PNL-SA-16875
CONF-891053--8

INTEGRATED COMPUTER-BASED TOOLS TO FACILITATE ENVIRONMENTAL MONITORING, ASSESSMENT, AND RESTORATION

PNL-SA--16875

DE90 002732

P. J. Cowley
J. C. Brown

October 1989

Presented at the
28th Hanford Symposium on
Health and the Environment
Richland, Washington
October 16-19, 1989

Work supported by
the U. S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

DISCLAIMER

Integrated Computer-Based Tools to Facilitate* Environmental Monitoring, Assessment, and Restoration

Paula J. Cowley
James C. Brown
Computational Sciences Department
Pacific Northwest Laboratory
P. O. Box 999
Richland, WA 99352

Abstract

The vast amounts of scientific and technical data gathered during the process of environmental monitoring, assessment, and restoration make the effective application of computer-based tools essential. This paper describes how an integrated set of automated resources is being applied at the Hanford Site to effectively manage the data gathered during site characterization and monitoring and to provide an environment that facilitates analysis and assessment. The Hanford Environmental Information System (HEIS) has upgraded our previous monitoring and characterization capabilities and represents significant advances in our ability to manage, retrieve, and display data. An integrated database, a geographic information system (that allows data to be displayed on a map), and support graphics allow the user to generate spatially related visualizations and perform data extractions. This allows the user to gain insight quickly and easily by obtaining a complete picture of the data of interest. HEIS also provides the capability to share the information with other software environments for further analysis and assessment.

Background

For many years, there has been an extensive site characterization and environmental monitoring effort at the 560-square mile Hanford Site in southeastern Washington State. Hanford is organized into 78 operable units with a total of over 1500 active or inactive waste sites. The types of samples taken at each operable unit may include atmospheric, sediment, sieve, soil-gas, biotic, surface water and ground-water samples. Each sample will be tested for numerous contaminants. Given this scope, data management which facilitates environmental monitoring, assessment, and restoration is a significant challenge. We recognized the need:

- for automation to support the well over 100 billion characters of data to be accumulated.
- to access data that already existed,
- and to provide new capabilities to view and analyze the data.

In response to these needs, we initiated work on the Hanford Environmental Information System (HEIS).

Functional Requirements

From the beginning, HEIS was intended to be an information system rather than simply a database. Although the database forms the heart of the system, there are additional capabilities including a geographic information system (GIS), support graphics, user-friendly access software, database query and report generation software, networking, distributed processing, data and system security, documentation, and user support. The system is intended to meet the needs of both the operational user, responsible for scheduling, data collection, data processing, and quality assurance, and the scientist who

will use the system to do environmental monitoring, assessment, and restoration planning. The HEIS database is centrally available as a quality assessed data source to promote the effective sharing of data by the entire user community.

Approach

Creation of an information system as large and complex as HEIS requires an interdisciplinary team of computer scientists and subject area specialists like hydrologists and geologists. The team must be committed to design and planning and have sufficient resources, time, and management support. We used prototyping in each of the subject areas we implemented. This has given us and the user community additional knowledge to implement the final, production system before large amounts of data start flowing into the system. We are committed to prototyping because it permits the iterative clarification of user requirements. We are also committed to computer-aided software engineering (CASE) tools. For example, the current HEIS data model and its documentation consumes several million characters of disk storage. This is too complex to keep track of manually.

The HEIS Database

Environmental restoration (ER) data is an important resource for remediation decisions. HEIS uses the Oracle (tm) (Oracle Corporation, Belmont, CA) database management system (DBMS) to store, manipulate, and retrieve the data. The relational database gives us the ease of use, ease of understanding, flexibility, and security necessary to effectively do data management.

The HEIS database integrates several existing databases as well as subject areas not previously accommodated. Figure 1 lists the major subject areas in the database. The technique of Entity - Relationship modeling is used to create data models of each known data subject area. Within each subject area the data is partitioned into entities. An 'entity' is a place, thing, event, or a concept about which one is interested in keeping information; (ie, site, well, facility, sample, analytical result). These entities are related to each other in different ways and for different reasons. This is indicated by 'relationship' lines from one entity to another. A 'relationship' defines the role that one entity has with respect to another (ie. there can be one-to-one and one-to-many relationships). Figure 2 is an example of an Entity - Relationship diagram showing the primary entities of interest for the vadose zone and geologic data subject area. The current HEIS Entity-Relationship model consists of 140 entities, 161 relationships and 913 attributes (fields). The management of the HEIS model is done using the KnowledgeWare (tm) CASE (Computer-Aided Software Engineering) tool (KnowledgeWare, Atlanta, GA).

