Capture and Management of Hydrometeorological Phenomena Using Sensor Web Concepts and Scientific Workflow

Siddeswara Mayura Guru[#]
CSIRO Marine and Atmospheric Research, Hobart, Australia. Email: siddeswara.guru@csiro.au

Chris Peters
CSIRO Tasmanian ICT Centre, Hobart, Australia. Email: chris.peters@csiro.au

ABSTRACT: Deployment of a large number of sensors across the world has provided several opportunities and challenges. One of the opportunities is the ability to conduct macro-scale experiments using heterogeneous data with wider temporal and spatial variability. This has potential to drastically increase the scientific outcomes and discover new knowledge about our environment. Challenges include dealing with heterogeneity of data for distribution in a seamless environment and management of tasks and processes that manipulate disparate data to discover new knowledge. In this paper, we have adopted Open Geospatial Consortium – Sensor Web Enablement (OGC-SWE) framework to standardise the description and access of hydrometeorological phenomena. This will ease the seamless discovery and sharing of data. We have also proposed Kepler inspired scientific workflow based hydrology workbench to compose heterogeneous hydrometeorological data services and processes for knowledge discovery. These concepts have been implemented on a test-bed of hydrometeorological sensors deployed in the South-Esk River catchment located in the north-east of Tasmania in Australia.

1 INTRODUCTION

Wireless sensor networks have gained prominence in the last couple of decades due to the advancement of hardware and communication technologies. Most of the research in wireless sensor networks is predominantly focused on in-network processing, energy minimisation and routing data to a gathering unit. The sensors are (semi-) autonomous, do continuous sensing; collect data of high temporal resolution and their knowledge is predominantly restricted to the network.

In contrast, several real-world application domains using wireless sensor networks can be complex and often experiments are conducted in macro-scale. Phenomena are measured at varied spatial and temporal resolution and large volumes of time-series data are collected. Complex analysis may require data from one or more network deployments where some of the sensors are completely autonomous. In these experiments, the major focus is not only to gather data from sensors, but also to effectively manage data to derive information for further knowledge discovery.

A vast deployment of sensors around the world has significantly increased the opportunity to conduct inter-disciplinary macro-scale experiments. To enable these experiments, sensor data needs to be shared among multi-disciplinary researchers across the world. The first step towards this is to create a "web of sensors" which consists of sensors of the world. This would help to discover and share sensor data and information in a seamless manner. However, there are several challenges for sensor discovery and data sharing in a distributed environment.

Several science domains like biology, chemistry and geology have used advance computing infrastructure to bridge the gap between raw data and knowledge discovery. We believe this will also play a key role in sensor networks in addressing data management and processing issues. This multi-disciplinary research is commonly called X-informatics where X stands for bio, chem and geo etc.

2 MOTIVATING APPLICATION: HYDROLOGY

Hydrology is a complex science of understanding the water cycle of the Earth. Traditionally, hydrologists use domain knowledge and mathematical models to solve water-related problems. The problems could be water availability, quantity, quality, usage

and distribution. Often these problems focus on single catchment but, with the understanding of the earth atmosphere and land surface, the problems can be scaled to larger areas (e.g., multiple catchment, river basin and continental scale). This is called macroscale hydrology (Lettenmaier, 2001). In recent times, new observation systems (e.g., satellites, in-situ sensors) are used to capture phenomenon, which leads to an increase in the volume of available data. This poses a challenge to the effective use of this data in addition to historical data to improve modelling capability of hydrology phenomenon. Sharing these data will have a huge impact on the science. However, there are difficulties in accessing data from different networks to answer complex queries. From data perspective, one of the problems is that there are no general standards to represent sensor data. This is an impediment to share machine readable data from different networks to answer complex query (Balazinska, 2007).

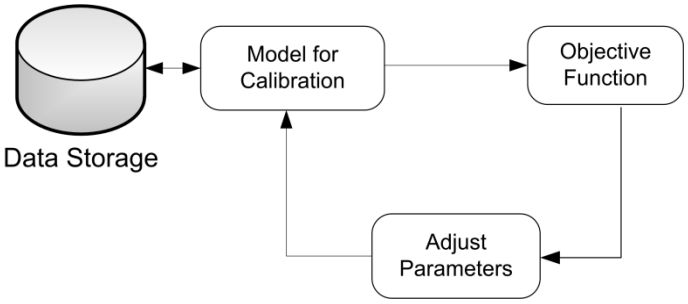

Figure 1. Model calibration process

2.1 *Hydrology Example*

Let us consider a simple scenario of modelling a river flow. Hydrologists need to identify the rainfall-runoff model suitable for the catchment, gather required data, calibrate and validate the model. The complexity of the model can vary considerably but the outputs of most models are surface runoff. However, most models attempt to replicate the signal in runoff of surface and groundwater processes. The data sets required for model calibration could be catchment characteristics, rainfall, evapotranspiration, river flow and other meteorological datasets. The data sets need to be checked for gaps in time-series and should be filled with necessary gap filling techniques. The model will be calibrated against the observed river flow and the model parameters are tuned to get the best result. For gridded models, the calibration process will be performed for each grid cells and the parameters are also tuned for every grid. Figure 1 show different processes involved in the calibration process and it is computationally intensive.

Calibration and flow forecasting processes look simple for a single catchment. However, to study the river basin, several catchments need to be considered to understand the water balance of individual hydrological response units through the interaction of climate, vegetation and soils (Mulligan, 2005). For example, the Murray Darling Sustainable Yields Project (MDBSY) (CSIRO, 2009) modelled all of the catchments within the Murray Darling river basin as grid cell of 5 × 5 km and the entire basin is made up of 4000 grid cells (Fitch, 2008). A rainfall-runoff model will run for each grid cell and the runoff of the catchment is the sum of the runoff of all the grid cells considered in the catchment.

To add to the complexities, the MDBSY project used different file formats and structures to store data. The initial data includes a mapping from a grid cell to a catchment, historical time series of observed rainfall and climate data. Hydrologists need to pre-process the data before calibrating a model, check calibration results and tune the parameters if required, verify the parameter set mapping for each grid cell and submit for simulation. Each step of the process is performed as a separate task and there is no direct interaction between the processes. This methodology requires constant intervention from hydrologists; lacks flexibility and is time consuming (Guru, 2009).

There are several challenges to accelerate the processes described. The first challenge is to deal with the data which are in different formats. The problem will be compounded if data needs to be distributed seamlessly in a sensor web environment. This challenge can be identified as *Data Integration and Dissemination*. The second challenge is managing tasks that manipulate data to discover new or intermediate knowledge. The manual composition of these tasks is painstaking and prone to error. Therefore (semi-) automatic orchestration of tasks is required. This challenge is called *Process Integration.*

In this paper, we will discuss and propose some of the solutions to the above challenges in the hydrology domain. We also describe an experimental test-bed that has been developed to test some of the concepts proposed. The test-bed demonstrates some of the complexities of the sensor network applications and acted as the motivation factor of the work described in this paper.

