Guidelines for Spectral Library Generation, Documentation, Dissemination, and Archiving

PDS Geosciences Node
Washington University in St. Louis

CRISM Spectral Library Working Group

Input from Mars Spectroscopy Workshop Participants, June 2002

October 31, 2002

DRAFT 3

1. PURPOSE AND SCOPE

The purpose of this document is to provide guidelines for the structure, content, and documentation of laboratory spectral libraries of solid materials that would maximize use of the data and provide archives that are PDS-compliant. A spectral library is defined to be a set of spectral observations of some set of targets or samples, together with documentation and correlative data that provide additional information of use in analyses of the primary observations. The primary observations could, for example, be acquired in transmission, reflection, or emission. The correlative data, including X-Ray fluorescence, X-Ray diffraction, and Moessbauer, and petrographic data provide added scientific benefit during analyses of spectral data. The overall objective of spectral libraries generated according to the PDS guidelines is to archive and maximize their long-term use by the planetary community.

This document begins with an overview for generation of spectral libraries, followed by example library user scenarios. The overview and scenarios are then used to provide guidelines for the generation, structure, and documentation of the libraries, followed by exemplary data base schema and user views that illustrate the advantages of delivery of spectral libraries using the guidelines.

2. OVERVIEW

The approach used in this document is to provide guidelines for generation and documentation of spectral libraries that are flexible, user-friendly, and PDS-compliant. The intent is to have PDS Geosciences Node personnel work with library suppliers to ensure that the guidelines are met. The spectral libraries can be either posted locally at suppliers institutions and linked to the PDS Geosciences site or posted directly that site. In either case, the libraries are archived at the PDS.

Uses of spectral libraries range from fundamental characterization of materials to comparisons with spaceborne and telescopic observations. For example, the Mars Express OMEGA and Mars Reconnaissance Orbiter CRISM instruments are both imaging spectrometers that will operate in the VNIR spectral range (~400 to 5000 nm). One mode of data analysis is to process the spaceborne data to spectral reflectance and to then compare the observational spectra (manually or automatically) to library spectra to find matches. Good library spectral matches to spaceborne data, based on least squares

or other approaches, provide a starting point for consideration of what materials are on the surface of Mars. To optimally conduct these comparisons, the normally higher-quality data in the spectral libraries would typically be degraded to match the observational data. Furthermore, the library samples must be well understood in terms of what controls their spectral properties, which implies well characterized samples (e.g, inclusion of major element analyses and Moessbauer spectra).

3. OPTIONS FOR CONTRIBUTING SPECTRAL LIBRARIES

Several options are available for delivering PDS-compatible spectral libraries:

- **PDS Node.** Data are provided in an agreed-upon format to the PDS Geosciences Node for archival in a central repository. Data provided in this manner will be stored centrally and distributed search mechanisms developed by the Geosciences Node. All data will be clearly attributed to the original donor. This option allows sites to contribute libraries without taking on responsibility for storage and dissemination.

- **Local node, PDS database.** Data are stored at the contributing site using a database designed and distributed by the Geosciences node. The product will include a user-friendly input interface developed by the Geosciences Node and designed in cooperation with contributing sites. A standardized search interface will also be provided. This option allows sites to retain control of their own dataset and make it publicly accessible on their own servers, with a minimum of technical effort.

- **Local node, local database.** Data are stored at the contributing site using a locally-developed database. Search interfaces must be PDS-compliant. This option allows sites to maintain full control over the storage and dissemination of their data, as well as the software used for storage. Sites choosing this option will need some local technical expertise in database design and development. Appendix 2 provides some support for this option, and the Geosciences Node will provide limited additional consultation with individual sites choosing this option.

Local nodes will be linked to the Geosciences Node so that users can easily find all spectral library sites within the node structure.

4. DATABASE DETAILS

Data may be provided for input into the database in various formats, including flat (text) files and Microsoft Excel spreadsheets in to-be-specified formats. A Web interface for input of additional data may also be available.

Once in the database, information will be searchable using a variety of criteria, using a standardized interface yet to be defined. Details of database search features will be determined in accordance with user requirements, with a goal of accommodating a range of different search methods.

