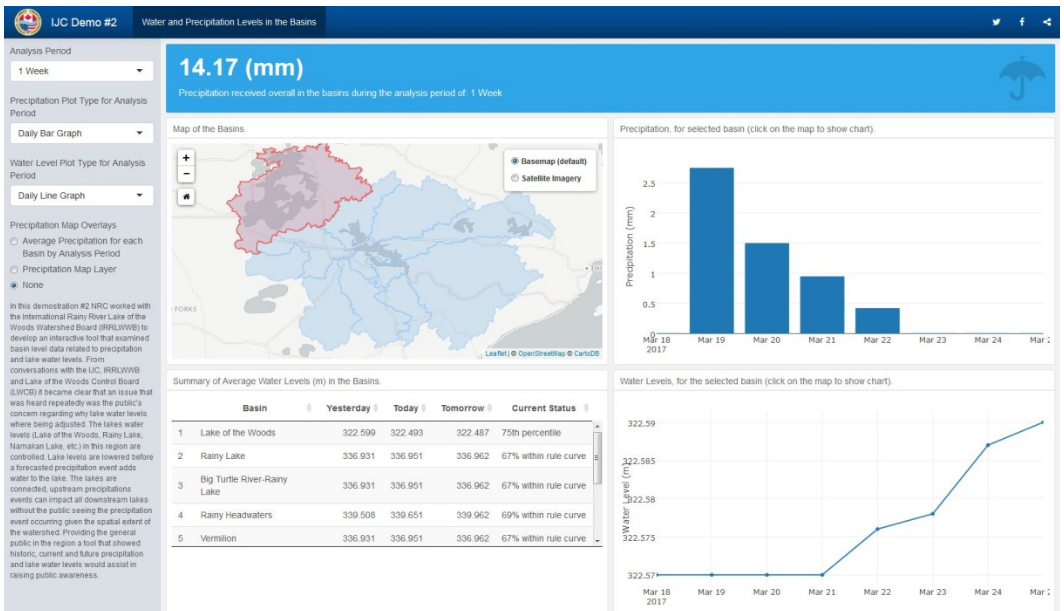


Figure 8: Screen capture of the dashboard developed for the IRRLWWB.

Instead of describing each panel, the discussion focuses on user needs and how these needs were met through the interactive visualization in the sections below

6.2.1 Display a dynamic, interactive map to identify the sub-basins.

The map, as illustrated in Figure 9:

- displays the Rainy River – Lake of the Woods basin and its sub-basins;
- is selectable by sub-basin; and
- provides a tool tip to indicate the name of the selectable sub-basin.

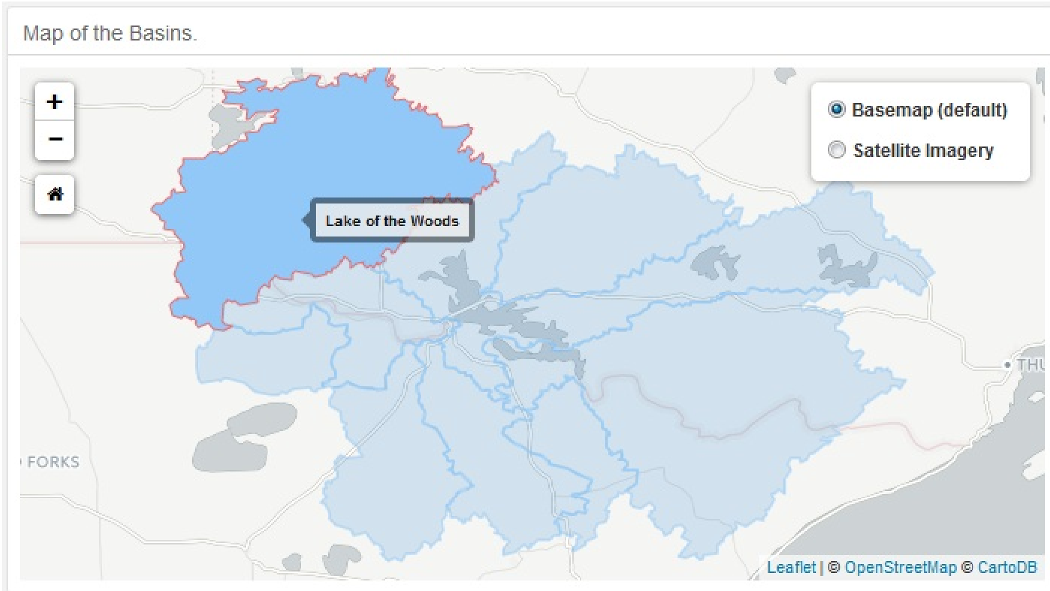


Figure 9: Screen capture of a dynamic, interactive map.

6.2.2 Display precipitation amounts on a dynamic, interactive map for a selected temporal period to visualize the differences between "wetter and drier" sub-basins.

The map, as illustrated in Figure 10:

- displays average precipitation amounts for the selected temporal period;
- provides a tool tip to indicate the mean precipitation for the selected sub-basin;
- provides a colour scale identifying the magnitudes associated with the precipitation levels; and
- has the ability to be toggled on or off.

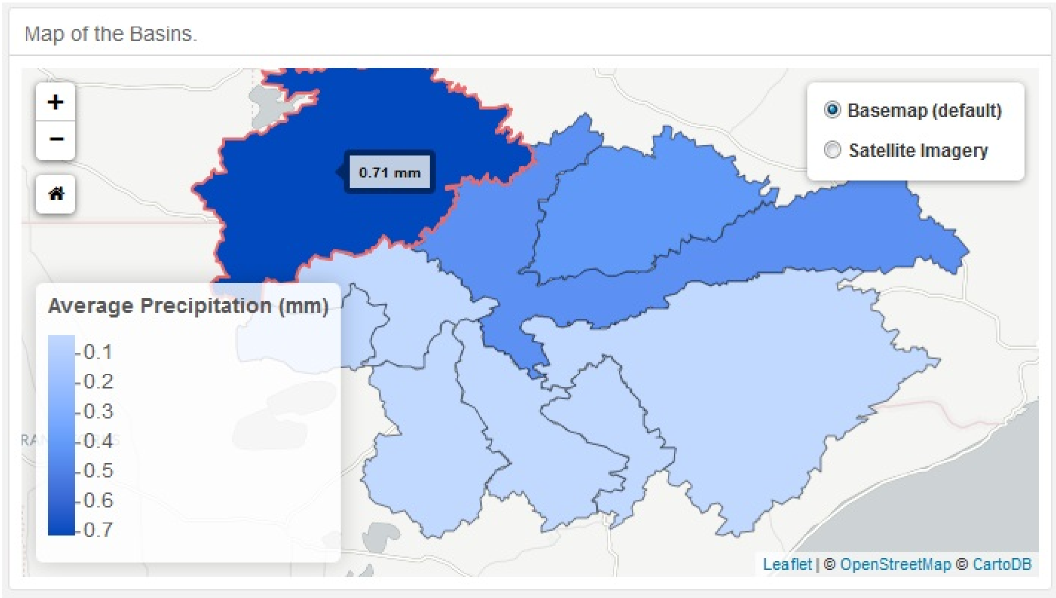


Figure 10: Screen capture of a dynamic, interactive map identifying average precipitation amounts by basin.

