Integrated Modeling for Water Resource Management

David A. Thurman[a], Randal Y. Taira[a], and Thomas D. Georgianna[b]

[a]Pacific Northwest National Laboratory
Battelle Seattle Research Center
Seattle, Washington
dave.thurman@pnl.gov
206-528-3221

[b]Department of Natural Resources & Parks
Water and Land Resources Division
King County, Washington
tom.georgianna@metrokc.gov
206-263-6569

We report on the development of a collaborative problem solving environment for hydrologists, water quality planners, and natural resource managers, all roles within a natural resource management agency and stakeholders in an integrated water resource management process. We describe the Integrated Water Resource Modeling System, under development by Pacific Northwest National Laboratory (PNNL) for the Department of Natural Resources and Parks in King County, Washington. This system will integrate a collection of water resource models (representing more than 60 watersheds, rivers, lakes, estuaries, etc.) to support scientific studies that will inform complex water, land use, and other natural resource management decisions in King County and the surrounding region. Here, we discuss the five-step process used to ascertain the (potentially opposing) needs and interests of stakeholders, describe the current status of system development, and how this integrated modeling capability extends the County's ability to address natural resource policy issues.

Keywords: integrated modeling, hydrology, problem solving, urban planning

Introduction

The Water and Land Resources Division (WLRD) within the Department of Natural Resources and Parks (DNRP) is King County's environmental monitoring arm tasked with monitoring the condition of freshwater resources in the region. WLRD and its predecessor, Metropolitan King County, have been collecting, analyzing and reporting on the condition of King County's lakes, rivers, streams and the Puget Sound since 1962. That history equates to a large amount of data resident in a variety of data bases, while ongoing monitoring projects continue to generate tremendous volumes of data on an annual basis.

Typically, these data have made their way into policy maker and public hands by way of annual reports. The slow process of analyzing and reporting in this fashion has created the well-known environment of an agency that is data rich and information poor. This situation was noted by King County (County) policy makers and WLRD was tasked with determining methods for better managing and using those data to address County needs (King County, 2001). One of the approaches taken by WLRD in addressing this problem is the application of computational models to predict the potential impact of various County policies. These models typically require large volumes of data as input and generate even larger amounts of data as output, often in formats accessible only via other software that can summarize and present the information in tabular or visual forms.

Having chosen the computational modeling approach, one immediate task was to determine how to link models of more than sixty watersheds, lakes, and rivers of potential interest. Moreover, since all models to be used could not be identified *a priori* the goal became to develop a flexible and extensible modeling framework that supported future integration of a variety of water body and risk models that would be identified in the future.

During the design of this overall integrating framework, the Integrated Water Resource Modeling System (IWRMS), the users of the end products were identified and their requirements of those products were catalogued. As a result, the end product that we are working toward is a system of integrated water quantity and quality models that can be launched in a variety of combinations to support studies required by WLRD planners. The WLRD planners represent a variety of projects and disciplines, such as salmon recovery, wastewater treatment, swimming beach monitoring, small lakes monitoring, and analysis of benthic data. The IWRMS consists of a number of components,

detailed in a subsequent section, designed to address the needs of a variety of users who support the WLRD planners. Key to effective system design was identifying these users and determining their role-related needs. In the remaining sections we describe the process used to identify potential users and collect their needs, the resulting system design, current status of the project, and our assessment of how the IWRMS will provide an important new capability for King County.

Stakeholder Involvement Process

As described above, the goal of this effort is to develop an integrated modeling *capability* that can be used by a range of users to address future land use and resource management scenarios and provide scientific support to decision makers. We choose to use the term *capability,* rather than *system,* to reflect the goal of creating not a single monolithic computer system that everyone must use to perform integrated modeling, nor even a single computing environment for all users, but rather a distributed set of technologies that facilitate communication, collaboration, coordination, and the exchange of information among a varied set of users, thus creating the integrated modeling capability within the organization. While such a capability is still a system *per se,* we found that use of the word "system" to be problematic. To many, it indicated a single, centralized computer system. Speaking in terms of building a "capability" addresses that problem and helps participants recognize that the integrated modeling capability will become resident within the organization through the development and provision of technologies appropriate to each user's needs.