A detailed list of data items for inclusion in the database appears in Appendix 1. The kinds of data to be incorporated are as follows:

- **Basic sample information**, including sample name, collection locality, classification (e.g., mineral, water/ice, or rock), along with additional classification-specific items (such as, for a mineral, formula and particle size).

- **Primary spectral information,** including details about the analyses performed, such as the methodology used and the name of the researcher who made the measurements. It will also include instrumentation information, relevant environmental parameters (e.g., temperature, pressure, and purge gas), and lighting and viewing geometry information. The spectral data will be contained in attached data files. If a single sample has been analyzed multiple times, details of each analysis will be retained.

- **Correlative analyses,** including other analyses (e.g., major element analyses, magnetic properties, and Moessbauer spectra) performed on the sample. Multiple analyses of a single type may be stored for a single sample. The results of these analyses will be viewable in flat-file format. Some summary information (such as magnetic properties) will also be present in searchable format. Full information from compositional analyses will also be searchable, since this information is considered particularly fundamental.

- **Analyst information.** Information about contributors, including name, institution, and contact information.

APPENDIX 1: Contents of a Spectral Library

This section is intended as a list of the data items that may be present in the database. It need not represent the format in which they will be entered or viewed. Within each category, items in boldface are required for all records; other items are optional except as noted.

General Information

(Information stored for every sample.)
Sample ID (automatically generated)
Sample Name (defined by supplier)
Origin (e.g., Earth, Moon, Meteorite, or Synthetic)
Classification (one of: mineral, rock, unconsolidated, ice)
Sample Collection Locality *(*e.g. "Mauna Kea, Hawaii", "Mojave Desert")
Supplier (e.g. R. Morris)
Sample Storage Location (e.g. Washington University)
Reference/Contact (primary publication reference, website, or name)

Classification-Specific Information

(*Each sample is classified into one of these categories and has the information associated with the category.)*

Classification

1. Mineral (pure or nearly pure, see standard definition of mineral)

2. Rock (a mixture of different components that are consolidated)   {Not all rocks are composed of minerals.}

3. Unconsolidated (a mixture of different components that are not consolidated)

4. Ice (solid volatile, pure or mixture)

Mineral Detail

Mineral Class (e.g. Native Element, Sulfide, Halide, Carbonate, Silicate)

Mineral Family (e.g. Metals, Neosilicate, Phyllosilicate)

Mineral Group (e.g. Olivine, Garnet, Pyroxene, Feldspar, Mica)

Mineral Species (e.g. fayalite, forsterite, biotite, muscovite)

Particle/Crystal Size Range

Mineral Variety (e.g. ferrian diopside, rose quartz)

Comments

Quality Flag

Rock Detail

Sample Name or ID

Rock Type (one of: volatile-rich, volatile-poor (igneous), volatile-poor (other))

Rock Subtype (e.g. mafic/basic, felsic/acidic, evaporite, clastic, biogenic)

Textural Modifier (e.g. porphrytic, granular, foliated, oolitic)

Description

Quality Flag

Unconsolidated Detail

Sample Name or ID

Material Group (e.g. igneous, sedimentary, metamorphic)

Material Subgroup (e.g. mafic, felsic, evaporite, secondary)

Material Species (e.g. tephra, ash, soil, Dirty Dick's favorite crud)

Material Variety (e.g. basaltic [tephra])

Particle Size Range

Particle Size Mean

Particle Size Standard Deviation

Particle Size Skewness

Particle Size Kurtosis

Description

Quality Flag

Ice Detail

Sample Name or ID

Ice Type (e.g. dry ice, water ice)

Ice Formula (e.g. CO_2, H_2O)

Description

Quality Flag

Primary Spectral Data

(*Primary spectral data are meant to be the core observations that define the library, e.g., a series of bi-directional reflectance spectra. These data are stored for each analysis performed on each sample.*)

Analysis ID

Facility

Acquisition Date

Reference/Contact

Instrument (e.g. Nicolet, ASD)

Measurement Type (*bi-directional reflectance, hemispherical reflectance, etc.*)

Calibration mode

Spectral Range

Spectral Sampling Interval

Spectral Resolution

Radiometric Resolution

Illumination type/source

Geometry (*not all geometric parameters will be defined for each analysis, since different viewing geometries may have different parameters associated with them; only those parameters relevant to the geometry used will be required*)

    Type (bi-directional/directional-hemispherical/etc.)