3 DATA INTEGRATION AND DISSEMINATION

Data integration is the process of combining different data. Data can be either from the same or different sources, homogeneous or heterogeneous. Heterogeneity of data is divided into syntax, structure and semantic (Sheth, 1998). Data that are not stored in databases are represented and stored in different ways. The use of XML as a uniform data exchange syntax can resolve syntactic differences, however, there are specialised file formats (e.g., netCDF, HDF, shape files) for data representation due to necessity and performance. In certain cases, similar data is represented using different schemas leading to structural differences. The schema mapping and integration is a well studied database topics but, there are still issues related to schema integration and mapping techniques between source schemas and derived schema. Traditional methods of storing data (e.g., database) don't capture the semantics of the data (e.g., origin, scope and context of the dataset). This is due to lack of expressiveness of the languages that define database schemas (Ludäscher, 2005).
In Australia, it is estimated that 260 organisations gather hydrological data and transfer it to Bureau of Meteorology (BOM, 2008). Each organisation uses different methods to gather information and several organisations gather the same data leading to duplication. Data is also stored in different formats (spreadsheets, database, text file etc.) and most organisations provide none or very little metadata. Therefore, data is heterogeneous from syntactical, structural and semantic perspectives. Apart from data, there will also be differences from system operations. Different technologies are used for data transfer (e.g., FTP, HTTP), remote invocation and platform representation.

The first step towards the seamless data integration is the adherence to the Open Geospatial Consortium–Sensor Web Enablement (OGC-SWE). The OGC-SWE provides an open standard framework to exploit web-enabled sensors and sensor systems (OpenGIS, 2007). Using these standards the sensor web can be realised by inter-connecting spatially distributed and heterogeneous in-situ and remote sensors. The framework, which is based on Service Oriented Architecture (SOA) enables the discovery, exchange and processing of sensor observations.

The advantage of OGC-SWE is its ability to store, query and publish information based on location, boundary and relationships among geographic features and phenomena. It provides standard framework to describe phenomenon, measurement types and data types. This will make automatic data discovery and sharing a relatively easy task. However, in the real-world, data is always measured and represented in different formats. In Australia, the temperature phenomenon is measured in Celsius whereas in USA, the same phenomenon is measured in Fahrenheit. The schema of data representation may also change between the agencies. The water level can be denoted as "water-level" in some agencies and "water level" (without dash in between) in others, even though both represent same observed value. All these make the interoperability harder and the standards play a key role. OGC-SWE tries to work to develop a framework of open standards to enable the discovery, exchange, and processing of sensor observations, as well as tasking of sensor systems.

The SWE standards are built primarily on the following specifications:

Observation and Measurement (O&M): provides a standard model based on XML schema to represent and exchange observation results (OM, 2006).

Sensor Model Language (SensorML): provides information model and encodings that enables discovery and tasking of Web-enabled sensors, and exploitation of sensor observations. SensorML defines models and XML Schema for describing any process, including measurement by a sensor system and post-measurement processing (SensorML, 2010)

Sensor Observation Service (SOS): provides a standard web service interface to request, filter and retrieve sensor observations. SOS can also provide metadata information about the associated sensors, platforms and observations. It is an important component of SWE and acts as intermediary between a user and a sensor system's observation repository (SOS, 2010).

Sensor Planning Service (SPS): provides a standard web service interface to assist in collection feasibility plans and to process collection requests for a sensor (SPS, 2010).

Sensor Alert Service (SAS): provides a standard web service interface to publish and subscribe to alerts from sensors. This will be soon replaced by Sensor Event Service (SES) (SAS, 2010).

Web Notification Services (WNS): Provides a standard web service interface for asynchronous message delivery. This service can be used in conjunction with other services like SAS and SPS (WNS, 2010).

In OGC-SWE, data is exchanged between the services in the XML format. The O&M specifications which represent observation in an XML format would resolve the syntactic differences to a certain extent. The structural differences arise when the same data is represented in a different form. For example, the river flow can be represented as the volume of water flow or the height of the river (the flow can be derived by the rating curve). Information Model of OGC-SWE comprising of O&M and SensorML are key for structural integration. O&M discuss the representation and exchange of observed data whereas SensorML describes the geometric, dynamic and observational characteristics of sensors. It has a capability to associate a measured and derived value of the phenomenon with a particular geospatial location at a particular time. Even though SensorML is part of OGC-SWE framework, it can be used independently of SWE. This will give the flexibility of using the same sensor description in different sensor web technology like SensorMaps (Kansal, 2006).

Some of the important features of SensorML are that it provides the sensor information for data discovery, process chain to derive new data products and archive fundamental properties and assumptions regarding sensors. However, the standards do not support reasoning of data therefore, the common ontology to describe the sensors and processes are required. For example, there should be a clear distinction between the tipping bucket rain gauge and acoustic rain gauge even though the phenomena measured is rainfall. However, processes used for measurement in both the instruments are different and need different process chain to get the derived observed data. Some of the ontologies that describe sensors are Ontosensor (ontosensor, 2010) and SWEET ontology (SWEET, 2010). But, they are not compatible with each other and there may be several ontologies (domain, sensor and process) required to automate the processing task. Recently, the semantic sensor network incubator group was formed under W3C to develop general sensor ontology to describe sensors (W3C, 2010). This ontology can potentially be used to describe any sensor in the world.

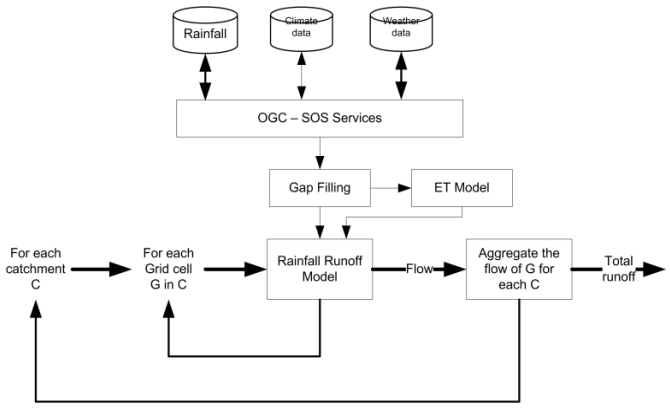

Figure 2. Conceptual workflow to simulate water flow in a catchment using gridded model

4 PROCESS INTEGRATION

There are very few systems that systematically construct and simulate hydrology systems and processes. Scientific workflow technology has gained popularity to describe, manage and share scientific experiments. Workflow represents tasks as components and data-flow between tasks. The composition of tasks provide meaningful outcome. Due to diverse services, resources, data-flow and process implementation in hydrology, scientific workflow will provide an environment to integrate heterogeneous resources in a common environment. To achieve this goal, the first step should be to develop a workbench for hydrologists based on scientific workflow. This would enable us to address the challenges of integration of data with models that represent hydrology phenomenon and process integration. Scientific workflows will be used to develop new tools based on the existing ones that would enable the integration of heterogeneous data resources with state-of-the-art hydrology models and visualisation tools. Workflows help scientists to share data and computation resources by using underlying *cyberinfrastructure*. Cyberinfrastructure is a term coined by the United States National Science Foundation (NSF): "*it consists of computing systems, data storage systems, advanced instruments and data repositories, visualisation environments, and people, all linked by high speed networks to make possible scholarly innovation and discoveries not otherwise possible*"(NSF, 2006). This helps scientists to concentrate more on discovering new scientific knowledge instead of worrying about accessing data, finding resources to run the experiments and transformation of different formats of data.