While several existing databases at Hanford contain information about the same topics, data did not always agree from one database to another. It might be possible to extract similar data from two databases and come up with two different answers. By integrating the data from these diverse systems into HEIS, we can provide one version of the truth to everyone. When necessary changes are identified and made, everyone has access to the changes.

Oracle provides protection from system failure and can prevent unauthorized access. System failures have the potential to destroy major portions of the database, particularly if the data most recently entered has not been backed up on disk or tape. Oracle provides facilities for backing-up and restoring the database. It also has facilities for logging entries to a journal file that can be used to recover data that has not yet been 'committed' to system storage.

Unauthorized access to data can result in accidental or malicious alteration or destruction of data. The DBMS provides protection at several levels: the database, relation (table), record, field, view, project, and account level. It also provides read and write access protection on the different levels.

Naming standards and conventions are major issues when integrating heterogeneous databases. Synonyms and homonyms for data element names existed because these databases were created by different groups and sometimes by different contractors. This makes the communication of information between the individual databases difficult. It also makes the integration of those databases more difficult. To address these issues a Data Administration Board sub-committee consisting of individuals from several programs and contractors was formed. The results from the sub-committee are being formalized and written policies and guidelines are being developed. Standard names, abbreviations, and naming conventions are being adopted that will be used in all ER work at Hanford.

The system can aid in meeting regulatory requirements and preparing for potential litigation by providing data access and traceability. HEIS has facilities for automatically generating the reports and tables required by

the regulators. The database contains directory information for supporting documentation and keeps traceability data such as when samples were taken and analyzed, who performed the analysis, the method used during analysis, and evaluations of the quality of the analytical result.

The Geographic Information System and Graphics

The geographic information system (GIS) is a very important component of HEIS since almost all aspects of the system are related to a location on a map. GIS can be thought of as a "smart map" that allows the user to start with a base map and select from a series of overlays to develop a customized computer-generated map. The HEIS base map consists of roads, trails, railroads, water bodies, and administrative boundaries. Overlays of Hanford's 1500 waste sites, ground-water monitoring wells, geologic boreholes, soil gas sampling locations, geophysical survey lines, and facilities are being developed. Over time, existing overlays will be enhanced and more overlays will be created. Using the GIS, the user can zoom in and out, generate contours, change colors and symbols, query the database, and have the results of the query displayed on the map. The GIS also provides the capability to create map overlays from analysis and remediation modeling. Since drawing a map involves much processing and output to a screen, GIS software is ideally suited to an engineering workstation where the workstation's processor is tightly linked to the display via hardware and is dedicated to either a single user or a small group of users.

In addition to displaying data on a map, users want to display data graphically in both two and three dimensions. Some graphics are oriented spatially while others show changes over time. Many graphics and drawing

software packages exist with their capabilities ranging from simple to sophisticated. HEIS will support several packages that can be integrated with other HEIS components. We are currently using DISSPLA (tm) (Integrated Software Systems Corporation, San Diego, CA), a set of graphics software subroutines that can be incorporated into software we develop. We have written software that retrieves data from the database and calls DISSPLA subroutines to generate plots. We are using AutoCAD (tm) (Autodesk, Inc., Sausalito, CA), a computer-aided design package, to generate lithology and well construction diagrams. Through its AutoLISP programming language, AutoCAD can automatically generate well diagrams from a file of data extracted from the database.

The HEIS Computing Environment

Adequate compute power is essential to store, organize, and manage the current estimates of well over 125 billion characters of HEIS data. We are using a Sequent S27 (tm) UNIX-based multiprocessor computer (Sequent Computer Systems, Inc., Beaverton, OR) and the Oracle database management system for the centralized HEIS database. A significant advantage of a multi- processor computer like the Sequent is its expandability - more processors, memory, and disk storage can be added as additional data, functions, and processing capabilities are added. Development in a UNIX-based environment also makes HEIS more portable.

