Metadata of spectral data collections

Hueni, Andreas; Nieke, Jens; Schopfer, Jürg; Kneubühler, Mathias; Itten, Klaus I

Abstract: Metadata is important for the interpretation of scientific data, quality assessment and long term usability of data sets. The sharing of spectral data collections among research groups is uncommon and one of the reasons for this is the missing standardisation of the sampling process. Appropriate metadata serves the purpose of detailing the sampling procedure and the surrounding conditions during data capture, thus providing necessary information for data sharing. Reliable data retrieval requires the organised storage of spectral and metadata. To this means RSL developed the SPECCHIO system which is based on a relational database and provides data input, query and output mechanisms that strive to minimize the manual data capture. SPECCHIO serves as a non-redundant repository and source for spectral signatures which can be retrieved by metadata queries. The system will be used in the level 2/3 processing of the APEX (Airborne Prism Experiment) product generation to support the classification of natural and manmade materials and landcovers.

METADATA OF SPECTRAL DATA COLLECTIONS

Andreas Hüni, Jens Nieke, Jürg Schopfer, Mathias Kneubühler and Klaus I. Itten

1. University of Zürich, Department of Geography, RSL, Zürich, Switzerland; ahueni@geo.uzh.ch

ABSTRACT

Metadata is important for the interpretation of scientific data, quality assessment and long term usability of data sets. The sharing of spectral data collections among research groups is uncommon and one of the reasons for this is the missing standardisation of the sampling process. Appropriate metadata serves the purpose of detailing the sampling procedure and the surrounding conditions during data capture, thus providing necessary information for data sharing. Reliable data retrieval requires the organised storage of spectral and metadata. To this means RSL developed the SPECCHIO system which is based on a relational database and provides data input, query and output mechanisms that strive to minimize the manual data capture. SPECCHIO serves as a non-redundant repository and source for spectral signatures which can be retrieved by metadata queries. The system will be used in the level 2/3 processing of the APEX (Airborne Prism Experiment) product generation to support the classification of natural and manmade materials and landcovers.

INTRODUCTION

Ground based hyperspectral signatures are collected for (a) calibration and validation of remote sensing imagery and its data products, (b) feasibility studies for airborne/spaceborne missions, (c) basic investigation of the relationship between physical or biochemical properties and the electromagnetic reflectance of objects and (d) definition of directional dependence of the reflectance of objects on the illumination and viewing geometry.

Since the advent of field spectroscopy with the first specifically built portable field instrument appearing in the late 1980's, e.g. PIDAS (34), a lot of research on the spectral properties in the VIS/NIR electromagnetic spectrum of natural and manmade objects has been carried out. At the same time considerably 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.

The comparison of spectral signatures between studies is complicated by the many different techniques for the capturing of spectral field data (19). Utilizing data from other studies requires an assessment of the data quality and suitability of the data set for the given task. Milton et al (20) state that accuracy depends on a clear definition of what is being measured and on the conditions under which is being measured, i.e. the description of the sampling experiment and of the sampling environment is of importance if the data quality is to be assessed. The factors that influence the spectral measurements taken in the field are detailed in (19), i.e. for a traceability of the measurement process these factors should be recorded and stored as metadata. Metadata support broad and long-term use and interpretation of scientific data (18). The lack of metadata can render previously collected data useless for new applications (6).

Spectral libraries are data collections that provide reference spectra for a number of procedures in remote sensing, e.g. spectral unmixing based on endmember spectra, landcover classification or atmospheric correction by the empirical line method (27). A number of public spectral libraries exist, e.g. the USGS spectral library (2), that contain high quality spectra of numerous targets but are mainly focused on minerals. Such libraries usually only contain first order statistical information, i.e. only one representative spectrum per target. This poses a serious restriction on the use as the variation described by second order statistics needed for e.g. classifications is not available (16). There is a need to include such information in spectral libraries to increase the matching accuracy of field spectra against library spectra (26). Furthermore, such libraries do often not account for the spatiotemporal variability of objects, e.g. plant phenology or intra species variability (23). Thus there is still a need to build customised spectral libraries that account for local variables affecting the spectral reflectance of objects.

A limited number of studies explicitly mention the building of a spectral library containing ground reflectances of targets which are subsequently used to derive products from imagery (11; 15; 25) while numerous studies based on field spectroradiometer data have been carried out but do not explicitly consider their spectral collections as libraries, e.g. (29; 31; 32; 33). However, most studies focus on the spectral characteristics of the targets and while the acquisition of field data is described, not much detail is given about the organisation and storage of these data and associated metadata in most cases.

Given the scenario outlined above, an organised 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. A relational database with appropriate interfacing software seems a natural choice of technology in this respect. RSL has implemented the SPECCHIO system which acts as a repository for spectral field campaign and reference signatures plus metadata (1; 12). A recent redesign of the data model and user interface has been based on an analysis of the metadata space and minimizes the needed user actions during data capture while offering added value to the researcher.

