



PNNL-17663

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Status Report on the Creation of a Preliminary Data Model and Dictionary for a New Petrologic Database

RD Mackley
JA Serkowski
GV Last

September 2008



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/ordering.htm>



This document was printed on recycled paper.

(9/2003)

Revised Status Report on the Creation of a Preliminary Data Model and Dictionary for a New Petrologic Database

Rob D. Mackley

John Serkowski

George V. Last

09/24/2008

1.0 INTRODUCTION

Petrologic, mineralogy, and bulk-rock geochemical data on geologic material of the Hanford Site have been collected for over 60 years. Typically, these data are organized in project-specific and partially-published sources (e.g. hard-copy reports). This information provides a foundation and technical baseline for understanding and defending conceptual models on the geologic framework, geochemical conditions and processes, reactive transport of contaminants, and subsurface remediation. Xie et al. (2003), using multivariate statistical analyses, found a distinction between Hanford and Ringold Formation sediments based on mineralogy and geochemistry. They recommended the continued collection, management, and use of these data for geologic characterization.

In fiscal year 2008 (FY08) Fluor Hanford, Inc. (FH) requested that the Pacific Northwest National Laboratory (PNNL) develop and update a prototype database for this information that would be suitable for transfer to a new Virtual Library module. This work is one of several inter-related activities being conducted under the Remediation Decision Support (RDS) function of the Soil and Groundwater Remediation Project to ensure maintenance of technical baseline data needed to resolve technical issues and support efforts to estimate contaminant migration and impacts.

This document describes the status of current efforts to include this geologic information into a relational database and presents the draft data structure of the prototype petrologic database. For the purposes of this document, this prototype database is called MinChem.

2.0 STATUS AND APPROACH

The original approach for this task was to create a set of data structures (tables and views) modeled after industry standards and structures currently present in HEIS (BHI 2002). These structures were envisioned to be implemented as a standalone database for transfer to a new Virtual Library module.

A preliminary/draft design of the data structures was presented to a petrologic database working group on June 16, 2008. Data users, generators, and technical experts in the field of database design and administration of Hanford Site data were in attendance. They provided important feedback that has helped to shape the current design. Chief among the comments and suggestions was to more closely model the petrologic database

after HEIS, to the extent that once complete, the data in the petrology database could be loaded directly into existing HEIS structures with minimal modification. Recent and potential changes to the sample tables and other attributes in HEIS (such as those incorporated for the borehole geophysical log data), may now allow for more direct mapping of the petrologic data into the new HEIS data structures.

In its current form, MinChem includes some structures just as they appear currently in HEIS, as well as new structures that could be incorporated into HEIS at the time these databases were merged. This will allow us to 1) test the compatibility of HEIS with petrologic data, and 2) facilitate efficient migration of MinChem into the HEIS tables (this is currently planned for FY09).

3.0 REQUIREMENTS

Information for samples/analyses from boreholes and non-borehole sites (e.g. outcrops) will be included. The analytical methods include, but are not limited to, optical and electron microscopy, x-ray diffraction, x-ray fluorescence, and electron probe microanalysis. Results can be presented as quantitative, semi-quantitative, ranking, order of abundance, and presence/absence. The database structure must be designed in such a way as to allow a direct interface with other databases such as HEIS and the Hanford Borehole Geologic Information System (HBGIS).

The petrology database structure must accommodate certain data associated with the collection and analysis of geologic samples from the Hanford Site. Sample information must include project information, such as sampling purpose and point-of-contact, a three-dimensional spatial location of the collection site, the stratigraphic horizon for the sample, if known, and the field details, such as name of person performing the sampling, sampling date, and methods used. The structure must be flexible enough to allow the inclusion of unlimited sample metadata that may vary from one study to the next.

For each sample, the structure will house one or more analytical results from a variety of mineralogical and petrological methods. Each result must be linked to a specific analytical laboratory, a laboratory preparation and analytical method, and a reference to where the data are published. A result may also have an uncertainty and quality level associated with it.