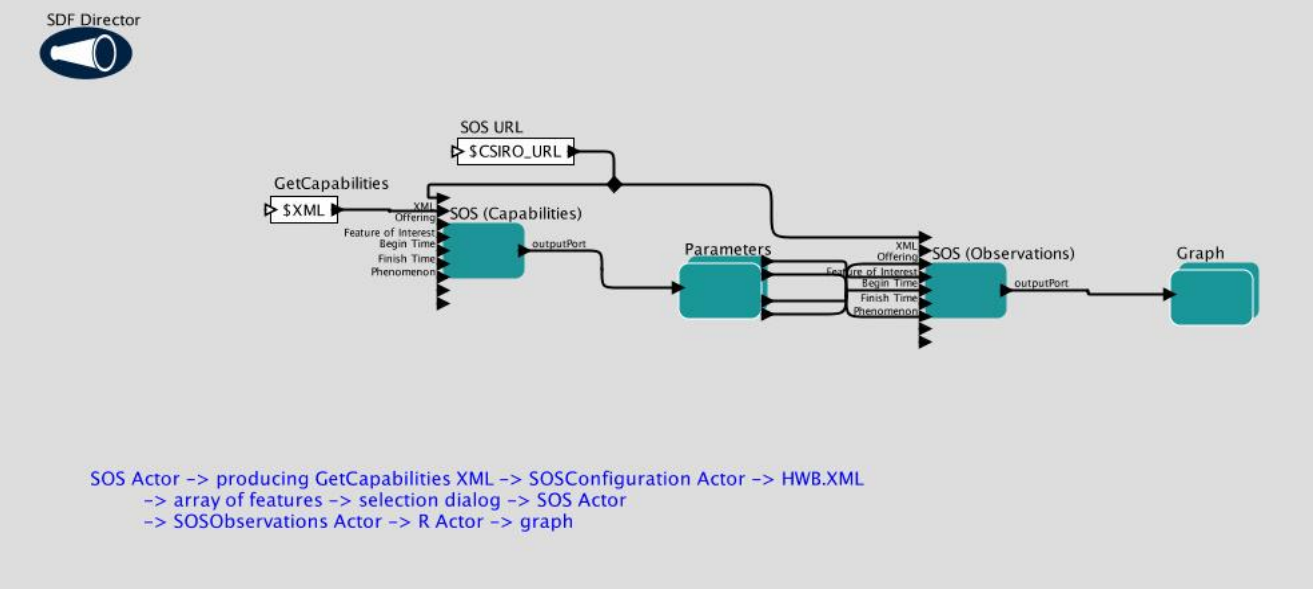

Figure 3. Sample workflow to publish data from the SOS in a graphical representation

Scientific workflows in hydrology should support a distributed infrastructure to enable hydrology research. It should enable the integration of data resources, processes and computation. Scientific workflows could leverage the services of grid or cloud computing infrastructure for data storage, computation and processing based on the needs.

To implement the river flow simulation explained in section 2.1, hydrologists need to gather required time-series data sets; restructure the data sets to the targeted rainfall-runoff model. For a Gridded rainfall-runoff model, the catchment needs to be divided into grid cells and the model is simulated for each grid cell. The aggregated runoff of the grid cells is the total runoff of a catchment. The conceptual level workflow to simulate water flow in a catchment is given in Figure 2.

To create a workflow of the above scenario, the hydrologist could consider each of the above steps as independent processes. Each process is modularised as components which perform certain tasks. The component implementation aspect is more often abstracted from the user (hydrologist). Running the above scenario consists of composing required components and configuring the parameters. This procedure does not require any programming skill but, needs domain knowledge to understand the requirements. The conceptual workflow of Figure 2 shows that sensor data is published through OGC-SOS and undergoes gap-filling. The most common input to rainfall-runoff model is historic rainfall data and EvapoTranspiration (ET). ET is the estimate of water that evaporates from soil and plant surfaces and the water plants lose through their leaves. ET is calculated using daily mean temperature, wind speed, relative humidity and solar radiation (Burba, 2006). All these inputs are provided by the OGC-SOS and fed into ET model after checking for gaps in data. Finally, the rainfall-runoff model is run through each grid of every catchment in a river.

There are several workflow tools and frameworks available. Most were initially developed for a particular domain but have since been extended to work in other domains. Workflow systems can be classified as task-based and service-based (Montagnat, 2006). In task-based systems, users specify the computing task to be executed with the location of the executable code, input data, dependencies and location of the output data that needs to be stored. The focus of a task-based workflow is to map and execute the workflow. Pegasus (Pegasus 2008), which uses DAGMan (Dagman, 2008) as a workflow engine uses this approach. In service-based environments, the application is wrapped around an interface. The workflow knows about the interface and accesses the application through it. The focus here is more towards the composition of a workflow. Taverna (Taverna, 2008), Triana (Triana, 2008), Kepler (Kepler, 2008) are examples of service-based workflow systems.

We have chosen Kepler due to its ability to use as a composition tool, support of web services, large collection of reusable component libraries called actors (Altintas, 2004). Most importantly, the model of computation is independent of the workflow composition (ludäscher, 2005a).

4.1 *Kepler workflow for Hydrology*

Kepler is a workflow environment that helps to create scientific workflows from different software programs. It is based on Ptolemy II (Brooks, 2008), which is a software infrastructure used for software design. Ptolemy II introduced the notion of domain polymorphism and modal models (Brooks, 2008a). Domain polymorphism is the design capability where the same component can be used in different domains and in modal models, finite state machines are combined hierarchically with other models of computation. Most computation models in Ptolemy II support actor-oriented design.

The actor-oriented design is often compared with object-oriented design. The difference lies in the way the components communicate. In actor-oriented design, data is passed from one actor to another through communication channels according to a message scheme. Actors interact with the channel instead of another actor. However, in object-oriented design, objects interact with another object by transferring control through function calls.

Directors, Actors and Parameters are the fundamental components of Kepler. Actors communicate with ports and parameters are used to configure the operations of an actor. Ports and relations help in the communication between the components. Components are represented in a GUI. Every workflow should have a director which orchestrates the workflow and acts as model of computation but, actors perform actual processing. Kepler uses Vergil, a visual editor to construct and execute the workflow.

