1.0

1.1

1.25

1.4

1.6

1.8

2.0

2.2

2.5

2.8

3.2

3.6

4.0

4.5

5.0

5.6

6.3

1.0.1

DISCLAIMER

IMPLEMENTING AND OPERATING THE HANFORD
ENVIRONMENTAL INFORMATION SYSTEM (HEIS)

P. J. Cowley[a]
M. R. Schwab[a]
R. D. Fox[a]

March 1994

Presented at the
Cost Efficient Acquisition and Utilization of Data in
the Management of Hazardous Waste Sites Conference
March 23-25, 1994
Washington, D.C.

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

Pacific Northwest Laboratory
Richland, Washington 99352

(a) Westinghouse Hanford Company
     Richland, Washington

PNL-SA-22430

Conf-940355--1

IMPLEMENTING AND OPERATING THE
HANFORD ENVIRONMENTAL INFORMATION SYSTEM (HEIS)

Paula J Cowley, PNL
Pacific Northwest Laboratory*
PO Box 999
Richland, WA USA 99352
Telephone: (509) 375-2282

Michael R. Schwab, WHC[b]
Richard D. Fox, WHC

ABSTRACT

In the process of performing environmental restoration at the 560-square mile Hanford Site in southeastern Washington State, vast amounts of scientific and technical data are being generated from sampling taking place all over the Site. This paper provides an overview of the lessons we have learned in designing, implementing, and putting into operation a computerized system named the Hanford Environmental Information System (HEIS), which is being used to manage the Site's environmental characterization sampling data. Topics discussed in this paper include helping the Site adapt to a data management culture, the advantages of electronic data over paper data, issues of data validation and defensibility, being a resource to the user community (including the regulatory community), managing and tracking data changes, integrating data from multiple programs, providing configuration control for data and software, setting priorities for software development, and developing a baseline for on-going funding to maintain the infrastructure for the information system.

DESIGNING AND IMPLEMENTING THE HEIS

The Hanford Environmental Information System (HEIS) has been designed and implemented to store, manage, and aid in the analysis of Hanford environmental sampling data. These data are crucial to Hanford Site's mission of environmental cleanup, both now and for the future. The HEIS is a repository for these data and includes tools for efficient data processing, viewing, reporting, and depiction.

As user needs and requirements change over time, additional capabilities are being added to the HEIS. The system currently handles data dealing with ground-water, surface water, soil, biota, well construction, tank characterization, and gaseous effluents. Capabilities for sample scheduling and tracking have been implemented. For each type of data added to the database, the HEIS software development team performs a user requirements analysis, a

formal data modeling process, an implementation and documentation phase, and internal and independent testing phases before the software is made available to the user community.

The core software development team for the HEIS consists of a group of Pacific Northwest Laboratory (PNL) computer scientists, some of whom have extensive experience in environmental information management. These scientists work closely with scientists and engineers in the various disciplines that are covered by the HEIS, such as geology and hydrology. Significant changes to the HEIS software or database structure, proposed by either the user community or the software developers, are presented to the HEIS Configuration Control Board (CCB) for approval. The board consists of senior staff from the Westinghouse Hanford Company (WHC) and PNL who represent the various programs involved in the HEIS. Invariably, the proposals emerging from the CCB are stronger than they were when originally presented. We believe that one of the strengths of the HEIS is the teaming of a strong computer science capability with strong subject area expertise.