The final production database must be simultaneously accessible to Hanford Site researchers using existing network connections. In the present prototype status, the database does not have a user interface or connection to allow access from beyond the PNNL firewall for end users. These will be developed in the future as part of the steps toward making this a production database.

4.0 DESIGN AND IMPLEMENTATION

4.1 Strategy

We looked at several strategies for tackling the storage of mineralogic, petrologic, and geochemical data. The most obvious solution is to incorporate it into an existing database. HEIS (BHI 2002), with data structures for storing sediment contaminant analysis, appears to be a suitable and appropriate repository. The data fits the basic HEIS model of sample, analysis method, and result. Even semi-quantitative data, such as abundance ranking, can be accommodated without undue creativity. However, our initial evaluation found that some data fields/attributes such as sub-sampling and sample preparation were not fully captured in the HEIS structure. However, recent changes to HEIS should remedy that situation. So, based on recent feedback from the working group, we developed a design centered around existing and proposed HEIS tables. Our intent is to test, populate, and exercise the preliminary data model outside of HEIS and then, after implementing any necessary modifications, migrate the data to HEIS.

Another strategy we had examined was to use an existing off-the-shelf or public domain system. While there are some references on the Internet to mineralogical and petrological databases and structures (e.g., <http://www.earthchem.org/earthchemWeb/index.jsp>), there do not appear to be any turnkey applications that meet the above-mentioned requirements. However, insights from the structures used by others have been incorporated into the design proposed in this document.

The design presented here addresses the requirements specified above and acknowledges the likelihood that MinChem might eventually be incorporated into HEIS. This latter condition, in many cases, dictates the table structures and key fields of the proposed design so that the transition to HEIS, if it occurs, will be straightforward.

We preserved HEIS field names whenever equivalences occurred, despite the fact that they do not always have consistent syntax. For example, the fields Method_Name, Lab_Code, and Con_ID in the HEIS RESULT table are all coded fields and, eventually, should all be consistently named with either a '_Code' or '_ID' suffix.

4.2 Design Overview

MinChem is a multi-user, relational database that resides on a networked, Windows-based computer. The database management system is Microsoft Access 2003. Data are stored in 14 normalized tables. The tables are divided into two categories: 1) structural clones of HEIS database tables, and 2) tables exclusive to MinChem. The intent of this division is to make future migration to HEIS straightforward. Access to the data is currently available only through the native Access interface and a basic data retrieval form. Additional interfaces will be developed with guidance from users.

4.3 Table Structure

Figure 1 shows the MinChem table structure. Tables with names including the prefix ‘mc’ (e.g., mcAttribute) indicate they are specific to MinChem; those with a prefix of ‘heis’ (e.g., heisResult) are structural clones of existing HEIS tables. The structure centers about the heisSample and heisResult tables, which house the sample information and analytical result data, respectively. The other tables provide additional information on and validation for fields in these two tables where codes are used for brevity. The heisValid_Codes table provides additional field value lookups.

