Migrating geoprocessing routines to web services for water resource management applications

Laura Díaz[1], Sergio Costa[2], Carlos Granell[1], Michael Gould[1]

[1] Centre for Interactive Visualization (CeVI)
Department of Information Systems
Universitat Jaume I, Castellón, Spain
{laura.diaz, carlos.granell, gould}@uji.es
[2] Divisão de Processamento de Imagens (DPI)
Instituto Nacional de Pesquisas Espaciais (INPE)
São José dos Campos, Brazil
scosta@dpi.inpe.br

SUMMARY

Spatial data infrastructures are beginning to support the most common requirements of spatial information users---discovery, access, and visual overlay of datasets-- however specialized users such as hydrological scientists require more advanced services for capturing, analyzing and processing huge volumes of data. To help migrate these scientists from desktop GIS to on-line service-data offers, we describe an approach to model and develop distributed geoprocessing services for water resource management applications built on top of basic SDI functionality.. We illustrate a GMES-funded scenario for predicting runoff in Alpine basins employing a framework which may serve to easily extrapolate our experiences to other specialized application fields like environmental and climate change applications.

KEYWORDS: *Geoprocessing services, environmental applications, WPS, e-Science*

INTRODUCTION

The Global Monitoring for Environment and Security (GMES) initiative was created by the European Commission (EC) and European Space Agency (ESA) to foster cooperation among data providers and users to make environmental and security-related information more readily available and useful. Within this remit is a requirement for harmonization of spatial data having different specifications, standards, and formats: data observed in situ, Earth Observation (EO) as well as traditional cartographic datasets that are already mandated (INSPIRE) to be distributed in Spatial Data Infrastructures (SDI). The workflows of critical environmental applications such as in hydrology (flood mitigation, surface water management, etc.) commonly require the fusion of these data sources: satellite imagery (e.g. snow or forest coverage area), data in situ from sensor stations (e.g. wind and temperature) and overlays with vector datasets (e.g. cadastre, transport network, etc.). Recent advances in the fields of web services, especially as applied to geoinformatics, can facilitate this access and data fusion workflow, including the wrapping of scientific algorithms for on-line parameterization and execution.

The field of hydroinformatics is rapidly gaining momentum due to both the increasing stress on global water resources (Cheng et al. 2006; Hughes and Forsyth 2006, Olivera et al. 2006, Soh et al. 2006) and the increasing offer of informatics tools accessible over Internet. This field integrates scientists' knowledge and understanding of water resources with applications and technology to improve decision-making in critical contexts such as runoff prediction in major drainage basins. These modelling, monitoring and prediction applications have special requirements different from

those of common geographic information users, and require advanced tools and information infrastructures for accessing huge volumes of data, and for its later storage, processing, analysis and visualization. In this new multidisciplinary process water scientists provide their domain knowledge and expertise, GIS researchers and programmers play an active role in facilitating access to geoprocessing services, models and components to better connect water scientists with their data, models, and services in an interoperable manner. This paper focuses on geoprocessing modelling services, components and on some of the key technical problems derived from the integration of these within current spatial data infrastructures: in a sense the intersection of SDI with e-Science (Ramamurthy 2006).

We present a web-based approach for supporting water resource management for predicting runoff for gauged drainage basins derived from the observed snowmelt data by using as inputs data from both in situ observations and EO. Compared to local-processing or desktop applications our approach is based on a distributed model composed of independent and specialized geoprocessing services that each take on, or assist, a part of the geoprocessing workflow. After introducing the basic concepts and existing relationships among spatial data infrastructures, web services and geoprocessing services, we present a case study for forecasting runoff in selected Alpine basins. This case study illustrates how distributed geoprocessing services can be included as helper applications in the hydrologist's workflow while calibrating and executing runoff models.

Spatial Data Infrastructures and OWS services

Connecting scientists in the GMES sector to their data, models, and services is an important challenge for SDI, which is beginning to facilitate access to distributed, heterogeneous geospatial data through a set of policies, common rules, and standards that together help improve interoperability. Traditional discovery and visualization-based SDI is evolving to a more service-based vision in which geoservices are used not only to access geospatial data, but also to transform them and process them, often in service chains. Many geoservices employ interfaces defined by the Open Geospatial Consortium (OGC, http://www.opengeospatial.org/) during the OWS Web Services specifications initiatives (http://www.opengeospatial.org/projects/initiatives/ows-4), such as those described by Anderson and Moreno-Sanchez 2003. Caldeweyher et al. 2006 have applied the basic services (WMS, WFS) yet these are shown to be insufficient to suit the processing and modelling requirements of water scientists. The ongoing OGC Web Processing Service (WPS) specification (Schut and Whiteside 2005) provides interfaces for accessing more complex services and also for wrapping existing off-line services as web services.