While the database is centralized, other components of the system are distributed as shown in Figure 3. HEIS is taking advantage of an extensive computer network allowing many of the terminals, personal computers, engineering workstations, minicomputers, and mainframes to communicate.

Distributed processing with networking puts computing power close to the user and allows that user to take advantage of the appropriate computing environment for the work being performed. For example, those doing data processing can use inexpensive terminals for data entry and running data processing jobs. Those doing geographic information system (GIS) and graphics work can use high-powered engineering workstations to transparently retrieve the most current data from the HEIS database computer. Those doing data analysis can use their favorite software and hardware to either access the HEIS database computer remotely or to download data sets for stand-alone processing.

While this type of computing environment provides greater flexibility, system integration must allow the distributed components of the system to work together in a manner transparent to the user. For example, the GIS requires access to the central HEIS database. Rather than the user having to go through a process of data extraction and downloading, the GIS accesses the database directly through the network. Taking advantage of the computer industry's significant strides in providing user-friendly access methods, HEIS provides forms-based data entry, menu-driven user access software, data browsing facilities, 'ad hoc' querying, and a multi-windowing environment on the engineering workstations. User support is provided through documentation, on-line help, training, and consultation.

Lessons Learned

Currently HEIS is oriented towards environmental monitoring and assessment. HEIS is not intended to meet computer requirements for project management, cost accounting, milestone tracking, or health effects tracking.

It is recognized that computer-based tools are required to meet these needs and planning is underway. Since HEIS provides a powerful and functional computational environment for these activities and integration with the scientific and technical data is extremely desirable, it is hoped that other automated systems for these activities will be developed in such a way that they can be integrated with the existing HEIS capabilities.

As our prototyping experience shows, flexibility is essential. The system must evolve as functions, data requirements, software, and hardware evolve. Although we are using the state-of-the-art now, we must keep pace with technological advances. The current computing environment makes expansion possible.

We expect that the system will continue to be expanded and updated to meet user needs as long as environmental restoration activities continue. Flexible data structures are required to support evolving needs. HEIS must be able to grow as required to accommodate the vast amounts of data being gathered.

Generation of high quality data to facilitate environmental monitoring, assessment, and restoration is expensive. Implementation of an automated information system adds to this cost. However, for a site as large as Hanford, the need for an effective, integrated information system has been recognized. HEIS is a necessity, not a luxury.

* This work was supported by the U. S. Department of Energy under contract DE-AC06-76RLO 1830.

