7th ISE & 8th HIC
Chile, 2009

THE HEALTH-E-WATERWAYS PROJECT:

DATA INTEGRATION FOR SMARTER COLLABORATIVE WHOLE-OF-WATER CYCLE MANAGEMENT

ABDULMONEM ALABRI, JANE HUNTER
School of ITEE, University of Queensland,
Brisbane, Queensland 4072 Australia

CATHARINE VAN INGEN
Microsoft Research
San Francisco, CA, USA

EVA ABAL
South East Queensland
Healthy Waterways Partnership
Brisbane, Queensland, 4001Australia

The Health-e-Waterways Project is a collaboration between the University of Queensland, Microsoft Research and the South East Queensland Healthy Waterways Partnership (SEQ-HWP) (a consortium of over 60 local government, state agency, universities, community and environmental organizations). The aim of the project is to develop a highly innovative framework and set of services to enable streamlined access to a collection of real-time, near-real-time and static datasets acquired through ecosystem health monitoring programs (EHMP) in South East Queensland. This paper describes the underlying water information management system and Web Portal that we are developing to enable the sharing and integration of the high quality data and models for SEQ water resource managers. In addition we will describe the interactive and dynamic ecosystem reporting services that we have developed and the WaterWiki that is being established to enable knowledge exchange between the online community of Queensland's water stakeholders.

INTRODUCTION

Climate change, urban development and population growth in South East Queensland are putting significant pressure on urban water supplies. Billion dollar investments in new water projects, such as the South East Queensland Water Grid, will enhance and secure water supply to the region. However the potential benefits of these massive investments in physical infrastructure will not be fully realized, without the investment in corresponding *cyberinfrastructure*. Scientists, urban planners and policy makers require access to integrated water information management systems that enable them to track water movement, consumption and quality across the "whole-of-water-cycle". They require access to high quality, complete datasets and accurate predictive models on which to base water allocation decisions. Understanding and satisfying the competing needs of water users, requires the integration of data and models that reflect the behaviour of many systems including: climate, rainfall, catchments, reservoirs, rivers, agricultural demand, industrial demand and urban demand. The linking of models requires a detailed understanding of: the questions being asked, the data that is available and the relationships between the datasets and models being employed. In parallel with the demand for more sophisticated querying and decision support interfaces, there is a trend towards wide-spread deployment of sensors capturing real-time information including volume flows, water quality, images and video. These real-time and near-real-time, temporal and visual data streams will require more sophisticated, high speed, data processing, indexing and archival services.

OBJECTIVES

The primary objective of this project is to enable and promote the sharing and collaborative integration and analysis of high quality information concerning water.

Figure 1 (below) provides an overview of the components and architecture of the proposed system.

