The spectral database SPECCHIO for improved long-term usability and data sharing


Remote Sensing Laboratories, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
European Space Research and Technology Centre (ESTEC), 2200AG Noordwijk, The Netherlands

Article history:
Received 28 August 2007
Received in revised form 12 February 2008
Accepted 8 March 2008

Keywords:
Metadata
Hyperspectral Signatures
Software
MySQL
Calibration

Abstract

The organised storage of spectral data described by metadata is important for long-term use and data sharing with other scientists. Metadata describing the sampling environment, geometry and measurement process serves to evaluate the suitability of existing data sets for new applications. There is a need for spectral databases that serve as repositories for spectral field campaign and reference signatures, including appropriate metadata parameters. Such systems must be (a) highly automated in order to encourage users entering their spectral data collections and (b) provide flexible data retrieval mechanisms based on subspace projections in metadata spaces.

The recently redesigned SPECCHIO system stores spectral and metadata in a relational database based on a non-redundant data model and offers efficient data import, automated metadata generation, editing and retrieval via a Java application. RSL is disseminating the database and software to the remote sensing community in order to foster the use and further development of spectral databases.

1. Introduction

Ground-based hyperspectral signatures are collected for the basic investigation of the relationship between physical or biochemical properties and the electromagnetic reflectance of objects and for the calibration, validation and simulation of remote sensing imagery and its data products.

Since the advent of field spectroscopy with the first specifically built portable field instrument appearing in the late 1980s, considerable research on the spectral properties in the VIS/NIR/SWIR (visible, near-infrared and shortwave infrared) electromagnetic spectrum of natural and manmade objects has been carried out. At the same time, much less effort has been spent on the issue of standardisation of the measurement process itself and the systematic collection and interpretation of ancillary data, the so-called metadata. Even less focus has been put on the issues of integrated spectral and metadata storage, efficient and automated methods for data input and formulation of data queries. There is a need for systems that support not only single reference spectra, but also handle the large amount of data resulting from hyperspectral field or laboratory measurement campaigns.

The comparison of spectral signatures between studies is complicated by the many different techniques used for the capturing of spectral field data (Milton, 2004) and the influence of the sampling environment on the measurement. The accuracy of spectral measurements depends on a clear definition of what is being measured and on the conditions under which it is being measured (Milton et al., 2006).

Utilising data from other studies requires an assessment of the data quality and suitability of the data set for the given task. The key factor for data sharing is thus the existence of metadata, which support the broad and
long-term use and interpretation of scientific data (Michener, 2000). The lack of metadata can render previously collected data useless for new applications. Given the scenario outlined above, an organised, shareable and non-redundant storage of spectral data and associated metadata is an important step towards better data quality, long-term usability and the possibility of data sharing between researchers. It is paramount to the success of such a storage system that the data input is highly automated, thus not deterring users from entering their spectral collections.

To this end the Remote Sensing Laboratories (RSL) have implemented the SPECCHIO system. A recent redesign of the data model and user interface has been based on an analysis of the metadata space and minimises the needed user actions during data input, while offering added value to the researcher (Hüni et al., 2007a,b).

In this paper, we review the existing spectral database systems in the remote sensing context; describe the concept of metadata space, the metadata set implemented in SPECCHIO, the referencing via timelines, the issues of automated, non-redundant data input, the data quality and the navigation in the metadata space and the technical implementation of the system.

2. State of the art of spectral databases

The organised storage of spectral data can be achieved via two principal methods: spectral libraries and spectral databases. The fundamental difference lies in the concept rather than the underlying technology.

Spectral libraries are data collections providing reference spectra for a number of procedures in remote sensing, such as spectral unmixing based on endmember spectra, landcover classification or atmospheric correction by the empirical line method (Richards and Jia, 2006). A number of public or commercial spectral libraries exist; for example, the USGS spectral library (Michener, 2000) or the SPECMIN package (SPECMIN, 2005), containing high quality spectra of numerous targets, mainly minerals. Such libraries usually contain first order statistical information only, i.e. one representative spectrum per target. This poses a serious restriction on their use for e.g. classifications, as the variation described by second-order statistics is not available (Landgrebe, 1997). There is a need to include such information in spectral libraries to increase the matching accuracy of field spectra against library spectra (Price, 1994). Furthermore, such libraries rarely account for the spatiotemporal variability of objects, for example, plant phenology or intra species variability (Pfaltzner et al., 2005), Spectral libraries are commonly available as static files. This has drawbacks such as low flexibility and low query performance (Bojinski et al., 2003), and thus spectral libraries are not suitable for the storage of spectral campaign data which exhibit a more dynamic nature.