Users of an integrated modeling capability range from hydrological modelers, to water quality planners, to natural resource managers and policy makers. While users share a common responsibility for managing King County's natural resources, they bring varied backgrounds and understanding of the technical issues inherent in developing a modeling capability. Thus, it is important that all user groups participate in identifying and defining system specifications, rather than allowing only those users with the most technical sophistication to identify requirements.

The project team developed a five-step requirements gathering process involving three categories of users: modelers, planners, and managers. The definition of requirements started with interviews and focus group sessions with representative users. Participants were selected with the goal of documenting a range of requirements and views, including views skeptical of the applicability of an integrated modeling capability. Prior to interviews, participants received background information on the purpose of the project and the requirements gathering process. Approximately 45 individuals were interviewed to determine their current use of computational models and discuss ways in which computational modeling could be more useful to them in performance of their job duties. Among other needs, modelers commented on the value of enhancing modeling to improve field monitoring, planners stressed the need for a system to address numerous "what-if" scenarios with which they are faced, and managers identified the need for the system to support decision making on scales ranging from county-wide scenarios down to individual basins and catchments.

The results of these interview sessions were consolidated to create potential use case scenarios, which were then used in focus groups for the purpose of identifying specific system capabilities that would enhance each user group's ability to perform their job duties. For each scenario, we sought information on the particular computational models that would be required to address the scenario, the data required as input to those models, the function requirements for organizing the results of modeling runs, the capabilities required for analyzing results, the data integration required among models when multiple models were required, and the capabilities necessary for facilitating communication of results to colleagues, clients, and stakeholders.

During the focus groups, it became apparent that some users' contributions were limited by a lack of awareness of computing technologies and their capabilities. In response, the project team conducted a *technology demonstration workshop* to present a number of currently available technologies, ranging from animation and visualization tools, other model integration efforts, data catalogs, and problem solving and decision support environments. The purpose of this workshop was to inform users about the current state of technology and to stimulate their thinking about how the demonstrated capabilities might be useful in their work.

After this workshop, the project team analyzed gathered information (more than 300 requirements statements from potential users) and generated eighteen "requirements bundles," statements that could be prioritized by representatives of all user groups. Via a web-based survey, County staff were then asked to prioritize the

requirements. Results of this activity were then presented back to the entire group in a final workshop to solicit additional input and agree on conceptual requirements for the integrated water resource modeling system. A more detailed description of this approach is presented in Thurman *et al.* (2004). From these results, PNNL developed a system design and implementation plan focusing on seven key technology areas to address identified needs. For components requiring user interface elements, we began by transforming identified requirements into design sketches that aimed to express how such functionality would be exposed to the user. Through the design process, we have sought active participation by all stakeholders, using a "participatory design" process (Kensing and Munk-Madsen, 1993) to engage potential users in the user-interface decision-making process. This process involves numerous iterative design sessions in which we use design sketches to allow potential users to work through typical scenarios indicative of expected system use and comment on likes and dislikes with regards to the interface design. Similarly, we have followed an iterative design/development/review process for other components of system functionality.

System Design

The IWRMS is being developed to provide an integrated modeling capability to support scientific investigations and planning efforts. While the core function of IWRMS is to integrate models, it is also intended to provide a collaborative working environment in which the target user population includes hydrologists, water quality planners, and DNRP management. As described by Dorow *et al.* (2004) and illustrated in Figure 1, the IWRMS system architecture consists of a collection of components to provide necessary identified functionality, including the identification and assembly of relevant data, model integration, definition and management of computational "studies," data analysis and visualization, and robust data management. A brief overview of the functionality of individual components is provided below.