OPERATING THE HEIS

The HEIS database contains over 2.8 million records stored in 160 tables. Major components of the system have been operational for three years. Users can query and access data from 178 different screens. The database was initially populated by converting data from an older database covering ground-water monitoring efforts since the 1950s. The conversion included data from the ground-water environmental surveillance program and ground-water data collected under the RCRA (Resource Conservation and Recovery Act of 1976)[1]. New capabilities were added to the HEIS to handle data on ground-water, soil, biota, and well construction, in accordance with CERCLA (Comprehensive Environmental Response, Compensation and Liability Act of 1980)[2]. Data are going into the system at an increasing rate with over 188,000 chemical and physical analysis ground-water records being added during calendar year 1992. Over 272,000 additional records were added in 1993. As of March 1994, the database contains over 1.1 million ground-water result records, 51,000 soil result records, 20,000 biota records, and 74,000 hydraulic head records. We are currently in the process of converting an existing surface environmental surveillance database to the HEIS. Data from 20 of the Hanford Site's 177 waste tanks will be entered during fiscal year 1994.

A number of lessons were learned during these years. In the rest of this paper, we present some of these lessons in the hope that others will be able to avoid some of the problems we have encountered.

Adapting to a Data Management Culture

We did not anticipate the huge culture change that data management would require of the Hanford Site. The scientists and engineers at the Hanford Site were used to gathering the data they needed, analyzing it, preparing a report from it, and storing the data in a file cabinet. If another person wanted to use these same data, he or she had to determine who had these data and then quiz that person on the assumptions that were made when the data were gathered. This led to an environment in which it was often easier and even more cost-effective to take new samples and generate new data. Because of the secrecy associated with the Hanford Site's production missions, data sharing was not encouraged. Now the Hanford Site has a single mission: to clean up the Hanford Site. With this mission and the openness associated with it, data sharing is encouraged and even expected.

However, it has been difficult to get the Hanford Site's scientists and engineers to accept the philosophy of treating data as an asset to be valued and shared, instead of as a commodity that is used once and then discarded. We have made significant progress, but it

has proven to be more difficult than anticipated to get people to accommodate the requirements of a data management infrastructure. One lesson learned is that people who generate data rarely understand the requirements of automated data management.

Implementing automated sample and data tracking in the HEIS is one area where we are making progress in changing the culture. When only a few hundred samples and their associated results were being tracked, a project manager could handle the tracking function manually or with a simple spreadsheet. When thousands of samples and data must be tracked, the task becomes overwhelming without the HEIS. We have implemented the capability to put sample information into the HEIS when the sample is taken and sent to the laboratory. When the results return from the lab and are entered into the database, the HEIS can automatically mark their receipt, so it is easy to determine what is still outstanding. This kind of data can also be used to determine laboratory performance. Some programs at the Hanford Site are already using similar sample scheduling and tracking mechanisms. The CERCLA program has utilized some sample tracking capabilities of the HEIS, but other portions have yet to be put into full operation.

The HEIS has the ability to generate a unique number for each sample regardless of the media. Users can request a block of sample numbers for an activity and use those sample numbers as needed. Unused numbers can be returned for reassignment. Initially, some staff resisted using the HEIS numbers. They were accustomed to assigning their own numbers. However, significant confusion arose when they were unable to keep their numbers unique. The use of unique HEIS sample numbers has now become a standard way of doing business, even for some programs that don't put their data into the HEIS.

Tracking down missing and/or incomplete data has been a difficult and time-consuming process. We worked hard to change the culture so that it has become part of the normal process for staff taking any field data to deliver those data for entry into the HEIS. There is significant interest in using computers in the field to capture data electronically rather than transcribing it later.

Electronic versus Paper Data

Most laboratories that analyze samples for Hanford Site can deliver analytical results in electronic data files on diskette. In our first years of operation, 10% of data we received was not in electronic form; but processing this paper-based data took roughly 70% of the data processing staff's time. Our experience has shown that it takes at least 20 times the staff resources to key in and verify non-electronic data, compared to loading and verifying data received electronically. Besides the extra time needed to manually key data into the system, we found that 100% verification of manually entered data was required, since 1-5% of the records entered manually contained errors. With electronic data loading, we merely check to make sure that the process ran to completion satisfactorily, and perform occasional spot checks when new versions of loader software which have been thoroughly tested are put into production. We are currently working with analytic laboratories to finalize a new format in which electronic data can be delivered for organics, inorganics, radchemistry, and wet chemistry analysis. This format, called the Format for Electronic Analytical Data (FEAD), was designed based on our experience in dealing with such electronic data deliverables.