Spectral databases on the other hand utilise a Database Management System (DBMS) to store spectra and metadata in relational tables. The DBMS offers functions for data definition and manipulation, but neither enforces data integrity nor removes redundancies. The latter two issues must be accounted for during the design of the data model.

In the remote sensing context, only three spectral database systems appear in literature: SPECCHIO (Bojinski et al., 2003), SpectraProc (Hueni and Tuohy, 2006) and the free online reference library for hyperspectral reflectance signatures by Ferwerda et al. The first version of SPECCHIO (Bojinski et al., 2003) offered web access and the data model included metadata, describing the sampling environment and geometry, spatial position, target type, landuse, sensor and campaign. SPECCHIO is used at RSL to store spectra and metadata in a central repository, which is accessible to all members of the laboratory. It serves as a spectral data source for various calibration/validation and simulation tasks and provides parameters for level 2/3 processing of Airborne Prism Experiment (APEX) hyperspectral imagery (Schlaepfer and Nieke, 2007). However, operational experience has shown that the success of such a spectral database system is highly dependant on its adoption by users. Many researchers were deterred from entering their data into the database due to suboptimal data capturing system interfaces, which necessitated redundant data entries. Furthermore, the redundancy was also inherent to the data model. A full redesign of the SPECCHIO system was undertaken to mend the existing deficiencies and include new requirements, such as the handling of instrument calibrations and reference panel performances.

The SpectraProc system (Hueni and Tuohy, 2006) is a solution for the storage, processing and analysis of hyperspectral signatures collected by ASD (Analytical Spectral Devices Inc., 2007) spectroradiometers. Data is stored in a relational database system and software written in C* serves as an interface, allowing the application of waveband filters, sensor convolutions, smoothing filters, derivative calculations and feature space transformations to data. SpectraProc is focused on hyperspectral signature processing and the data model, therefore, contains only minimal metadata. Still, some data model structures used in SpectraProc were included in the new SPECCHIO design. The SpectraProc system package can be downloaded from http://www.geo.unizh.ch/rls/research/SpectroLab/projects/spectraproc_index.shtml.

The free online reference library by Ferwerda et al. was constructed to facilitate data sharing. The data model includes spectra and metadata, the latter being organised flexibly enough to handle diverse metadata parameters. Web interfaces allow data browsing, geographic selections

and data export. The system has been put online, but is still under development and currently lacks queries on metadata and import. Thus users cannot upload their own spectral collections at this point of time.

3. Concepts

3.1. Metadata space

Metadata are essentially descriptive data about a resource. In the case of spectral data, the resource is the spectral response of an object and the metadata contains further information about the object and the sampling environment at the time of data capture. Metadata spaces are n-dimensional spaces defined by descriptive dimensions and most efficiently described by orthogonal vectors (Wason and Wiley, 2000).

Metadata spaces provide an analogy for thinking about, describing and creating effective metadata systems (Wason and Wiley, 2000). The descriptive quality of a metadata space can be defined via the notions of precision, resolution and repeatability. Precision is the degree of accuracy with which a resource can be represented. Resolution is the ability to differentiate between two similar items. Repeatability is the ability to have the same resource described the same way on two or more occasions (Wason and Wiley, 2000).

3.2. Data types of dimensions

The metadata vector of a spectral resource contains four types of variables: quantitative, categorical (qualitative), alphanumeric string and pictorial.

Quantitative variables are gained from measurements of quantitative features of the sampled object or the surrounding environment, e.g. spatial position or ambient temperature.

Categorical variable values are assigned to objects on the basis of a priori knowledge. Examples for such qualitative data are landcover type or species.

Alphanumeric strings are used to hold textual descriptions. String dimensions are searchable via full text search or can be parsed and indexed previous to queries.

Pictorial variables can hold supplementary information about the sampled object or its environment in the form of images, for example, photos of the sky (hemispherical), sampling setup or target.

3.3. Metadata of spectral data collections

The metadata variables implemented in the SPECCHIO system are based on Bojinski et al. (2003) and Pfitzner et al. (2005, 2006).