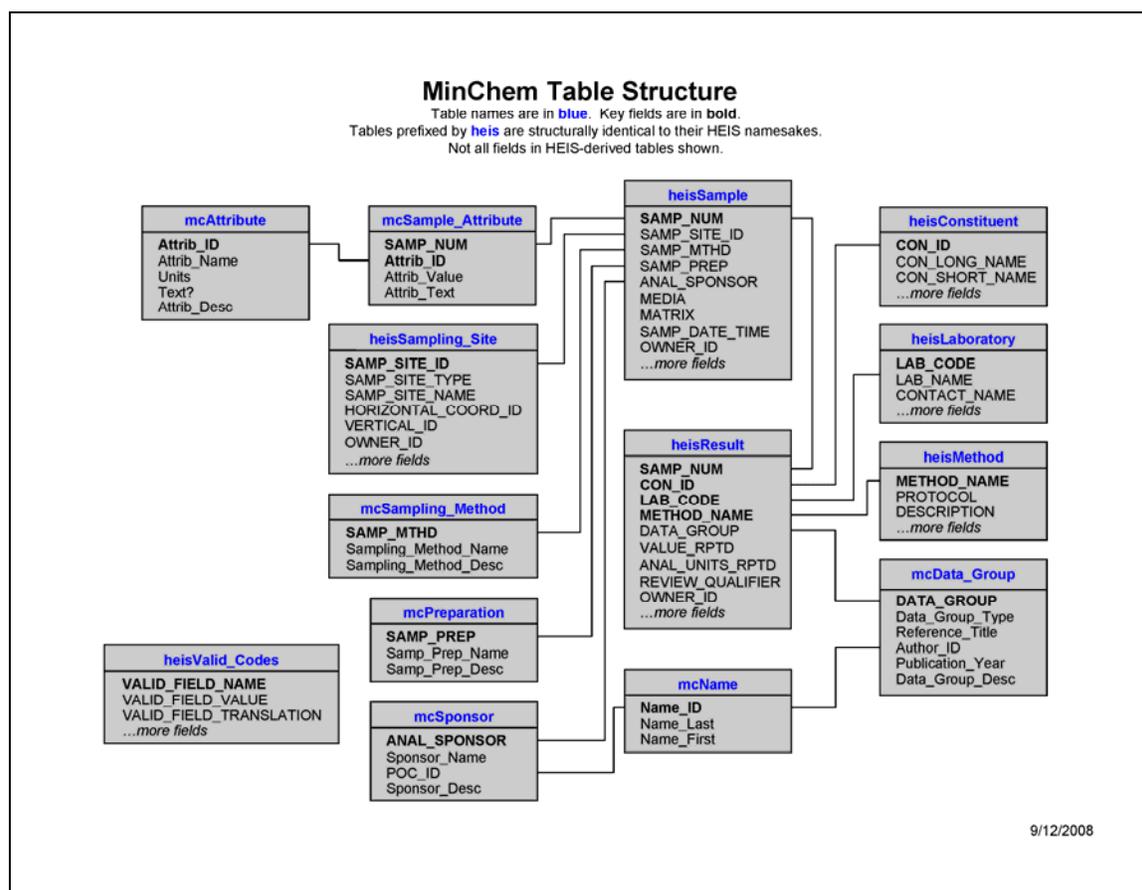


Figure 1. MinChem table structure.

The mcSample_Attribute table provides a mechanism to attach an indefinite number of attributes to any sample. An attribute can be any numeric or textual value associated with the sample, other than an analysis result. We provide this capability in anticipation of the inevitable bits of descriptive information that certain projects will find useful but are impossible to predict ahead of time. For example, a researcher may wish to record something that is not normally relevant to most projects (e.g., the wind speed at the time of sample collection) and then later have the ability to query their data on this information.

The table heisSampling_Site is based on a prototype for a revised version of the existing HEIS SAMPLING_SITE table. Currently, samples in HEIS are associated with either a WELL_NAME or, if not collected from a well, a SAMP_SITE_ID. In this proposed HEIS structure, which has not been formally documented yet, every sample will have a SAMP_SITE_ID associated with it. Moreover, sampling sites may have hierarchical associations with other sampling sites. For example, a sample may be extracted from the deepest few centimeters of a well core sample collected during drilling. The several-centimeter-long sampling site is assigned a unique code and described in terms of its location with respect to the drill core. The entire core sample, which may be a meter long, may be assigned to another sampling site that defines the depth range from which the core was obtained in the well. The core is associated with the sampling site for the entire well, and the well, in turn, may be associated with a sampling site defining the area of a waste facility. Thus, the small rock sample may be associated to the drill core, the well from which the drill core was extracted, and the waste facility at which the well is located. This permits very powerful location-based queries for samples.

4.4 Data Dictionary

Table 1 is an explanation of the fields in the data dictionary tables. Tables 2 through 7 are the data dictionary for each new MinChem table (those with table names prefixed with 'mc'). Key fields, those that uniquely define a record, are shown in **bold**. The complete data dictionary for the HEIS tables may be found in BHI 2002.