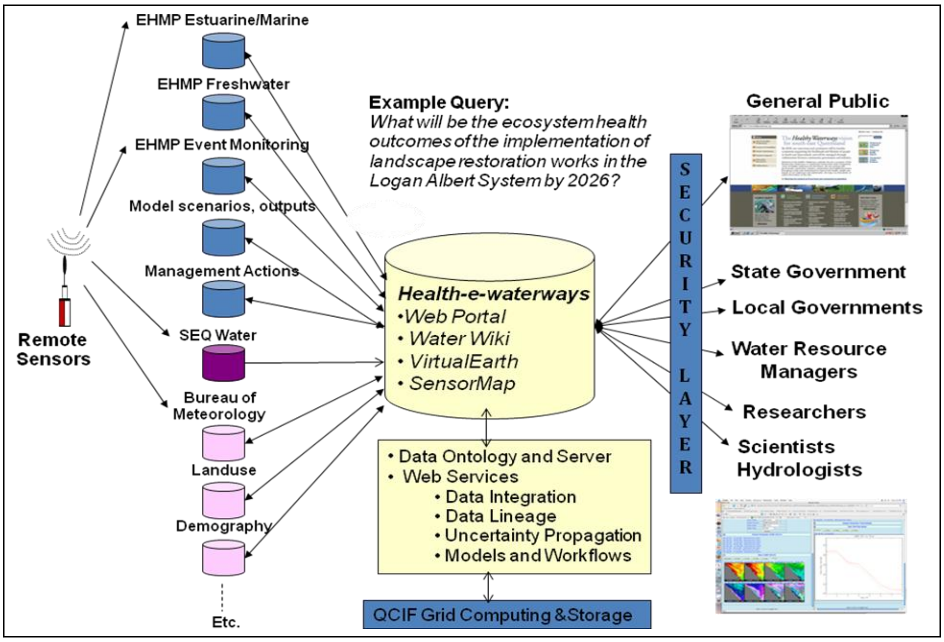

Figure 1: System Overview and Architecture

The project will provide a Web portal to the latest relevant data and models for Queensland scientific, education, government, community and commercial stakeholders with an interest in water research, conservation, and management. It will develop and apply leading edge ICT to satisfy the anticipated, unprecedented demand for sophisticated water information integration, analysis and management services globally.

It will enable private, secure but integrated access to the vast amounts of water-related data being generated and shared regionally, nationally and globally. It will deliver improved data management services to the partner organizations of SEQ-HWP [1] which in turn will lead to improvements in the security and sustainability of our water supplies in Queensland. It will provide valuable cyber-infrastructure for leading water scientists and hydrologists who are developing innovative water-recycling and water monitoring technologies. It will generate software services and tools for enhanced water information integration, analysis and management. Although the focus is on the requirements of South East Queensland stakeholders initially, the services that are generated will be designed to enable easy deployment and application to other regions.

METHODOLOGY

The research plan comprises six key components which build on top of an innovative combination of semantic web technologies and web services to access, process and integrate data and models held within both centralized and distributed hydrological databases:

1. Identification and prioritization of the key stakeholder user requirements, queries and datasets;
2. Development of the common data models and ontologies to integrate both static and real-time data streams, visual, spatial and temporal data, legacy databases and newly generated datasets;
3. Design and implementation of the semantic interoperability layer on the scientific data server;
4. Development of a Web-based querying, visualization and presentation interface – that integrates GIS technologies (e.g., ARC Hydro) with VirtualEarth-type interfaces, ontology-based querys and SensorMaps - to display the latest integrated data sets through a mapping interface;
5. Development of a secure Web Portal and *WaterWiki* that provides different levels of collaborative access to data, models, services, storage and compute power. Standardized authentication and access protocols will enforce controlled access to resources and software;
6. Development of a model registry and scientific workflow tools. This component will enable users to upload and share models as web services through a common model registry, link them using scientific workflows and execute them over grid computing infrastructure.

DATABASE SURVEY, MODELING AND INTEGRATION

The first phase of the project involved identifying the currently available databases that are most critical and relevant to the water stakeholders of South East Queensland. In the long term the stakeholders want to be able to integrate the following datasets:

- Ecosystem Health Monitoring Program (EHMP) Data:

- Climate data – rainfall, temperature, humidity from the Bureau of Meteorology;
- Vegetation data – from the Australian Herbarium;
- Biodiversity data – from the Atlas of Living Australia;
- Remote Sensing Satellite Imagery – available from NOAA – measuring surface temperature and soil moisture;
- Atmospheric and Air quality data – available via FluxNet

However in the immediate short term, they want to be able to integrate and analyse the data being captured through the Ecosystem Health Monitoring program (EHMP) [2]. This program involves sampling 30 freshwater indicators at 100 sites twice a year and 250 estuarine/marine sites every month. The number of sampling sites and size of the datasets is expected to rapidly expand in the near future, with the planned installation of sensor networks in Moreton Bay. The EHMP data comprises the following databases:

- Ecosystem Health Monitoring Program (EHMP) – Freshwater (Dept NRW)
- EHMP Estuarine / Marine (EPA)
- EHMP Monitoring (Ambient and Event) (DNRW)
- Management Action Database (MAD) (SEQ-HWP)

Our first activity was a detailed analysis of these databases in order to develop a common data model that will support the integration of these datasets through a common search interface. A detailed report describing the databases and recommending the best approaches for integration is available [3].