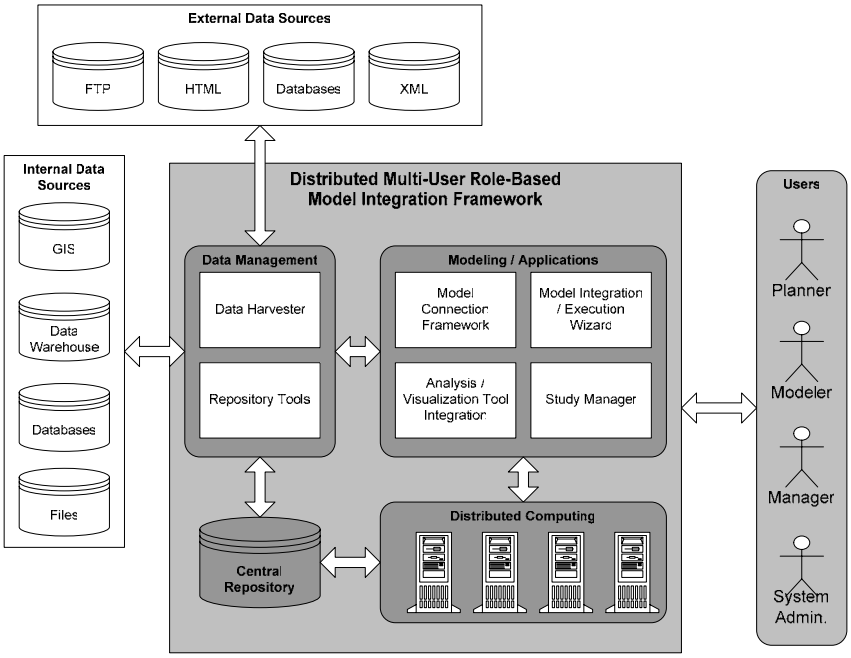

Figure 1 – IWRMS System Architecture

<u>Model Connection Framework</u> – The core functionality of IWRMS is the ability to integrate medium-specific models (e.g., watershed, lake, and river models) in a seamless fashion such that any appropriate sub-collection of models can be used together as part of an integrated modeling approach. Two key requirements drive the design of this component: (1) While the County has selected the models it will use in near-term scientific studies, it does not want to be precluded from using different models in the future, and (2) the County needs the flexibility to add different "types" of models (e.g., estuary or health risk assessment models) to the system in the future. Thus, the

goal of the model connection framework is to enable any consuming model (e.g., a lake model) to accept data from any appropriate producing model (e.g., watershed model), but to do so without "hardwiring" any two models together. That is, no changes are made to any specific model to facilitate data exchange with any other model. Instead, the model connection framework provides a collection of standardized data exchange specifications (i.e., descriptions of data types and formats) between model 'types.' Thus, where appropriate specifications have been produced between any two model types, all models of those types can exchange data so long as they produce and consume data according to the standardized data exchange specifications. In designing the system in this way, IWRMS users can upgrade and switch models and also extend the types of models used in the system, without necessitating changes to any of the other models used in the system.

Data Harvester – The models require data from multiple internal and external sources that include, but are not limited to, stream and rainfall gauge data, metrological data, habitat data, and biological data. These data are retrieved in various forms (e.g., from databases and websites) and with varying frequencies (e.g., daily to annually). Retrieving and reformatting data is extremely time-consuming and expensive. Using the Data Harvester, users can identify external data sources from which to retrieve data and set up transformation steps to map the retrieved data into County-specific formats. Extracted data are stored in an SQL database, from which they can be retrieved to support QA/QC operations and other processing needed to prepare the data for use by the models.

Model Integration Wizard – Integrating models typically requires the development of unique software applications to transform each model's input and output data into formats that can be passed between models. The IWRMS Model Integration Wizard enables users to "register" a model in the system, guiding the user through a process of graphically mapping the model-specific input and output files to the system-defined data specifications. From this mapping, the wizard can then generate the necessary data transformations to create the required model input files and convert the model output files during subsequent model executions.

Data Repository with Pedigree – The Data Repository is the central storage facility for data that is imported into, and produced within, the IWRMS, and for the models that are integrated into it. It is comprised of a SQL database, a source code management system, and a file management system. It is the data store for all elements in the system including: data transformations, data specifications, modeling scenarios, registered models, data sets (either from data sources or model run results), reports, and archived data. In addition, metadata for each data set in the system will be produced to provide a "pedigree" that uniquely identifies its source and the processes which produced it, which is vital to results tracking and reproducibility.

Analysis/Visualization Tool Integration -- The DNRP has specific data analysis and visualization tools that it employs to study and analyze modeling results (e.g., ArcGIS™ and MATLAB™). This component will extend IWRMS capability to export modeling results to said tools, using functionality similar to that employed in the Model Integration Wizard.

