SPECCHIO: a free spectral data management and processing system

Hueni, A ; Kneubühler, M

DOI: https://doi.org/10.1117/12.824695

Originally published at:
Hueni, A; Kneubühler, M (2009). SPECCHIO: a free spectral data management and processing system. In: SPIE Optics and Photonics 2009, San Diego, US, 2 August 2009 - 6 August 2009, 74570A.
DOI: https://doi.org/10.1117/12.824695

SPECCHIO: A Free Spectral Data Management and Processing System

A. Hueni* and M. Kneubuehler
Remote Sensing Laboratories, Department of Geography, University of Zurich, Switzerland

ABSTRACT

The management and storage of spectroradiometer data are important issues, especially in regards of long-term use, data quality and shareability. The SPECCHIO spectral database system developed at the Remote Sensing Laboratories (RSL) provides a solution for the organized storage of spectral data and associated metadata and for the spectral processing based on interactive, customizable and generic processing chains. Optimized data structures and graphical user interfaces combined with intelligent file parsing routines enable the efficient entry of spectral data and metadata. The system can be operated in a heterogeneous computing environment, offering multiuser access to a centralized database and enabling easy data sharing within and even across research groups.

Keywords: Spectral database

1. INTRODUCTION

Spectroradiometer data are collected for the calibration, validation and simulation of imaging spectrometer data and thus form an important part of a complete observing system. Usage of field or laboratory spectra with such a system requires an organized storage and appropriate documentation to retain the data value over longer time and make information shareable among a group of researchers. These requirements are fulfilled by spectral databases, such as the SPECCHIO system [1].

The first version of SPECCHIO [2] was based on an Oracle database for the storage of metadata and a file server holding the actual spectral files. User access was possible via a TCL/TK web interface or by shell script. Operational experience however showed, that the input mechanisms offered by the system were too cumbersome to use and consequently, the amount of spectra being entered into the database remained marginal.

SPECCHIO underwent a total redesign in 2006, enhancing the data model considerably, simplifying the architecture by storing all information in a MySQL database [3] and allowing for more complex software implementations by changing the language of the SPECCHIO application to Java [4]. The main goals of the software design aimed at (a) automating the import of spectral data files by providing special readers, (b) creating metadata automatically by parsing the input files and filling the according tables in the database, (c) speeding up the manual input of metadata by allowing parallel modification of multiple spectral entries, (d) allowing flexible searches in metadata space by generic, real time generation of SQL queries based on user defined constraints, (e) easing export to electronic files for data transfer into other software packages. In a further development step, multiuser capability was added to allow the easy sharing of data within and across research groups while retaining the integrity of data. SPECCHIO remains under constant further development and more recent additions include: (a) the graphical definition of processing chains for operations on the spectral data, (b) special interfaces for handling the administration of instrumentation by system administrators, including instrument and reference panel calibrations and (c) sophisticated import/export functionalities for the data exchange of spectral campaigns between SPECCHIO database instances.

SPECCHIO is freely accessible online on www.specchio.ch with automated user account generation for test and data sharing purposes. The Remote Sensing Laboratories (RSL) provide installation packages for local installations to interested parties.

* ahueni@geo.uzh.ch

2. ARCHITECTURE

The core of the SPECCHIO system is a MySQL [3] database hosted on a database server. Figure 1 shows the setup of the SPECCHIO online database. Database access via the SPECCHIO application is possible by internet/intranet connection to a specific port, given that no firewall restricts the access on the user side. An Apache web server running PHP offers access to various resources and services such as the database user creation, which is automated via dynamic web pages written in PHP. The SPECCHIO application is implemented as a Java2 [4] application, allowing full flexibility on local file system operations and utilization 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.

