

SAHRA Geo-Database: Design and Implementation

Matej Durcik, Hoshin V. Gupta, and Yuqiong Liu

SAHRA, Department of Hydrology & Water Resources, University of Arizona, Tucson, AZ

Introduction

The SAHRA Geo-Database (SGD) is designed to facilitate storage of and access to data via the Internet. Data access refers to viewing and downloading data from the SGD. The data are of two kinds – spatial and tabular data (including time-series data) – and will be maintained in the form of a relational database. The goal is to provide support to SAHRA researchers, non-SAHRA researchers and the general public, with appropriate levels of data access defined in the Data Policy. The SGD description and activities can be accessed from http://sahra.arizona.edu/research_data/SAHRAGeoDB/.

SAHRA's primary geographical focus is on three river basins: the Rio Grande/Rio Bravo, the Upper San Pedro and Salt-Verde. We also engage in supportive work at the regional scale, in the southwestern U.S. Data stored in the database covers whole south-west region with detailed focus on the studied basins, see Fig.1.

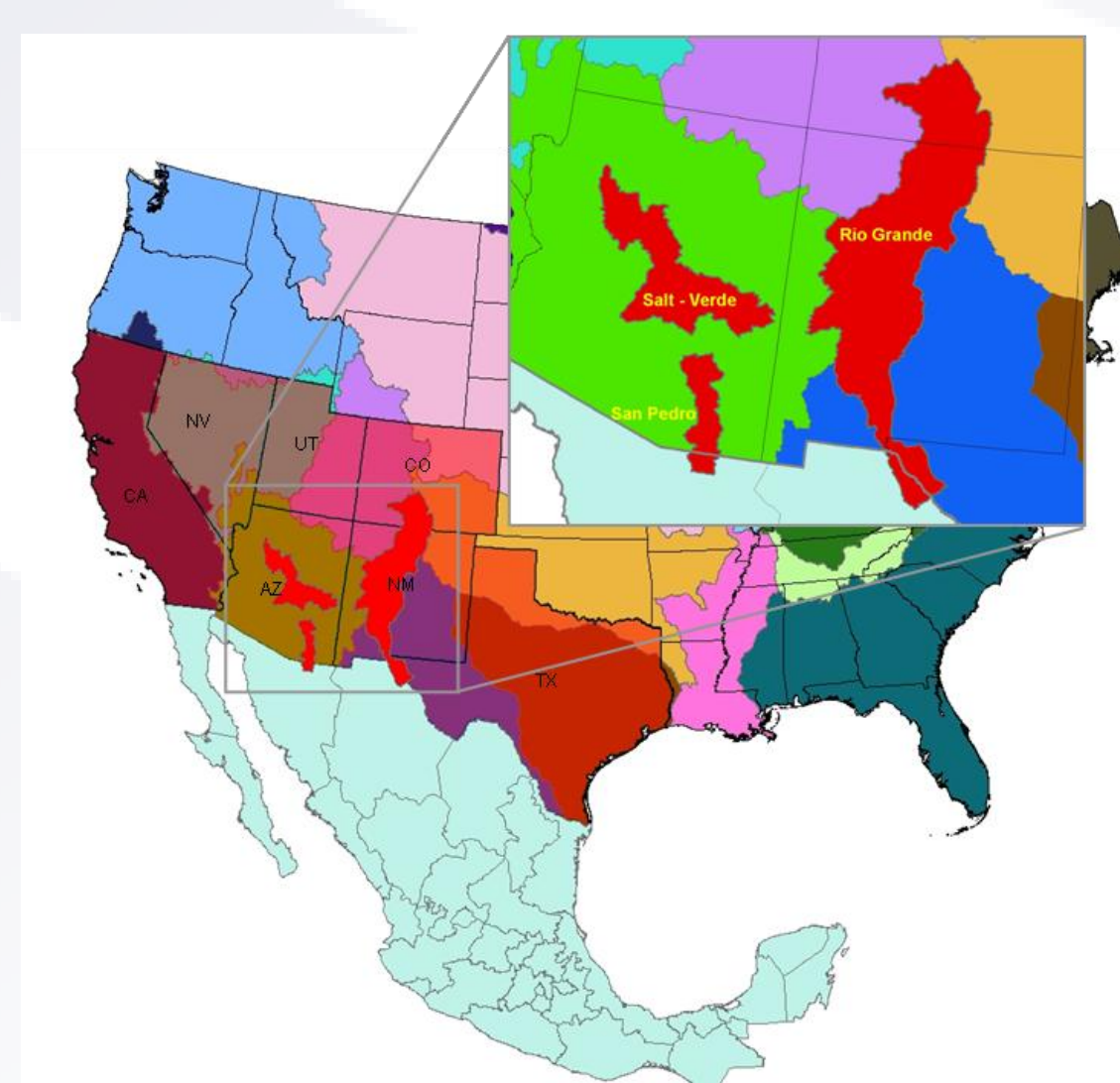


Fig.1 SAHRA's geographical focus

Tab.1 Development time table

	2005	2006
Database design (prototype in MS Access)	█	
Database implementation (Oracle)		█
Data access - graphical interface		█
Data retrieval interfaces		█
Spatial, attribute and TS data collection, QC and loading		█

Development Plan

Database development plan is divided into following activities: conceptual and physical design, database implementation which includes data collection and quality control, and data access tools. Overview of the development timeline is shown in Tab.1. The communication and discussion with SAHRA modeling groups is an important part of database design and implementation. This helps to determine their specific data needs, access and usage. It has two phases: first, to determine which data types have to be included in the database (database survey http://sahra.arizona.edu:8082/DB_Survey/ performed at the end of last year) and second, to discuss specific data which have to be obtained and loaded to the database, the priorities of downloading and storing these data, the formats of generated and measured data by SAHRA researchers, and data access mode which researchers plan to use.

Database Design

Overall view of SGD system is shown in Fig.2. SGD is design to:

- Support all areas of SAHRA hydrologic research that incorporates multi-resolution integrated modeling of basin-scale processes, socio-economic and behavioral modeling and decision support
- Distribute data to SAHRA researchers and public users over the internet through the database interface
- Allow direct data access for SAHRA researchers using database connection in ArcGIS tools or customized applications.
- Store spatial (vectors and grids) and non-spatial (tabular) data together including time-series and attribute tables in a relational database
- Stored metadata together with data in the database.