    Incidence angle

    Emission angle

    Phase angle

Environment:

    Pressure

    Temperature

    Purge Gas

    Purge Duration (length of time or until no change)

Comments

Spectrum data filename

Correlative Data

(For each correlative analysis performed on a sample, one of these sets of data—that corresponding to the analysis type--will be stored.)

Correlative Data Types

1. Magnetic Properties

2. Images

3. Mössbauer Spectroscopy

4. Raman Spectroscopy

5. X-Ray Diffractometry

6. Elemental Composition

> *6a. X-Ray Fluorescence Spectroscopy*

> *6b. Electron Microprobe*

> *6c. Other (e.g., ICP)*

8. pH and Eh

9. Thermal Analysis

Magnetic Properties Detail

Analysis ID

Date

Facility

Reference/Contact

Results (data table or image)

Comments

> ...

Imaging Detail

Analysis ID

Date

Facility

Reference/Contact

Results (data table or image)

Comments

> ...

ETC.

Suppliers

(*Intent is to provide contact information for people referenced in the database.*)

Name

Location

Contact address

APPENDIX 2: DATABASE DESIGN DETAILS

The PDS Geosciences Node is in the process of designing a database structure, or schema, that is tailored to spectral libraries. Details are provided here for remote sites that may wish to create their own databases. Remote sites are not obliged to follow this scheme, as long as the chosen format contains the information specified in Appendix 1.

A2.1: OVERVIEW OF RELATIONAL DATABASES

This section gives an overview of the basic concepts associated with relational databases. It may be useful in understanding the concepts and terminology in Section 2.1.

A relational database is structured as a set of interrelated **tables** of data. Figure 1 below shows an example table, containing the **fields** "name," "phone," and "email":

Figure 1: Example Table: People

Name	Phone	Email
Ray Arvidson	(314)935-5679	arvidson@wunder.wustl.edu
Laura Back	(314)935-8555	back@wunder.wustl.edu
Susan Slavney	(314)935-5493	slavney@wunder.wustl.edu

In a database table, some field or set of fields must be specified as the **key**. Every entry in a table must have a unique key. (When the key is composed of multiple fields, the combination of values contained must be unique.) In the example above, "Name" might be declared a key, if we expected only one entry per person (and no two people with the same name) in the "People" table.

A relationship between two tables is based on a field (or multiple fields) appearing in each table, so that values between two tables can be matched.   The relationships we are concerned with may be one-to-one or one-to-many; that is, there may be only one item in the second table corresponding with an item in the first table, or there may be several. Continuing with the above example, we may have a second table titled "Papers" with title, author, and publication of various research papers. The "Name" field in the "People" table could have a one-to-many relationship with the "Author" field in the "Papers" table, with the value "Ray Arvidson" appearing once in the former table and repeatedly in the latter.

A2.2. SPECTRAL LIBRARY DATABASE STRUCTURE

The database will consist of the following tables:

1) **General sample information.** Table 1 contains information available for all samples, such as name and classification, and is keyed on the sample name. (This requires that no two samples have the same name.)

2) **Classification-specific tables.** Since each classification (e.g., mineral, rock, water/ice) may be associated with different specific information ("particle size," for example, is relevant to a mineral sample but not an ice sample), separate tables are needed for the different sets of fields used for each classification. One table (e.g., Table 2a, Table 2b, etc.) will be used for each classification. Each sample appearing in Table 1 will appear once in one of these tables. These tables will be related to Table 1 by the sample name.

3) **Spectral analysis information.** Because multiple spectral analyses are possible for a given sample, there is a one-to-many relationship between Table 1 and Table 3, based on the sample name. (Note that "many" means there may be multiple values in the latter table; it is also permissible for there to be only one.)   The key for this table would be a combination of sample name and spectral analysis number.

4) **Correlative analysis information.** Again, multiple tables (Table 4a, Table 4b, etc.) will be used, since different types of analysis will be associated with different fields. A one-to-many relationship will exist between Table 1 and each of these tables, based on sample name. The key for each table will be a combination of sample name and an analysis number (e.g., XRD analysis number, wet chemistry analysis number, etc.).