Table 1 lists the metadata variables and their data type as currently implemented in the SPECCHIO data model. Data types are abbreviated as follows: Categorical (C), Quantitative (Q), String (S) and Pictorial (P). The ‘A.’ column lists the possibility for automated retrieval or calculation with the data sources coded as: Spectral File (SF), Weather Station (WS), Calculation (CA) and File System (FS). Mandatory variables, according to the definition of metadata quality in SPECCHIO, are denoted with an asterisk.

3.4. Referencing based on timelines

Spectral data can be tied to instrument calibrations (Hu"ni et al., 2007b) and reference panels via temporal information. The handling of the latter is elaborated hereafter.

White reference panels are required to obtain reflectance or absorbance values from radiance measurements. It is important to calibrate the reference panel over time (Pfitzner et al., 2005). This can be achieved by comparing the field panel to a non-contaminated laboratory panel. Based on such measurements, a wavelength-dependent ratio can be calculated which subsequently can be used to correct field spectra to the ‘true’ white reference standard. The laboratory reference itself should be calibrated against some national or international standard on a regular basis. This procedure will again yield correction ratios.

It is possible to link spectra to the correction ratios automatically by maintaining a history of field and laboratory references in the database. This linking function reduces the amount of input, as it requires only the selection of the reference panel used in the field campaign.

Fig. 1 illustrates the concept using timelines. At time \( t_1 \), a new laboratory reference panel is acquired and calibrated against a national reference standard. Just before starting field campaign 1 at time \( t_2 \), the field reference panel is calibrated against the laboratory panel, yielding the FLPR(\( t_2-t_3 \)). The spectra collected during campaign 1 (\( S_1-S_4 \)), all refer the field reference panel and consequently the correction ratios. At the end of the campaign, the field panel is again calibrated against the laboratory standard. The performance of the panel during the campaign can thus be assessed.

3.5. Non-redundant and automated data input

Based on experience with the first version of SPECCHIO (Bojinski et al., 2003), it has become clear that in order to be successful, a spectral database system must minimise the manual data input as much as possible by removing data redundancies and offering automated metadata generation.

Redundancy is avoided in two ways: (a) the database model is in third normal form, which by definition contains no data redundancies (McFadden and Hoffer, 1988) and (b) the interface software that is used to feed data into the system is flexible enough to support the relational model by offering group updates.