Study Manager – The Study Manager provides a robust, usable human interface to the integrated modeling capability, enabling users to create and track a collection of modeling scenarios or studies. New studies are generated via interaction with the New Study Wizard that enables users to specify modeling scenarios, select geographic regions of interest, and assign relevant tasks to team members. Envisioned as supporting a collaborative process in which team members use the wizard to walk through the creation of a new study, planners and modelers will work together through the scientific problem being addressed, identifying necessary resource issues, assessment methods, and task assignments. Once a study is underway, the Study Manager also provides task status information to support effective work management across the group of planners and modelers. In addition, it can also be used to search for previously executed studies that may be executed again or used as a template for a new study. Finally, the Study Manager provides a reporting capability to summarize status and results of any previous or current study.

Distributed Computing Environment – The distributed computing subsystem provides the ability to run the models in IWRMS across a collection compute nodes within a dedicated cluster to increase the speed with which many studies can be completed. It provides location-transparent model execution on any number of machines within a

---

™ ArcGIS is a registered trademark of Environmental Systems Research, Inc. (ESRI), Redlands, CA.
™ MATLAB is a registered trademark of The MathWorks, Inc., Natick, MA.

local area network. This component enables the County to undertake a wider range of modeling scenarios and perhaps implement stochastic analyses that are not currently practical because of time and resource constraints.

Development Status

The software development phase of the project began in January 2004. Our project management plan called for progress "demonstrations" at six-month intervals in addition to monthly reports and regular interaction between PNNL and County staff. The progress demonstrations provide County staff with an update on project progress and demonstrate the current state of component functionality. Two such demonstrations were held in July 2004 and January 2005. The next demonstration will be held in July 2005, at which time we plan to demonstrate the majority of intended system functionality. Official delivery of the initial version of IWRMS to the County is scheduled for December 2005, with on-going support provided for at least twelve months following delivery. The sections below provide a brief overview on the development status of each system component identified previously.

Model Connection Framework – One of our primary project goals is to utilize existing technology where feasible in order to save development time and expense. The concept of a system to integrate models is not new and existing systems were surveyed as part of the IWRMS development process (see Taira *et al.*, 2003). Through that research, it was determined that the Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) best met IWRMS requirements on the basis of cost, model integration paradigm, system knowledge, application experience, and access to code authors. Developed by PNNL in partnership with the U.S. DOE, U.S. EPA, U.S. DoD, and the U.S. NRC, FRAMES provides a platform for the integration of medium-specific models (Whelan *et al.*, 1997) such as the environmental models that will be used by the DNRP. A majority of the IWRMS model integration requirements are met by the existing system. Extensions to FRAMES are being made, however, to support simultaneous use by multiple users and to support interaction with a central data repository instead of local files resident on the computer running FRAMES. Software development of this component has been completed and it is in the testing phase. A screen capture of the FRAMES user interface is shown in Figure 2, illustrating the conceptual connections between a collection of source, transport media, and risk assessment models.