6.2.3 Display latest precipitation layers provided by an external weather service.

The map, as illustrated in Figure 11:

- displays near real-time data currently provided by a weather service;
- displays a colour scale identifying the magnitudes associated with displayed precipitation; and
- has the ability to be toggled on or off.

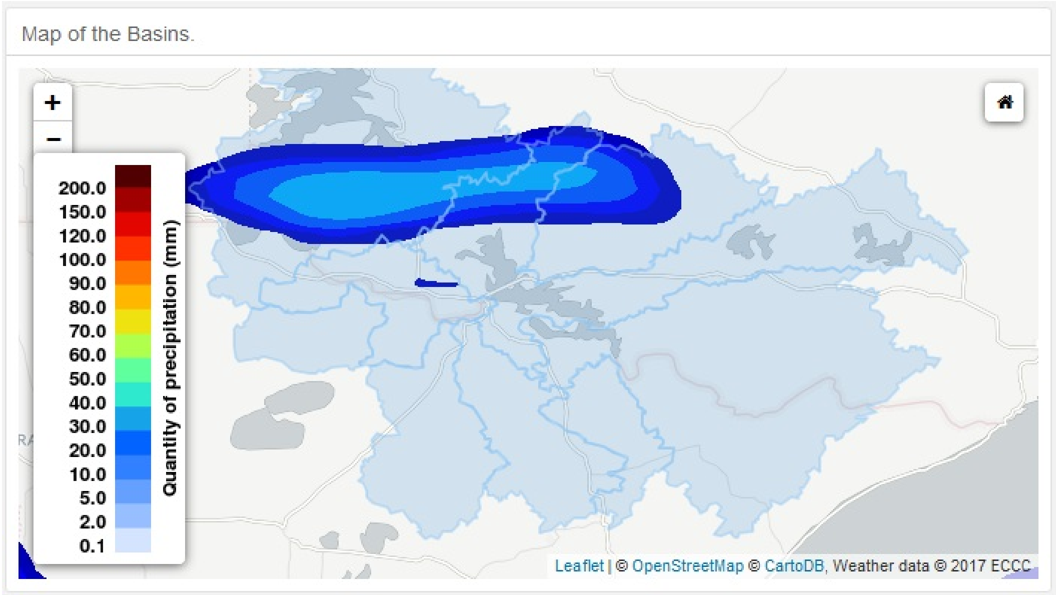


Figure 11: Screen capture of a dynamic, interactive map overlaid with external data

6.2.4 Display a dynamic, interactive map to provide some familiar spatial context of the location.

The map, as illustrated in Figure 12:

- displays colour satellite imagery from ESRI web services; and
- has the ability to be toggled on or off.

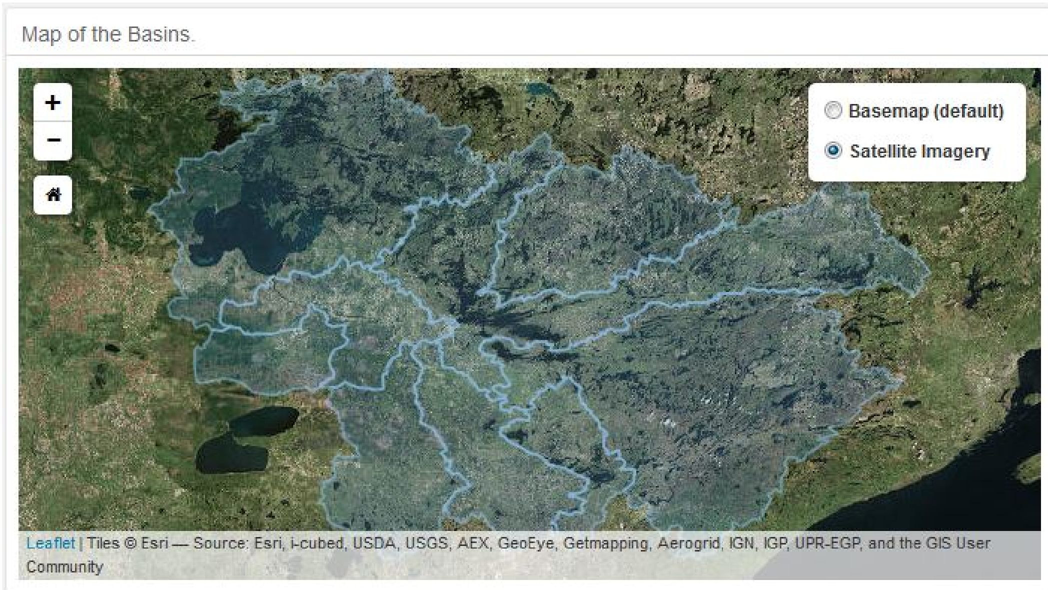


Figure 12: Screen capture of an interactive map using ESRI web services.

6.2.5 Display precipitation and water levels on dynamic, interactive charts, by selected sub-basin, to see how these parameters are related.

The charts are provided in:

- a graph view stacked on top of one another displaying the same x-axis (which represents time), a tooltip for more information as illustrated in Figure 13, with an additional cumulative precipitation not shown here; and
- a stacked calendar view displaying the same calendar dates and similar colour scales identified with the magnitudes for each parameter as illustrated in Figure 14.