Issues of Data Validation and Defensibility

The HEIS user community includes the Washington Department of Ecology (WDOE) and the Environmental Protection Agency (EPA). These two organizations are the "regulators" who perform an oversight function to assure that the U. S. Department of Energy (DOE) and its contractors are working in accordance with the *Hanford Federal Facilities Agreement and Consent Order*[3] commonly known at the Tri-Party Agreement (TPA), which was signed by the Department of Energy, WDOE, and EPA in 1992. The regulators have access to the HEIS database and use it as a tool in performing their oversight function. The database is set up so that all users have access to the data, while only those responsible for the data have the ability to change it. A special security scheme has been put in place to allow different programs to "own" their data so that only they can insert and modify it.

Until recently, the regulators had access only to "validated" data. Validated data have been reviewed against a set of criteria using a systematic process including data editing, screening, checking, auditing, verification, certification, and review. The validation process takes time to perform, and at first we did not receive the validated results electronically. Consequently, a large backlog of data accumulated to which the regulators did not have access. We had to work extremely hard to make this data available to the regulators. For the last six months of 1993, we eliminated the backlog by making available to the regulators 60,000 backlogged records per month. We are now receiving most of the validated data electronically, and getting validation results into the database in a timely manner. Our inability to get validation results into the database in a timely manner was the main motivation for granting the regulators access to unvalidated data.

We are also rethinking the value of data validation. The validation process does not look at the analytical results in the context of the location from which the sample is taken and what the results say about contamination at that location; but only in the context of what happened to the sample at the laboratory. During the site characterization phase, we are only trying to quantify how contaminated a site is, not trying to prove that the site is clean enough. Future plans call for us to perform the costly data validation process only when we are trying to establish that a site has been cleaned up, not when we are trying to establish how contaminated it is.

We are still struggling with the issue of what constitutes "legally defensible" data. We have tried to design the system so that data can be used in a court of law. We know that a decision to "close" a site and consider it cleaned up will be based largely on our data. But what is sufficient and necessary? What extra steps must be taken when data must remain available for the 30 years we anticipate it will take to clean up Hanford Site? We know all data must be of known quality, and we include a significant amount of "metadata" (i.e., data about the data) that describes how the sample was obtained and analyzed. We also know data must be traceable to its source. Once again, metadata has been included to support this requirement. In addition, "change records" are generated to support traceability whenever a result record is modified. Some of these changes are the addition of more metadata. The database will undoubtedly evolve as we come to better understand the legal issues.

Being a Resource to the User Community

In addition to accurately capturing data for current and future use, it is important that the user community have easy and user-friendly access to data. This is particularly important for a system like the HEIS where data are highly technical, naturally complex, and difficult to interpret. We have tried to store reference data (e.g., regulatory limits and laboratory contractual limits) in the database to aid the user in understanding the data. We have

discovered that users must receive two kinds of training in order to utilize the database effectively:

- They must be trained in the mechanics of using the database. This includes logging on, launching applications, and exercising those applications.
- They must be trained in the meaning of data, how data are organized, why the database is organized the way it is, and what to do and not do to correctly interpret any data they access.

We originally tried to take a relatively informal approach to training because we felt the HEIS software was easy to use. We have learned that what is simple to a computer scientist or a highly experienced user is not so simple to an inexperienced user. We are finding that the users require more training than anticipated and have developed a more aggressive, formal training program, particularly as we bring new software online.

We had expected that users would do data extractions themselves, but many have neither the time nor the inclination to extract data and format reports. We have developed a cadre of experienced staff who receive user requests, perform data extractions, and deliver finished reports. We expect that our training program will encourage users to do more of this work themselves, but are prepared to continue to support them if needed.