THE COMMON MODEL/ONTOLOGY

The analysis of the existing databases enabled us to develop an Entity-Relationship model that incorporates all of the key classes, properties and relationships. Figure 2 below illustrates graphically the model that has been determined by identifying the common top-level entities from the 5 key databases that we have analysed. The advantage of developing a model that is object-oriented and based on a common set of top-level entities is that it is extensible. As new databases become available, it will be relatively easy to incorporate these by refining and extending the existing model. Once the EHMP model was determined and agreed, we represented it as an OWL ontology.

During the development of this common model, we drew on a number of pre-existing water information models of relevance but refined and adopted to accommodate local terms and parameters:

- The CUAHSI Community Observations Data Model (ODM) [4]
- The OpenGIS Observations and Measurements Model [5]
- The SEEK Observation Ontology (OBOE) [6]
- WRON-RM (Water Resources Observation Network Reference Model) 0.1 [7]

We will continue to monitor developments towards the establishment of national and international standards for documenting water information and models and ensure that our ontology and approach is compatible with open standards in this discipline.

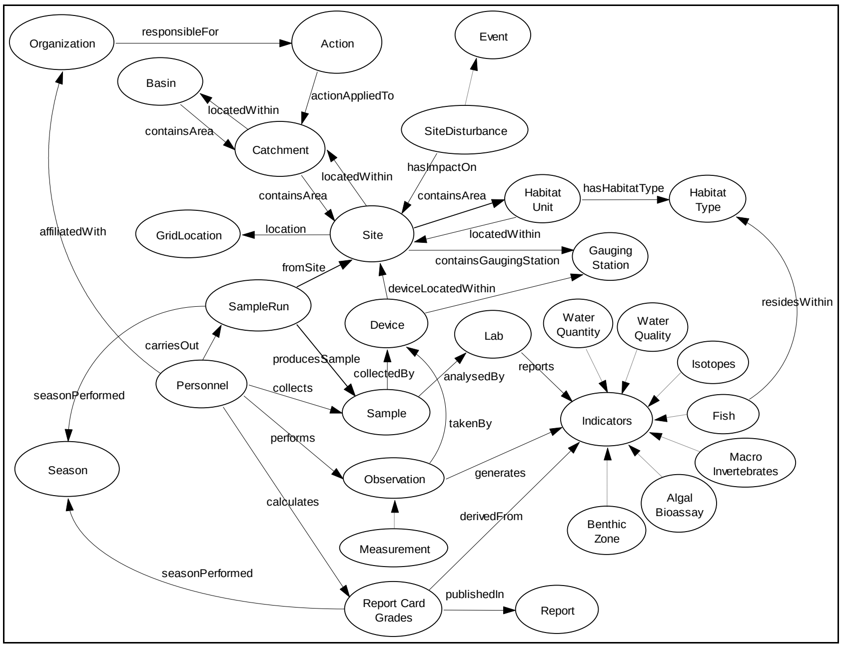

Figure 2: Overview of the EHMP Ontology

Data Access and Integration

The next step involved resolving data access agreements with the custodians of the individual databases. This was a very time-consuming process that required numerous meetings and discussions with stakeholder groups to gain their trust and support. We are currently working with the OAKLaw project to identify licenses that could be applied to enable much easier and streamlined access to the databases via web services interfaces.

Once the data agreements were resolved, we were given copies of the datasets for local storage and re-use. We then developed XSLT scripts to map the data from the legacy structure to the common model (Figure 2). We then loaded the mapped data into a common datastore on a SQLServer database.