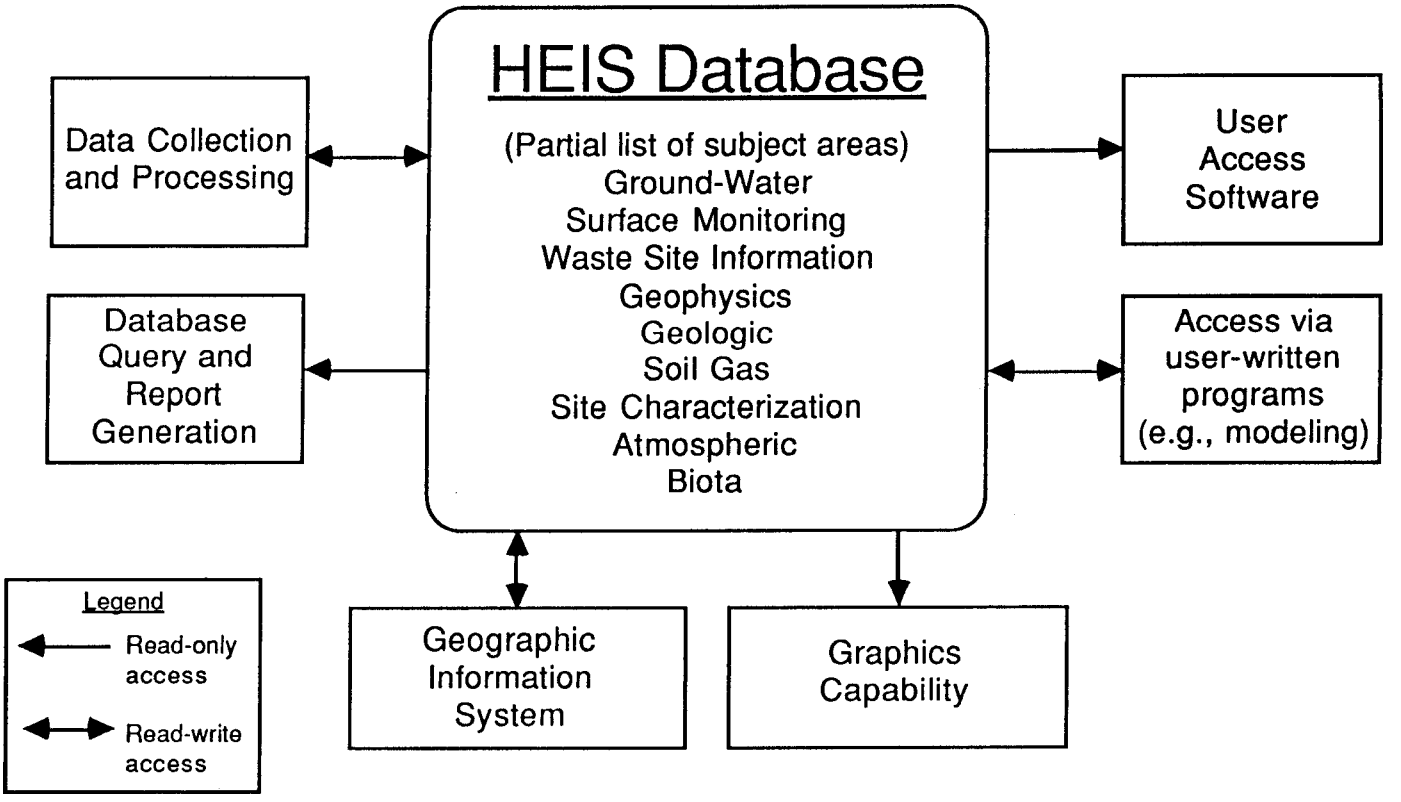

Figure 1: Software Components of HEIS

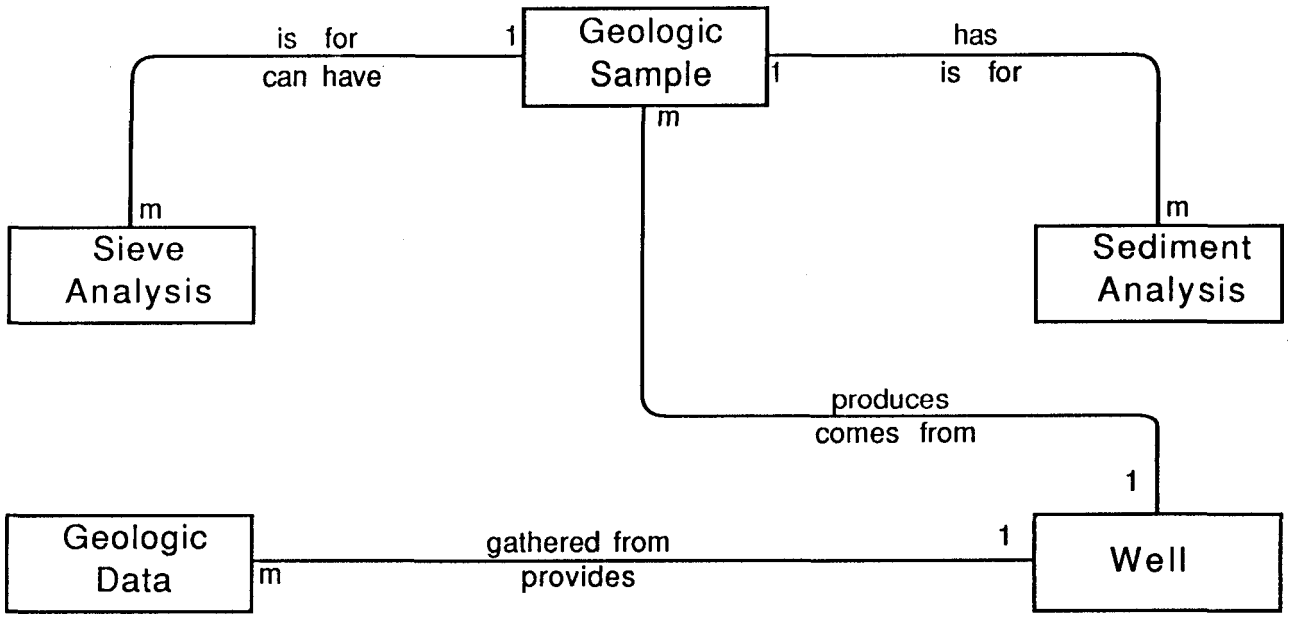