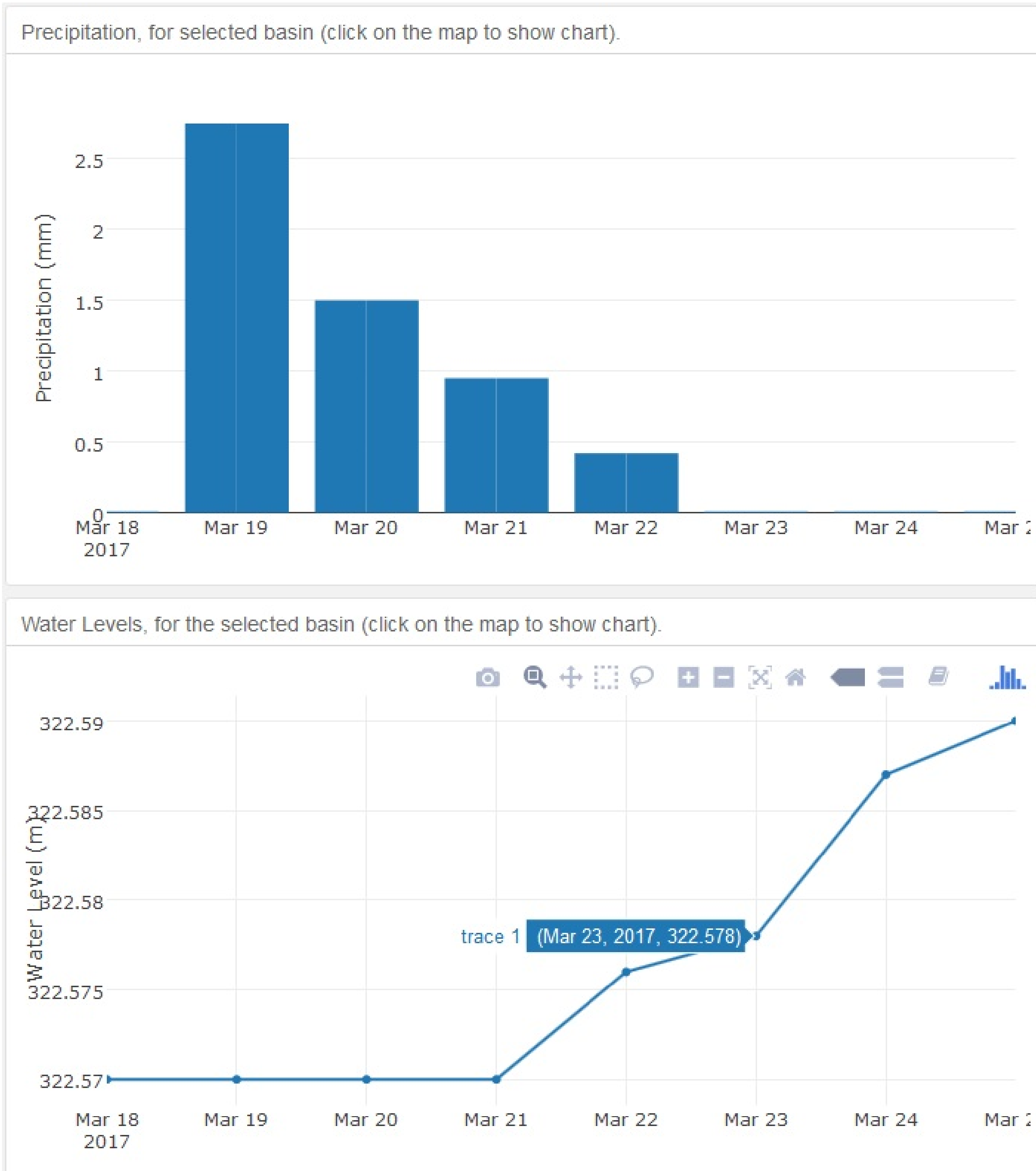


Figure 13: Screen capture of interactive bar and line charts indicating the amount of precipitation that has fallen and the water level of a select sub-basin.

