

Principal Investigators: Deborah Dent, Information Technology Laboratory, 601 634-3455

Title: Information Architecture & Supporting Tools for RSM

Topic Area: D1: Database Tools for Data Storage and Mining.

Objective: This work unit will provide the underlying information technology architecture for databases, data transport mechanisms and data manipulation tools to support RSM Informatics and establish a data warehouse for use by all program work units.

Benefits: Proper development of the RSM Information Architecture is crucial to the production of easy to use systems that provide end users accurate predictive tools that support effective decision making with acceptable response times [Need 18 - Topic Area B3]. Proper development begins with the establishment of data standards that facilitate the flow of information between physical databases and systems [Need 21]. The RSM database will include both geospatial and temporal data that will provide input to knowledge discovery tools and RSM models [Need 52 - Topic Area D2 and Need 56 Topic Areas B3 and D2]. Additionally, the infrastructure will support the interagency sharing of sediment data [Need 55]. The infrastructure will support the generation of problem lists and predict systemic impacts of these problems [Need 58 - Topic Area D2] and will provide predictive tools for the assessment of erosion or deposition [Need 61 - Topic Area A1]. This Work Unit provides the information infrastructure support for other work units and therefore must be functionally tested, deployed and evaluated in conjunction with other RSM work units. Therefore, requirements of Task E Technology Transfer and Insertion must be planned, tested and evaluated in cooperation with other work units.

This work produces new tools and methods for the USACE and nation. It is an integral part of the Regional Sediment Management Research Program, and thus contributes primarily to support of the USACE's navigation, flood/storm damage reduction, and environmental protection and quality missions. It supports all 8 Civil Works strategic goals and 7 of 9 Listening Session objectives identified by HQUSACE as R&D priorities. With companion work units, it employs active technology transfer and insertion.

Work Description: A conceptual information architecture will be designed based on a RSM requirements analysis. A sound database system, based on the Spatial Data Standards for Facilities, Infrastructure, and Environment (SDSFIE), will be developed to manage the volumes of data that will be required to support RSM. This system will draw on information in multiple remote databases through the LMS framework as well as leverage proposed enterprise geospatial strategies ongoing in the CW Geospatial R&D program. Data mining tools will be provided to facilitate discovery of trends and interdependencies among RSM datasets, as well as to provide input to Decision Support Systems (defined in Topic Area D2). Additionally, a RSM data warehouse will be established to facilitate the mining and archiving of temporal data.

Initially, work will concentrate on the identification of required and existing sources of RSM information. This requirements analysis will include interviews with model developers and managers, field users, and decision makers from all aspects of sediment management including planning, design, construction, and operations. Data to support alternative analyses techniques developed in Topic Area D2 will also be identified. The analysis will include identification of required data content, types, scales, and resolutions. A scalability context mechanism will be developed that allows for the storage of data at multiple levels of specificity (i.e. regional and local contexts) and allows the user to switch between the different contexts. The result of this effort will be a conceptual information architecture.

Based on the requirements of the conceptual information architecture, a correlation with the SDSFIE will be performed and documented, noting missing information. Appropriate SDSFIE enhancements will be performed based on the correlation, resulting in a standard database schema integrated with the SDSFIE. Where appropriate, an existing toolbox of applications which accompanies the SDSFIE will be injected into this effort. For example, a Filter Maker Tool was developed to allow user communities to automatically create subsets of the SDSFIE. This tool will be used to create a RSM filter which will be used in conjunction with other SDSFIE tools. Ultimately, this effort will result in an enhanced version of the SDSFIE which will provide the data standard for RSM.

Interface/connectivity tools developed within the LMS program, the XMS (GMS, SMS, WMS) program, and the Civil Works R&D Geospatial framework will be leveraged to provide connectivity between RSM decision support tools, computational models, local data, and the data warehouse. Where required, interface agreements between ERDC (or the Corps of Engineers when appropriate) and organizations responsible for other efforts will be developed. Presently, the interfaces and therefore responsible organizations are undefined, but could consist of any or all of the efforts identified under "Connections to other Efforts" in the Task Description.

An efficient method for collecting and archiving the data to a RSM data warehouse will be developed. The data warehouse will be the primary source of information for all of RSM advanced decision support techniques. The data warehouse effort will identify the subset of RSM information needed to support the decision support process. Extract, translate and load (ETL) utilities will be used to transfer the needed information from the various RSM data sources into the warehouse. The warehouse will be implemented based on the data warehouse product provided in Oracle 8i. A set of access utilities will be developed within the warehouse to provide basic decision support capabilities. Furthermore, user accessibility to the warehouse information will be provided via a web interface.

After the completion of the warehouse, a coordinated effort with Topic Area D2 will investigate the high-level decision support requirements for RSM. As a result, an investigation of data mining tools will be conducted in an effort to identify the data mining technology that best fits the problem. The investigation will consider predicted value types (discrete or categorical), the types and uncertainty of the input information used by the mining techniques, the amount of information available to mine, and the

output representations that are acceptable to the user. For example, are the users comfortable with a probabilistic distribution or would they be interested in a result with a simple certainty assessment? The objectives of the mining process will be closely coordinated with efforts in D2.

Documentation will be prepared in the form of help files, technical reports, and technical notes. Training materials will be prepared and workshops will be conducted to transfer this technology to the users. The developed database and tools will be published via web technology to users (members of a field evaluation group) for review and comment. Modifications suggested by the field evaluation group will be implemented in accordance with configuration management procedures established for the project. Technology transfer and insertion efforts within Task E will be leveraged. Other leveraging will include: 1. Civil Works R&D Geospatial Technologies Program - Framework Technologies. 2. Civil Works LMS Framework, ERDC. 3. CADD/GIS Technology Center Project: "Spatial Data Standards for Facilities, Infrastructure, and Environment (SDSFIE)", 2001, ERDC, ITL, Vicksburg, MS. 4. CADD/GIS Technology Center Project: "Dredging-Related Standards and Application Interfaces to Support Enterprise Wide Dredging Software", 2001, ERDC, Vicksburg, MS. 5. CADD/GIS Technology Center Project: "Development of a SDSFIE to GMS Interface", 2000, ERDC, Vicksburg, MS.

Collaborators in this work effort include Barry McCleave (ITL, 601-634-2599) and Denise Martin (ITL, 601-634-4574)/.

Products and Schedule:

The primary products of this work will be the RSM database architecture and tools and an RSM data warehouse.

<u>Product</u>	<u>Scheduled</u>
1. RSM Information Requirements	3Q/02
2. RSM Information Model	4Q/02
3. Tech Notes on Information Model	4Q/02
4. Enhanced SDS (based on the findings in product 1)	2Q/03
5. RSM Data Warehouse Prototype Design	4Q/03
6. TN: Warehouse Design	4Q/03
7. RSM Data Warehouse	3Q/04
8. TR: Information Model, Warehouse	3Q/04
9. RSM Data Mining Tool Prototype Design	4Q/04
10. TN: Data Mining Tool Prototype Design	4Q/04