There are five different types of directors (Brooks, 2008b):

Process Networks (PN) director models actors as network of processes that communicate with each other by passing message through unidirectional FIFO channel. Each actor waits for data and executes once it receives data. This director is similar to Unix pipes.

Synchronous Data Flow (SDF) Director is used for sequential workflow. The simple example can be gathering data from database and transforming data to a visual graph in a sequential order.

Discrete Event (DE) Director is used for workflows where events occur at a discrete time. Actors communicate when event occurs and an event is a data value with a timestamp.

Continuous Time (CT) Director is used in workflows that predict how system evolves as a function of time. Typically, this director is used to model systems with differential equations.

Dynamic Data flow (DDF) Director is a superset of SDF Director. Actors consume or produce data based on certain constraints. An example is a conditional if-else statement.

A workflow in Kepler is a composition of different actors. There are different actors to perform different tasks. Two or more actors can be combined to form a composite actor to perform complex operations. The actors are initialised through parameters. In a complex workflow, it is possible to have different Directors at different stages. Kepler has a large library of actors that can be reused for any application domain.

Ports in an actor are used to input and output data. An Actor can have a single port or multiple ports. Ports are categorised as input, output and input/output. Relations are used to branch the data flow and send the same data to multiple actors. Resource allocation is performed in Kepler by configuring the parameters of actors in a workflow. If the resources are web services, they are searched using a registry which will be updated regularly to include new services.

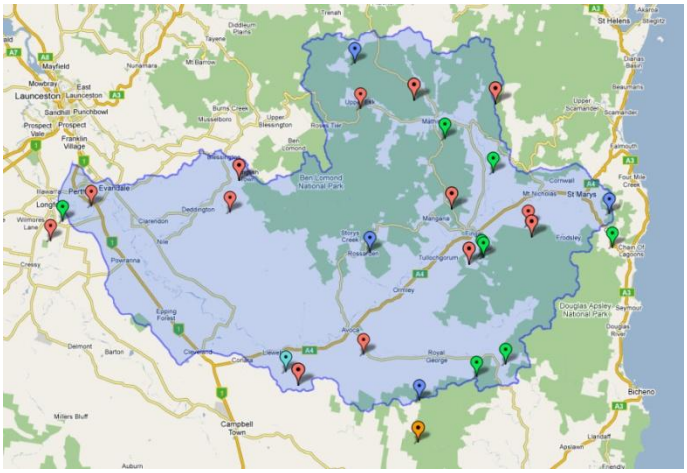

Figure 4. Map of South Esk river catchment. The pin represents sensors deployed from various organisations in a catchment.

Kepler uses its own run-time engine to execute a workflow. It supports grid and web service resources that are represented as actors in a workflow. Figure 3 represents Kepler workflow to publish data from the SOS. A user can select featureofinterest and phenomenon to display observed data. The display in Figure 3 is a composite actor with the combination of actor to read data in XML format and actor to display as graph.

5 EXPERIMENTAL TEST-BED

The experimental test-bed was setup to test some of the integration ideas. The test-bed is setup in the catchment of the South-Esk River in Tasmania. The catchment covers an area of approximately 3350 square kilometres. The rainfall in the catchment var-

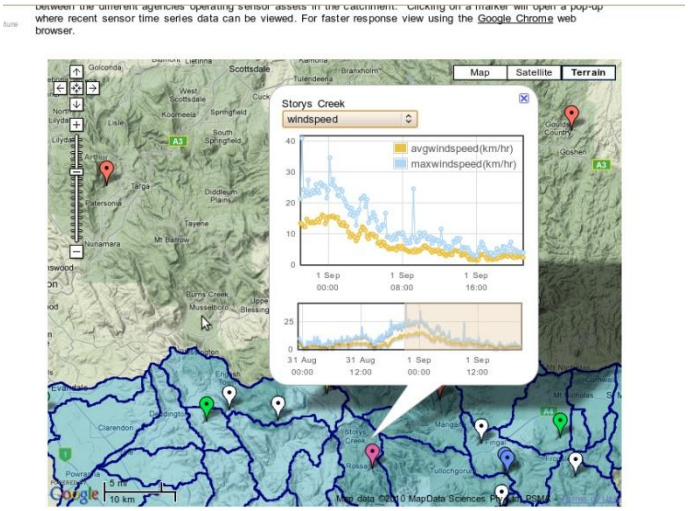

Figure 5. Measurement of wind speed phenomenon in a feature of interest, Stony Creek.

ies from 500 mm in the low lying areas and up to 1500 mm in the highlands. The total catchment yield at Longford is around 43% of the total water input. This means around 57% of water is evaporated, transpired or moved into the ground water system (DPIWE, 2007).

The understanding of the groundwater movement in the South Esk catchment is minimal. From the business perspective, currently, the water allocation is solely based on surface water information. However, it is a common knowledge in the catchment that the river flow is predominantly groundwater-driven. Therefore, there is a need to understand the surface and groundwater connectivity in this catchment. With this knowledge, there is an opportunity to improve water planning and management through continuous monitoring and forecasting of river flow.

The water accounting in a catchment cannot be possible without understanding the landscape of different sub-catchments, ground water retention and the spatial-temporal pattern of different measured phenomenon in a catchment. Until the end of 2009 the CSIRO Tasmanian ICT Centre, as part of CSIRO Water for Healthy Country Flagship project initiative has deployed 11 rain gauges, 4 Automatic Weather Stations (AWS). These sensing instruments became part of existing 5 water level (to measure stream flow) sensors and 4 rain gauges (to measure rainfall) of Tasmanian Department of Primary Industries, Parks, Water and Environment (DPIWE), 7 water level sensors from Bureau of Meteorology (BoM), 1 water level sensor from Hydro Tasmania and 1 AWS from Powercom. Figure 4 shows the map of South Esk river catchment with deployed sensors shown as button.

The catchment is now a sensor-rich environment with an opportunity to study the hydrology phe-

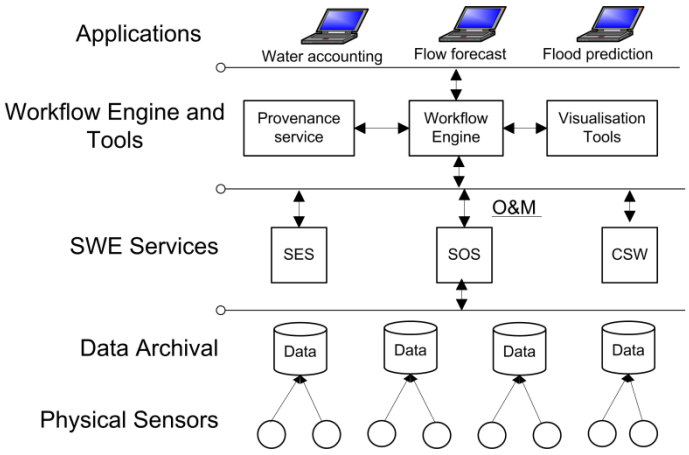

Figure 6. High level architecture of the proposed hydrometeorological sensor web.

nomenon in a catchment. The test bed system attempts to incorporate sensors, models and data from different organisations all operating within the same catchment. The hydrometeorological sensor web will explore how environmental sensors, hydrological models and decision support tools can be combined to determine the continuous flow forecasting. The sensor web would enable different stakeholders like the DPIWE, BoM and Hydro Tasmania to share their data among each other and rest of the world using common interfaces.