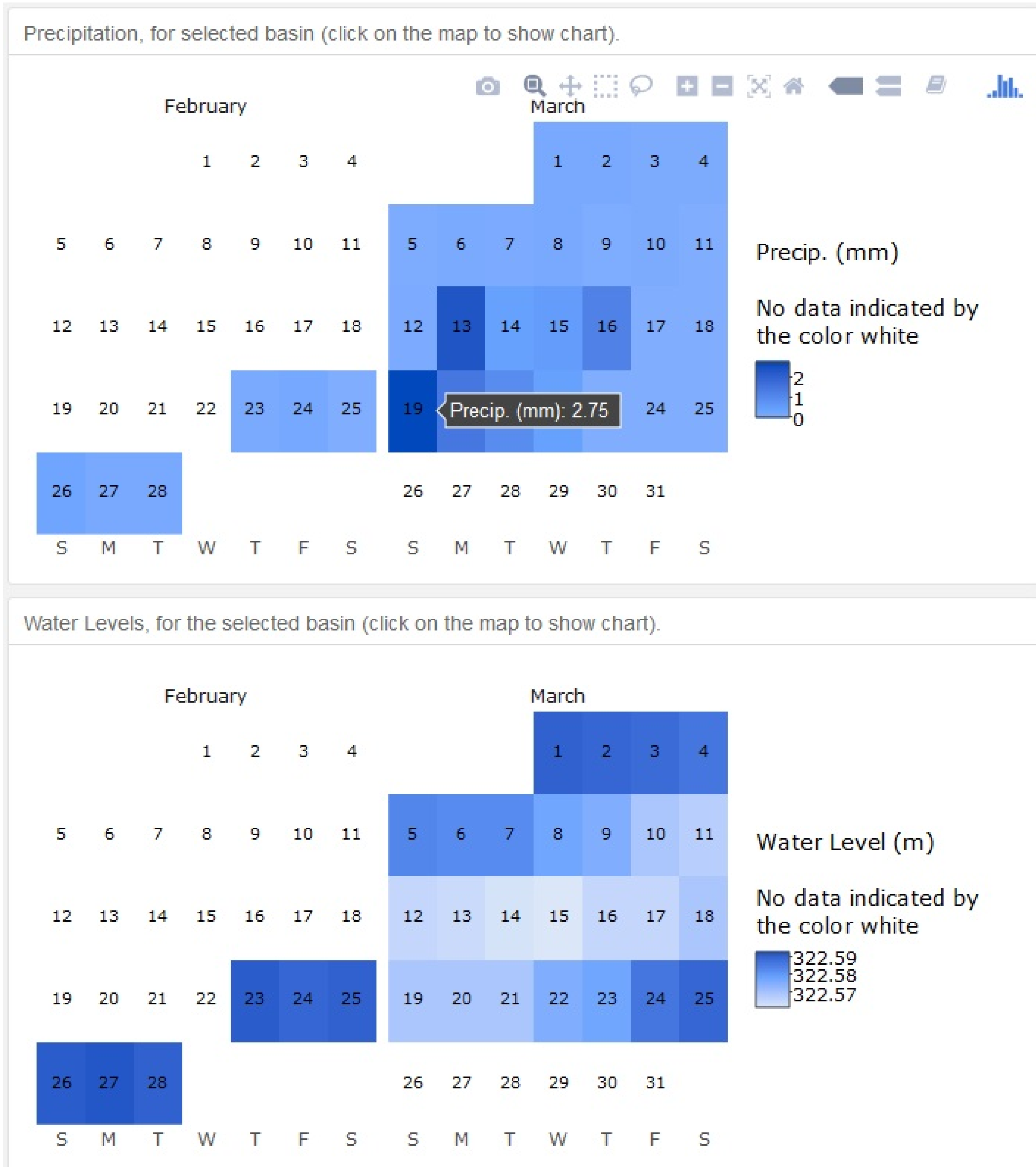


Figure 14: Screen capture of interactive calendar charts indicating the amount of precipitation that has fallen and the water level of a select sub-basin.

6.2.6 Provide a selection of time periods available to view the difference between recent and historic data.

The tool provides a drop-down menu for an analysis period that extends between 1 week, 2 weeks and 1 month.

6.2.7 Provide a table summarizing beneficial information.

The interactive table as illustrated in Figure 15:

- provides a list of all of the sub-basins with latest water level conditions and their status; and
- has the ability to sort.

Summary of Average Water Levels (m) in the Basins.

	Basin	Yesterday	Today	Tomorrow	Current Status
1	Lake of the Woods	322.599	322.493	322.487	75th percentile
2	Rainy Lake	336.931	336.951	336.962	67% within rule curve
3	Big Turtle River-Rainy Lake	336.931	336.951	336.962	67% within rule curve
4	Rainy Headwaters	339.508	339.651	339.962	69% within rule curve
5	Vermilion	336.931	336.951	336.962	67% within rule curve

Figure 15: Screen capture of the interactive summary table.

6.3 Examples of Other Prototypes

The NRC reviewed a number of technologies during this project. The goal was to understand the capabilities and constraints of various technologies and how they pertain to the IJC. Development experience assisted in gaining a deeper understanding of issues that may not be immediately apparent by simply reviewing documentation or other resources (e.g. level of effort required to learn and for development, background tools that are required to work with the technologies, availability of documentation and support, etc.).

6.3.1 D3

The D3 technology based on JavaScript provides an impressive capability to customize interactions and aesthetics of a visualization tool. The initial learning curve is steep, but numerous examples (i.e. program code) exist and many of these permit further modification. Ample tutorials and resources are available to assist during development as well.

The prototype illustrated in Figure 16 below is an example of a time series graph indicating natural flow, Canada/US share flows and annual cumulative flows for SMRTWG data discussed in Section 6.1. The main features of the tool illustrate the following capabilities:

- Two interactive graphs. The bottom graph displays the available period of record and is used to select a targeted period of interest via a window (shaded in gray below) which is automatically drawn by the top graph.
- Drop-down list to select the language of preference to update plot elements (e.g. axis labels, legend, etc.).
- Drop-down list to choose the preferred units of measure (e.g. metric and/or imperial).
- Mouse over allows the user to read the *y*-value based on the closest *x*-value.

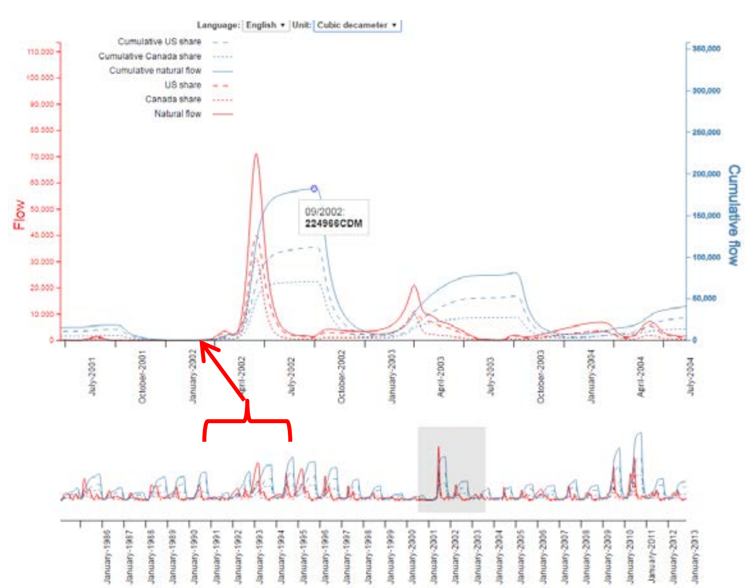


Figure 16: Screen capture of the original graphing tool prototype using D3 technology.

6.3.2 DC

DC technology based on JavaScript helps simplify program code and effort required to link D3 and cross-filtering capability to allow for highly efficient and interactive exploration of datasets. For example, when a dataset is filtered on a particular graphic (e.g. map, table or chart), all of the other graphics linked to the same dataset will automatically update to the requested filter. A particularly useful function in DC enables linkages between map polygons to charts. A steep learning curve is associated with DC but a number of functions help reduce the amount of program code required compared to D3. Developer documentation is useful, as are many examples and tutorials available online.

The prototype illustrated in Figure 17 below allows the user to select a period of interest automatically updating all of the included elements (charts and table). SMRTWG data (further described in Section 6.1) was used in this prototype as well.