SPECCHIO is planned to be an integral part of the APEX (Airborne Prism Experiment) (22) level 2/3 processing chain and is foreseen to be used for pre and post classification of hyperspectral image cubes (28).

In this paper we describe the general concept of metadata space, its application to field spectroradiometer metadata, the metadata set implemented in SPECCHIO, the conceptual integration of SPECCHIO into the APEX level 2/3 processing and the user interfaces of the SPECCHIO system.

METHODS

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. The space is most efficiently described by orthogonal vectors, i.e. the dimensions are independent of each other (35).

Metadata spaces provide an analogy for thinking about, describing and creating effective metadata systems (35). 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 (35). This concept is illustrated by an example from the remote sensing context: a spectrum of a deciduous tree can be described by its landcover using a fixed vocabulary. If this vocabulary only contains three classes (water, forested land, non-forested land) then the spectral signature of the tree can be described by ‘forested land’. This is however a low precision. Suppose a spectrum of a coniferous tree is added to the system. The vocabulary does not permit to distinguish the two items, thus the resolution is low. If two different people were to describe the resource again it is very likely that both will pick ‘forested land’, thus the repeatability is high.

The definition of vocabularies is no straightforward task; an increase in precision and resolution decreases the repeatability. The classes assigned by the cataloguers have a variability that can be described by a probability distribution. A very restricted vocabulary results in a high probability that the same descriptive value is assigned to the same resource by different cataloguers. The uncertainty of the description of a resource by metadata is termed the ‘fuzziness of the system’ (35). Vocabularies must therefore strive to balance precision, resolution and repeatability in order to minimize the fuzziness. If a given vocabulary is not sufficient in discriminating two similar objects in metadata space then some other features must be utilized, i.e. more dimensions are needed to describe objects unambiguously.

Data Types of Dimensions

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

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

Categorical variable values are assigned to objects on the basis of a priori knowledge. Examples are: landcover type, species, arbitrary sampling site number or sampling location name.

Alphanumeric strings are used to hold textual descriptions. They do not contain information in a structured way but can help the user in understanding the data. In the context of metadata space alphanumeric string variables neither form clusters nor do they group data in any organised way. String dimensions are searchable via full text search or can be crawled and indexed previous to queries.

Pictorial variables can hold supplementary information about the sampled object or its environment in form of images, e.g. photos of sky (hemispherical), sampling setup or target. Pictorial variables have the potential of yielding quantitative or qualitative data if subjected to image analysis or image indexing techniques. This is however not further investigated at this point.

Metadata of Spectral Data Collections

The metadata model detailed in this section is based on Bojinski et al (1) and Pfitzner et al (23; 24) and reflects the current implementation status of SPECCHIO. Possible extensions to this model are discussed in the conclusion.

Table 1 lists the metadata variables, their data type and the potential automation as currently included in the SPECCHIO data model. The data types are abbreviated as follows: C (Categorical), Q (Quantitative), S (String) and P (Pictorial). The 'Automation' column lists the possibility of automated retrieval or calculation: SF (Spectral File), WS (Weather Station) and CA (Calculation).

For the following metadata variables detailed explanations are given in three subsequent sections: (a) sensors, instruments and instrument calibrations, (b) sampling geometry and (c) data structuring information.

Table 1: Metadata variables contained in the SPECCHIO data model

Metadata variable	Type	Automation
Auto number	C	SF
User comment	S	SF
Capturing date and time	Q	SF
Spectral file name	S	SF
Number of spectra averaged internally by the instrument	Q	SF
Sensor	C	SF
File format	C	SF
Instrument	C	SF
Instrument calibration number	C	SF
Foreoptic	C	SF

Illumination source	C	
Sampling environment	C	
Measurement type (single, directional, temporal)	C	
Measurement unit (Reflectance, DN, radiance, absorbance)	C	SF
Target homogeneity	C	
Spatial position (latitude, longitude, altitude)	Q	SF
Landcover (based on CORINE land cover (8))	C	
Cloud cover (in octas)	C	
Ambient temperature	Q	WS
Air pressure	Q	WS
Relative humidity	Q	WS
Wind speed (Qualitative description)	C	WS
Wind direction (categories in 45 degree steps)	C	WS
Sensor zenith angle	Q	CA (Goniom.)
Sensor azimuth angle	Q	CA (Goniom.)
Sensor distance	Q	
Illumination zenith angle	Q	CA (Sun pos.)
Illumination azimuth angle	Q	CA (Sun pos.)
Illumination distance	Q	
Spectrum names	C	
Target type	C	
Goniometer model	C	
Pictures	P	
Data structuring information	C	Gleaned from folder structures
Investigator	C	
File path to spectral data on file system	S	
Campaign comments/description	S	
Campaign name	S	

Sensors, Instruments and Instrument Calibrations

The SPECCHIO data model contains definitions of sensors, instruments and instrument calibrations. While the terms sensor and instrument are sometimes interchanged in everyday scientific language it is important to understand that sensors and instruments are two different dimensions in metadata space.