QUERYING, REPORTING AND VISUALIZATION SERVICES

In the long term we aim to provide interactive generic querying services over the aggregated data sets. However, initial stakeholder meetings identified the immediate need for a more efficient approach to the production of the annual EHMP Report Cards, published by SEQ HWP [8].

Currently hard copies of the EHMP Report Cards (that report the ecosystem heath grades for catchments in SEQ) are printed and published annually. Politicians and planners from 20 different agencies (including the EPA), 4 universities, 18 local councils and CSIRO, use the report cards to make decisions with respect to land use, water quality, allocations and investments in water recycling plants etc. The report cards provide a standardized method for understanding and comparing the health of catchments between regions and over time. To date, these report cards are largely produced manually, via the following process which is documented in [9]:

1. Values for each of the 5 indicators (physical, chemical, nutrients, ecosystem processes, acquatic macroinvertebrates and fish) are calculated from the 16 indices recorded for each site and for two seasons (spring and autumn);
2. Index values are compared against Ecosystem Health Guideline values to derive standardised scores that range from 0 (unhealthy) to 1 (healthy);
3. Standardised scores are averaged across combinations of indices, sites and seasons to provide intelligible summaries for catchments and seasons;
4. Graphical summaries of results are produced to enable the comparison of indices and indicators across reporting areas and between seasons: box and whisker plots, horizontal bar charts, and Ecosystem Health plots (EcoH plots).
5. Summarisation of the EHMP results culminate in a report card grade from A to F.

Currently the production of the EHMP Report Cards process takes about 5 months, with the most time-consuming step being the manual generation of the graphs and EcoH plots using Photoshop Illustrator. The results for July-June of each year, are published in the following November.

For the past 6 months, we have been working with the SEQ-HWP staff who are responsible for generating the annual ecosystem report cards. We have been developing software services that will enable the report cards to be generated dynamically via a Web-based Map interface that overlays the database containing the spatio-temporal monitoring data. Figure 3 shows the results of our collaboration – the Web-based user interface on the left hand side and the architectural components of the system on the right hand side.

The GUI enables users to specify:

- Spatial regions of interest (e.g., particular catchments or sites) through either a GoogleEarth or Microsoft VirtualEarth interface;
- Concepts or indicators of interest through the EHMPOntology (a localized and extended version of the ODM ontology developed by CUAHSI HIS);
- Seasons or years of interest through a timeline.

A Report Card Grade is generated for the specified catchment and period. Clicking on a grade, displays the corresponding EcoH plots, dynamically generated from the 5 indicators in the underlying SQL Server database. Clicking on an EcoH plot, displays the actual raw data (16 indices) used to generate the indicators and plots. The combined use of RDF (Resource Description Framework), Microsoft Silverlight and SQL Server enables very fast aggregation of datasets and the dynamic generation of graphical mashups.