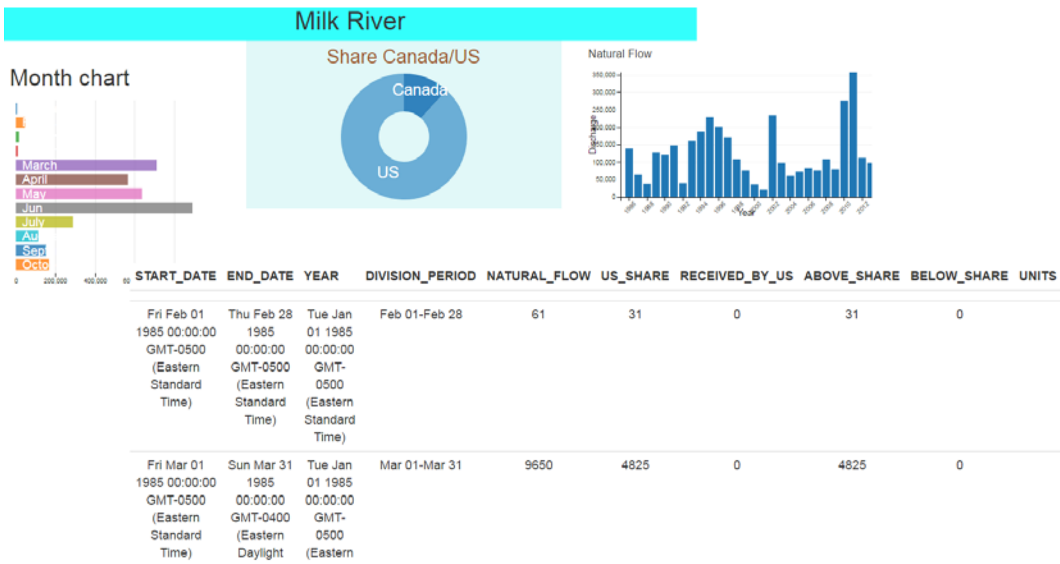


Figure 17: Screen capture of the graphing tool prototype using DC technology

6.3.3 Leaflet for JavaScript

Leaflet for JavaScript provides an excellent web mapping library. It is well supported and actively maintained. A number of associated libraries exist to add functionality to the map. Building a basic web map requires minimal knowledge of JavaScript, but skills are essential to fully utilize Leaflet's capabilities. Numerous resources and learning materials are available for developers.

The prototype illustrated in Figure 18 demonstrates IJC's transboundary basins on an interactive web map. The prototype enables the user to pan, zoom and view tooltips when the mouse hovers over a basin.

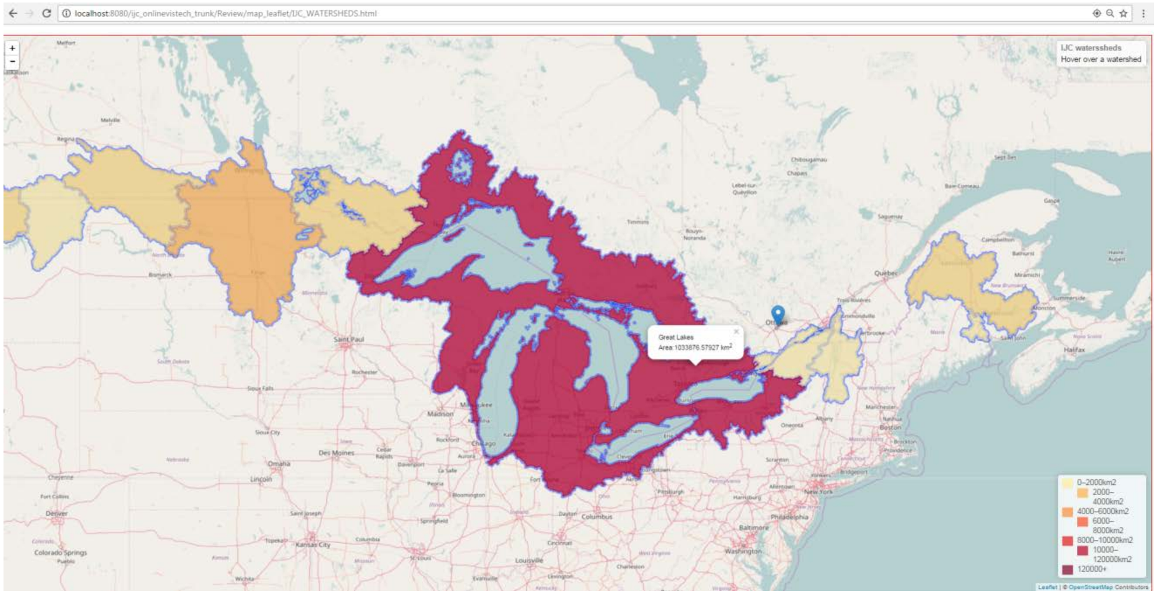


Figure 18: Screen capture of an interactive map example using Leaflet technology.

6.3.4 Bokeh

Bokeh library based on Python provides excellent charting and interactivity that allows developers to read, process and visualize data in a single language. The learning curve for Python developers is minimal. The developer documentation is excellent and the availability of learning resources is numerous.

The prototype illustrated in Figure 19 below demonstrates an aggregated view of SMRTWG data (further described in Section 6.1) of annual above and below shares for the US. The main features of the tool include:

- tooltip to read the *y*-value based on the closest *x*-value;
- zooming and panning of the graph; and
- exporting of the graph as an image.

The prototype illustrated in Figure 20 is a continuation of Figure 19 demonstrating how individual plots can be combined into a dashboard to improve communication. Two plots were generated as standalone HTML files. An HTML dashboard template was used to display each of the two individual plots. The main features of the tool include:

- all of the features listed for Figure 19;
- the HTML dashboard template resizes based on users' screen size; and
- drop down menus on the dashboard page to link to additional HTML pages or documents.

Figure 19: Screen capture of a bar graph in Bokeh.

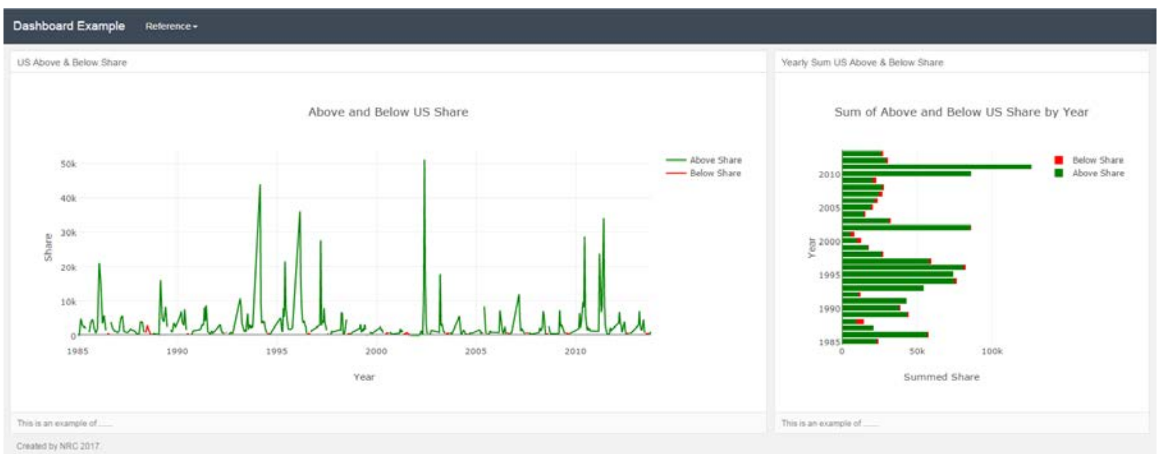


Figure 20: Screen capture of a simple dashboard in Bokeh.