In this way a new concept of geoservice is born: web services that go beyond the traditional OGC realm and which may be defined generically as by W3C, as "software application identified by a URI, which interface and links can be described and discovered as XML artefacts" (http://www.w3.org/2002/ws), not only to discover or present spatial information but to provide geoprocessing functionality, and which are also independent of data and context, thereby providing high reusability. OGC web processing services provide three methods to provide the functionality of a certain geoprocessing services by first using the *getCapabilities* method, common in other OWS services, in order to know the methods offered. The WPS define input and output parameters in a very detailed way by providing a *describe process* method. Finally, the *execute* method actually invokes the geoprocessing service. In our opinion, OGC WPS go beyond providing unique geoprocessing routines because it can be an interesting specification with which to wrap both spatial and non-spatial processing services, leading to increased interoperability between OWS services and general purpose web services.

CASE STUDY –APPLICATION FOR PREDICTING RUNOFF IN ALPINE CATCHMENTS

The following case study is being carried out in the framework of the AWARE Project (A tool for monitoring and forecasting Available WAter REsource in mountain environments, http://www.aware-eu.info) in order to offer EO-based geoprocessing services and tools to monitor and forecast the water resources derived from a specific quantity and distribution of Alpine snowmelt.

River discharge is important to monitor because of its significant influence on environmental systems and potentially on human lives in the case of flood and landslide management. The first task for supporting these monitoring applications is to investigate and identify the people who would use the software and the sequence and nature of tasks. AWARE is developing a user friendly application for running runoff forecast models, permitting not only expert users (hydrologists and other scientists) to run models but also end users (non-experts users) who often are the providers of local data necessary for the models. This implies usability requirements in our system design leading to a flexible, easy-to-use prototype that features intuitive interfaces and wizards assisting non-experts with the complex tasks of runoff forecast modelling and interpretation of the results. Expert users often are more comfortable directly handling and analyzing data, and feel that the meaning and quality of data provided should be accurately investigated by them.
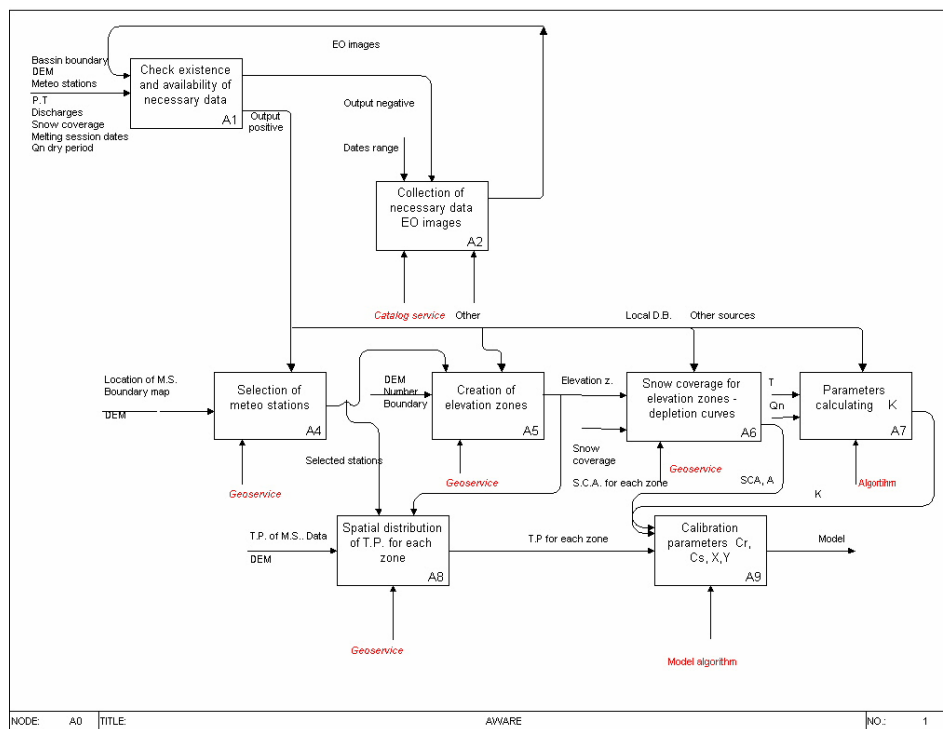

Figure 1: Data flow of the task of SRM model.