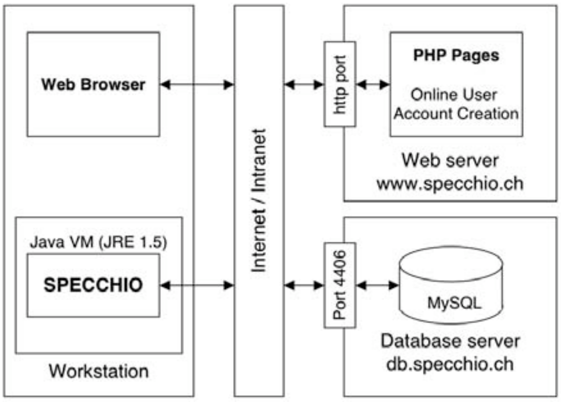

Figure 1: SPECCHIO system architecture of the SPECCHIO online instance

3. MAIN FUNCTIONS

3.1 Data Input

Data input in SPECCHIO consists of two major steps: (a) loading of spectral files and (b) editing of metadata. Spectral files are loaded using specialized reading functions that extract metadata from the input files. A prerequisite is the structuring of the data on the file system. As a minimum, data must be ordered by spectral campaign, but hierarchical structures of any depth are supported. Figure 2 show an example of a campaign storing spectra of New Zealand native plants with spectral files being structured by species and spatial sample location.

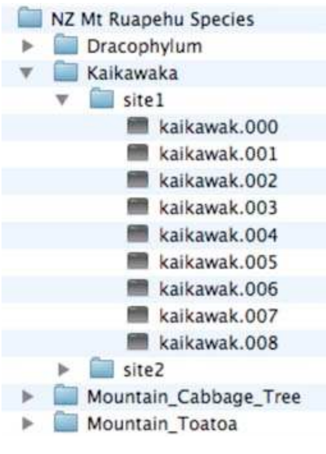

Figure 2: Campaign data structure example

Data of a whole spectral campaign can be loaded into the database by a single operation. The loading mechanism starts by navigating through the structure below a specified campaign directory, parsing all spectral files and storing the folder structure in the metadata. The amount of metadata parameters per primary resources can be defined by the Metadata Space Density (MSD) [5]. For the example of ASD spectroradiometer files [6], MSD figures of around 15 are reached by the automated import.

Metadata not covered by the parsing process have to be edited manually using the SPECCHIO Metadata Editor. It is a powerful tool that allows entering metadata in a streamlined way, thus minimizing the required user input. Metadata values can be applied to several primary resources with a single operation, performing a multiple rows update in the database. Selection of the spectra to be edited is performed using the Spectral Data Browser, a component displaying the hierarchical structure of the data (see Figure 3).

This allows the researcher to edit data efficiently by basing on well-known data structures. Multiple updates are carried out by selecting a hierarchy; all metadata parameter set in this way will apply it to all spectra below the chosen hierarchy.