Sensors define the physical setup of sensors, i.e. number of channels, centre wavelength and FWHM (Full Width at Half Maximum) per channel. Instruments on the other hand are existing instances of a certain sensor type. There can be several different instruments that are all of one sensor type. Instruments also have a history of calibrations which account for the change in instrument characteristics over time.

Consider the example of a GER 3700 instrument: this instrument is an instance of a GER 3700 sensor. The sensor defines the average wavelength per channel. As long as no calibration for the instrument has been entered into the database the channels defined by the sensor will be used for plotting and exporting spectral data. When calibrations are entered for instruments they override the sensor specifications. Calibrations are tied to spectra either by explicit calibration numbers available in input files or by temporal information. In either case, spectra will refer to the newest calibration in respect to their capture time. Figure 1 illustrates the concept of spectra referencing the sensor or calibrations based on the timeline. Time t_0 denotes the start of the system. At time t_1 a sensor X is defined in the database. Spectrum S_1 refers to the sensor default specification for plotting and exporting as at this time no calibration information is available. At t_2 the instrument is calibrated and a first calibration data set added to the database, thus the spectra S_2 and S_3 captured after t_2 refer to the Cal 1 entry. At time t_3 the calibration data set 2 is created and loaded. Spectrum S_4 therefore references this latest entry (Cal 2).

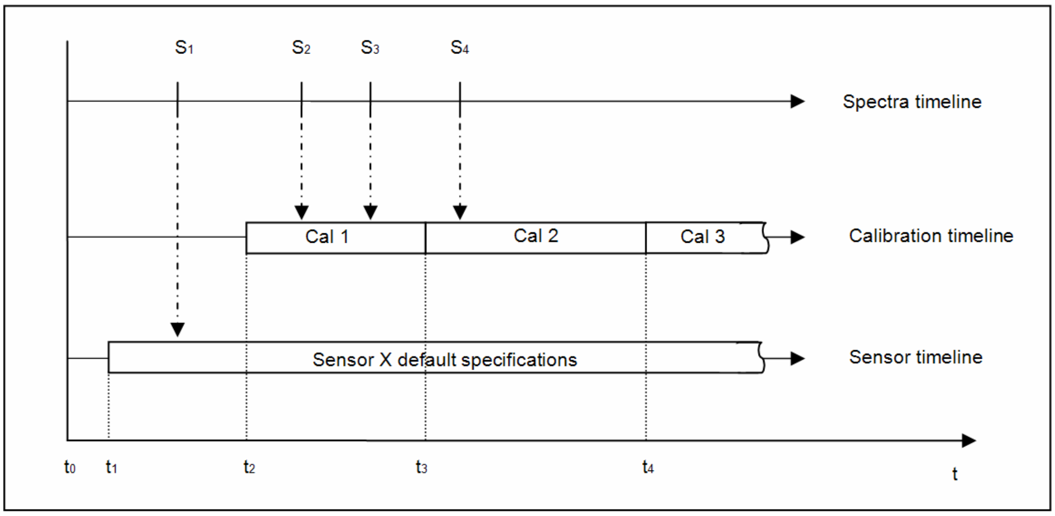

Figure 1: Referencing of sensor and calibration entities by spectra

Sampling Geometry

The sampling geometry is defined by the zenith and azimuth angles and the distance to the sampled object of both illumination source and sensor. The illumination source of field experiments is usually the sun and its geometry can be calculated if the spatial position and the capture time and date are known. The latter two variables can be captured automatically using the internal clock of the field laptop or a GPS unit connected to the system respectively.

For goniometer measurements the sensor geometry can be calculated automatically if the sampling process adheres to a well defined sampling protocol as is the case for the FIGOS and LAGOS goniometers (30).

Data Structuring Information

Spectral sampling campaigns yield spectral signatures of objects. Physically, spectroradiometers produce files containing digitized spectral signatures of the sampled objects with usually one file being created per reading. The sheer number of files resulting from sampling campaigns requires an organised method of storage. One possibility is the creation of a hierarchical folder structure which is designed and implemented before the start of a campaign and used to store the data files during acquisition. The concept of structuring of spectral sampling campaign data is elaborated in (12). Such structures reflect the way the user thinks about the data and are in this respect important metadata. The structural information can be gleaned from the folder structure by the creating agent, i.e. the SPECCHIO data loading process, and subsequently utilized for data handling operations in graphical user interfaces.

Metadata Quality

Assessment of the data quality is a prime issue when it comes to utilizing spectral collections from other scientists. Within SPECCHIO we define metadata quality by the descriptive power of the metadata space. If the metadata is non-existent the spectral data is not described and thus rendered useless to persons not having intricate knowledge of the dataset. The more metadata is recorded the higher is 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. Based on the available metadata parameters a subset (cf. Table 2) was defined that is mandatory if the data should be useful to other scientists.