Currently, sensor data is published via a SOS and can be viewed on Google Maps. A SOS has three important operations: GetCapabilities, GetObservations and DescribeSensor. A user consuming data can obtain sensor observations from a sensor-centric and observation-centric approach. In a sensor-centric approach, users already know about the sensor existence and invoke GetCapabilities operation to get the capabilities of the SOS. The capabilities documents consist of list of offerings, procedures of the offering (all the sensors or sensor systems measuring a phenomenon), name of the observed properties (Phenomenon measuring) and feature of interest (it is a domain (site) where the observations are measured). The Sensor description which is stored in a SensorML can be extracted by querying DescribeSensor operation using sensor system identifier (procedure id). GetObservation gives the observation of a sensor in an O&M encoded XML format. The user has to know the offering id to perform any GetObservation operations. The offering id can be extracted from the GetCapabilities document.

The GetObservation can extract results based on the combinations of featureofinterest, phenomenon, spatial, temporal and scalar capabilities. This gives an opportunity for users to query SOS for a particular phenomenon in a given spatial region for a certain time period. These types of queries can also be called observation-centric. Currently, CSW are not

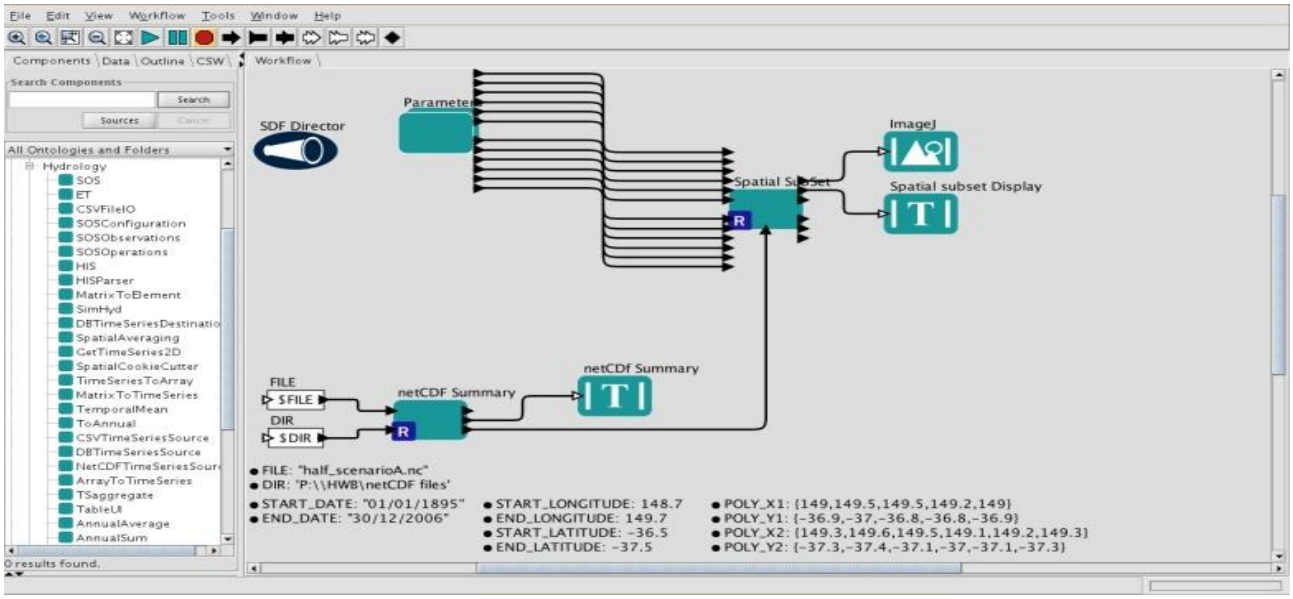

Figure 7. Hydrology workbench with the workflow at the right pane and hydrology related actors in the left pane.

used to catalogue all the services. However, in the future all the services will be catalogued so that users can discover services that suits their requirements and query the SOS instances (Jirka, 2009).
The GetObservation operation provides an output in an XML format based on O&M encoding. This is not user friendly and difficult to visualise. Using Google Map based user interface to the SOS GetObservation operations, the SOS instances are visible on the map; the user can click on any instance to select the phenomena to see the observed measurement as a graph in a sliding window of 6 days. The website is made public to view data from different stakeholders in the South Esk river catchment[1]. Figure 5 shows an observation of wind speed phenomenon in a Stony Creek site.

The high-level architecture of the hydrological sensor web is shown in figure 6. It is a Multi-level architecture where the end-users are completely abstracted from sensor details.

In the test-bed, sensors collect observed data every 15 minutes and telemeter to Hobart once every 24 hours. The data is stored into PostGIS-enabled PostgreSQL database. When users perform GetObservation operation of a SOS, the observation data from the database is dynamically encoded into O&M schema and published. Since all the services in OGC-SWE are loosely coupled, workflow is used for service choreography (Gil, 2007). The workflow tool will give an integrated support to extract data and metadata from SOS, transform data for further manipulation, process data with third-party models and visualise results. We are using Kepler scientific workflow tool as a workbench environment. A user will compose the workflow and execute them using Kepler workbench. The workflow can also be stored for later use.

Intermediate data may be useful for different workflows. For example, some of the hydrology data collected from a field may have several deficiencies such as large gap, unusual spikes. Data cleaning step will always be performed before it is used for analysis. The cleaned data set can be reused as long as it is tagged with metadata and complete provenance information is available. It is a challenging task for workflow system to determine what intermediate derived data to store when the storage space and data transfer rate are constraints. Provenance management is an important issue to efficiently reuse data, processes and workflows.

Provenance by definition is a record of the history of ownership. From scientific workflow perspective, provenance is classified into data and process (Simhan, 2005). Data provenance shows an evolution of created data. It consists of data, processes used, date of creation and intermediate steps of creation. Process provenance provides the origin of derived processes. Provenance helps to make a judgement about the derived data quality, validity and also to reuse the data and processes with confidence. Provenance framework is supported both by Kepler (Altintas, 2006) and Pegasus (Kim, 2007).

6 KEPLER IMPLEMENTATION

Kepler gives a unique advantage of re-usable actors in a workflow. Kepler workflow was developed primarily for ecology domain therefore; several actors can be reused for hydrology domain. Actors can represent any process independent of their im-

[1] http://www.csiro.au/sensorweb/au.csiro.OgcThinClient/OgcThinClient.html.