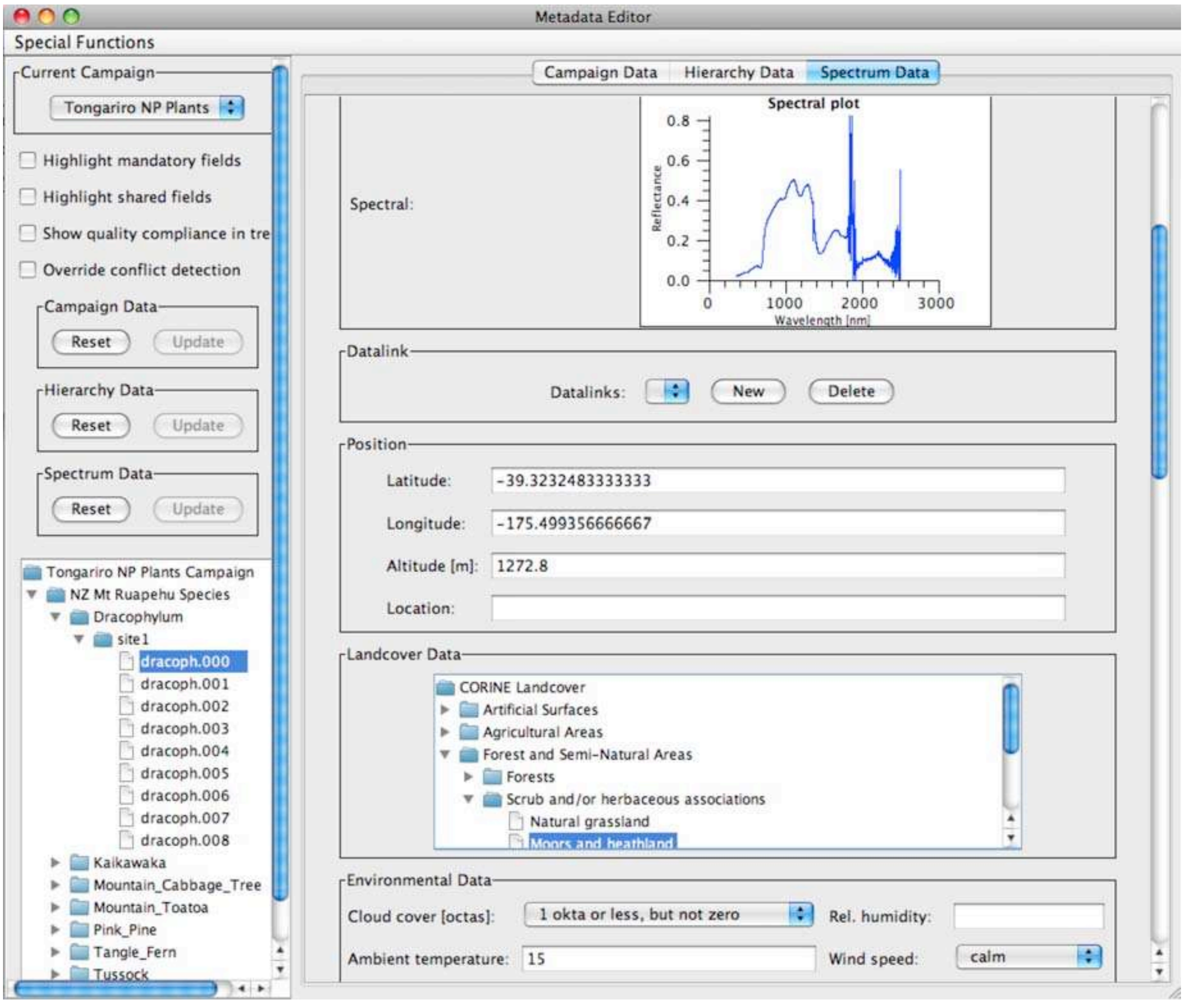

Figure 3: SPECCHIO Metadata Editor with the Spectra Data Browser (bottom left) showing the structure of the 'Tongariro NP Plants' campaign

3.2 Data Retrieval

Data are retrieved from the system by forming SQL select statements based on constraints given by the user. The SPECCHIO Query Builder is used to define constraints in two ways: by utilizing the Spectral Data Browser or by defining explicit constraints per metadata space dimension. In both cases, SQL select queries are automatically built and executed in real-time to provide the number of involved spectra. Figure 4 shows the Query Builder in the 'Query conditions' mode, with 37 spectra being the result of the following specified constraints: campaign name, landcover and spatial position. The definition of constraints performs what is known as subspace projection. The spectra contained within the subspace may then be used as input for further functions, which are: (a) showing an online report, (b) exporting the spectra to a file, (c) loading the spectra into a Space for use in the Space Network Processor and (d) calculating the mean and the standard deviation Metadata Space Density (MSD) for the selected spectra.