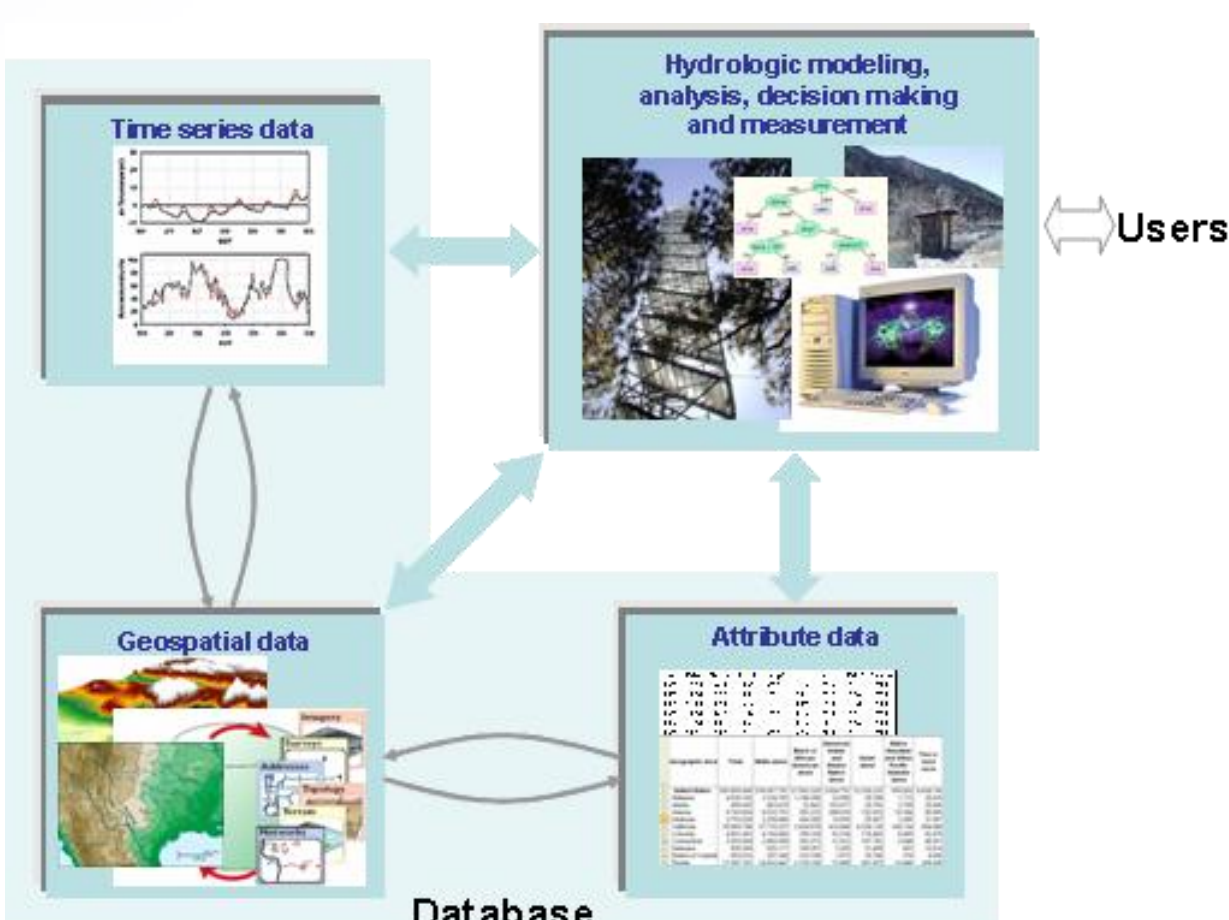


Fig.2 SGD system: overall view

Database Implementation

SGD consists of two partitions: public and private. It is implemented in the Oracle Relational Database Management Systems (RDBMS) which allows multi-user data access and editing. Geographical data are managed in RDBMS using ArcSDE which provides functionalities to store and retrieve vector and grid data in the database and makes data available to clients such as ArcCatalog, ArcMap, ArcIMS and other applications that serve data through the Internet. An ArcSDE two-tier Application Server Connect hardware configuration was used to setup client-server connection. A GIS client system runs the client application and a Database Server runs the ArcSDE service and houses the GIS database. The ArcIMS service runs on the same server as ArcSDE service. Fig.3 shows the database architecture.

The Arc Hydro data model (Maidment 2002) was used to build a public part of the database to store hydrologic and hydrographic data. Arc Hydro partitions surface water data into four components and the time-series model:

- **Network** - Connected set of points and lines showing pathways of water flow
- **Drainage** - Drainage areas and stream lines defined from surface topography
- **Channel** - A 3D representation of the shape of river and stream channels
- **Hydrography** - The base data from topographic maps and tabular data inventories
- **Time series** - Tabular attribute data describing time-varying attribute values for any database feature.