We continue to look for ways to make data more accessible to users. We have created special data extraction routines, purchased user-friendly report generation software, and are investigating how to apply the new generation of client-server computer technology (e.g., to allow access to data from a personal computer running special applications under Microsoft Windows 3.1[4]).

We are constantly reassessing how we can better address the needs of these inexperienced users. What makes this particularly challenging is the fact that the users are unable to articulate their needs until they have had some experience with the system.

Managing and Tracking Data Changes

Most database activity involves adding new records to the database. The majority of the records are generated when new samples are taken and analytical results are received. Once records are in the database, they may change as more data become available or mistakes are corrected. The HEIS has the ability to track changes to the records, where legal defensibility of these data is likely to be an issue. Whenever one of these records is changed, we generate a copy of the record as it existed before the change, and we record the reason for the change. These copies are stored in separate database tables, so there is no chance that a user will confuse a previous version of a record with the current version.

Integrating Data from Multiple Programs

It is common at the Hanford Site for different organizations to generate similar data. For example, ground-water data is generated for the CERCLA program, the RCRA program, and the site-wide environmental surveillance program. Of course, each program has its own priorities and requirements. The HEIS database is intended to serve as an integrated source of environmental data for all Hanford Site programs. In order to achieve the data integration we are striving for, we have appointed a representative from each program to the HEIS Configuration Control Board (CCB) as a voting member. The CCB sets HEIS priorities, reviews database and software designs, and makes decisions on system directions. The CCB members have worked effectively to develop solutions to their common problems and to accommodate the needs of the various programs. In several cases, a proposal presented by one program has been adopted by others. By openly communicating and working towards the

common goal of making the HEIS a useful data repository, we have increased communication among the programs.

Providing Configuration Control for Data and Software

Data Entry Verification. Configuration control has been handled largely with administrative controls of the HEIS data entry process. Electronic data entry is verified by spot checks against the hard copy. Hard copy data entry is verified with a 100% self-check, using specialized database printout forms. These data entry verification forms are generated by HEIS, and present the data in the same format as the data entry forms received from the lab or field. A 100% data entry peer review is also performed after the self-checks are completed, using the same data entry verification forms.

Data Change Control. We are implementing a data configuration control procedure that mirrors our software configuration control procedure for the HEIS. A data change request tracking system, with Data Change Request (DCR) forms, has been in operational testing. A team was established as an interim HEIS data configuration control board, to allow operational testing of the configuration control requirements prior to formalizing the process. The team, comprised of data custodians for the major programs within the HEIS, is trying to establish the number and qualifications of those who are allowed to make changes to HEIS data. There appear to be some differences of opinion on the need to limit the number of data owners. We are currently defining what constitutes a suitable set of "data subject material experts" to review proposed "global" data changes in the HEIS database. We are also establishing consensus data definitions within a data dictionary. We expect to complete the requirements for a data configuration control procedure this year, and to implement a data configuration control board within the HEIS data owner community.

Software Configuration Control. For the first several years of development, we used a software code management system called RCS (Revision Control System) to track all versions of our software. When we began converting our screen management software to Uniface[5], we discovered that, while Uniface presented many advantages over our old software, it offered no way to track changes to individual pieces of software. Therefore, we developed an on-line method of tracking changes without tracking individual pieces of code. Not only has the on-line system reduced the amount of paperwork, it also serves as a tracking mechanism for design changes. All software developers believe that the new system has reduced a significant tracking burden that was not particularly cost-effective, and has not deprived us of any important capability. We are continuing to track some software using RCS, particularly algorithms that are used to calculate summary values since those algorithms can change over time.