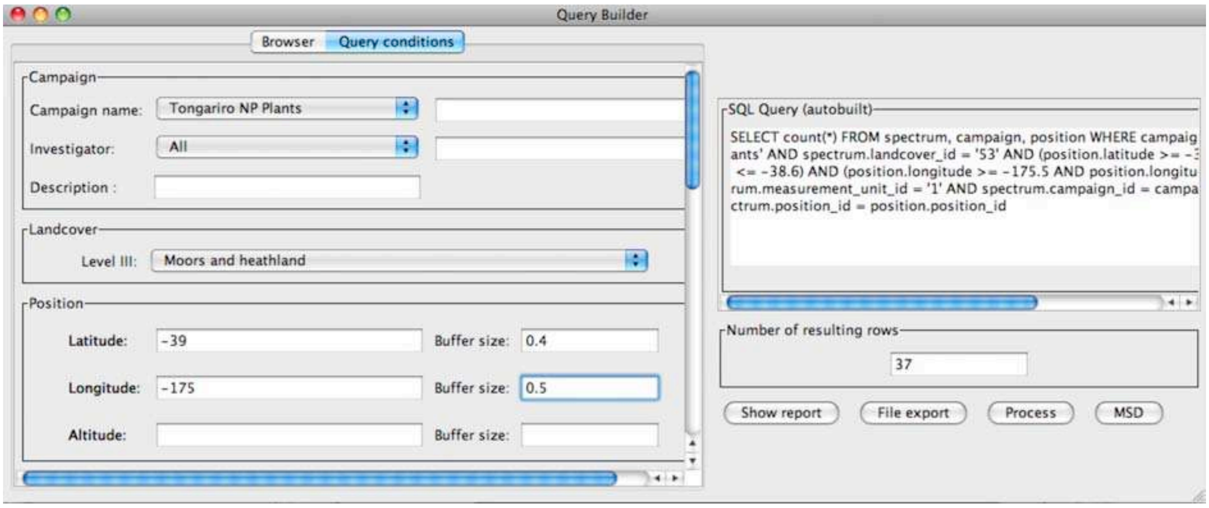

Figure 4: SPECCHIO Query Builder

3.3 Data Processing using the Space Network Processor

Starting with version 2.0 SPECCHIO offers interactive, configurable data processing. The concept is based on the feature spaces [7] and complex process flows can be realized by building networks consisting of spaces and processing modules [8]. Spaces are used throughout the system for processing, visualization and file output. In all these cases, vector data must be related to spectral dimensions; this information is held by the space.

Interactive, flexible and configurable data processing is based on the concept of the Space Network. Such networks consist of processing modules and data sinks/sources, connected by directed edges.

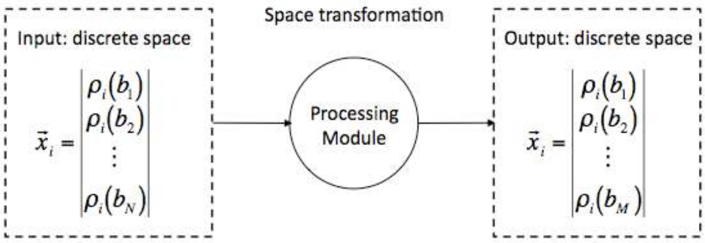

Figure 5: Transformation into a new space by a processing module

Processing modules are effecting a transformation on a space, i.e. the spectral data vectors of the input space are transformed to an output space. The algorithm of the processing module defines the dimensionality of the resulting space. This is illustrated in Figure 5 with an input space of dimensionality N being transformed into another discrete space of dimensionality M. Although processing modules tend to have singular input/output in most cases, they may have multiple inputs and generate multiple outputs.

The starting spaces are created by queries in the Query Builder. Attaching processing modules to a space automatically creates a new space, holding the output of the respective module. Visualization and file export modules may be attached to any space, allowing the detailed monitoring of processing stages. The advantage of processing data in the Space Network processor is the capability of using information contained in the Metadata, such as the Spectralon panel used during a sampling campaign. Figure 6 shows an example of a simple processing chain that selects and plots the correction factors for the Spectralon panel used during data capture of the spectra contained in the input space.

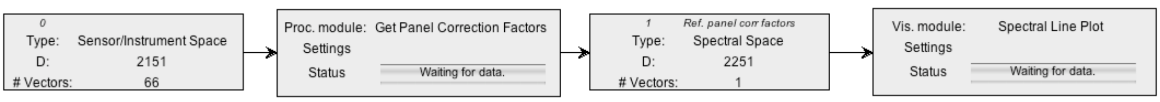

Figure 6: Processing chain selecting the correction factors and plotting them as a spectrum.

3.4 Data Output

Data contained in a Space resulting from selections in the Query Builder or from a processing module in the Space Processing Network can be output in several ways: (a) as spectrum report, (b) as simple plot, (c) in a special GUI for data exploration or (d) as electronic file.

Figure 7 shows examples of a spectrum report and a spectral plot generated from a Space Network Processor module. The spectrum reports comprise a spectral plot plus a list of all metadata parameters for the spectrum.