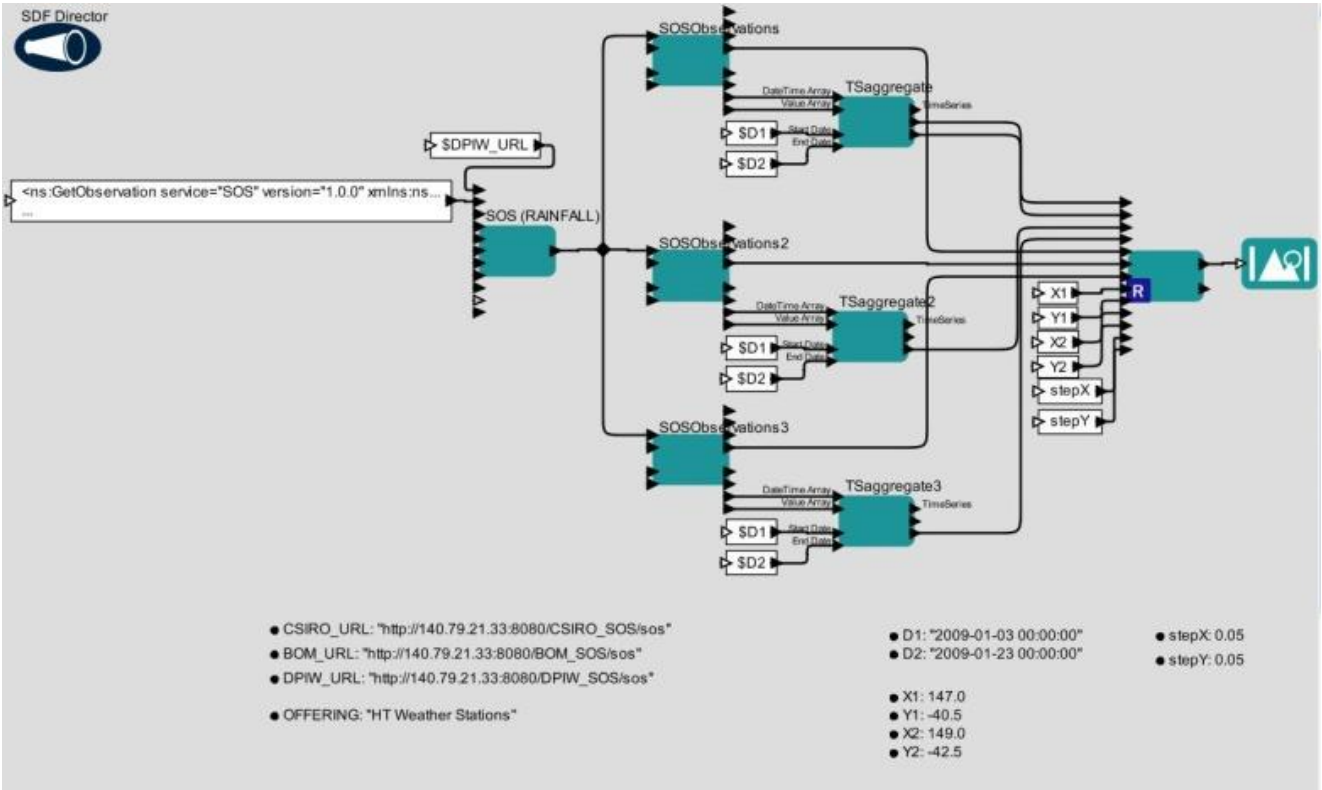

Figure 8. Kepler workflow to produce gridded rainfall dataset.

plementation programs. Therefore, actors can be developed to expose Matlab expressions, R scripts, Python/Jython scripts and Java implementation. The .NET implementation can also be exposed using Java-.NET interpreter.

Several actors are developed to incorporate functionalities required in hydrology science. Some of the important actors developed based on functionality are:

- SOS actor which is used to invoke OGC-sensor observation services.
- ET actor is used to compute Evapo-Transpiration (ET) based on Penman-Monteith equation.
- Data manipulation actor to gap-fill time-series data using interpolation techniques.
- Actors to resize data based on space and time.
- Actors to access data from specialised file format like NetCDF.
- Actors to access rainfall-runoff models from TIME Toolkit[2] which is a .NET implementation.

Figure 7 shows a Kepler workbench with the available actors on the left-pane. Even though any actor can be used to create a workflow, each actor is associated with a domain to make it easy for searching purpose. An actor can associate itself with two or more domains. All the actors related to hydrology are implemented and stored under the hydrology domain in the left pane. Composition of a workflow is achieved by dragging and dropping actors from the left to right pane, interconnecting actors, setting actors parameters and selecting a director based on the model of computation. In Figure 7, SDF Director (blue icon on the top left-hand corner of the right pane) is used to orchestrate a workflow in the right pane. The workflow will create a subset of dataset based on space.

To create an executable Kepler workflow for a conceptual workflow given in Figure 2, several processing steps need to be performed. One of the tasks is to create a gridded dataset from the time-series data. The workflow used to generate the gridded rainfall dataset from the historical rainfall data is given in Figure 8. The raw rainfall time-series data is exposed via SOS between the temporal window given by parameter D1 and D2. The "TSaggregation" actor is used to produce a time-series dataset with uniform temporal intervals between the range D1 and D2. The R actor executes a script to predict a grid over the catchment covering the area specified by the coordinates (latitude/longitude) of the parameters (X1, Y1, X2, Y2). The grid size is specified by the parameters stepX and stepY in decimal degrees and kriging is used to interpolate rainfall data across the grid cells. The interpolated data is output into a suitable file format for ingestion into a rainfall-runoff model.