Figure 1 illustrates the different tasks involved in the Snowmelt-Runoff Model (SRM) (Martinec et al. 1983), one of the runoff forecast models used in the AWARE project. Each box in the figure represents a model task implemented by several geoprocessing services as described in the next section. It is out of the scope of this paper to describe each one of these steps however it is important to determine and analyze the data types necessary for the model. As box A1 in figure 1 shows, many

datasets provide the expert user a more complete view and context of the hydrologic process: stream gauge datasets; temperature and precipitation data (collected from weather stations at different locations); the location of all streams, weather stations, the geographic boundaries of the watershed; the Digital Elevation Model (DEM) of the basin; and satellite imagery representing the snow coverage area of interest. Analyzing these disparate datasets together is a tedious task and hence is often performed inefficiently resulting in incomplete and inaccurate results. Furthermore our prototype should facilitate the data collection and integration the many types and characteristics of watersheds across multiple data sources.

Many expert users own the local data necessary to run the model and so the first choice is to allow users to feed the model with the local data they possess. Still, a goal of the AWARE project is to be compliant with the INSPIRE initiative so that other non-expert users might also discover and access geodata via SDI services. Any user might be interested in searching catalogues for appropriate satellite imagery for the study basin (geographic constraint) and during the snowmelt station (temporal constraint).

In summary, the design of a web-based tool for runoff forecast models should take into account usability, utility and flexibility. Our tools are being developed according to these criteria as described in the following section.

DESIGN AND IMPLEMENTATION

The AWARE geoservices architecture follows a multi-tier model that integrates a geoportal (or entry point) as the main user interface, session management and user access control, and distributed geoprocessing services grouped into modules. The application allows expert users to perform geoprocessing functions in a distributed way, many of the processes being previously wrapped as web processing services on top of the datasets and resources provided by a SDI.

Figure 2 shows the multi-tier architecture. Typically users are validated and authenticated by accessing the geoportal in the presentation tier (top of figure 2). Distributed geoprocessing services form the middleware tier (centre figure 2) whereas geographic data and other resources are stored in the data tier (bottom of figure 2). Components in each conceptual tier interact with others component in the following way:

The **presentation tier** provides the geoportal user interface, which permits expert users to select and perform a water resource model for a concrete watershed of study. The main component is the *Data flow handler* that guides expert users through such models by providing user-friendly interfaces both to access data –either from catalogues or provided by users–, feed such data to models, and execute them, and also for common web–based functionalities such as user authentication and session tracking. The latter refers to the possibility to permanently store the current user session so that the session can later be continued. This feature is necessary because some tasks of these models may last from seconds to hours or days, thus it is necessary to monitor and guard the user session state. The *Data flow handler* also captures the data needed for a given task, so that can be invoked by the corresponding geoprocessing service through the *WPS client handler*, which transforms user requests from the presentation tier into an OGC WPS request addressed to a geoprocessing service. Once results come back to the *WPS client handler*, it transforms this response to be properly delivered to the presentation tier (users).

We also find an open source mapping client (map viewer) embedded in the user interface in the presentation tier, which integrates common visualization functions (pan, zoom, etc.) through a simple toolbar. We prefer that the middleware tier uniquely serve geoprocessing functions separately from the common visualisation capabilities (Lu 2006). Recent advances in web technologies permit the creation of powerful client-side applications, for example using AJAX technology as does Google Maps (http://www.google.com/apis/maps/documentation/), to provide an interactive user interface for visualization functions, all computed on the client side rather than in the server (Sayar et al. 2006).

The **middleware tier** consists of several components that encapsulate all required tasks to access, analyze, and transform data into useful information and then to forward the results to the presentation tier. This tier groups all of geoprocessing services necessary to perform both basic, general geoprocessing routines and those spatial and non-spatial particular to our case study requirements.