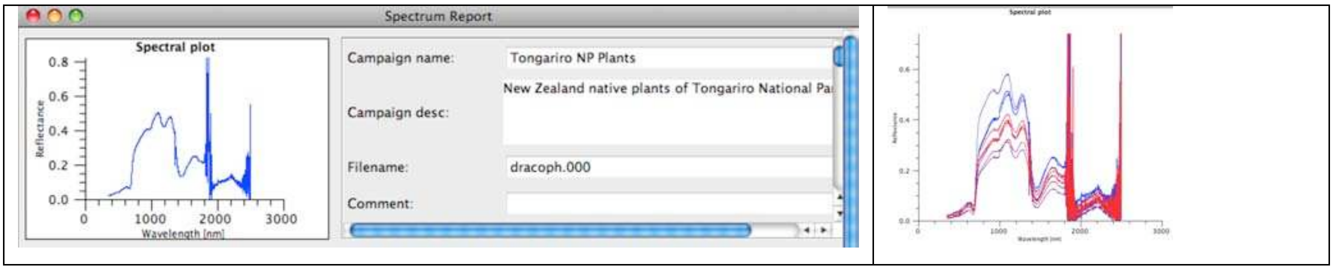

Figure 7: Spectrum report (left) and spectral plot generated from a Space Network Processor module

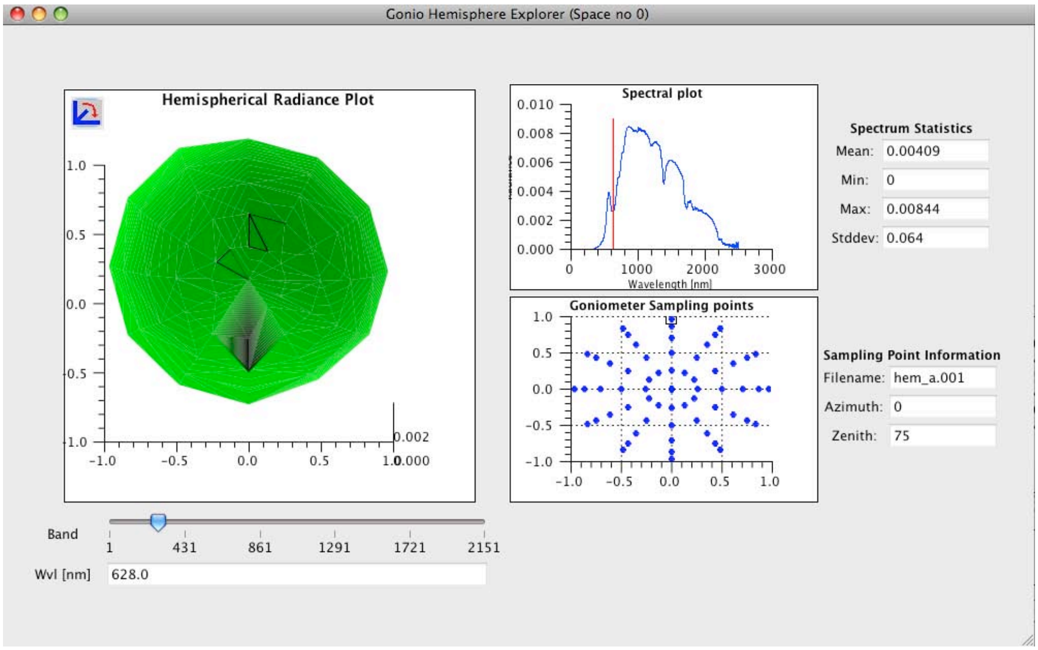

Figure 8: Goniometer Hemisphere Explorer

Data exploration modules are interactive graphical user interfaces written to allow the examination of data. An example is the Goniometer Hemisphere explorer module. The Hemisphere Explorer allows the interactive exploration of a spectrodirectional dataset, typically acquired by a goniometer system. Figure 8 shows an explorer window displaying a LAGOS (Laboratory goniometer system) dataset [9]. The explorer window comprises a number of components interacting with each other. The hemispherical plot displays an interpolated surface of a specific wavelength, which can be modified with a band selection slider. The sampling points plot shows the sampling point positions projected onto a 2d Cartesian coordinate system. Information about the currently selected point is shown in the text fields on the right of the plot: filename of the respective spectrum, azimuth and zenith angles of the observation geometry. Changing the selected sampling point changes the spectrum displayed in the spectrum plot automatically. The spectral plot component displays the spectrum of the selected sampling point. A red, vertical line indicates the current wavelength as selected by the band selection slider. The text fields on the right of the plot display spectral statistics of the current spectrum.