This database structure was expanded by adding components which describe ground water data (Hydrogeology and Modeling) (Strassberg et al. 2004) and data resources relevant to hydrologic modeling such as administrative boundaries, basic maps, SAHRA areas, grids and monitoring:

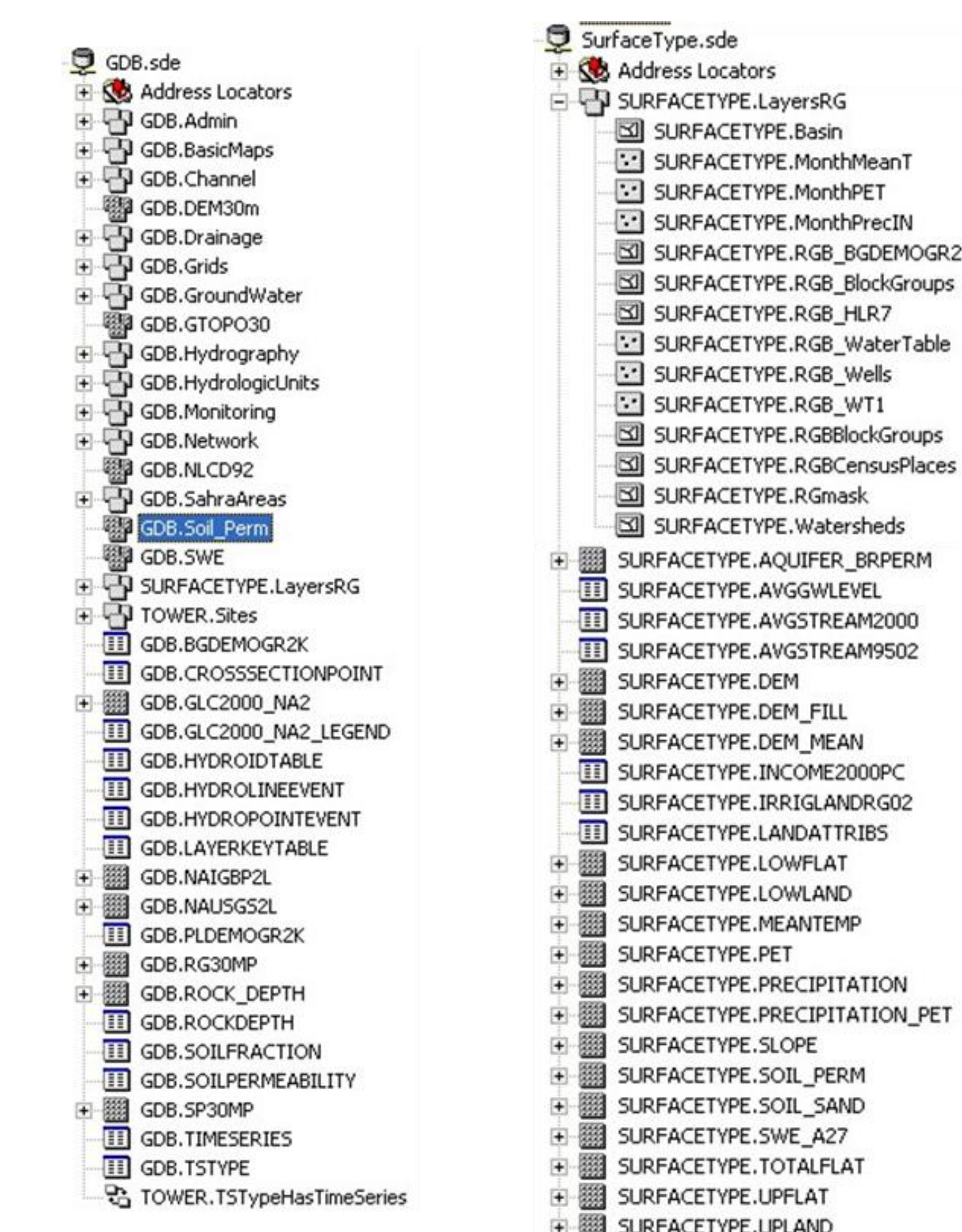


Fig.4 SGD structure

Fig.5 SurfaceType database structure

- **SurfaceType** - The collection of feature classes, grids and tabular data used in the conceptual framework modeling activities, see Fig.5.
- **Tower** - Stores time-series data collected from the Mt. Bigelow micrometeorological tower. This database partition was implemented as time-series part of the Arc Hydro data model, see Fig.6.
- **SWE** - Grid time-series implemented as raster catalog (generated by Roger Bales) , see Fig.7.