7 Recommendations

Based on the review of IJC's needs and available technologies for the development of visualization tools, and following testing of a number of prototypes, three key recommendations were developed by NRC. Each recommendation matches a specific IJC need and/or target audience for data visualization tools (as identified during the review) to the appropriate technology.

All three recommendations, if implemented, would result in the development of high-quality user-friendly interactive web-based visualization tools. Each tool would be identified by its own unique web address and embedded into an IJC webpage.

***Recommendation 1**: For applications where high quality aesthetics and fully customizable interactivity are a priority (e.g. to convey crucial messages, and / or information from high impact reports to the public and wide audiences) - Use JavaScript and associated libraries (e.g. Leaflet, D3, DataTables) to develop online visualization tools.*

Key aspects of the approach, and features of the tools that would be developed through the implementation of Recommendation 1 include:

1. JavaScript's Leaflet for interactive maps, D3 for interactive charts and DataTables for interactive tables would be used to develop the visualization tools. All three libraries have high interactivity options available that the IJC can fully utilize at no cost.
2. The user's browser would render and process simple data that is not computationally intensive.
3. Users would have to reload the website to update near real-time data in the visualization tool.
4. Third-party Windows OS server (which would incur costs based on selection of provider, server specifications and required system administration) would be used to:
    - host all program code required for rendering of the visualization tools,
    - host additional packages that will be integrated such as ArcPy and PyEnSim,
    - perform computationally intensive data processing,
    - access external data sources,
    - perform any other additional processing required, and
    - host minimal but necessary historic and processed data required by the visualization tools.
5. Data hosted on the server would be saved in simple data formats (e.g. CSV, JSON, GeoJSON, etc.) or in simple databases such as SQLite.
6. Hosted data would only be retained for short periods as needed by the visualization tools. The intent is to create datasets that can be used by multiple visualization tools and thereby avoiding repetitive processing.
7. Python would be installed to manage and coordinate the server activities as described in Section 5.3.3 as well as processes in item 4 above. Python is an open-source programming language that is available at no cost.
8. Existing ESRI technology would be integrated within the visualization tools for spatial processing when possible and applicable.

***Recommendation 2**: For applications where aesthetic quality and interactivity are important but where there is a potential need for periodic and / or rapid modification of the tools (e.g. periodic or routine dissemination of data to the public and / or IJC boards) – Use Python and associated libraries (e.g. Bokeh, Plotly, Pandas) to develop online visualization tools.*

Key aspects of the approach, and features of the tools that would be developed through the implementation of Recommendation 2 include:

1. Python and its libraries (e.g. Bokeh, Plotly, Pandas) would be used to develop visualization tools. Python and its libraries are open-source packages and are available for use at no cost.
2. HTML files would be generated and updated at some predefined interval on the server before being passed to the user's browser for viewing.
3. Users would have to reload the website to update near real-time data in the visualization tool.
4. Third-party Windows OS server (which would incur costs based on selection of provider, server specifications and required system administration) would be used for processes as provided in item 4 of *Recommendation 1* above.
5. Data hosted on the server would be saved in simple data formats or in simple databases.
6. Python would also be used to manage and coordinate the server activities as described in Section 5.3.3 as well as processes in item 4 of *Recommendation 1* above.
7. Existing ESRI technology would be integrated within the visualization tools for spatial processing when possible and applicable.

***Recommendation 3**: For applications where technical users are the target audience requiring seamless integration with other program code (e.g. dashboards integrating existing in-house statistical tools) as well as increased development flexibility and user options, or for tools that the IJC may ultimately wish to be developed / managed by boards and technical groups (e.g. rapid dashboard development) – Use R and Shiny to develop and integrate online visualization tools.*

This recommendation offers increased flexibility in development and accessibility, but requires additional infrastructure (with incurred costs) to implement. Key aspects of the approach, and features of the tools that would be developed through the implementation of Recommendation 3 include:

1. Leaflet for R for interactive maps, R Plotly for interactive charts and DataTable for R for interactive tables would be used to develop visualization tools.
2. Two servers will be required:
    a. Shiny Server with a paid subscription to a third-party server known as shinyapps.io would be used for rendering visualization tools on the user's browser. Visualizations would be updated automatically as near real-time data becomes available.
    b. Third-party Windows OS server (which would incur costs based on selection of provider, server specifications and required system administration) would be used for processes as provided in item 4 of *Recommendation 1* above.
3. Data hosted on the server would be saved in simple data formats or in simple databases.

4. Python would also be used to manage and coordinate the server activities as described in Section 5.3.3 as well as processes in item 4 of *Recommendation 1* above.
5. Existing ESRI technology would be integrated within the visualization tools for spatial processing when possible and applicable.

8 Conclusions and Next Steps

The IJC is moving to embrace the capabilities and advantages that web technologies can provide in terms of data visualization (e.g. efficient access to content with only a web browser; displaying information in new, innovative ways; engage in interactivity such as panning and zooming on a web map). Web technologies evolve at a rapid pace, requiring a strategic approach towards implementation ensuring that risk to investment is minimized. A well-defined approach will ensure a lasting product that will provide value for the IJC and its stakeholders.

The NRC conducted a desk-based review to establish where user-friendly interactive web-based visualization tools could be used to the benefit of the IJC, and to identify the current technologies available to create these tools. Several potentially useful tools were demonstrated using prototypes. The following is a list of criteria that were applied during the review and prototype development:

- The visualization tools must have a high quality aesthetic and a visual appeal catering to a broad audience.
- The software supporting the tools must be well established and employed by other organizations to reduce risk of obsolescence.
- The overall system design must ensure components can be interchanged and reused easily in the future.
- The visualization tools must be fully automated with only minimal maintenance required by the IJC, its boards and committees.
- The majority of input data to the visualization tools must be compiled from available online resources, requiring minimal and transient data storage by the IJC.
- Simplicity must be exercised to ensure for a quick response time to the end users
- The system must embrace technology that the IJC already has access to, most notable being the ESRI tools.

Although a number of boards and committees answer to the IJC, each with distinct responsibilities, priorities and interests; five common themes emerged: (i) data informing the management and operation of control structures, (ii) water quality, (iii) streamflow, (iv) biology data and (v) precipitation, snow and climate data.