Table 2: Mandatory metadata parameters for minimum metadata quality

Investigator	Sampling environment
Sensor	Spatial position
Instrument	Cloud cover
Foreoptic	Sensor zenith
Landcover	Sensor azimuth
Target homogeneity	Illumination zenith
Measurement unit	Illumination azimuth
Measurement type	Target type

Queries in Metadata Space

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. I.e. the specification of query conditions puts restrictions on metadata space dimensions and the resulting subspace contains the queried data sets (35). 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 equivalent SQL (Structured Query Language) syntax for a query on the SPECCHIO database is defined by:

SELECT SPECTRUM FROM <all_involved_tables> WHERE <all_constraints> AND < required_PK_FK_combinations>;

Where the values in the <> brackets are of the following structure (defined in EBNF (Extended Backus-Naur Form)(14; 36)):

all_involved_tables = table_name, {‘,’, table_name};

all_constraints = ‘(‘, dimension_constraints, ‘)’, {‘ ‘,‘AND’,’ ‘,’(‘, dimension_constraints, ‘)’};

dimension_constraints = dim_constr_part {‘ ‘, ‘OR’,’ ‘,dim_constr_part };

dim_constr_part = ‘(‘, constraint, [‘ ‘, ‘AND’, ‘ ‘, constraint], ‘)’;

constraint = table_name.column_name, operator, allowed_value;

operator = ‘>’|’<’|’>=’|’<=’|’=’|’like’;

required_PK_FK_combinations = PK_FK_combination, {‘ ‘,‘AND’,’ ‘,PK_FK_combination};

PK_FK_combination := table_name.PK, ‘=’, table_name.FK;

Integration into the APEX Processing Chain

The level 2/3 processing of the APEX processing chain will include algorithms for spectral albedo calculation and classification. Conceptually, spectral albedo is generated by the BRDF process which utilizes spectral directional information contained in the spectral database 'SPECCHIO'. For the classification two potential approaches have been identified: (a) to use the spectral database as a library where unknown spectra, i.e. the image cube pixels, are looked up using an approach similar to the Tetracorder system (3) or (b) to build or train classifiers on datasets selected from the database as the result of a query in metadata space, e.g. by restricting the spatial, temporal and landcover dimensions it is possible to select a training dataset which is subsequently used to build a spatiotemporal, landcover optimized classifier. Thus the latter approach is closer related to end user requirements (landcover maps, species maps, spatiotemporal aspect, etc) due to the retrieval of spectral data based on higher level information contained in its metadata.

RESULTS

Results of the definition of a metadata space for spectral signatures are demonstrated hereafter on (a) examples of analyses in metadata space using visualization, (b) the graphical user interface (GUI) for the definition of metadata, (c) the GUI for the building of metadata queries and (d) the resulting query report.

Analysis in Metadata Space

A better understanding of the spectral data can be developed by analysing its metadata. An appropriate form is the visualisation of a two or three dimensional subspace that can be built by any combination of metadata dimensions. Figure 2 illustrates this by a plot showing the spatial position of every sample taken (left side) and of the number of sample sites per species (right side). Note that because both species name and site number are categorical variables all spectra concentrate into four positions in this subspace. In the given example the site number was gleaned from the folder structure used during the data acquisition.

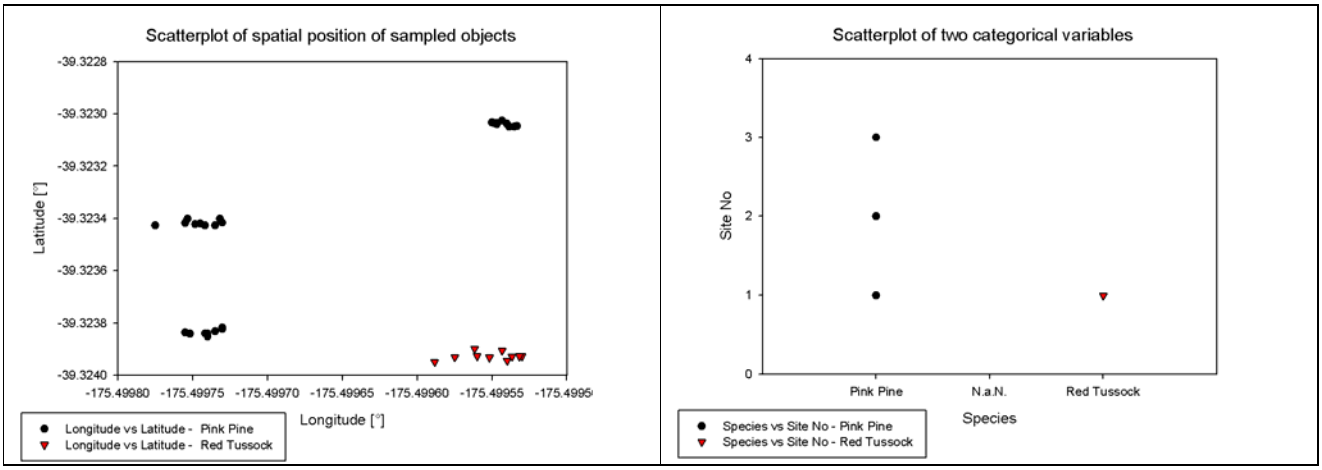

Figure 2: Scatterplots of spatial position of sampled objects (left) and sample sites per species (right)

