AFRI AND CERA: A FLEXIBLE STORAGE AND RETRIEVAL SYSTEM FOR SPATIAL DATA

1. Introduction
The exploration of the earth has lead to a worldwide exponential increase of geo-referenced data. Data collection by satellites, global change investigation, and measurements all over the planet result in amounts of data that require modern techniques for handling and storage. As the different data are highly inhomogenous, it seems to be impossible to build-up a consistent common data model. So there is vital necessity for a coordinated construction of a network accessible meta database (MDB) to store information about the available data. Moreover, scientists shall be able to retrieve the underlying data directly, no matter where it is stored physically.

For this purpose at Potsdam Institute for Climate Impact Research (Potsdam-Institut für Klimafolgenforschung e.V., PIK) a meta database (Climate and Environmental Data Retrieval and Archiving System, CERA) was developed and implemented in cooperation with the Deutsches Klimarechenzentrum (DKRZ, Hamburg), the Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI, Bremerhaven), and, in the first phase, Forschungszentrum Karlsruhe (FZK).

Fig. 1: Interaction between AFRI and databases.
Additionally, a Flexible Retrieval Interface (AFRI) for spatial data has been developed. It is platform independent and capable for data retrieval in the Internet as well as locally without the need to set up an Internet server. Furthermore, it can easily be adapted to any in-house database in geosciences.

2. The CERA Database Structure

Some years ago, several German geoscience institutes have joined for a collaboration on development, implementation, and linkage of the meta databases at their sites to give their meta data a common structure and to make it mutually accessible.

The main aims of the common MDB are:
- to make accessible all geographical and other global change relevant data held inhouse to all employees and collaborators of the institute,
- to give transparent access to the corresponding data of the other involved institutes, together with quality information about it, by online-linkage of the databases,
- to enable the investigator to survey all available data and to assess its quality and reliability,
- to coordinate the storage of data and to avoid redundancies,
- to enable metadata interchange with national and international institutions by means of a distributed MDB,
- to lower the budget for development, tools, user interfaces, installations.

To meet the needs of very different types of institutes the highest possible flexibility of the data structure is required. The configuration has to be open, as the evolution of necessities will require an evolution of the meta data model. Furthermore, it has to meet the main exchange standards for spatial meta data (NASA-DIF, FGDC-CSDGM etc.) as well as functional standards (IEEE).

Fig. 2: CERA is divided into core, modules and local extensions.
Fig. 3: The CERA Blocks

- **PUBLICATION**: Any publication related to the data can be stored in this block, together with the form of presentation of the data. Publication types include, e.g., the publication of the data itself or papers that make use of the data.

- **COVERAGE**: This block contains information on the volume of spacetime, covered by the data.

- **METADATA ENTRY**: This is the main CERA Block, providing information on:
  - the entry's title
  - its type and relation to other entries, if implemented
  - the project the data belong to
  - a summary of the entry
  - a list of keywords related to the data
  - the DIF title of the data

- **ATTRIBUTES**: The content of this block describes the topic, the data gives information about, its unit, and its aggregation in space (e.g., data given per country, per square degree) and time (weekly, daily data).

- **SPATIAL INFORMATION**: Information on the coordinate systems used.

- **CONTACT**: Information on contact persons and contact institutes related to the data, as e.g., distributor, investigator, and owner of the copyright.

- **STATUS**: Here some status information is given as, e.g., on data quality and processing that has been applied to the data.

- **DISTRIBUTION**: Information on how data is distributed is provided in this block, including access restrictions, data format, and fees.
In 1996 and 1997 the common data structure of the MDB was developed and released as CERA 2.3. The CERA-2 database concept is highly flexible, as the necessary relational schemes (CERA Core = 42 tables) are separated from those, only used by few institutes (fig. 2). The latter are contained in modules (two modules, 16 tables up to now) that can be attached to the CERA Core table group, if they are applicable for the respective institute. The module table groups contain detailed information on, e.g., the way to access the data or the order in that the data are stored. Institute specific tables can be supplemented at any site, but in a way, that does not interfere with the basic CERA structures and definitions. Furthermore, local extensions to CERA can contain site specific information, not provided for external internet access.