5) **Compositional information.**   For each type of correlative analysis that produces composition information, a table (Table 5a, etc.) will be included to store the analysis results.   Each analysis on a sample produces a set of material/percentage pairs (this one-to-many relationship necessitates the separation of composition information into separate tables), and each such pair will be stored as a separate entry in these table. The tables will be related to the tables for the correlative analyses producing this information (one analysis table entry for multiple composition table entries), based on sample name and analysis ID. The key for each of these tables will be sample name, analysis ID, and material. Not every table under (4) will correspond with a table here, since not all correlative analyses produce composition information.

6) **Suppliers**. Each contributing researcher will appear once in this table. The key will be researcher name, and the table will have a one-to-many relationship with the various analysis tables (which will contain, for each analysis, the identity of the researcher who performed it).

Figure 2 is a schematic illustration of the design described above. Each table is shown as a list of fields. (Only a few representative fields are shown; see Appendix 1 for the complete list of information to be included.) Values in italics are keys; arrows

represent relationships between fields of different tables. For simplicity, only one of Tables 2a, 2b, etc., one of Tables 4a, 4b, etc., and one of Tables 5a, 5b, etc. are shown.
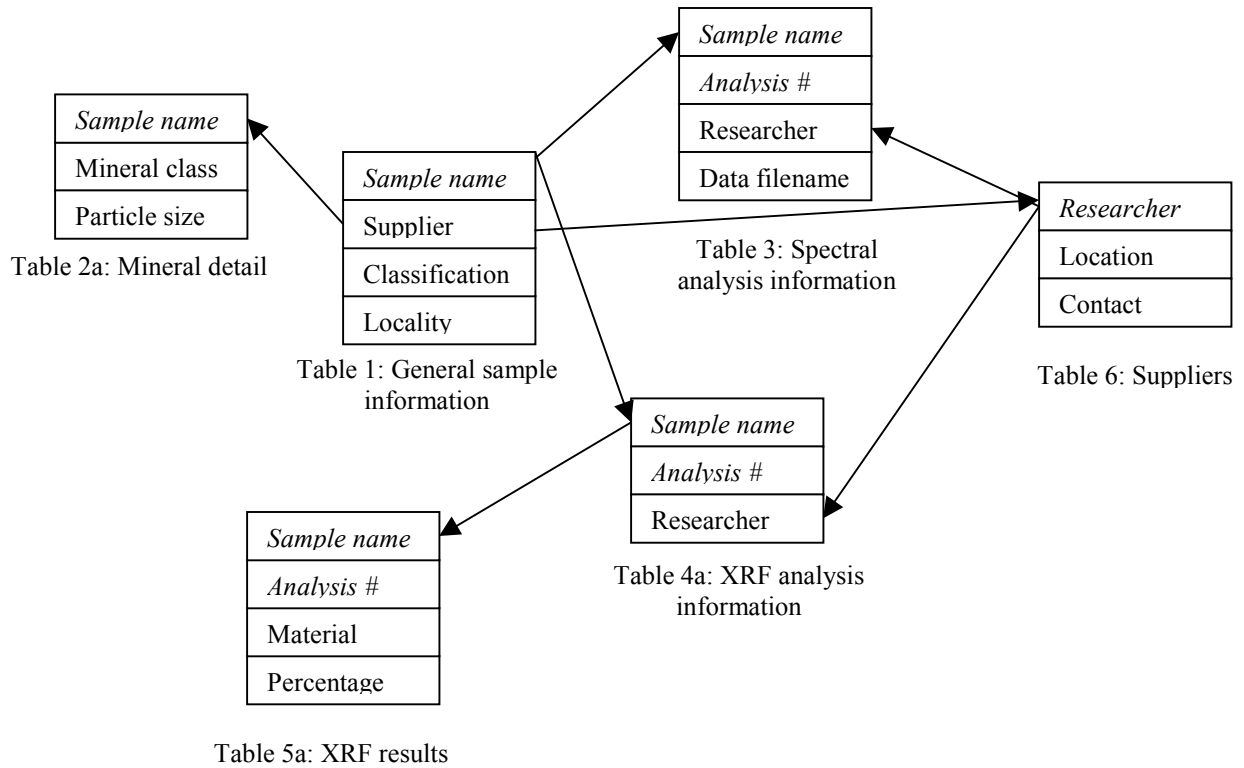

Figure 2: Illustration of Database Schema