Definition of Metadata

Metadata are entered into the SPECCHIO database by means of the Metadata Editor (cf. Figure 3). The editing of metadata is campaign based, i.e. only one campaign can be edited in the editor at one time. The structure of the campaign as gleaned from the directory structure by the creating process is visualized by a tree structure (lower left in Figure 3). Selection of the data to be edited happens via this tree. Three tabs (right side in Figure 3) hold the metadata fields of the campaign, the hierarchy and the spectrum. The content of the fields reflects the current selection in the tree. Multiple updates are possible by selecting multiple hierarchies and/or spectra. This feature can speed up the metadata capturing process significantly as identical data have to be entered only once per spectral group. A metadata conflict detection is executed for multiple selections and only non-conflicting metadata parameters can be updated, e.g. if every spectrum in a selection already refers to a different spatial position editing will be disabled for the position. The conflict detection can however be overridden and thus common updates of differing metadata fields of a spectral group are possible.

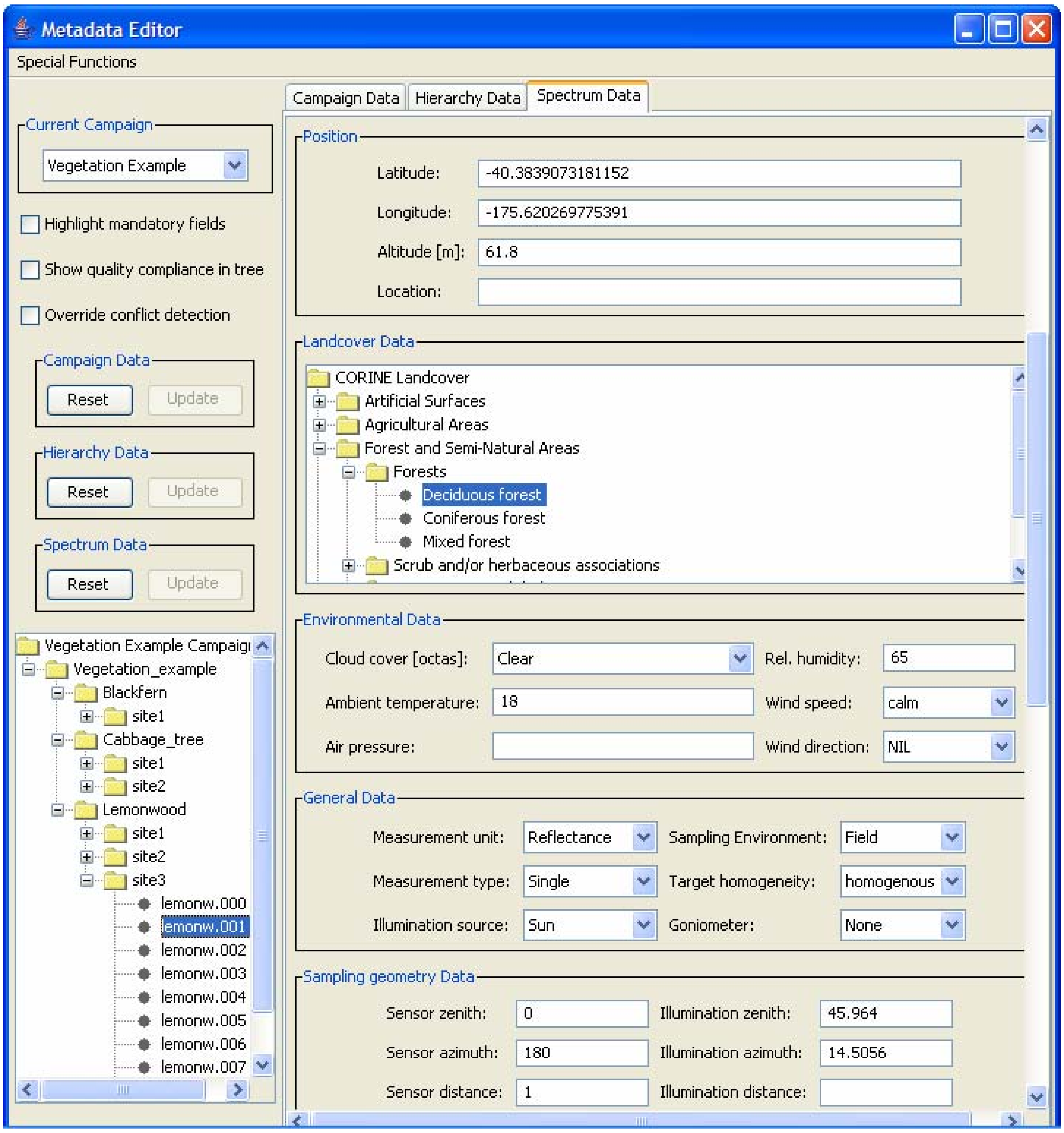

Figure 3: SPECCHIO Metadata Editor

Categorical variable values are either selected from combo boxes or from hierarchical representations which are both pre-filled from the database. This ensures that categorical variables comply to the defined vocabulary of the respective metadata space dimension. Quantitative variables are entered into fields restricted to numerical values.