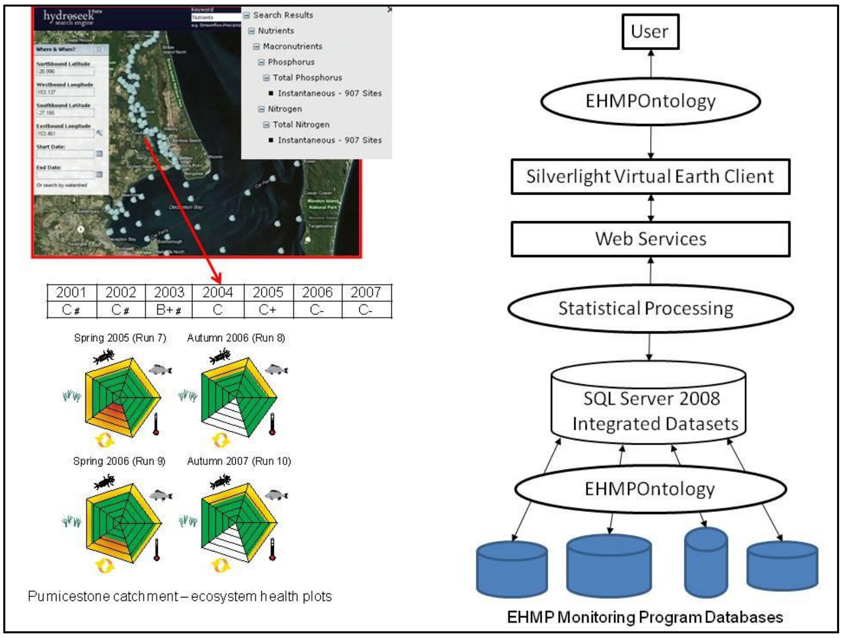

Figure 3: The User Interface to Dynamically Generated Report Cards (LHS) & A high-level view of the System Architecture (RHS)

FUTURE WORK

We are currently only into the 3rd Phase of the 6 phases described in the Methodology section and have significant further work to do in order to evaluate and refine the system based on user feedback. More specifically, over the next 12 months we are planning to focus on the following activities:

- Add ontology-based conceptual querying (based on SPARQL and SQL server) to the Virtual Earth interface;
- Investigate periodic automated web service retrieval of data updates from agency databases;
- Identify new datasets to be incorporated into the system e.g., near-real-time sensor data, vegetation data and satellite imagery;
- Incorporate dynamic MatLab analysis and reporting tools into the IE Browser;
- Investigate animated visualizations of modelled water quality/quantity data;
- Develop simple services to encourage uploading and quality control of community data – the SPIRE project's Spotter software [10] is of particular interest.
- Hold stakeholder meetings and usability studies to acquire feedback from SEQ-HWP users and refine the system;
- Establishment of a Water Wiki for exchange of information between the SEQ-HWP partners.

CONCLUSIONS

Numerous state, national and international agencies are advocating the need for standardized frameworks and procedures for environmental accounting. The Health-e-Waterways project provides an ideal model for delivering a standardized approach to the aggregation of ecosystem health monitoring data and the generation of dynamic, interactive reports (that incorporate links back to the raw data sets). The system we have described here will not only save agencies significant time and money in generating environmental accounts, but it can be used to guide regional, state and national environmental policy development, based on high quality evidential data. Although significant further work is required to fully develop the Health-e-Waterways system, we believe that it provides an exemplary and extensible framework for delivering water monitoring data in the most efficient and accessible way.

REFERENCES

[1] South East Queensland Healthy Waterways http://www.healthywaterways.org/

[2] Ecosystem Health Monitoring Program (EHMP) http://www.ehmp.org/

[3] J. Hunter and B.Mauger, "Health-e-Waterways Database Survey Report", May, 2008 http://www.health-e-waterways.org/reports.php

[4] D.Tarboton, J.Horsburgh, D.Maidment, B.Jennings, "CUAHSI Community Observations Data Model Version 3", http://www.cuahsi.org/his/odm.html

[5] Open Geospatial Consortium (OGC) Observations and Measurements http://www.opengeospatial.org/standards/om

[6] SEEK KR Ontologies http://seek.ecoinformatics.org/Wiki.jsp?page=KROntologies

[7] R. O'Hagan, R. Atkinson, S. Cox, P. Fitch, D. Lemon, G. Walker, "A Reference Model for a Water Resources Observation Network", MODSIM07, New Zealand, 2007 http://www.mssanz.org.au/MODSIM07/papers/18_s56/AReferenceModel_s56_O'Hagan_.pdf

[8] Ecosystem Health Monitoring Program Annual Report Cards, http://www.ehmp.org/annual_report_cards.html

[9] M.J. Smith and A.W.Storey, "Design and Implementation of Baseline Monitoring (DIBM3): Developing an Ecosystem Health Monitoring Program for Rivers and Streams in Southeast Queensland". Report to the South East Queensland Regional Water Quality Management Strategy, Brisbane, 2001.

[10] Spire: Ontology Tool (SPOTTER) http://spire.umbc.edu/spotter/