Two output formats are implemented: (a) CSV (Comma Separated Value) files that can be read by statistical and spreadsheet applications and (b) ENVI Spectral Library files that are primarily a data format used by ENVI [10] but can be read by other remote sensing packages as well.

4. ADMINISTRATOR FUNCTIONS

4.1 Instrumentation Management

SPECCHIO supports the management of instrumentation from version 2.0 onwards. The SPECCHIO schema can store information about instruments and references, which can be linked to spectra of various campaigns. Definition of instrumentation data is restricted to system administrators, as removing or changing existing instrumentation may affect all campaign related information in the database.

Instruments have defined owners, serial numbers and names and may also be described by pictures. The serial numbers can be used during input file parsing to automatically set the used instrument in the metadata. Future extensions will also allow the definition of calibration information similar to the current implementation for references.

References can be of a certain type (e.g. Spectralon), have defined owners and identification numbers and may have linked pictures as well. Calibration information can be loaded into the database using a CSV file format, defining the correction factors and the associated uncertainty in respect to a primary standard. The correction factors can then be applied during processing in the Space Network Processor with their automated selection from the database based on timeline information.

Both instruments and references are managed using the Instrumentation Metadata Editor (Figure 9). It allows the easy definition of new and the update of existing instrumentation.

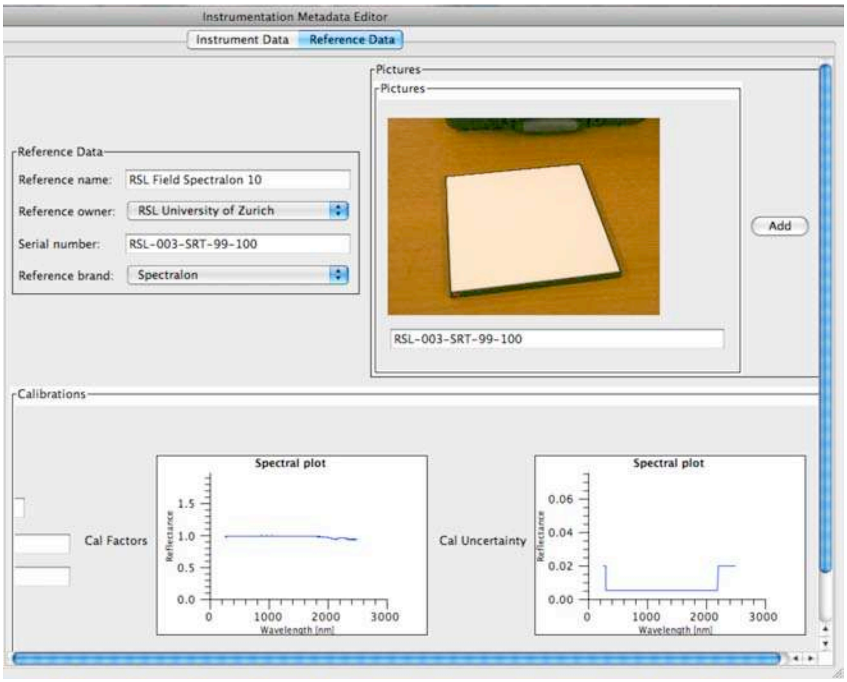

Figure 9: SPECCHIO Instrumentation Metadata Editor showing the metadata of a Spectralon reference panel

4.2 Data Export/Import

SPECCHIO was originally envisaged as a centrally hosted entity; however, experience has shown that some users favour individual installations, both internally and externally. The advantage of decentralized databases is their use independently of network connections, such as under field conditions, and the full control over the data access. Nevertheless, the goal remains to eventually store these separate datasets in a central database, where all users may gain access. Furthermore the transfer of datasets between different SPECCHIO databases is of interest for the purpose of data sharing within or across groups without involving a central server. For some users, a full export of all their data may not be wished for and hence only selected data collections will be available for import into the central database.

To enable such data transfers, partial database import/export functions have been implemented in SPECCHIO to allow the selective data exchange between distributed spectral databases while retaining the full data context in its relational form [5]. The metadata context even extends to system entries, such as instrumentation information, thus, if the exported spectra refer to a reference panel, all relevant calibration factors will be included in the export.

The actual transfer of data is based on XML style files, which can be loaded into the target system, creating identical copies of the source dataset.