The **data tier** manages, stores, and controls data although our interest here is not data management but discovery and access of data needed to run our geoprocessing services just like any SDI's user.
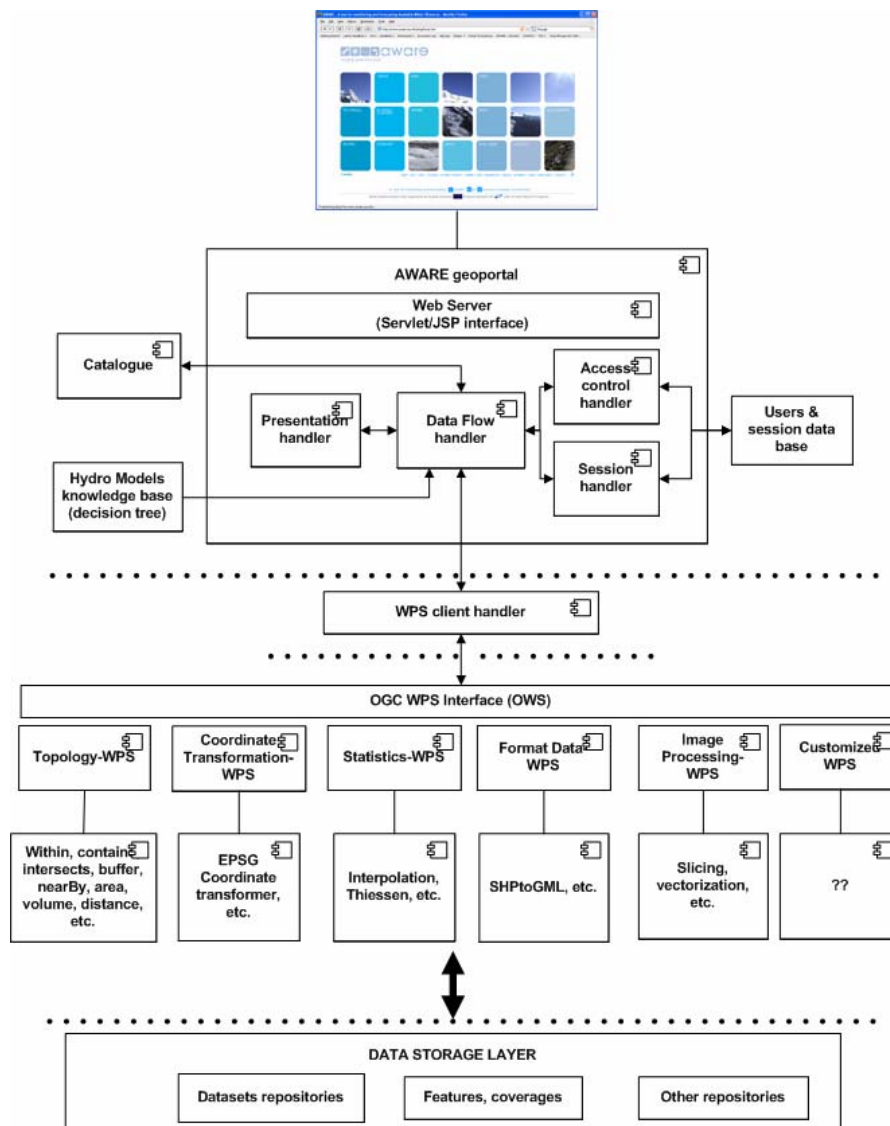

Figure 2: Multi-tier architecture for the AWARE Information System.

Service design principles

In order to provide useful geoprocessing services that suit the concrete requirements of the water resources models, we have identified basic functions shared among the analysis tasks. The ultimate goal is to create a geoprocessing service library in which customized and elaborated functions relies on other much more simple, atomic and well-tested functions. In this way, the reuse of geoprocessing services is fostered because the process of creating new customized geoprocessing services is possible by mainly reusing already available geoprocessing services from shared libraries (Granell et al. 2005).

For this reason, our service design strategy consists of identifying the atomic functions required for the use case. Initially we studied the model requirements and then made a top-bottom analysis in order to find the suitable set of geoprocessing services. In our context, we consider a suitable geoprocessing service one which performs a basic function, can be easily tested and is domain-independent enough to be applicable to other contexts. Once identified, the basic functionalities are grouped into modules with similar functionality. Transforming the design into executable WPS is straightforward: each module is a geoprocessing service (WPS) whereas each basic function is implemented as a process served by the WPS *DescribeProcess* interface.