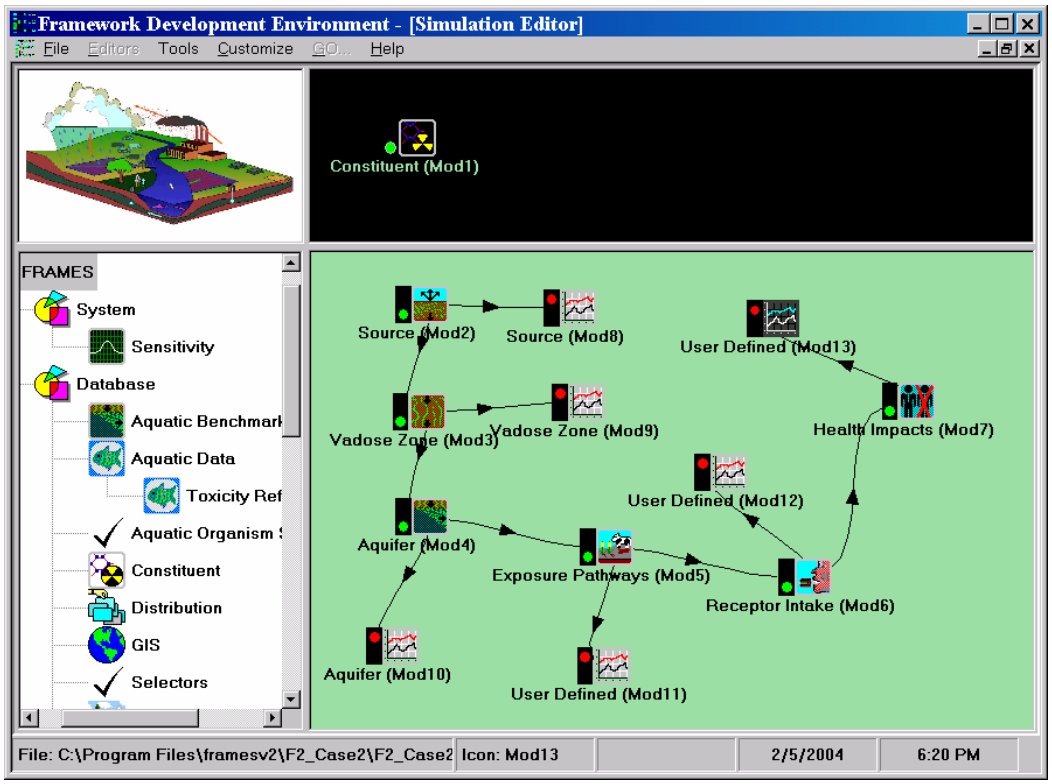

Figure 2 – FRAMES User Interface

Study Manager – An iterative design and development process was used to develop the Study Manager, shown in Figure 3 below. Numerous demonstrations were provided to County staff over a 12-month development period to ensure their familiarity with the Study Manager concept and provide opportunities for design input. Development of this component is complete and it is currently undergoing integration testing with the Central Data Repository.

Data Harvester – As described above, the Data Harvester provides the ability to extract data from four data source types: web sites, FTP servers, databases, and XML data sources. Currently, the development of the web site harvester is nearly complete, providing the ability to schedule a one-time or recurring extraction. It handles both raw data and forms-based web sites, downloads raw data according to the preferred schedule, and provides the means of parsing the data for storage in an SQL database. This functionality is current being extended to support other data source types. The core of the software for the database harvester has been completed, but it has not yet been tied into the user interface for scheduling data extractions and parsing the downloaded data. Subsequent development efforts will address FTP servers and XML data sources.