Groups are sets of spectra that are projected to a common subspace by fixing the values of some of their metadata properties. Such a grouping is shown in Fig. 2, where the spatial positions of the spectral samples of two species are plotted. In this two-dimensional (2D) metadata subspace, the samples form clusters which can be treated as groups. A definition of the plant name for all the
<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>Description</th>
<th>Data type</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>General/Campaign</td>
<td>Campaign name</td>
<td>Name of the sampling campaign</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Campaign description</td>
<td>Textual information about the campaign</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Investigator*</td>
<td>Person responsible for the campaign</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>File path</td>
<td>File system path to the spectral campaign data</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Spatial and temporal</td>
<td>Capturing date and time</td>
<td>Date and time of the sampling in UTC.</td>
<td>Q</td>
<td>SF</td>
</tr>
<tr>
<td>information</td>
<td>Latitude*</td>
<td>Spatial sampling position</td>
<td>Q</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Longitude*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Altitude*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target information</td>
<td>Target homogeneity*</td>
<td>Homogenous or heterogeneous</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Landcover type*</td>
<td>Based on CORINE land cover (European Commission DG XI, 1993)</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectrum names</td>
<td>Scientific and common names of the target</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Target type*</td>
<td>RSL internal designation of target types, e.g. snow, pasture</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pictures</td>
<td>Images depicting the target. May also be used to document the sampling environment.</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Sampling geometry</td>
<td>Sensor zenith angle*</td>
<td>Sensor zenith angle measured from nadir, i.e. nadir = 0</td>
<td>Q</td>
<td>CA</td>
</tr>
<tr>
<td></td>
<td>Sensor azimuth angle*</td>
<td>Sensor azimuth angle relative to the illumination angle, i.e. 180° for the principal plane opposite of illumination source</td>
<td>Q</td>
<td>CA</td>
</tr>
<tr>
<td></td>
<td>Sensor distance</td>
<td>Distance of the sensor to the target</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Illumination zenith angle*</td>
<td>Illumination source zenith angle measured from nadir</td>
<td>Q</td>
<td>CA</td>
</tr>
<tr>
<td></td>
<td>Illumination azimuth angle*</td>
<td>Absolute illumination source azimuth angle measured from geographic North</td>
<td>Q</td>
<td>CA</td>
</tr>
<tr>
<td></td>
<td>Illumination distance</td>
<td>Distance between the illumination source and target</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>Measurement details</td>
<td>Number of averaged spectra</td>
<td>Number of spectra averaged internally by the instrument</td>
<td>Q</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>White reference</td>
<td>White reference panel used</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sensor*</td>
<td>Sensor model</td>
<td>C</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Instrument*</td>
<td>Specific instrument identified by a serial number</td>
<td>C</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Instrument calibration number</td>
<td>Number of the instrument calibration</td>
<td>C</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Foreoptic*</td>
<td>Additional optic that changes the field of view (FOV) in degrees</td>
<td>C</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Illumination source</td>
<td>Type of illumination source, e.g. sun, Hg lamp</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sampling environment*</td>
<td>Field or laboratory</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Measurement type*</td>
<td>Single, directional, temporal</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Measurement unit*</td>
<td>Reflectance, digital numbers, radiance, absorbance</td>
<td>C</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Goniometer model</td>
<td>Name of the goniometer used</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Environmental conditions</td>
<td>Cloud cover*</td>
<td>Amount of clouds covering the sky defined in octas</td>
<td>C</td>
<td>WS</td>
</tr>
<tr>
<td></td>
<td>Ambient temperature</td>
<td>Air temperature in degrees</td>
<td>Q</td>
<td>WS</td>
</tr>
<tr>
<td></td>
<td>Air pressure</td>
<td>Air pressure in hPa</td>
<td>Q</td>
<td>WS</td>
</tr>
<tr>
<td></td>
<td>Relative humidity</td>
<td>Relative humidity as percentage</td>
<td>Q</td>
<td>WS</td>
</tr>
<tr>
<td></td>
<td>Wind speed</td>
<td>Qualitative description of the wind speed: calm, breezy, windy, stormy</td>
<td>C</td>
<td>WS</td>
</tr>
<tr>
<td></td>
<td>Wind direction</td>
<td>Direction classes in 45° steps, measured from geographic North</td>
<td>C</td>
<td>WS</td>
</tr>
<tr>
<td>File information</td>
<td>Auto number</td>
<td>Automatic, consecutive number assigned by the spectroradiometer capturing software</td>
<td>Q</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>User comment</td>
<td>Comment added by the user</td>
<td>S</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>Spectral file name</td>
<td>Name of the spectral file</td>
<td>S</td>
<td>FS</td>
</tr>
<tr>
<td></td>
<td>File format</td>
<td>File format of the spectral file</td>
<td>C</td>
<td>SF/FS</td>
</tr>
<tr>
<td></td>
<td>Data structuring information</td>
<td>Hierarchical structure as gleaned from folder structure</td>
<td>C</td>
<td>FS</td>
</tr>
</tbody>
</table>
samples contained in this subspace is then reduced to two

Table 1 lists the automation possibility and the data
source for every metadata variable. The files produced by
the spectroradiometer data capturing software usually
include, by default, a wealth of information that can be
easily extracted and inserted into the database.

3.6. Metadata quality

Assessment of the data quality is a prime issue, when it
comes to using spectral collections from other scientists.
Within SPECCHIO, we define metadata quality by the
descriptive power of the metadata space. If the metadata
are non-existent, the spectral data are not described and,
thus rendered useless to persons, not having intimate

Fig. 1. Referencing of white reference correction ratios by spectra and calibration of panels against standards.

Fig. 2. An example of spectra grouped (clustered) by their spatial properties.
knowledge of the data set. The more the metadata are recorded, the higher the chance that a sampled object can be discriminated in metadata space. Utilisation of all dimensions of the metadata space enables the user to assess the sampling circumstances in great detail, and thus decide if the data can be trusted. Other researchers are provided with a mandatory, minimal subset of metadata parameters (Table 1), allowing for an assessment of the data.