Figure 2: Entity / Relationship Diagram

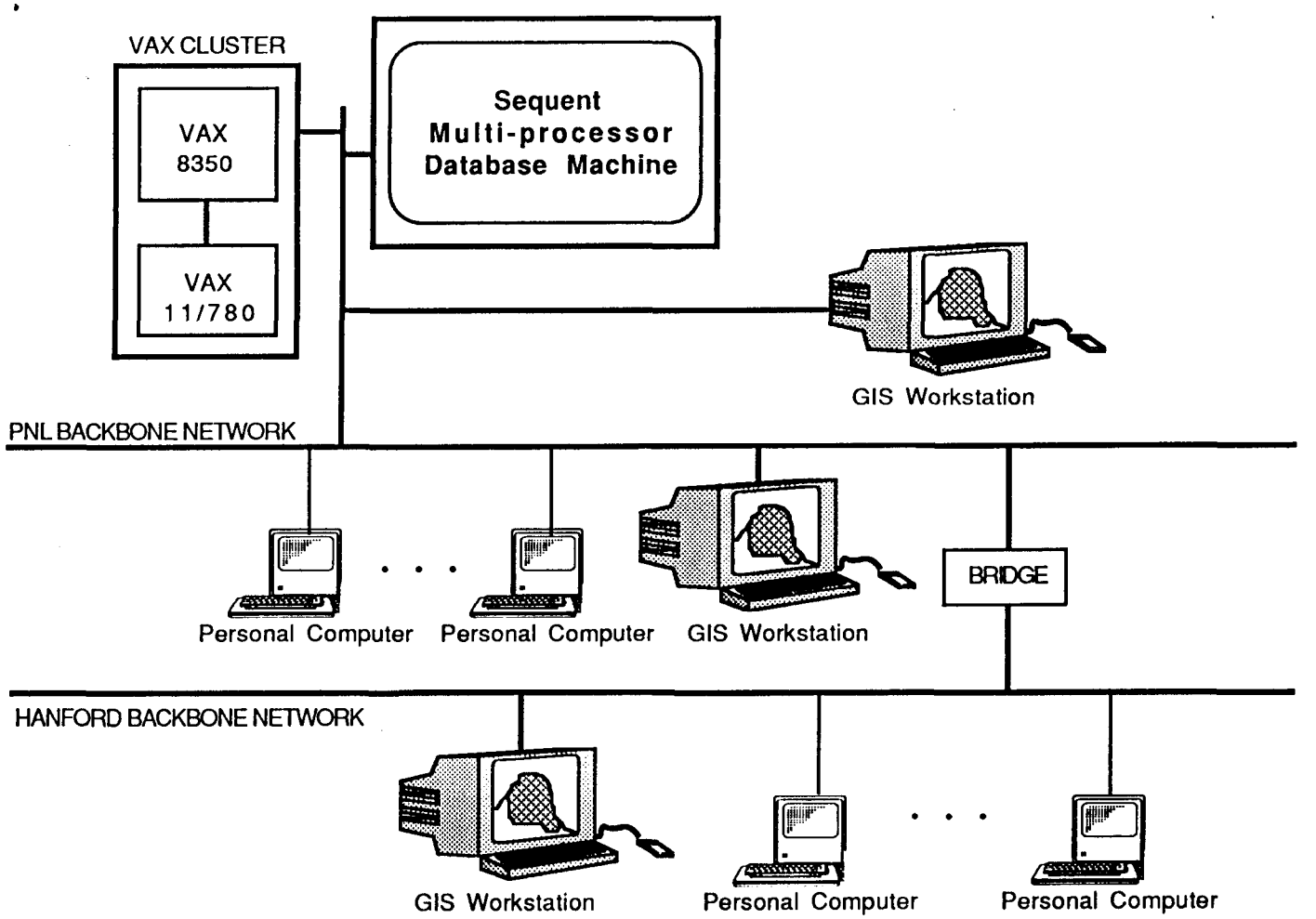

Figure 3: HEIS Hardware & Networking Configuration