Each of the three tabs has associated reset and update buttons which will restore the previous values or commit the changes to the database respectively.

Metadata Queries

Data is queried by using the Query Builder (Figure 4). Two operational modes are supported: (a) direct selection (browsing) of records by using the tree structure as defined by the structural information dimension and (b) specification of query conditions (metadata space constraints).

Both modes support the building of SQL queries on the fly. The resulting, automatically built query and the number of records contained in the selected subset are displayed instantaneously in the right side of the Query Builder interface. The current version only supports single constraints per dimension.

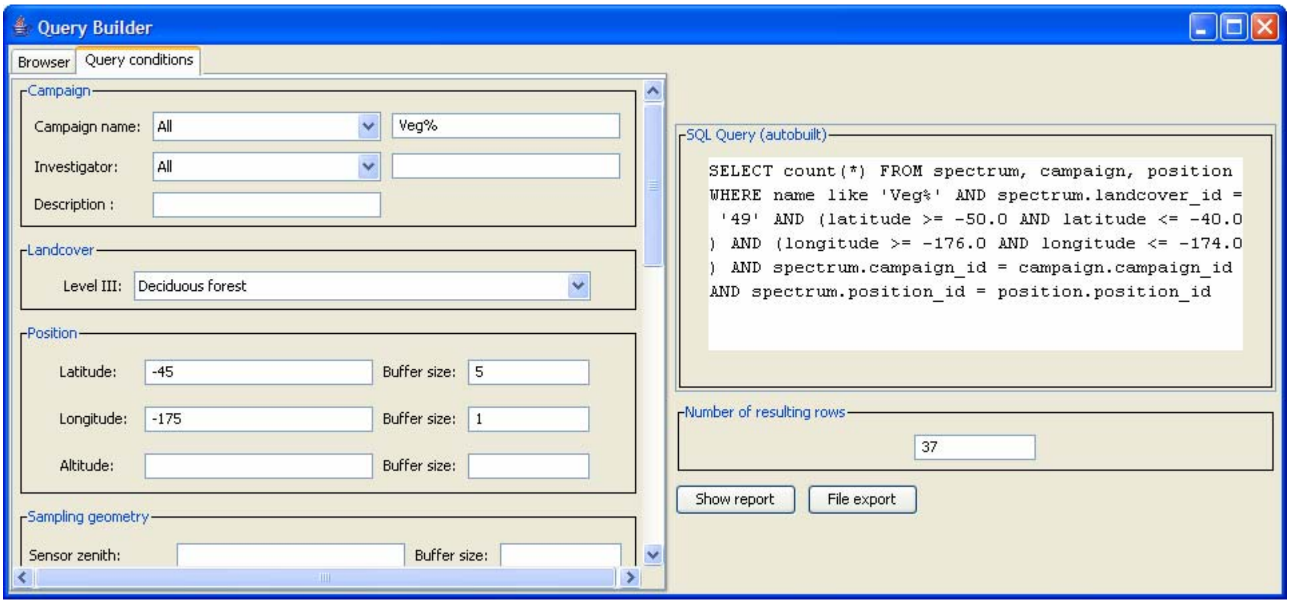

Figure 4: SPECCHIO Query Builder

Query Report and Output to Files

Queries built in the Query Builder can be applied for online reports or file outputs. Figure 5 shows a spectrum report window with a spectral plot on the left side and metadata attributes listed on the right side.

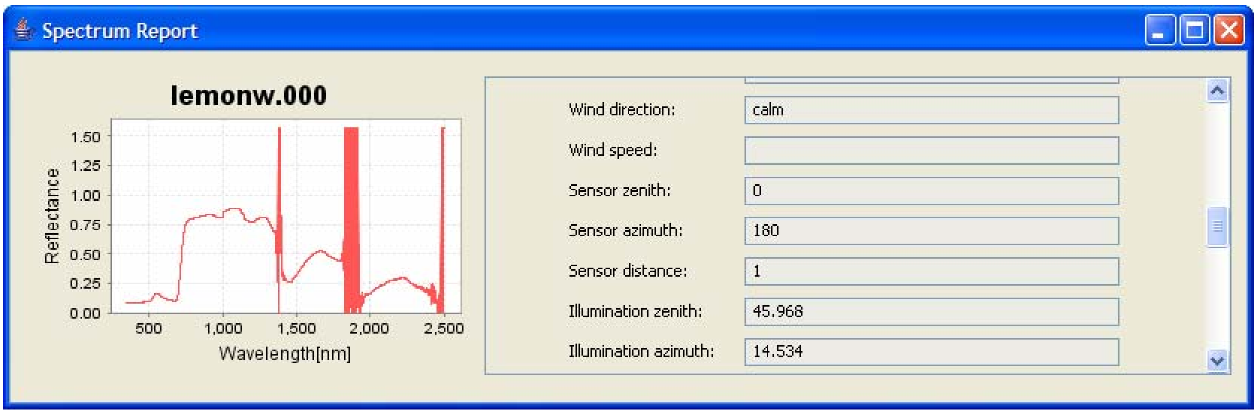

Figure 5: SPECCHIO Spectrum Report

Data can be written as CSV (Comma Separated Values) or ENVI Spectral Library files. During file output the data is grouped by sensors/instruments and calibrations and written to separate files. This accounts for the need of having the sensor central wavelengths defined only once per file.