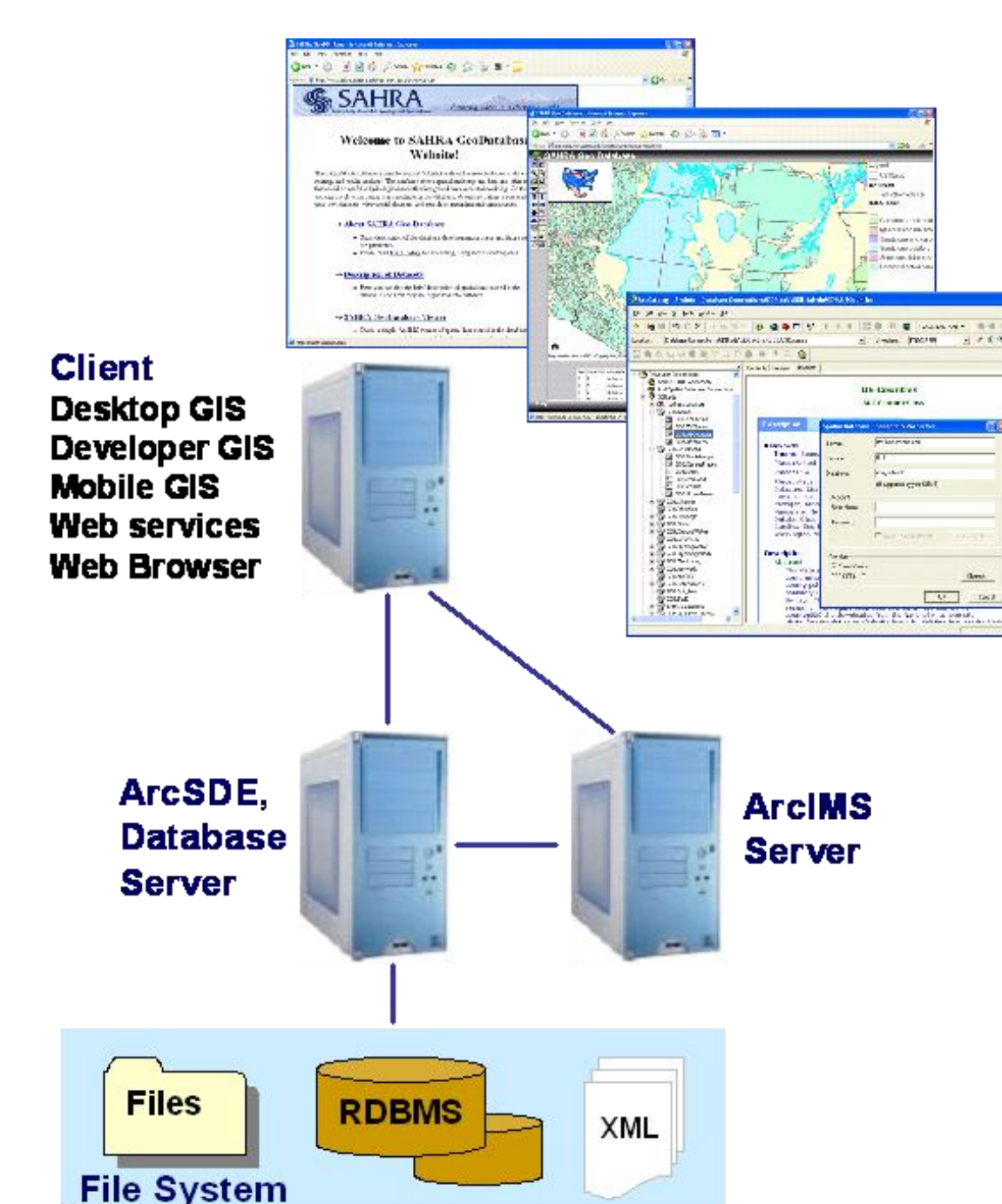


Fig. 3 Two-tier Database architecture

- **Hydrogeology** - A set of feature classes that describe the hydrogeology of an aquifer system
- **Modeling** - A set of feature classes that represent common modeling objects
- **Admin** - Boundaries of countries, states and counties
- **BasicMaps** - A set of feature classes that describes transportation network, cities and census areas
- **HydrologicUnits** - High resolution Hydrologic Units boundaries
- **SAHRAareas** - SAHRA research areas and monitoring sites
- **Grids** - Climatologic modeling grids as feature classes
- **Monitoring** - A set of feature classes that represent the spatial distribution of surface water, monitoring
- **Raster Data** - DEM, land cover, SWE, soil data, etc.

The SGD design in Fig.4 shows the structure of feature and raster datasets, raster catalog and tabular data.

The private part of SGD consists from thematic areas where the data are generated. To date, there are three areas covers by data: tower, snow water equivalent (SWE) and surface type data:

Data Policy

The purpose of this SAHRA data policy is to establish guidelines and agreements that govern data submission, access to and usage of data stored in the SGD. By doing so, we hope to facilitate the widest possible accessibility to SAHRA relevant data while protecting the data from unauthorized access and use, and maintaining a desirable level of data quality.

The data policy defines three categories of data: private, semi-private and public and three different levels of users based on the data categorization. It specifies a set of rules for data access, usage and submission. These rules will be implemented in the database interface which will include data access tools, metadata guidelines and templates for creating metadata, the data submission protocol for submitting data through the database interface and data format standards for data submission.

The SAHRA Data Policy can be accessed on the SAHRA website at http://www.sahra.arizona.edu/research_data/.

Database Access

Data access is free, but all users are required to fill out a registration form the first time they enter the database, to gain access. Users will be assigned privileges based on affiliation and relation to SAHRA. The database can be accessed in two different ways (see Fig.3):

- **Public Interface** - this browser-based mode enables any user to access public data and to submit request for private data. It will also enable authorized users to access semi-private data.
- **Direct Access** - this network-based mode (administrator-controlled) enables SAHRA users to submit and gain access to private and semi-private data, and also access the public data. Data can be access using ArcGIS Desktop application such as ArcMap and ArcCatalog, Developer GIS and web services.