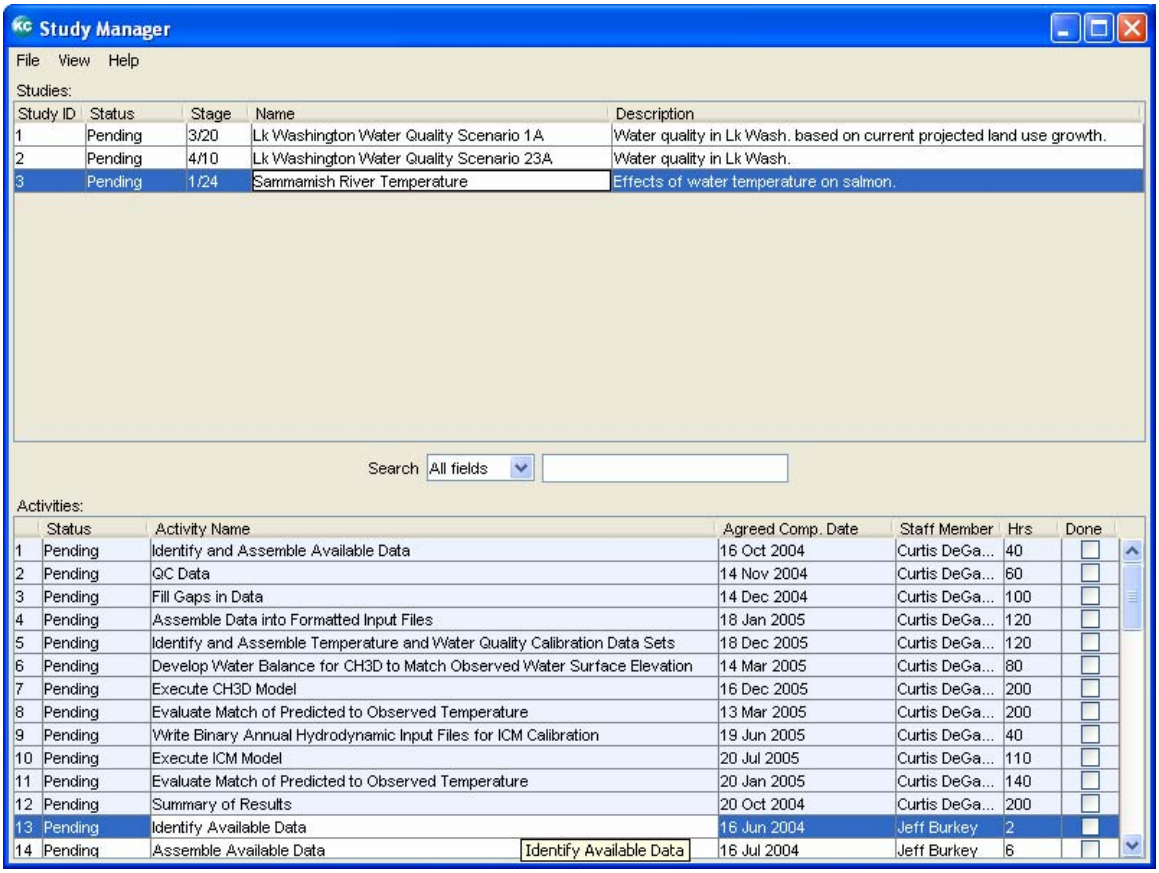

Figure 3 – IWRMS Study Manager

Data Repository with Pedigree – The database component of the repository has been implemented and connections with the Study Manager and Model Connection Framework are nearing completion. Integration with other IWRMS components (e.g., data harvester, model integration wizard, and analysis/visualization tools) is on-going, while the development of a file server component and version control system is just beginning. The development of a data browser, which enables users to search the repository using standard query tools is currently underway.

Distributed Computing Architecture – In collaboration with County network administrators, it was decided in December 2004 to implement the system on a dedicated computing cluster opposed to the original vision of using

available resources on the local area network. PNNL developed the system specification for this system and purchased the initial components for configuration and testing. Work remains to be completed on the development of software to support the execution of IWRMS models and data transformation utilities on the cluster.

Model Integration Wizard – The user interface that enables the user to enter model information and to register a model executable and associated files has been completed. It also enables the user to select system data specifications, in the form of FRAMES dictionary files, to be associated with the model's input and output files as shown in Figure 4. Initial work has begun on the interface that allows the user to map parameters in the specifications to data in the input and output files. Basic parsing tools have been implemented; however, the code to do more complicated recursive and conditional parsing is still in development. The code needed to communicate with the data repository, and substitute data from the repository into model input files, is also under development.