Table 1. Key to data dictionary .

| Column | Description |
|-------------|--|
| Contents | Brief description of what is in the table |
| Field | Name of table field |
| Type | Data type of field: T = text, N = numeric, I = index (long integer), B = boolean |
| Description | Description of field – contents and limitations |
| Example | Examples of data in this field |
| Lookup | Name of VALID_FIELD_NAME or lookup table if coded field |

Table 2. Table mcAttribute

| Contents: Each record is a unique attribute type that may be associated with a sample. Attributes are sample descriptors for which there is no structure in table heisSample. In particular, attributes are parameters that a user might wish to use as selection criteria in a query. An attribute can have either a text value or a numeric value, but not both. | | | | |
|---|------|--|-----------------|----------|
| Field | Type | Description | Example | Lookup |
| Attrib_ID | I | Code for sample attribute name | 1, 2 | |
| Attrib_Name | T | Unique identifier for the sample attribute | “Size Fraction” | MC_UNITS |
| Units | T | Units of attribute | “Rank”, “% Vol” | |
| Text? | B | True if attribute is text; false if number | True, False | |
| Attrib_Desc | T | Description of attribute | “sample block” | |

Table 3. Table **mcData_Group**

| Contents: The data group is a reference to the source of the analysis results. This might be a published document or a computer file. | | | | |
|--|------|---|-------------------|-----------------|
| Field | Type | Description | Example | Lookup |
| DATA_GROUP | T | Code of form: MCREF $nnnnnnnn$, where $nnnnnnnn$ is a unique number. | “MCREF0000001” | DATA_GROUP_TYPE |
| Data_Group_Type | T | The type of media on which source data reside | “Document” | |
| Reference_Title | T | Title of document or name of file | “The Rock Report” | |
| Author_ID | I | Code for name of primary author or file owner | 1, 2 | |
| Publication_Year | N | Year in which document was public or file created | 1998 | |
| Data_Group_Desc | T | Description of reference | “PNNL-17663” | |

Table 4. Table **mcName**

| Contents: This table contains the first and last names of people referred to in the database. | | | | |
|--|------|--------------------------------|--------------|--------|
| Field | Type | Description | Example | Lookup |
| Name_ID | I | Unique ID for name | 1, 2 | |
| Name_Last | T | Person’s last name | “Serkowski” | |
| Name_First | T | Person’s first name or initial | “John”, “AJ” | |

Table 5. Table **mcPreparation**

| Contents: Sample preparation is the modification of a sample between the time it is collected and the time it is analyzed. There may be some overlap with sample collection and sample analysis, as both these activities may also modify the sample to some degree, so preparation is reserved for those actions not addressed in the sampling method or analysis method. Typically, procedures such as sieving, sorting, drying, crushing, etc., are considered sample preparation. | | | | |
|--|------|--|----------------|--------|
| Field | Type | Description | Example | Lookup |
| SAMP_PREP | T | Unique short name for sample preparation | “Sieve2” | |
| Samp_Prep_Name | T | Long name for sample preparation | “Sieve 2 Mesh” | |
| Samp_Prep_Desc | T | Description of sample preparation | “sieved ...” | |

Table 6. Table **mcSample_Attribute**

| Contents: This table contains attribute data for a sample. An attribute record can have either a text value or a numeric value, but not both. | | | | |
|--|------|--|-------------|-------------|
| Field | Type | Description | Example | Lookup |
| SAMP_NUM | T | Sample number associated with attribute | “MC0000001” | heisSample |
| Attrib_ID | I | Code for sample attribute name | 1, 2 | mcAttribute |
| Attrib_Value | N | Numeric value if attribute type is Text? = False | 3.14 | |
| Attrib_Text | T | Text string if attribute type is Text? = True | “sG” | |

Table 7. Table **mcSponsor**

| Contents: The sponsor is the project associated with the sample. This is usually the funding project. | | | | |
|--|-------------|--|--------------------|---------------|
| Field | Type | Description | Example | Lookup |
| ANAL_SPONSOR | T | Code of form: MC <i>nnnn</i> , where <i>nnnn</i> is a unique number. | “MC0001” | |
| Sponsor_Name | T | Descriptive name of sponsor | “BWIP” | |
| POC_ID | I | ID for sponsor point of contact | 1, 2 | mcName |
| Sponsor_Desc | T | Description of sponsoring project | “Basalt Waste ...” | |

4.5 Interfaces

A formal user interface is not specified for MinChem at this time. Data retrieval, processing, and output is handled using the native Access interface, which includes a complete set of query, report, and programming tools.