CONCLUSIONS

Metadata supports the interpretation of scientific data in general, helps to ensure long-term usability and provides 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. The generation of metadata in the system has been optimized by (a) automated gleaning of metadata from spectral input files and containing data structures and (b) 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.
The CORINE landcover scheme (CLC) (8) 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 (10). Alternative landcover schemes include the Core Service Land Cover (CSL) (10) 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 (a) the metadata editor user interface implementation is complicated due to the needed dependency checks and (b) queries yield no datasets when contradicting restrictions are put on correlated dimensions.

Although the spectrum names are listed as a categorical variable the current data model implements them as alphanumeric strings. 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 of further research.

The data model may be extended to support further important metadata which include: (a) the complete history of white references and their spectral performance compared to e.g. laboratory or national standards, (b) tying of spectral data to reference panel readings, i.e. spectra can be corrected for deviations of the used reference panel from the laboratory or national standard, (c) the documentation of the illumination source over time by the use of e.g. sun photometer data (d) storage of chemical or biophysical measurement values which are connected to the spectrally sampled object and are subsequently used for e.g. the generation of inversion models and (e) flags that help to assess the data quality of the spectrum. The spectrum data quality could be assessed by (a) estimation of the SNR where a low SNR would indicate low quality and vice versa, (b) detection of spectral misregistrations between VNIR and SWIR detectors and (c) data screening procedures based on reference spectra as defined by Zhang et al. (37). These screenings are designed to identify and exclude outliers in spectral datasets. Zhang et al. (37) list three tests to assess the so called ‘spectral data quality’: (a) check of 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.

Metadata should comply with some widely and internationally accepted standards (17). 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 (13) such as ISO 19115, (b) the FGDC Content Standard for Digital Geospatial Metadata defined by the U.S. Federal Geographic Data Committee (FGDC) (7; 9) and (c) the OpenGIS standards Sensor Observation Service (SOS) (21), Geography Markup Language (GML) (5) and Observations and Measurements (O&M) (4). The provision of a standardised data interface to SPECCHIO requires further investigation of the potential standards.

RSL is currently considering different options to make SPECCHIO available to the spectroscopy user community. For further information please refer to the SPECCHIO website: http://www.geo.unizh.ch/rsl/research/SpectroLab/projects/specchio_index.shtml.

REFERENCES

1 Bojinski, S., Schaepman, M., Schlaepfer, D. & Itten, K., 2003. SPECCHIO: a spectrum database for remote sensing applications. *Computers & Geosciences,* 29, 27-38.

2 Clark, R. N., Swayze, G. A., Gallagher, A. J., King, T. V. V. & Calvin, W. M., 1993. The U. S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 microns, *U.S. Geological Survey Open File Report*, Vol. 93, pp. 1340.

3 Clark, R. N., Swayze, G. A., Livo, E. K., Kokaly, R. F., Sutley, S. J., Dalton, J. B., et al., 2003. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysic Research,* 108, E9.

4 Cox, S., Botts, M., Robin, A., Davidson, J. & Falke, S., 2006. Observations and Measurements. http://www.opengeospatial.org/standards (accessed 10th April, 2007)

5 Cox, S., Daisey, P., Lake, R., Portele, C. & Whiteside, A., 2004. Geography Markup Language. http://www.opengeospatial.org/standards (accessed 10th April, 2007)

6 Curtiss, B. & Goetz, A. F. H., 1991. Field Spectrometry: techniques and instrumentation. http://www.asdi.com/Field%20Spectroscopy-screen.pdf (accessed 2nd April, 2007)

7 Di, L. 2003. *The Development of Remote-Sensing Related Standards at FGDC, OGC, and ISO TC 211.* Geoscience and Remote Sensing Symposium, Toulouse, France, pp. 643-647.

8 European Commission DG XI. 1993. *CORINE land cover*: European Commission Directorate-General Environment, Nuclear Safety and Civil Protection, Office for Official Publications of the European Communities, Luxembourg.

9 FGDC. The Federal Geographic Data Commitee homepage. http://www.fgdc.gov/ (accessed 11th April, 2007)

10 Geoland Consortium. 2006. Geoland Core Service Land Cover Findings - Rationale for the Geoland Core Service Land Cover -. http://www.gmes-geoland.info/upload/files/CSL-0350-TN-05_CoreServiceRationale-I1.0.PDF (accessed 11th April, 2007)

11 Herold, M., Roberts, D. A., Gardner, M. E. & E., D. P., 2004. Spectrometery for urban remote sensing - Developement and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment,* 91, 304-319.

