Enabling Assessment and Propagation of Information Quality in the South Esk Hydrological Sensor Web

Heiko Mueller, Qing Liu, Quan Bai, Corne Kloppers, and Andrew Terhorst

Tasmanian ICT Centre, Australian Commonwealth Scientific and Research Organization (CSIRO), Hobart, Australia (heiko.mueller@csiro.au)

Over 200 organisations are required to provide water information to the Bureau of Meteorology (BOM) under the Australian Water Act 2007. Many of these organisations have developed their own proprietary systems for collecting and maintaining water information. Understanding the quality of information maintained by different organisations is a challenge.

The Sensor Web is an emerging technology concept that promises to solve heterogeneity problems induced by proprietary systems and revolutionise the way water information is collected and disseminated in Australia. Based on the Open Geospatial Consortiums (OGC) Sensor Web Enablement (SWE), the Sensor Web provides an interoperability layer over existing sensing infrastructure that not only allows uniform access to sensor description and sensor observations, but also to information derived from predictive models.

A Hydrological Sensor Web has been established in the South Esk river catchment in North-Eastern Tasmania based on the OGC's SWE standards. The South Esk Hydrological Sensor Web aggregates sensor assets owned and operated by multiple agencies. Observations from the aggregated sensor assets drive a rainfall-runoff model that predicts river flows at key monitoring points in the catchment. River managers use the flow predictions to manage water restrictions.

One crucial aspect to consider when providing access to raw and/or derived information is quality assurance. Users who have to make decisions based on the information need to know how accurate or reliable the information is. Three areas of research that previously emerged in the context of data curation can help in dealing with information quality and uncertainty in environmental systems: provenance, annotation, and versioning:

- Provenance information describes the entities and processes involved in the derivation of an information product. In the Hydrological Sensor Web, provenance allows the identification of the sensor observations, data transformations, and predictive models that were used to derive an information product.

- Annotation allows one to attach structured information to individual data items. Such annotations can for example describe the accuracy of individual observations, or known issues with sensing devices.

- Versioning allows one to keep track of changes to the data, like updates to the location of a sensing device, its predicted accuracy, or changes in ownership. Keeping track of previous data versions ensures that provenance information refers to the data as it was at the time a derivation process was run.

Provenance allows retrieval of information about the sensing devices, like sensor type, sensing accuracy, sensor location, or sensor calibration history. Knowing the derivation history of an information product furthermore allows one to retrieve annotations attached to entities or processes involved in its derivation. Together, this information enables a Hydrologist to assess the quality of derived information. Individual steps in the derivation process can further be augmented with models describing how these steps influence the accuracy or uncertainty of input and output values. Together with provenance and annotation information we are thereby enabled to compute (the predicted) accuracy of derived information.

Researchers at the CSIRO Tasmanian ICT Centre are currently incorporating techniques for provenance, annotation, and version management into the South Esk Hydrological Sensor Web. These extensions are based on technologies for the Semantic Web, namely RDF and SPARQL. The Resource Description Framework (RDF) provides a graph-based data model that is well suited to handle data, derivation information, and annotations. Domain ontologies describe domain specific concepts like observation accuracy and water balance models. These ontologies form the basis for annotations attached to entities following linked data paradigm. The SPARQL Query Language

for RDF allows efficient queries over a provenance-annotation graph. Versioning is currently not supported by RDF. Techniques that have recently been developed for archiving curated databases may fill this gap.

Together, these techniques will provide better insight into information quality and will build confidence in information used by decision-makers in the water resource management domain.