The core tables of CERA are divided into eight CERA Blocks (fig. 3), each containing the information of a certain theme, such as access, coverage, quality, or distribution information. About half of the tables are value lists to fill pull down menus. This facilitates for the user the understanding of the tables’ meaning and helps to keep the entries consistent.

At PIK a loading layer of tables was installed to buffer the input to CERA until it has been checked for clearness and consistency.

A variety of SQL tools, that are running on CERA, are available by download from the internet. User interfaces to read, update, and enter data on basis of ORACLE Forms can be accessed in the same manner. Easily existing data (files, tables in DBMS) can be integrated into the CERA scheme.

3. The AFRI Retrieval Interface

To integrate the CERA MDBs and other geoscience databases of PIK and collaborating institutes into one flexible and easy-to-use information system, AFRI, A Flexible Retrieval Interface for spatial data written in Java, has been developed at PIK since 1997. AFRI provides intuitive retrieval of different databases as well as vivid presentation of query results, offering features like network ability, platform independence, and flexible configuration features, as well as a comfortable and dynamic graphical user interface including an Interactive Digital Atlas (IDA).

3.1 Flexibility

One of the key aspects of AFRI is the demand to maximize usage bandwidth and to minimize porting costs and maintenance burden. AFRI is designed to be both platform independent and network enabled to allow the system to run on different hardware and operating systems in a location independent manner.

Written entirely in Java, AFRI can be downloaded as applet via the World Wide Web (WWW) and run on the the client's machine, offering the user an increased level of client side interaction. Being World Wide Web (WWW) enabled, AFRI can take advantage from this wide-spread net infrastructure and provides scaleable access from world wide use (Internet) to inhouse use (Intranet). Alternatively, AFRI can be run locally, thus avoiding the need for an internet server.

The communication between applet and database(s) is designed as three-tier-system (fig. 4). The applet sends retrieval requests to a communication server that manages the required database interactions and transfers the query results to the applet. Like AFRI, the communication server is written in Java, allowing to be run platform independent and to address different database management systems just by including the relevant JDBC (Java Database Connectivity) drivers.

Facing the need for the integration of new databases, as well as of changes in existing databases' structures, and for generating an appropriate graphical user interface (GUI) for each
database included, AFRI incorporates far-reaching server side configuration abilities, allowing to handle these tasks without any need for reprogramming. A database table containing information about databases to be queried, relevant attributes, and about the desired retrieval interface is used to set up AFRI's appearance and behaviour; the system looks up all relevant structural information at runtime and provides an adequate GUI.

Dynamically reflecting the state of the underlying databases, AFRI ensures that changes in database structure will also be present in the user interface. To increase loading performance, AFRI also can be switched to run in 'trust' mode, extracting all relevant information from an automatically generated setup file.

3.2 Comfortable and Dynamic Retrieval GUI
Since a software's acceptance highly depends on its human-computer interface, AFRI provides an intuitive GUI (fig. 5) to relieve the user from entering tedious retrieval requests in order to specify queries.
The system offers a collection of comfortable query components, that form a suited user interface for each included database according to the setup mechanism described above. Each query component, allowing comfortable use of well known graphical input components like checkboxes, buttons, menus, sliders or textfields, can be regarded as a filter, defining a subset of the underlying database's data, including thematic, temporal and spatial aspects, as well as the possibility to specify values for all attributes not addressed by the other components.

Dynamic browsing functions for selectable attributes provide with information on the available data items, thus supporting the definition of meaningful queries. The thematic query component provides a n-staged hierarchy of selectable keywords, which - depending on selections made on higher stages - are looked up directly in the current state of the database.