12 Hüni, A., Nieke, J., Schopfer, J., Kneubühler, M. & Itten, K. 2007, submitted. *2nd Generation of RSL's Spectrum Database "SPECCHIO".* ISMPSRS, 12-14 March 2007, Davos, Switzerland.

13 ISO TC 211. ISO TC 211 homepage. http://www.isotc211.org/ (accessed 11th April, 2007)

14 ISO/IEC. 1996. *Information technology -- Syntactic metalanguage -- Extended BNF (14977).* Geneva, Switzerland, ISO/IEC Copyright Office.

15 Kutser, T., Miller, I. & Jupp, D. L. B., 2006. Mapping coral reef benthic substrates using hyperspectral space-borne images and spectral libraries. *Remote Sensing of Environment,* 70, 449-460.

16 Landgrebe, D., 1997. On Information Extraction Principles for Hyperspectral Data. West Lafayette, Purdue University.

17 Lanz, A., Brändli, M. & Baltensweiler, A. 2007. A Large-scale, Long-term view on Collecting and Sharing Landscape Data. In F. Kienast, O. Wildi & S. Gosh (Eds.), *A Changing World - Challenges for Landscape Research* (pp. 93-111): Springer Netherlands.

18 Michener, W. K. 2000. Metadata. In W. K. Michener & J. W. Brunt (Eds.), *Ecological data: Design, management and processing* (pp. 92-116). Oxford, UK: Blackwell Science.

19 Milton, E. J., 2001. Methods in Field Spectroscopy. www.soton.ac.uk/~epfs/methods/spectroscopy.shtml

20 Milton, E. J., Fox, N. P. & Schaepman, M. 2006. *Progress in Field Spectroscopy.* Geoscience and Remote Sensing Symposium, Denver, CO, US, pp. 1966-1968.

21 Na, A., Priest, M., Cox, S., Botts, M., Robin, A., Walkowski, A., et al., 2006. Sensor Observation Service. http://www.opengeospatial.org/standards (accessed 10th April, 2007)

22 Nieke, J., Itten, K., Debryun, W. & and the APEX Team. 2005. *The Airborne Imaging Spectrometer APEX: From Concept to Realisation.* 4th EARSeL Workshop on Imaging Spectroscopy, Warsaw, pp. 73-80.

23 Pfitzner, K., Bartolo, R. E., Ryan, B. & Bollhöfer, A. 2005. *Issues to consider when designing a spectral library database.* SSC 2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute, Melbourne, pp. 416-425.

24 Pfitzner, K., Bollhöfer, A. & Carr, G., 2006. A standard design for collecting vegetation reference spectra: Implementation and implications for data sharing. *Journal of Spatial Science,* 51, 2, 79-92.

25 Powell, R., Roberts, D. A., Dennison, P. E. & Hess, L. L., 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, 106, 253-267.

26 Price, J. C., 1994. How Unique Are Spectral Signatures? *Remote Sensing of Environment,* 49, 181-186.

27 Richards, J. A. & Jia, X. 2006. Interpretation of Hyperspectral Image Data. In *Remote Sensing Digital Image Analysis* (4th ed., pp. 368). Berlin; New York: Springer.

28 Schlaepfer, D. & Nieke, J. 2007 (in preparation), 23-25 April. *Optimizing the Workflow for APEX Level2/3 Processing.* EARSeL Workshop on Imaging Spectroscopy, Bruges, Belgium.

29 Schmidt, K. S. & Skidmore, A. K., 2003. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment,* 85, 92-108.

30 Schopfer, J., Dangel, S., Kneubühler, M. & Itten, K. 2007. *Dual Field-of-View Goniometer System FIGOS.* ISPMSRS, 12-14 March 2007, Davos, Switzerland.

31 Thenkabail, P. S., Enclona, E. A. & Ashton, M. S., 2004. Accuracy assessment of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment,* 91, 354-376.

32 Thenkabail, P. S., Smith, R. B. & De Pauw, E., 2000. Hyperspectral Vegetation Indices and Their Relationship with Agricultural Crop Characteristics. *Remote Sensing of Environment,* 71, 158-182.

33 van Till, M., Bijlmer, A. & de Lange, R., 2004. Seasonal Variability in Spectral Reflectance of Coastal Dune Vegetation. *EARSel eProceedings,* 3, 2, 154-165.

34 Vane, G. & Goetz, A. F. H., 1988. Terrestrial Imaging Spectroscopy. *Remote Sensing of Environment,* 24, 1-29.

35 Wason, T. D. & Wiley, D., 2000. Structured Metadata Spaces. *Journal of Internet Cataloging,* 3, 2/3, 263-277.

36 Wirth, N., 1977. What can we do about the unnecessary diversity of notation for syntactic definitions? *Communications of the ACM,* 20, 11, 822-823.

37 Zhang, X., Liu, S., Wang, J., Mao, D., Yang, S. & Chen, W. 2004. *Studies on Methods for Quality Assessment of Crop Spectral Data.* Geosience and Remote Sensing Symposium IGARSS 04.