3.7. Navigation in metadata spaces

The position of every spectrum in metadata space is given by its descriptive vector. The space can be projected to a subspace by fixing the value of one or more dimensions. Thus, the specification of query conditions puts restrictions on metadata space dimensions and the resulting subspace contains the queried data sets (Wason and Wiley, 2000). Restriction in several dimensions is achieved by a logical AND of the constraints per dimension. Multiple restrictions on one dimension, i.e. several allowed classes for categorical variables, several value intervals for quantitative variables or several matching patterns for alphanumeric string variables are combined by a logical OR.

The concept of subspace projection is illustrated in Fig. 3, where the values of target type and spatial sampling position, given as latitude and longitude, are fixed to a certain class (pasture) or value range, respectively (longitude $\geq 10^\circ$ AND $\leq 15^\circ$ and latitude $\geq 45^\circ$ AND $\leq 47^\circ$). The subspace, shown as dark little cube (Fig. 3, right) contains all spectra that represent pastures being sampled at a geographic location limited by the above coordinates.

The structure of subspace projections can be directly translated into Structured Query Language (SQL). The definition of the appropriate SQL syntax in Extended Backus-Naur Form (EBNF) (International Organization for Standardisation, 1996) is contained in Hüni et al. (2007b).

4. Implementation

4.1. Architecture

The core of the SPECCHIO system is a MySQL database (MySQL AB, 2007) hosted on a database server (cf. Fig. 4). The SPECCHIO application was implemented as a Java 2 (JavaTM, 2006) application which allows full flexibility on local file system operations. The Java technology keeps the software independent of the operating system, thus allowing its use in a heterogeneous computing environment. The application runs on any machine with a Java Virtual Machine (VM) installation and connects to the database via TCP/IP on a configurable port. Connection to the SPECCHIO database is, therefore, possible via the Internet, enabling the sharing of data between research groups worldwide. The application can also be run remotely from a terminal on a server by the use of the X11 protocol.

The spatial aspect of data sets offers the possibility for direct linkage with a GIS system. In the case of ArcGIS (ArcGIS, 2006), a database connection is established via Open Database Connectivity (ODBC).

4.2. Database

The database was implemented on a MySQL Version 5 database server. The schema comprises 42 tables and views. Starting with version 5, MySQL provides views and access to the information schema containing table structure information. This allows for the dynamic and generic building of SQL statements in the client application for e.g. retrieving primary and foreign key column names of related tables.

Multiuser support is an important issue as the system is designed as a platform for spectral data exchange. Users can upload, modify and delete their own data and are allowed to browse and download all data in the database. This is achieved using individual database user accounts, views, triggers and the granting of rights. All tables of the SPECCHIO schema are available for select operations.
Delete, update and insert operations are only granted on the views, where the views include a restriction based on the current user id. Therefore, users can modify only their own data. The update of the underlying tables with the user id upon inserts is handled via triggers, thus keeping the data consistent, irrespective of the client application used to send insert statements.

Data modification rights for system tables like sensor, instrument or calibration are only granted to the administrator of the system.

4.3. Client application

User interaction with the database is handled by the SPECCHIO client application based on graphical user interfaces (GUI). The main functions are: creating and loading of spectral campaigns, metadata editing, data querying, visualising and exporting. Fig. 5 shows the SPECCHIO metadata editor GUI.

The current version of SPECCHIO supports the following spectral signature files as data input formats: ASD binary (Analytical Spectral Devices Inc., 2007), GER signature (Spectra Vista Co., 2005), ASCII tab separated and MFR OUT (Yankee Environmental Systems Inc., 2000). Support for other spectroradiometer input file formats depends on user demands. According requests should be directed to the first author.

Further file formats are: sensor specifications in a proprietary format for the definition of sensors in the database and GER calibration files to maintain calibration histories of GER instruments in the database.

Two output formats are implemented: Comma Separated Value (CSV) files that can be read by statistical and spreadsheet applications and ENVI Spectral Library files that are primarily a data format used by ENVI (ENVI, 2005), but can be read by other remote sensing packages as well.

5. Discussion

The implemented metadata space comprises 41 variables. The suggested metadata parameters by van der Meer and de Jong (2001) and Pfitzner et al. (2005, 2006) sum up to a total of 57 parameters. Most of the additional variables not accounted for in the SPECCHIO data model are connected with enhanced target information, such as ground cover, soil, phenology or plant height. These are in some cases very specific variables that may not be suited for a generic data model. The validation of the current metadata definition is an issue for future work. The data model may be extended to support further important metadata which include: (a) the documentation of the illumination source over time, by the use of sun photometer data, (b) storage of chemical or biophysical measurement values, which are connected to spectrally sampled objects and are subsequently used, for e.g. the generation of inversion models and (c) flags that help to assess the data quality of the spectrum.