Basic geoprocessing services

Figure 2 shows the basic WPS modules identified up to now, such as *Topology-WPS*, *Image Processing-WPS, Coordinate Transformation–WPS, Format Data-WPS* and *Statistics-WPS*. The *Topology-WPS* module deals with topological relations: properties like connectivity, adjacency and intersection among geospatial objects, which are important for developing algorithms for computing geospatial proximity and association commonly used on hydrological data. This WPS module has been developed for modelling spatial properties such as *intersect with, within, cross, contains,* etc. Methods for calculating geospatial proximity or geographical distances among geospatial objects and spatial operations like buffer and volume are also included in this module, all of them well-documented in (Ryden 2005). This module exposes clearly what is considered as a basic geoprocessing service: an atomic, general purpose function that is easily implemented, tested, and context independent.

The *Image Processing-WPS* contains common raster algorithms like the slicing process in which each cell of the raster image is classified into categories according to the cell value. The module *Coordinate Transformation-WPS*, implemented by the AWARE partner ICC (Institut Cartogràfic de Catalunya) transforms coordinates among different coordinate reference systems, whereas the *Format Data-WPS* converts files in shape format into GML. Most geoprocessing services implemented in our geoportal require GML format for inputs and outputs, because GML is a standard, well-structured specification of widespread use for vector data. Similarly, the *Statistic-WPS* groups similar functions for interpolation computations or implemented well-known algorithms like the Thiessen interpolation technique.

Customized geoprocessing services

So far we have described geoprocessing services capable of invoking basic functions that are not closely related to tasks involved in water resources models as explained in our case study (see Case Study section). In fact, previous geoprocessing services form the background in which more elaborated, customized, independent-domain geoprocessing services can be built up in order to create a distributed geoprocessing service library. This section describes how a customized geoprocessing services is modelled by combining and reusing other geoprocessing services already created.

As part of the earlier use case, we consider a specific processing task that calculates the percentage of snow coverage area in an initial zone in relation to a larger region (basin), using satellite imagery. This task is used for the depletion curves calculation task (box A6 in Figure 1). In

figure 3 we can see the architecture and the workflow of this customized WPS (Snow cover percentage calculation WPS). Suppose that a snow coverage area (SCA) satellite image for the study area is already acquired. Then, a raster-vector geoprocessing service in the *Image Processing-WPS* module is able to transform the SCA image, containing a range of values (for example snow, clouds, water, and land), and return vector data (GML or SHP format) representing uniquely snow values. Steps 3 and 4 in the figure request the *Format Data-WPS,* if needed. Next, (steps 5 and 6) by using a spatial intersection service available in the *Topology-WPS* module, is computed the intersected region between the initial zone (polygon) and the vector data representing the snow coverage area (polygon set). Steps 7 and 8 will perform the execute of *Area* process in the *Topology-WPS* to calculate the area of the intersected region, finally a non-spatial processing service in *Statistics-WPS* returns to the client application the percentage of snow coverage area as a relation between the area of the intersected region and the area of the initial zone.
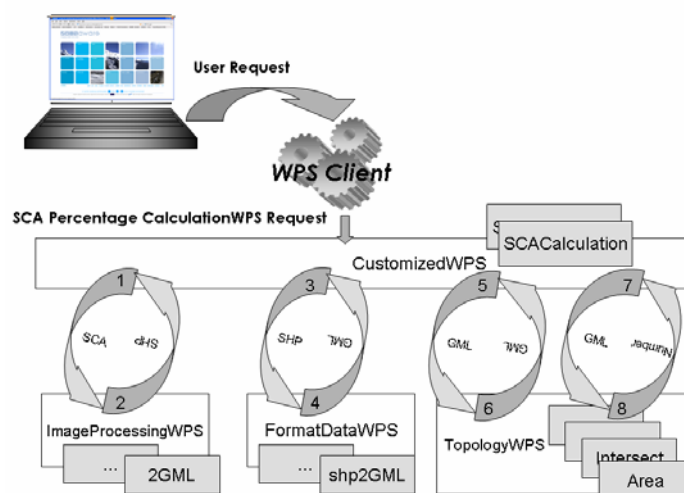

Figure 3: Snow cover percentage calculation WPS workflow.