[2] http://toolkit.ewater.com.au/time

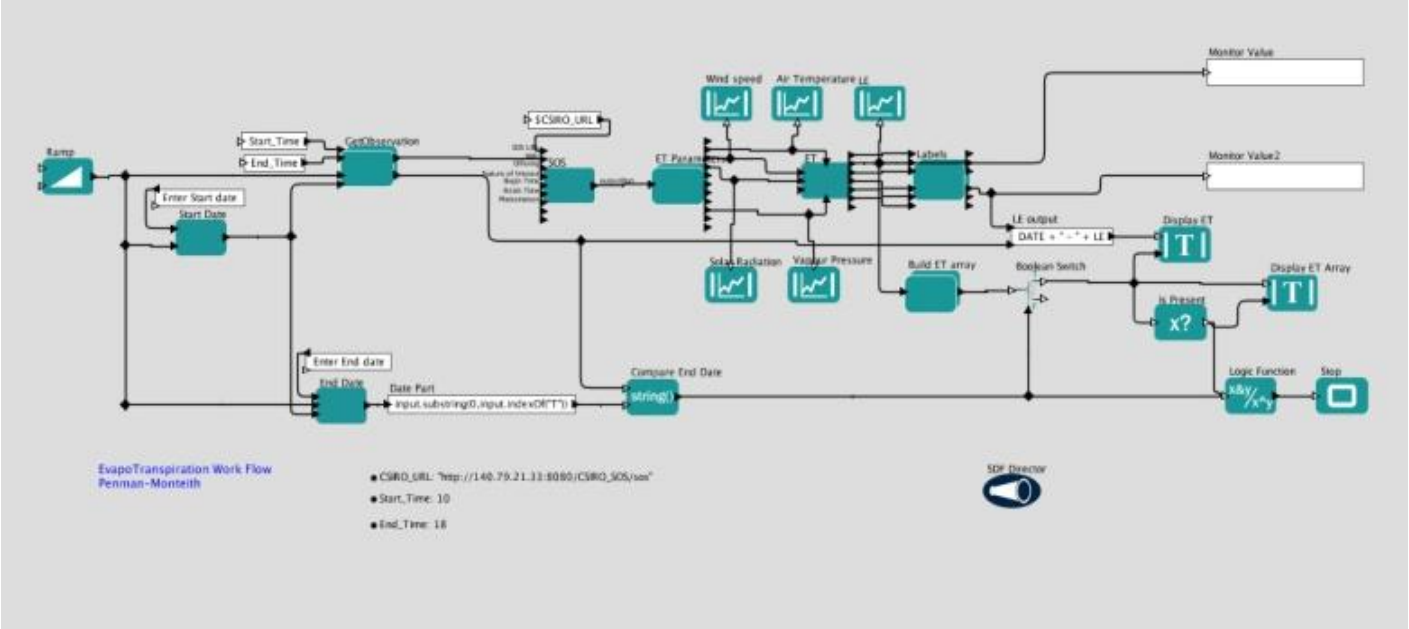

Figure 9. Kepler workflow to generate daily ET dataset for a given time interval.

The workflow in Figure 9 generates a daily ET dataset for a specified period. The Penman-Monteith equation is used to compute daily ET with the input data available via a SOS. Java programming language is used for the implementation of ET actor and Jython scripts are used to make the workflow interactive. For example, a user can dynamically select date interval to compute ET with graph showing the progress of the computation.

The workflow in Figure 10 simulates river runoff in the South Esk using the SimHyd Rainfall-Runoff model. The SimHyd model is implemented in the .NET environment and is part of the TIME toolkit. Third-party Java-.NET bridge program is used to pass parameters between Java and .NET environment. The input to the model is the daily rainfall and ET. The calculation of ET is part of the workflow and the rainfall data is captured from a SOS instance. The aggregation and gap filling is performed by a "TSaggregate" actor and the "RExpression" actor consists of scripts to display the output in a graphical and textual format. This workflow integrates actors with different implementation programs and environments under a common Kepler workbench.

7 CONCLUSION AND FUTURE WORK

We have provided a different perspective to the challenges facing the application domain of sensor networks. As more sensors are deployed in an environment, new challenges like data management, seamless data integration and managing processes that manipulate data will become paramount. Building and adhering to standards and exploiting advanced computing technology can address several challenges. But, building a consensus to develop standards is a difficult task. The cyberinfrastructure projects funded by US-NSF has played key role in promoting advance computing technologies to resource and data intensive tasks in a scientific applications.

This paper discussed the adaption of OGC-SWE framework to store, query and publish hydrometeorological data. The adaptation is motivated by the benefits that can be obtained by global sharing of datasets with minimal difficulty. We also proposed the application of the Kepler scientific workflow to orchestrate loosely coupled OGC-SWE services in order to achieve common objective. A high-level architecture that uses the scientific workflows within a SWE framework was proposed. This architecture reduces the difficulties in managing heterogeneous data and processes in order to derive the required knowledge. The usefulness of standards and scientific workflow is demonstrated in the reusable example workflows.

Currently, we are using SOS as the data publishing service. In future, we intend to use the complete suite of OGC-SWE services to discover and manage sensors across the catchment. We also intend to run Kepler as Web processing Service (WPS) so that users can create and expose workflow as a service instead of as a stand-alone application.

8 ACKNOWLEDGEMENT

Part of this work was conducted when Siddeswara was a Post-Doctoral Fellow in Tasmanian ICT Centre.
This project is jointly funded by the CSIRO Water for a Healthy Country Flagship and the Tasmanian

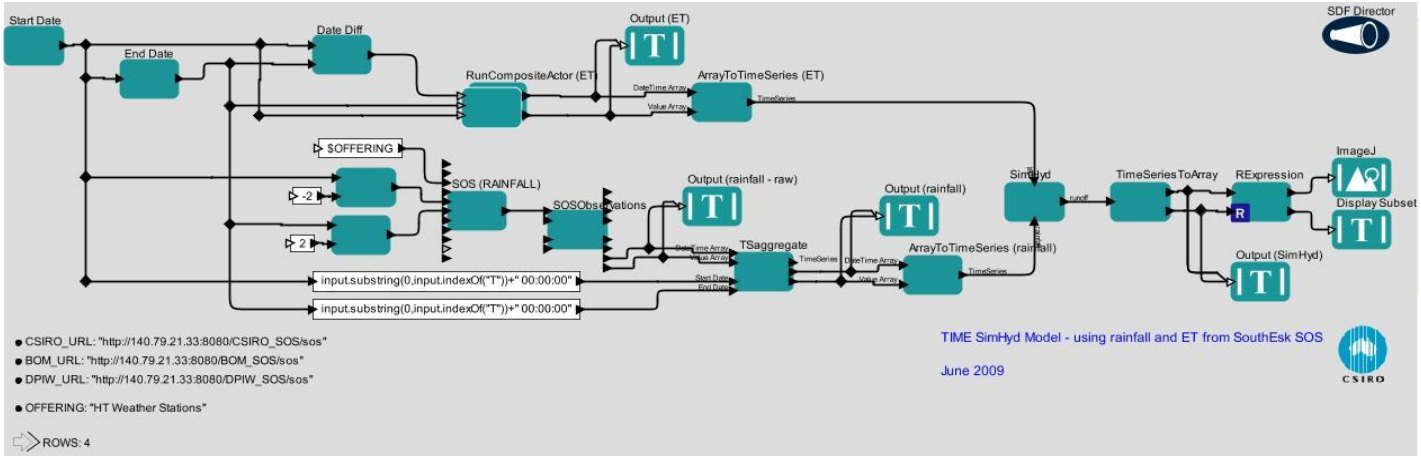

Figure 10. Kepler workflow to forecast river flow using SymHyd rainfall-runoff model.

Government. The CSIRO Tasmanian ICT Centre is jointly funded by the Australian Government through the Intelligent Island Program and Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO). The Intelligent Island Program is administered by the Tasmanian Department of Economic Development and Tourism.