The IJC has a wide ranging audience who would benefit from the use of visualization tools made available on the IJC website. The audience includes technical experts who are directly involved as members of IJC's boards that must meet a number of regulatory needs, technical and non-technical stakeholders who rely on the provided information and are affected by the decisions made by the IJC and its boards, the local community who will benefit by the information provided, and IJC's staff and partners (e.g. communications personnel, scientists and engineers working on IWI projects) who can use the tools to further engage the public and provide key messages using simple visualizations.

The review of the IJC's technical capabilities has emphasized that the organization is fully embracing modern technologies that effectively communicate the organization's efforts and contributions to their official mandate as well as to the IWI projects. This experience will be beneficial in support of the effective development of data visualization tools.

A number of technologies were reviewed and discussed and many of these were used towards the development of prototypes. Based on IJC's identified needs and objectives, appropriate technologies were shortlisted. Final recommendations were developed and are provided in Section 7.

8.1 Next steps

The following section provides a list of next steps in order to meet the overall objective of providing a web-based interactive visualization technology for the IJC:

1. Determine the best way forward for the IJC provided the recommendations in Section 7 of this document. The recommendations are meant to be implemented in parallel answering to different needs of the organization. In short, *Recommendation 1* is best suited to provide clean and modern looking visualization tools to the audience of the IJC's website for high impact messaging; *Recommendation 2* is best suited to provide timely and up-to-date information for day-to-day messaging at a minimal cost; while *Recommendation 3* is best suited to provide visualization tools whose development environment is very dynamic and flexible, and where IJC boards and affiliated groups manage and share information dynamically.

2. Establish a web server that will house the following important components of the system: the programming code of the visualization tools, the server management software as described in Section 5.3.3, and the storage space for transient data and any other data that are not accessible online through the IJC or another provider. In parallel, a choice of browsers must be identified that will be supported by the visualization tools. The smaller the number of browsers selected, the easier it will be for the IJC to support the tools in the future.

3. Develop server software programming code that will perform the following capabilities:
    a. Communication to third-party data providers for downloading and updating necessary data;
    b. Transposing of third-party and IJC's internal (as discussed in Section 4.5) data to agnostic data formats;
    c. Storing of transient data;
    d. Additional computations and processing of input data for further analysis (e.g. processing of spatial data to provide sub-basin averages). This step will also take into consideration additional existing analysis code developed by the IJC and its partners that can be reused to support the visualization tools (as discussed in Section 5.5.2).

4. Develop visualization tools that answer to the needs of five themes identified in Section 3.1. The tools should be data agnostic, simple, transferrable across a multitude of needs, and fully automated. The NRC will require IJC's direction with respect to which users as identified in Table 1 should be targeted, if not all. Wherever possible, implementation of

the ESRI tools should be embraced. This step is further separated into sub-tasks in chronological order:

a. Develop interactive prototypes that focus on the interface.
b. Build the application and prepare for design change requests.
c. Test the tool with various users and scenarios.
d. Test the tool with programming code of step 3 above.

5. Conduct regular stakeholder meetings throughout the development process to ensure that the tools remain relevant to the intended story line and key messages.

9 References

Continuum Analytics (n.d.). *Why Python for big data?*. [online] Available at: < https://www.continuum.io/why-python > Accessed: April 30, 2017.

Fisheries and Oceans Canada (2016). *Three web services providing access to official water level data.* [online] Available at: < http://tides.gc.ca/eng/info/WebServicesWLD > Accessed: April 12, 2017.

Github (n.d.). *Python Bindings to the NOAA National Digital Forecast Database (NDFD).* [online] Available at: < https://github.com/rconradharris/python-noaa > Accessed April 30, 2017.

High Scalability (2011). *6 Lessons From Dropbox – One Million Files Saved Every 15 Minutes.* [online] Available at: < http://highscalability.com/blog/2011/3/14/6-lessons-from-dropbox-one-million-files-saved-every-15-minu.html > Accessed April 30, 2017.

IJC. (2014) *2014 Activities Report*. International Joint Commission, Canada and the United States. [online] Available at: < http://www.ijc.org/files/publications/IJC-activities-report-2014-English-final.pdf > Accessed: April 11, 2017.

Python (n.d.). *Python Success Stories.* [online] Available at: < https://www.python.org/about/success/usa/ > Accessed April 30, 2017.

Quora (2016). *What is python used for at Google?.* [online] Available at: < https://www.quora.com/What-is-python-used-for-at-Google > Accessed April 30, 2017.

The R Projects (n.d.). *What is R?*. [online] Available at: < https://www.r-project.org/about.html > Accessed April 30, 2017.

U.S. Geological Services (2016). *USGS Daily Values Site Web Service*. [online] Available at: < https://waterservices.usgs.gov/rest/DV-Service.html > Accessed: April 12, 2017.

W3Counter (2017). *Browser & Platform Market Share April 2017*. [online] Available at: < https://www.w3counter.com/globalstats.php > Accessed May 2, 2017.

Appendix A – Evaluation of data availability at IJC

The information provided in Table below was collected by IJC's advisors. To gather the data the advisors reviewed a number IJC reports and products produced by the IJC and its boards and communicated with other IJC members and liaisons of the IJC boards. The intent of this exercise was to improve understanding of the data that is of importance to the IJC and its stakeholders. This is by no means an in-depth scope of IJC's current initiatives and data.

Preliminary evaluation of data availability at IJC

Watershed Board	Data Type (description)	Data Source	Visualization Options/Ideas
St. Croix	Water Quality (multiple parameters)	Milltown - USGS	Timeseries Plot of constituents, vs guidelines
	Biology Data (Fish Counts: alewife, sucker, bass, etc.)	Milltown - ASF?	Multi-year timeseries
	WQI Score	Milltown, Forest City – Board	Timeline vs Rating
	Control Structure Data (Water Levels, discharges)	Forest City, Milltown - ECCC	Time series vs. min/max water levels
Souris River	Water Quality (multiple params)	Sherwood and Westhope – ECCC, USGS, ND – Excel Sheets	Line with exceedences Box whisker Current vs. historical.
	Control Structures (Multiple Water Levels, inflows, outflows)	SK-WSA, NPS	Current vs. historical
	Natural Flow Calculations	ISRB	Current vs. historical
	Runoff Potential Map	MB	Current Current vs. Historical
	Stream/River flow	USGS/EC	Time series, flow-weighted concentrations for parameters of concern (e.g. Cl, SO4, TDS)
	Soil moisture		Heat map
Red River	Streamflow	MB – Emerson	Timeseries - Current vs historical
	Water Quality (multiple params)	Emerson - ECCC Other WQ Data collected by agencies throughout the basin Sulphate in Sheyenne River / Red River	Line with exceedences? Box whisker? Current vs. historical. Stations Map
	Biology data (Benthic Invertebrate)	Benthic Invertibrate Sampling – Emerson Sherwood	Report BII against historical (timeline?)
	Control Structures (Water Levels, Flows)	Devil's lake	