Likewise, there are no specialized data entry programs. Data may be loaded by direct manual input into tables, cut-and-paste operations, and native Access data import processes.

4.6 Implementation

MinChem table structures and relationships are stored in the Microsoft Access 2003 relational database management system. This software was chosen because it is relatively easy to use, has multi-user capability, interfaces with HEIS, and is widely available to Hanford Site contractors.

The data tables reside in a single Access MDB file on a networked, Windows-based server computer. For those within the PNNL firewall, access to the database is made by linking to the networked tables through a local Access database. In addition to the table links, each user’s local Access database may include queries, forms, reports, and programs for retrieving, processing, and outputting the data. These local tools may be developed by the user, or provided by the MinChem development team.

Once the prototype database has been tested and the need for the production database arises, easy access to the database for all Hanford end users will be implemented (likely by incorporating the MinChem data and structures into HEIS).

5.0 SECURITY AND CONFIGURATION CONTROL

As a prototype, MinChem has more lenient configuration control than a full production database.

Programs and interfaces written by MinChem developers will be tested and all tests documented. Changes to existing programs will require retesting of relevant parts. All programs will be internally documented and numbered (using the standard major.minor scheme) to distinguish versions.

Data loading will only be done by MinChem development team members. All other users will only have read access to the database. All analysis result data will be loaded with a “U” (unvalidated) REVIEW_QUALIFIER until a validation procedure is established and the data reviewed.

Database backup will be done daily using PNNL’s Connected DataProtector service.

6.0 USER GUIDE

As user interaction with MinChem data is currently limited to the native Access interface, this user guide focuses on connection to the database and a brief overview of some useful Access commands.

6.1 Obtaining Access to MinChem

Staff wishing to use MinChem must contact the MinChem technical administrator, Rob Mackley, at 371-7178. They will receive a static (most-recent snapshot of the database) copy of the MS Access (.mdb file).

6.1.1 Viewing and Querying Data

To view data in a table, double click its name in the Table Objects window. You may scroll through the table using scroll bars at the right or sort the records on a field by

clicking on a field name and pressing the  or  button. An easy query is done by selecting a field value on which you wish to query and pressing the  button. This retrieves all records containing the selected field value. To show all the records again, press the  button.

More advanced queries can be made by creating a Queries Object. See Access help for details.

7.0 SUMMARY

A preliminary database has been developed that will allow mineralogy and bulk-rock geochemical information to be managed under configuration control and facilitate electronic querying. The database is currently developed in Microsoft Access. Each field and table has been described in a data dictionary. The current state of the prototype database reflects the original intent to deliver a standalone petrology database for transfer to a new Virtual Library module. Recent and upcoming developmental changes taking place in HEIS data structures might make it more feasible than previously recognized to include the petrology database directly into HEIS rather than create a separate Virtual Library module.

8.0 FUTURE PLANS

Efforts planned for FY09 include:

- 1) Create a plan and timeline for the migration of MinChem to HEIS.
- 2) Develop a data validation procedure to assure the quality of data entered into MinChem.
- 3) Continue entering data from the Excel data file of Xie et al. (2003) into MinChem, as well as legacy and newer data. These data will come from both electronic and non-electronic (hardcopy) sources.
- 4) Design a graphical user interface (GUI) that will serve as a front-end portal to querying and downloading data. Such an interface would use existing HEIS GUIs wherever practical.

9.0 REFERENCES

- BHI, 2002, *Hanford Environmental Information System – Detailed Design Specification*. BHI-01639 Rev. 0, Bechtel Hanford Inc., Richland, WA.
- Xie YL, CJ Murray, GV Last, and RD Mackley. 2003. *Mineralogical and Bulk-Rock Geochemical Signatures of Ringold and Hanford Formation Sediments*. PNNL-14202, Pacific Northwest National Laboratory, Richland, WA.