The current implementation defines data quality via the descriptive power of the metadata space. It would, however, be desirable to evaluate the spectral data quality as well. This could be assessed by the estimation of the SNR, where a low SNR would indicate low quality and vice versa, detection of spectral misregistrations between VNIR and SWIR detectors and data screening procedures based on reference spectra, as defined by Zhang et al. (2004). These screenings are designed to identify and exclude outliers in spectral data sets. Zhang et al. (2004) list three tests to assess the so-called ‘spectral data quality’: (a) checking the existence and position of spectral
characteristics of a measured spectrum against a reference spectrum, (b) testing the shape similarity by calculating correlation coefficients between the measured and the reference spectrum and (c) building upper and lower thresholds for the intensity, by defining a so-called spectrum zone around the mean using standard deviations of the reference data set.

The CORINE landcover scheme (CLC) (European Commission DG XI, 1993) has been chosen for the current implementation of SPECCHIO. However, analysis of the precision, resolution and repeatability of the CORINE vocabulary suggests that other schemes should also be considered. One of the identified problems with the CORINE scheme is that some classes tend towards a description of landuse rather than of pure landcover. Alternative landcover schemes include the Core Service Land Cover (CSL) (Kuntz) which comprises 21 thematic classes compared to the 44 classes of CLC. This reduction in classes may decrease the precision and resolution, but should provide better repeatability.

An optimal metadata space should be orthogonal; however, the SPECCHIO metadata model contains the sensor, instrument and calibration dimensions, which are correlated. The implications of this are an increased complexity of the metadata editor user interface implementation, due to the needed dependency checks and the possible creation of queries yielding no data sets when contradicting restrictions are put on correlated dimensions.

Although the spectrum name is listed as a categorical variable, the current data model implements it as an alphanumeric string. This approach was chosen due to simplicity, however, having a well-defined vocabulary based on e.g. known plant taxonomies, would increase the repeatability and precision of this variable. The problem of combining different taxonomies into one hierarchical vocabulary is an issue for further research.

Metadata should comply with some widely and internationally accepted standards (Lanz et al., 2007). For data sharing purposes, other file formats or database access interfaces should be considered. However, such standards should be generic enough to accommodate all metadata that are contained in the current SPECCHIO data model. Formats and definitions to be considered include: (a) the geographic information/geomatics standards developed by ISO TC 211 such as ISO 19115, (b) the FGDC Content Standard for Digital Geospatial Metadata defined by the US Federal Geographic Data Committee (FGDC) (Di, 2003) and (c) the OpenGIS standards Sensor Observation Service (SOS). Geography Markup Language
(GML)\textsuperscript{7} and Observations and Measurements (O&M).\textsuperscript{8} The provision of a standardised data interface to SPECCHIO requires further investigation of the potential standards.

6. Conclusions

Metadata support the interpretation of scientific data, in general, help to ensure long-term usability and provide a basis for the assessment of data quality and possibility of data sharing between scientists. The recently updated SPECCHIO system is a repository for spectroradiometer data and associated metadata, thus providing a platform for spectral signature data exchange. The generation of metadata in the system has been optimised by automated gleaning of metadata from spectral input files and containing data structures, and by providing group updates on spectral sets. Spectral data sets are retrieved by the means of metadata space queries, which put restrictions on metadata dimensions and thus create a subspace containing the required data sets.

A Java application is used for the interaction with the database, enabling the use of the system in a heterogeneous computing environment with a server hosting the database.

RSL maintains an online version of the SPECCHIO database and interested parties can acquire a database account for testing and data sharing purposes. The SPECCHIO system installation package allows local installation and is intended for users requiring access control over their data. In-house databases may also offer higher performance than the online version. RSL distributes the SPECCHIO system package free of charge. Expressions of interest are welcome and should be directed to the first author. For further information, please refer to the SPECCHIO website.\textsuperscript{9}

References


Landgrebe, D., 1997. On Information Extraction Principles for Hyperspectral Data, Purdue University, West Lafayette, IN, USA, p. 34.


\textsuperscript{9} \url{http://www.specchio.ch}