Maintaining the HEIS Infrastructure

A system as large and complex as the HEIS, with its Configuration Control Board, extensive user manuals[6], requirements analysis, and formal software development process, is expensive to establish and maintain. The system itself grows as more user requirements are identified. Hardware and software maintenance costs are on-going. There are also the on-going operational costs of keeping the HEIS-specific software up to date as advances are made in computer technology. Strong management support is required if a system like the HEIS is to be successful. The HEIS's multi-contractor project team is constantly working to ensure that our resources are being used effectively for the benefit of the entire HEIS user community. Management has become increasingly aware of the role of the HEIS in cleaning up the Hanford Site. We feel much more comfortable that the on-going resources the HEIS

requires will be forthcoming, because we now have the management support we need, unlike our first years when management had a hard time understanding the value and role of the HEIS.

CONCLUSIONS

The environmental restoration mission at the Hanford Site is projected to require from 30 to 50 years to complete. We anticipate creating and managing many gigabytes (a gigabyte is one billion characters) of data. The HEIS has been designed to support this huge data management task by integrating diverse software tools and data structures into a computer system that meets the challenges of the Hanford Site's environmental mission. We have made significant progress in changing the data management culture for the Environmental Restoration Program. Written procedures have been changed to require the use of HEIS sample numbers, to request laboratory results electronically, and to deliver field data to the HEIS team for data entry. Validation results are being delivered electronically and in a timely manner. HEIS is being used routinely to generate reports and answer questions. Many staff have been trained. Funding for HEIS is secure because we have management support. We have taken major steps to meet our user community's needs, but the job is not yet complete. Our focus has been on capabilities to store and manage data. We are now beginning to turn our attention to the end-user component of our user community.

Our objectives are to provide the entire user community with a resource that brings the right data in the right form at the right time to the right place at the right cost. We have found that the HEIS has been the most effective means of implementing this data management vision.

REFERENCES

1.   "Resource Conservation and Recovery Act of 1976," 42 USC 6901 et seq.

2.   "Comprehensive Environmental Response, Compensation and Liability Act of 1980," 42 USC 9601 et seq.

3.   ECOLOGY, EPA and DOE, "Hanford Federal Facility Agreement and Consent Order," 2 vols., Washington State Department of Ecology, U.S. Environmental Protection Agency, and U.S. Department of Energy (1992).

4.   Windows 3.1 is a trademark of Microsoft Corporation.

5.   Uniface is a trademark of Uniface B.V., Amsterdam, the Netherlands.

6.   DOE-RL, 1994a, *Hanford Environmental Information System (HEIS): Volume 1 User's Guide*, DOE/RL-93-24-1, U.S. Department of Energy, Richland Operations Office, Richland, WA.

DOE-RL, 1994b, *Hanford Environmental Information System (HEIS): Volume 2 Operator's Guide*, DOE/RL-93-24-2, U.S. Department of Energy, Richland Operations Office, Richland, WA.

DOE-RL, 1994c, *Hanford Environmental Information System (HEIS): Volume 3 Subject Area Reference Manual*, DOE/RL-93-24-3, U.S. Department of Energy, Richland Operations Office, Richland, WA.

DOE-RL, 1994d, *Hanford Environmental Information System (HEIS): Volume 4 Well-based Subject Areas*, DOE/RL-93-24-4, U.S. Department of Energy, Richland Operations Office, Richland, WA.

DOE-RL, 1994e, *Hanford Environmental Information System (HEIS): Volume 5 Surface-based Subject Areas*, DOE/RL-93-24-5, U.S. Department of Energy, Richland Operations Office, Richland, WA.

DOE-RL, 1994f, *Hanford Environmental Information System (HEIS): Volume 6 Soil Subject Area*, DOE/RL-93-24-6, U.S. Department of Energy, Richland Operations Office, Richland, WA.

DOE-RL, 1994h, *Hanford Environmental Information System (HEIS): Volume 8 Tank Characterization Data Subject Area*, DOE/RL-93-24-8, U.S. Department of Energy, Richland Operations Office, Richland, WA.

END

DATE FILMED

9/10/94