To support highly intuitive spatial queries, as well as the representation of spatial query results (see 3.3), AFRI makes an Interactive Digital Atlas (IDA) available. IDA provides comfortable and flexible usage through complete mouse control and interactively displays area names and geographic coordinates, as well as browsing informations - like the number of measurement station entries stored in the selected database, according to the map area the mouse points on.

Fig. 5: Screenshot of AFRI's spatial query component including IDA
Furthermore, IDA allows to include different map hierarchies, i.e., administrative and river basin hierarchies. The user is able to navigate towards higher level and more detailed maps as well as to zoom into a map. Geographical areas can be selected simply by clicking or by freely defining an area, using a rubber band; the bounding coordinates of a selected area are automatically transferred to AFRI's spatial query component.

AFRI leaves it completely up to the user, which query aspects he wants to refine. The different query components' input can be flexibly combined to define the desired query, supported by client side input controls and user guidance.

3.3 Intuitive and Interactive Result Presentation

Query results can be presented in different ways depending on the semantic of the data retrieved. All textual results are displayed by the AFRI ResultViewer, a scrollable spreadsheet allowing sorting, re-ordering of columns, and mouse-driven selection of rows.

![Fig. 6: Screenshot of the SpatialResultViewer](Image)

For the interactive visualisation of query results that have a spatial representation, the AFRI SpatialResultViewer (fig. 6), incorporating IDA, is made available. For example, a query resulting in a set of measurement stations is represented on the appropriate map using different graphic symbols for different kinds of measurement stations. Since IDA is used to
provide the maps for the graphical result display, features like zooming, navigating towards other maps, as well as interactive feedback on area names and geographic coordinates, are supplied. Furthermore, the AFRI SpatialResultViewer allows to select items for up to three attributes - e.g., type of station, measurement variables, measurement frequency - and shows or hides the relevant station symbols; moving the mouse over a station's symbol on the map displays further information.

Stations can be selected by clicking their graphical symbol or by using a rubber band; the relevant symbols are highlighted. Since AFRI establishes a communication mechanism between the spreadsheet and the SpatialResultViewer, selections made in one tool are also present in the other. This feature allows to easily recover the relevant entries of selected station symbols in the spreadsheet, as well as to select stations with certain properties - e.g., an altitude between 100 and 150 meters -, in the spreadsheet and find them highlighted on the map.

To employ increased flexibility in point data visualisation, AFRI again takes use of it's setup mechanism and allows to associate the values of up to three attributes of each database with a collection of graphical symbols, different colors and sizes. AFRI looks up these informations at runtime, compares it to the retrieval result and generates the appropriate symbols on client side.

Prototypically, an interface for visualisation of measurement data, using web enabled serverside visualisation tools, has been developed. Future developments will focus on the integration of a more general interface for data representation, including client side and server side visualisation, HTML-documents and other applets.

4. Conclusions
Currently, the development and installation of the first stage of AFRI and CERA at PIK have come to an end and the practical inhouse usage has started. The coupling of the different databases is proceeding. The development of better visualization facilities for numerical data is planned as well as automated control of data integrity, as most data are spread over the different working groups of the institute. In the future, flexible tools will enable us to access and survey better the increasing amounts of data. Coming versions of AFRI and CERA will go on this way.

5. References
CERA-2 Central Page: http://www.pik-potsdam.de/dept/dc/e/SDM/cera
Toussaint F., M. Lautenschlager & M. Reinke: CERA-2 - ein raumbezogenes Daten- und Metadatenmodell; in Ralf Kramer & Friedel Hosenfeld (eds.): Heterogene, aktive Umweltbanken; GI-Workshop Vilm 1998, Metropolis-Verlag, Marburg 1999
Wrobel M.: AFRI/IDA - Ein flexibles Retrieval-Interface für heterogene, raumbezogene Daten; in Ralf Kramer & Friedel Hosenfeld (eds.): Heterogene, aktive Umweltbanken; GI-Workshop Vilm 1998, Metropolis-Verlag, Marburg 1999