DISCUSSION AND CONCLUSION

This paper presented the system design for modelling distributed geoprocessing services for water resource management application on top of basic SDI functionality. Our system includes analysis and GIS functions commonly used by environmental scientists. Some stand-alone and web-based applications for hydroinformatics are available (Cheng et al. 2006, Hughes and Forsyth 2006, Olivera et al. 2006, Soh et al. 2006), yet they work either on distributed but static data or necessarily on local data. A different approach is taken here, similar to Friis-Christensen et al. 2006 and Kiehle et al. 2006 because WPS is also used, where users perform analysis processes with distributed data and processing services. In this paper we have explored the possibility of migrating desktop-based algorithms into web services to better connect scientists with their data, models, and services in an interoperable manner. We are also interested in enabling semantics properties (Lemmens et al. 2006) in geoprocessing service descriptions to ease and assess the reusability of such geoprocessing services in different contexts.

BIBLIOGRAPHY
Anderson, G., Moreno-Sanchez, R.: Building Web-Based Spatial Information Solutions around Open Specification and Open Source Software. Transaction in GIS 7 (4) 447-446 (2003)

Caldeweyher, D., Zhang, J., Pham, B.: OpenCIS–Open Source GIS-based web community information system. International Journal of Geographical Information Science 20 (8) 885-898 (2006)

Cheng, Q., Ko, C., Yuan, Y., Ge, Y., Zhang, S.: GIS modeling for predicting river runoff volume in ungauged drainages in the Greater Toronto Area, Canada. Computers & Geosciences 32 (8) 1108-1119 (2006)

Friis-Christensen, A. et al.: Building Service Oriented Applications on top of a Spatial Data Infrastructure – A forest Fire Assessment Example. In: Proc. of 9th AGILE International Conference on Geographic Information Science, pages 119-127 (2006)

Granell, C., Gould, M., Gronmo, R., Skogan, D.: Improving Reuse of Web Service Composition. In: Proc. of the EC-Web Conference. Springer LNCS 3590, pages 358-367 (2005)

Hughes, D.A., Forsyth, D.A.: A generic database and spatial interface for the application of hydrological and water resource models. Computers & Geosciences 32 (9) 1389-1402 (2006)

Kiehle, C.: Business logic for geoprocessing of distributed geodata. Computers & Geosciences 32 (10) 1746-1757 (2006)

Lemmens, R., Wytzisk, A., de By, R., Granell, C., Gould, M., van Oosterom, P.: Integrating Semantic and Syntactic Descriptions to Chain Geographic Services IEEE Internet Computing Magazine, 10(5): 42-52 (2006)

Lu, X.: GIS-T Web Services: A New Design Model for Developing GIS Customized ITS Application Systems. In: Proc. of ICCSA Conference. Springer LNCS 3980, pages 875-884 (2006)

Martinec, J., Rango, A., Major, E.: The Snowmelt Runoff Model (SRM) User's Manual. NASA RP-1100. Greenbelt, Maryland (1983)

Olivera, F. Koka, S., Nelson, J.: WaterNet: a GIS Application for the Analysis of Hydrologic Networks Using Vector Spatial Data. Transactions in GIS 10 (3) 355-375 (2006)

Ramamurthy, M.K.: A new generation of cyberinfrastructure and data services for earth science education and research. Advances in Geosciences 8 69-78 (2006)

Ryden, K. (ed.): OpenGIS Implementation Specification for Geographic information - Simple feature access – Part 1: Common architecture. Open Geospatial Consortium. http://www.opengeospatial.org/standards/sfa (2005) (accessed 17 January 2007)

Sayar, A., Pierce, M., Gox, G.: Integrating AJAX Approach into GIS Visualization Web Services. Proc. of the Advanced International Conference on Telecomunications and International Conference on Internet and Web Applications. IEEE Computer Society, pages 169 (2006)

Schut, P., Whiteside, A. (eds.): OpenGIS Web Processing Service 0.4.0. Open Geospatial Consortium http://portal.opengeospatial.org/files/?artifact_id=13149&version=1&format=pdf (2005) (accessed 17 January 2007)

Soh, L.-K., Samal, A., Zhang, J.: A Task-Based Approach to User Interface Design for a Web-Based
Hydrologic Information System. Transactions in GIS 10 (3) 417-449 (2006)