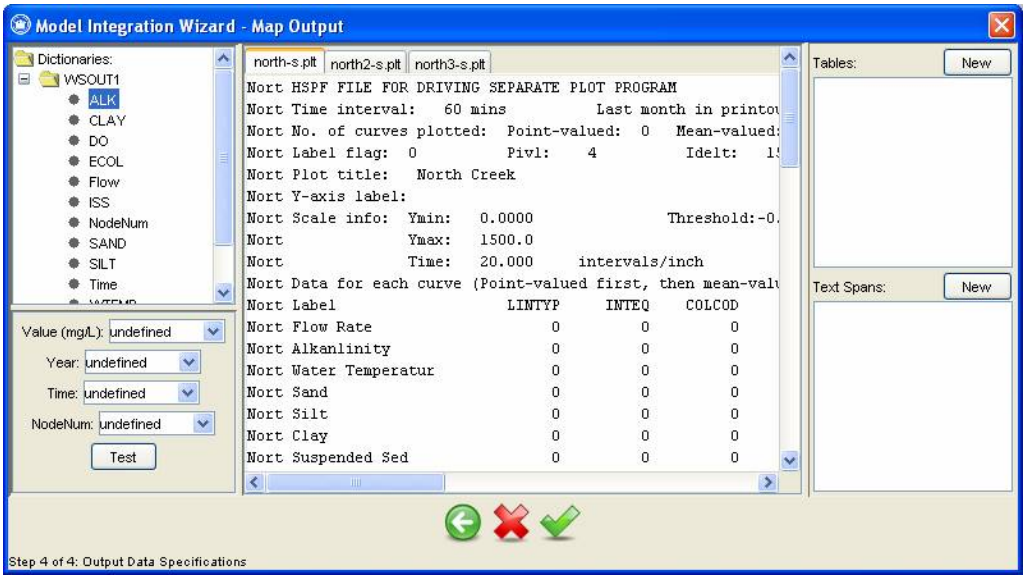

Figure 4 – IWRMS Model Integration Wizard

Assessment & Conclusions

The more obvious results of the IWRMS system will be a much greater speed in producing usable information from large quantities of data. A large part of the increased speed will stem from the ability to automatically move data from its initial residence in a database into the data repository. The data repository will have a memory as to the format requirements of the various models and will be able to make necessary conversions in an automated way. Data security is an important issue in the production of information that is used for public policy determination. The data repository will provide a moderately high degree of security for all data that is used in model runs.

In addition to increased speed of data access, we expect these data to be usable in new ways. As an example, land use is an input to many of the models. The IWRMS will allow WLRD to ask questions about development options by varying land use data. The same is true for environmental management policies that are reflected in the ways that land is used. Alternative population forecasts can be evaluated in terms of their impacts on water resources and the interaction between population and policy scenarios can be seen.

There is a spatial coverage issue that is dealt with by the modeling system. The current method of sampling and reporting is limited by time and budget constraints. There is a limited number of sites in lakes and streams that can be sampled, dictating the selection of sampling sites that represent a finite set of characteristics for lakes and streams. In other words, we can't cover everything. Using models that incorporate knowledge of the hydrodynamics of lakes, streams and rivers, however, provides us with the ability to disperse knowledge of water

characteristics at one point to many points. Thus, we come closer to universal coverage both spatially and over specified periods of time.

The fact that the mathematical models are linked, or integrated, provides a positive framework in that we now have a cohesive structure. The cohesive structure derives from the fact that since data is passing from one class of models to another, the rationale and the set of assumptions (as well as the data) that underlie the models have to be portable from one model to the next. That cohesive framework travels even further down the system to the data and the sampling. The sampling is carried out by several different programs and by individuals representing several different disciplines. In an unconstrained business environment there is a tendency for those programs and disciplines to develop their own rationales. The linkage of the models delivers some cohesion to those programs and disciplines in that the model data requirement must be met, and those data requirements are determined by the agreed upon rationales for model production.

Acknowledgements

King County staff involved in this project include Jonathan Frodge, Larry Jones, Jim Simmonds, Randy Shuman, Jeff Burkey, Curtis DeGasperi, Doug Henderson, Kevin Schock, and a host of others. We gratefully acknowledge their continued patience with the IWRMS design and development process.

PNNL staff leading efforts within the IWRMS design and development process include Kevin Dorow, Ranata Johnson, and Andrew Cowell. Other staff making key contributions include Liam McGrath, Eric Robinson, Nathan Tenney, Justin Almquist, Bonnie Hoopes, Karl Castleton, Alex Donaldson, Julie Dunkle, and Sharon Eaton.

References

Dorow, K.E.; Gorton, I., and Thurman, D.A. (2004). A distributed multi-user role-based model integration framework. In *Proceedings of the 2nd International Conference on Prevention, Assessment, Rehabitation and Development of Brownfield Sites.* Siena, Italy. June 2004.

Kensing, F. and Munk-Madsen, A. (1993). Participatory design: Structure in the toolbox. *Communications of the ACM, (36)* 6, pp. 78-85.

King County, Department of Natural Resources and Parks, Water and Land Resources Division Strategic Plan, May 2001.

Taira, R.Y., Johnson, D.M., and Thurman, DA. (2003). An Evaluation of Model Integration Frameworks for the Integrated Water Resource Modeling System. PNNL Technical Report #PNNL-Draft. Pacific Northwest National Laboratory, Richland, Washington.

Thurman, D.A., Cowell, A.J., Taira, R.Y., and Frodge, J. (2004). Designing a collaborative problem solving environment for integrated water resource modelling. In *Proceedings of the 2nd International Conference on Prevention, Assessment, Rehabitation and Development of Brownfield Sites.* Siena, Italy. June 2004.

Whelan, G., Castleton, K.J., Buck, J.W., Gelston, G.M., Hoopes, B.L., Pelton, M.A., Strenge, D.L., & Kickert, R.N. (1997). Concepts of a Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES). Report #PNNL-11748. Pacific Northwest National Laboratory, Richland, Washington.