Watershed Board	**Data Type (description)**	**Data Source**	**Visualization Options/Ideas**
	Stream/River flow	USGS/EC	Time series, flow-weighted concentrations for parameters of concern (e.g. Cl, SO4, TDS)
	Soil moisture		Heat map
St. Mary – Milk	Natural Flow Calculations	SMM-TWG (USGS/ECCC)	
	Flow Allocations and Deficits	SMM-TWG	
	Control Structures (Water Levels, discharges, Time series, with historical stats and supply volume thresholds Reservoir Storage)	SMM-TWG	
Rainy-Lake of the Woods	Control Structures (Lake levels, Inflows, Outflows)	LWCB	Rule Curve Plots. w/ historical
	Water Quality (multiple parameters)	MOE, MPCA Chlor., TP, - Basins - Rivers	Time series Box plots
	Precipitation, Temp	NOAA/NWS/Can.	Historical heat map
Osoyoos	Streamflow	Similkameen – USGS	Timeseries – current vs historical
	Control Structures (Water Levels, Inflows, Releases)	Okanagan Lake	Timeseries – current vs historical
Kootenay	Control Structures (Water Levels, Inflows, Releases)	Kootenay Lake	Timeseries – current vs historical
	Hydraulic Control Condition	Kootenay Lake	Timeseries
All Basins	Heating Degree days	Historical Forecast Data?	
	Basin precipitation (historical and forecast)	ECCC CaPA / MSC Forecast, NOAA Reanalyiss and Forecast	
	Temperature (historical and forecast)	Historical Forecast, Gauges	
	Snow Depth /snow cover/ SWE		
Lake Superior	Lake Superior Outflow (monthly) + forecast	ILSBC (ECCC and USACE)	Timeseries, current vs. historical
	Great Lakes water levels (daily means, monthly means)	Coordinating Committee (individual gauge data from CHS and NOAA)	Timeseries, current vs. historical, spaghetti plots, statistical representations (e.g., box plots?)

Watershed Board	Data Type (description)	Data Source	Visualization Options/Ideas
	Great Lakes monthly water level forecast	Coordinating Committee (published by CHS and USACE)	Timeseries, range/band of future conditions
	St. Marys Rapids flow (monthly) + Compensating Works gate setting?	ILSBC (ECCC and USACE)	Time series, current vs. historical, spaghetti plots, statistical representations (e.g., box plots?)
	Water supplies, precipitation, runoff, evaporation, temperatures, snow water equivalent, etc.	Various, obtained through Coordinating Committee	Time series, current vs. historical, spaghetti plots, statistical representations (e.g., box plots?)
Lake Ontario – St. Lawrence River	Lake Ontario Outflow (hourly daily, weekly) + forecast	ILOSLRB (ECCC and USACE)	Timeseries, current vs. historical, spaghetti plots, statistical representations (e.g., box plots?)
	Great Lakes water levels (daily means, monthly means)	Coordinating Committee (individual gauge data from CHS and NOAA)	Timeseries, current vs. historical
	Great Lakes monthly water level forecast	Coordinating Committee (published by CHS and USACE)	Timeseries, range/band of future conditions
	Water levels and flows throughout Lake Ontario – St. Lawrence River system (hourly, daily, weekly) + forecast at some locations	ILOSLRB, OPG, NYPA, CHS, NOAA, ECCC, ORRPB, DFO, CEHQ	Timeseries, current vs. historical (+linked to and displayed in an online map/GIS application?)
	Water supplies, precipitation, runoff, evaporation, temperatures, snow water equivalent, etc.	Various, obtained through Coordinating Committee	Time series, current vs. historical, spaghetti plots, statistical representations (e.g., box plots?)

Appendix B – Data Providers

NOAA data providing methods

- File transfer protocol (FTP) access, downloading data in ASCII format, recommended for large volumes of data
- Search tool, web-based data access system.
- Map tool, access through the National Centers for Environmental Information (NCEI) GIS map services.
- Web services, programmatic access to NCEI data and use custom and standard implementations https://www.ncdc.noaa.gov/cdo-web/webservices/ncdcwebservices)
- ISD Lite FTP Access, a derived product from the full dataset, for general research and scientific purposes.
- Daily summaries, FTP access to daily summarized data from more than 9,000 stations.

API v2 and API Legacy are two currently available climate data online APIs offered by NCEI. API v2 provides access to near real-time climatic data and allows searching and accessing datasets and station, and/or location data in JSON format. API Legacy provides station time-series data in XML and CSV formats.

ECCC

ECCC provides a free data service which includes hourly outputs from numerical climate models for North America, a variety of radar data, data from Canadian site observations, etc. ECCC provides a number of free data through a service known as the Datamart in formats of XML, CSV and GRIB2. Some of the data types provided here include:

- Forecast and observations of air quality;
- Weather warnings from ECCC;
- Forecasting monthly to multi-seasonal climate conditions;
- Regional Deterministic Precipitation Analysis (RDPA);
- Global Deterministic Prediction System (GDPS);
- North American Ensemble Forecast System (NAEFS);
- Regional Ensemble Prediction System (REPS);
- High Resolution Deterministic Prediction System (HRDPS);
- Regional Deterministic Wave Prediction System (RDWPS);
- Marine weather forecasts;
- Real-time hydrometric data for hourly and daily water level across Canada;
- Official forecast values issued by ECCC;
- Global Ice Prediction System (GIOPS);
- Scribe Nowcasting Matrices;
- Observation data of the Canadian meteorological stations;
- Real-time observed weather data produced or distributed by ECCC and partners; and
- Images from the Canadian weather radar network.

A cost recovery data service is available to those needing meteorological data on a 24/7 operational basis.

GeoMet Geopspatial Web Services provide access to raw numerical weather predictions, model data layers and the weather radar mosaic in WMS and KML formats.

HYDAT is a quarterly updated database containing daily mean, monthly mean, annual maximum and minimum daily mean and instantaneous peak water level and discharge data, as well as sediment data (for sediment stations). The data are delivered in MS Access or SQLite formats.

CHS

CHS web services use a communication protocol known as SOAP. The provided data are in XML format. Each service has an XML description which is useful for automatically generating code to use the service.