Authors would like to thanks Hydrology workbench and South-Esk project team members for suggestions and discussions.

REFERENCE

Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B. and Mock, S., "Kepler: an extensible system for design and execution of scientific workflows", Proc. 16th International Conference on Scientific and Statistical Database Management, 2004.

Altintas, I., O. Barney, E. Jaeger, "Provenance Collection Support in the Kepler Scientific Workflow System". International Provenance and Annotation Workshop, 2006, pp. 118-132.

Balazinska, M., Despande, A ., Franklin, M. J., Gibbons, P. B., Gray, J., Nath. S., Hansen, M., Liebhold, M., Szalay, A., and Tao, V. "Data Management in the worldwide Sensor Web", IEEE Pervasive Computing, Vol. 6, No. 2, 2007, pp. 30-40.

BOM, Bureau of Meteorology- Water regulation 2008 Commenced 30th June 2008, viewed 6th March 2009, <http://www.bom.gov.au/water/regulations/water-2008.shtml>.

Brooks, C., E. A. Lee, X. Liu, S. Neuendorffer, Y. Zhao, and H. Zheng, "Heterogeneous Concurrent Modelling and Design in Java (Volume 2: Ptolemy II Software Architecture)", EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2008-29, 2008.

Brooks, C., E. A. Lee, X. Liu, S. Neuendorffer, Y. Zhao, and H. Zheng, "Heterogeneous Concurrent Modeling and Design in Java (Volume 1: Introduction of ptolemy)", EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2008-29, 2008a.

Brooks, C., E. A. Lee, X. Liu, S. Neuendorffer, Y. Zhao, and H. Zheng, "Heterogeneous Concurrent Modeling and Design in Java (Volume 3: Ptolemy II domains)", EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2008-29, 2008b.

Burba, G., Hubbart, J., Pidwirny, M., Weis, J., "Evapotranspiration." In: Encyclopedia of Earth. Eds. Cutler J. Cleveland (Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment). [First published in the Encyclopedia of Earth October 19, 2006; Last revised August 3, 2010; Retrieved August 31, 2010]http://www.eoearth.org/article/Evapotranspiration

CSIRO, Water for Healthy Country Flagship – Sustainable Yields Project, viewed 12 January, 2009, http://www.csiro.au/partnerships/MDBSY.html.

Dagman, DAGMan, viewed 15th November, 2009, <http://www.cs.wisc.edu/condor/dagman>.

DPIWE, "Surface Water Hydrology of the South Esk River Catchment", Technical Report, Department of primary Industries Water and Environment, WA 07/02, 2007.

Fitch, P., J. Perraud, and A. Dijik, "Technological Integration for Water Resources Assessment", CSIRO water for healthy country flagship, Canberra, 2008.

Gil, Y., Deelman, E., Ellisman, M., Fahringer, T., Fox, G., Gannon, D., Goble, C., Livny, M., Moreau, L., Myers, J., "Examining the challenges of scientific workflows", Computer, vol. 40, no. 12, 2007, pp. 24-32.

Guru, S. M, Kearney, M., Fitch, P., and Peters, C., "Challenges in Using Scientific Workflow Tools in the Hydrology Domain", 18th World IMACS MODSIM Congress, 2009, pp. 3514-3520.

Jirka, S., Broring, A., Stasch, C., "Discovery mechanisms for the sensor web", Sensor, 2009, pp. 2661-2681.

Kansal, A., Nath, S., Liu, J., and Zhao, F., "SensorWeb: An infrastructure for shared sensing", IEEE Multimedia, vol. 14, no. 4, 2007, pp. 8-13.

Kim, J., E. Deelman, E., Y. Gil, G. Mehta, V. Ratnakar, "Provenance Trails in the Wings/Pegasus System", Journal of Concurrency and Computation: Practice and Experience, 2007, pp. 587-597.

Lettenmaier, D. P., Macroscale Hydrology: Challenges and Opportunities. In Matsuno, T and Kida, H. (eds.), Present and Future of Modeling Global Environmental Change: Toward Integrated Modeling, Terrapub, 2001, pp. 111-136.

Ludäscher, B. and Lin, Kai. and Bowers, S. and Jaeger-Frank, E. and Brodaric, B. and Baru, C., "Managing Scientific Data: From Data Integration to Scientific Workflows", GSA Today, Special issue on Geoinformatics, 2005.

Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, I., Jones, M., Lee. E., Tao, J., and Zhao, Y., "Scientific workflow management and the kepler system", Concurrency and

Computation: Practice and Experience, vol. 18, No. 10, 2005a, pp. 1039-1065.

Montagnat, J. and T. Glatard, and D. Lingrand, "Data Composition Patterns in Service-based Workflow", Workshop on Workflows in Support of Large-Scale Science (WORKS' 06), Paris, France, 2006.

Mulligan, M., Modelling Catchment Hydrology. In Wain-Wright, J and Mulligan, M. (eds.), Environmental Modelling Finding Simplicity in Complexity, John Wiley & sons, 2005.

NASA SWEET, http://sweet.jpl.nasa.gov/ontology/, 2010.

NSF, "NSF's Cyber infrastructure vision for 21st Century Discovery", NSF Cyberinfrastructure Council, 2006

OpenGIS, "OGC Sensor Web Enablement: Overview and High Level Architecture", 2007, pp. 1-14.

OM, "Observation and Measurement Specification, document 07-022r1", <http://www.opengeospatial.org/standards/om>, 2006.

Ontosensor, http://marinemetadata.org/references/ontosensor, 2010.

Pegasus, Pegasus Workflow Management, Viewed 15th November, 2009, <http://pegasus.isi.edu>.

SAS, "Request for comment on Sensor Alert Service", http://www.opengeospatial.org/standards/requests/44, 2010.

SensorML, "OpenGIS Sensor Model Language (SensorML)", http://www.opengeospatial.org/standards/sensorml, 2010.

Sheth, A, Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics In Goodchild, M. and Egenhofer, M. and Fegeas, R. and Kottman, C. (eds), Interoperability Geographic Information Systems, Kluwer, 1998, pp. 5-30.

Simhan, Y. L. and B. Plale, and D. Gannon, " A Survey of Data Provenance in e-Science". SIGMOD Record, vol. 34, no. 3, 2005, pp 31-36.

SOS, "Sensor Observation Service Specifications", http://www.opengeospatial.org/standards/sos, 2010.

SPS, "Sensor Planning Service Specification", http://www.opengeospatial.org/standards/sps, 2010.

Taverna, Traverna Project Website, viewed 15th November, 2009, http://taverna.sourceforge.net

Triana, Triana – Open Source Problem Solving Software, viewed 15th November, 2009, <http://www.trianacode.org>.

W3C Semantic Sensor network Incubator Group, http://www.w3.org/2005/Incubator/ssn/, 2010.

WNS, "Web Notification Service", OpenGIS Project Document: OGC 03-008r2 Version 0.1.0, 2003.