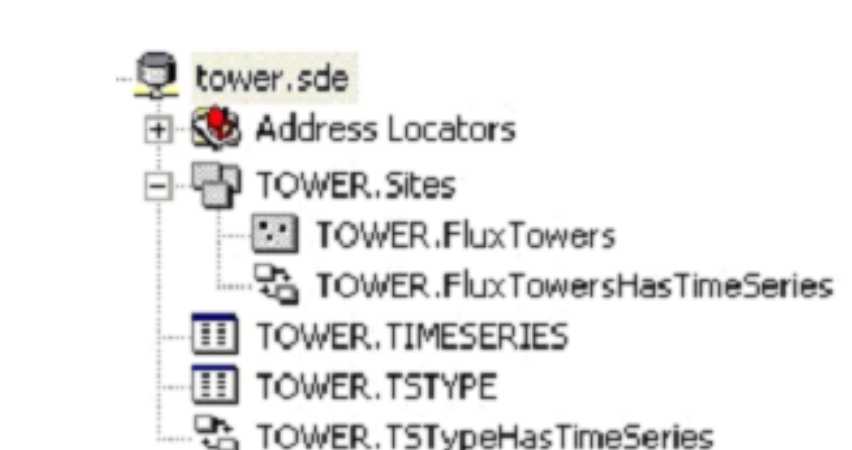


Fig.6 Tower database structure

Fig.7a SWE raster time-series table

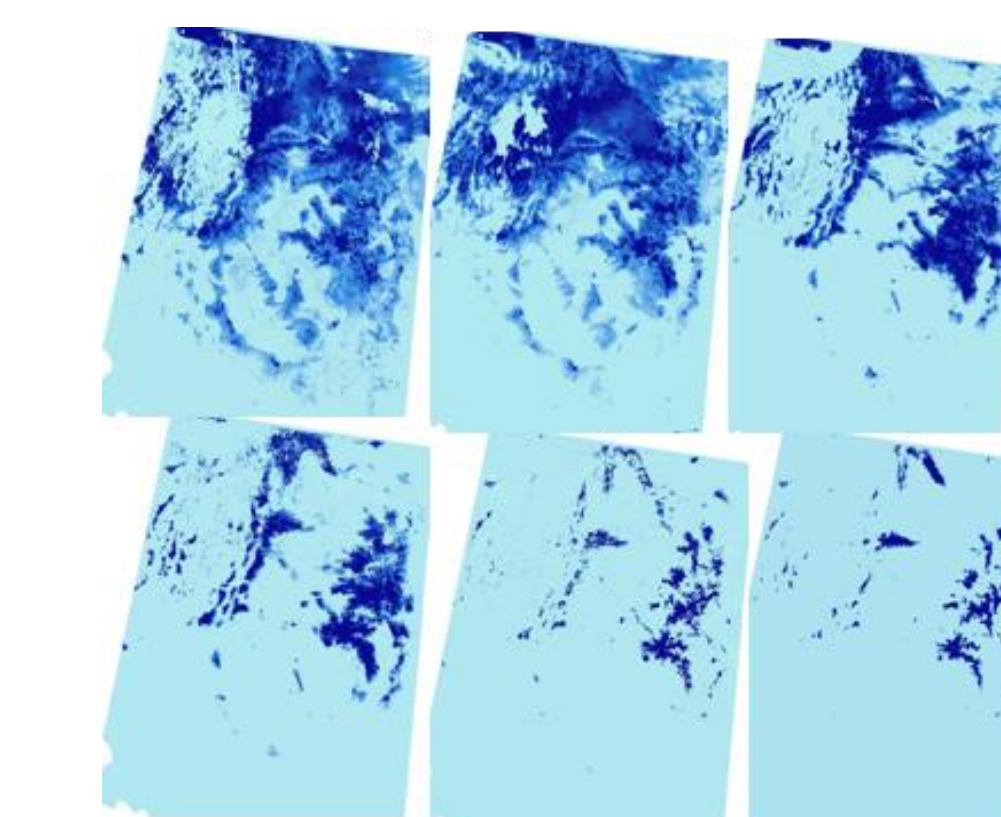


Fig.7b SWE raster time-series view from Jan95 to Jun95

Next Steps

The main focus in database development is to implement data policy in database access, use and data submission. It consists from the following activities:

- **Discussion with macro themes/modeling groups** about data needs, setting priorities for downloading and storing necessary data in the database, and data types and formats generated by modeling groups
- **Specify/adopt metadata core elements** (in progress)
- **Setup new database server**, tune database and load data from old server
- **Load data** generated by research activities and necessary data obtained from public sources
- **Improve Geo-database web interface** and implement data policy for public data access and make it publicly available
- **Develop database functionalities** for data and metadata search and access, data extraction and download, data submission, metadata template, and direct data access
- **Delineate/adopt river network** for the focused basins (in progress)
- **Generate TIN models** for the focused basins

References:

- D.Maidment (2002) Arc Hydro GIS for water resources. ESRI Press, Redlands, CA.
- G.Strassberg, D.Maidment and N.Jones (2004) Arc Hydro Groundwater data model. <http://www.cwrw.utexas.edu/gis/gishydro04/>.

Acknowledgements:

Primary support for this study is provided by Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA) under the STC Program of the National Science Foundation (NSF) Agreement No. EAR-9876800.