5. GIS CONNECTIVITY

The SPECCHIO database can be connected to GIS systems such as ArcGIS [11] via ODBC (Open Database Connection). Any metadata can then be displayed in connection with the spatial position of the spectra. Figure 10 shows the creation of a map in ArcMap, using a direct connection to RSL's inhouse SPECCHIO database. The 'Attributes' window shows part of the position records, which were joined with the spectrum table to get the latitude, longitude and location name for each spectrum.

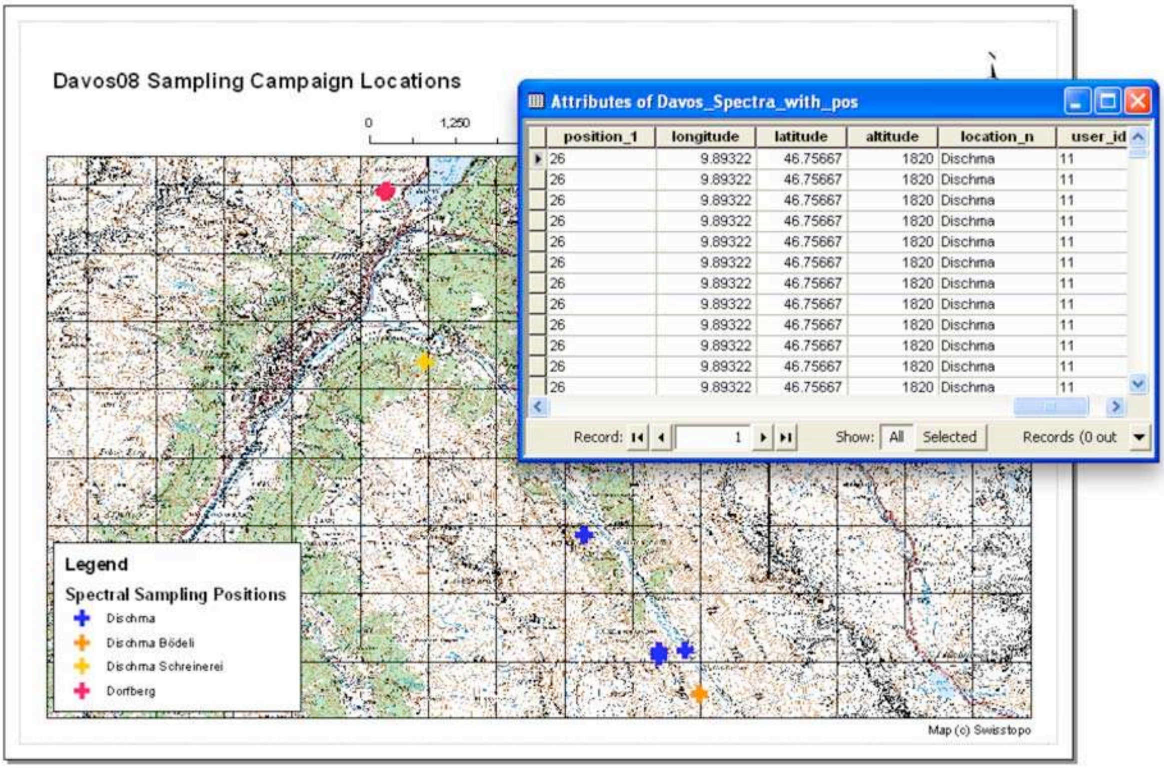

Figure 10: Creation of a map in ArcMap showing the spatial positions of spectral samples

6. SPECCHIO ONLINE

RSL is providing public access to the SPECCHIO system on the Internet and around 60 researchers from all around the world have already obtained accounts for the online system (See Figure 11 for a spatial representation of user numbers per country). The main purposes are to give users easy possibilities to test the system and to create a platform for sharing of spectral data. Access to the online system requires a user account, which can be created online using the respective forms on www.specchio.ch. User name and password are automatically sent to applying user via email. For the actual database access it suffices to download the SPECCHIO Java application from the website, requiring no further installation steps.

Two database instances are currently available to the public on the db.specchio.ch database server: (a) specchio and (b) specchio_test. The former is the productive server and holds well described data to be shared. The latter is provided for users wanting to work through the provided tutorial session or run some tests of their own.

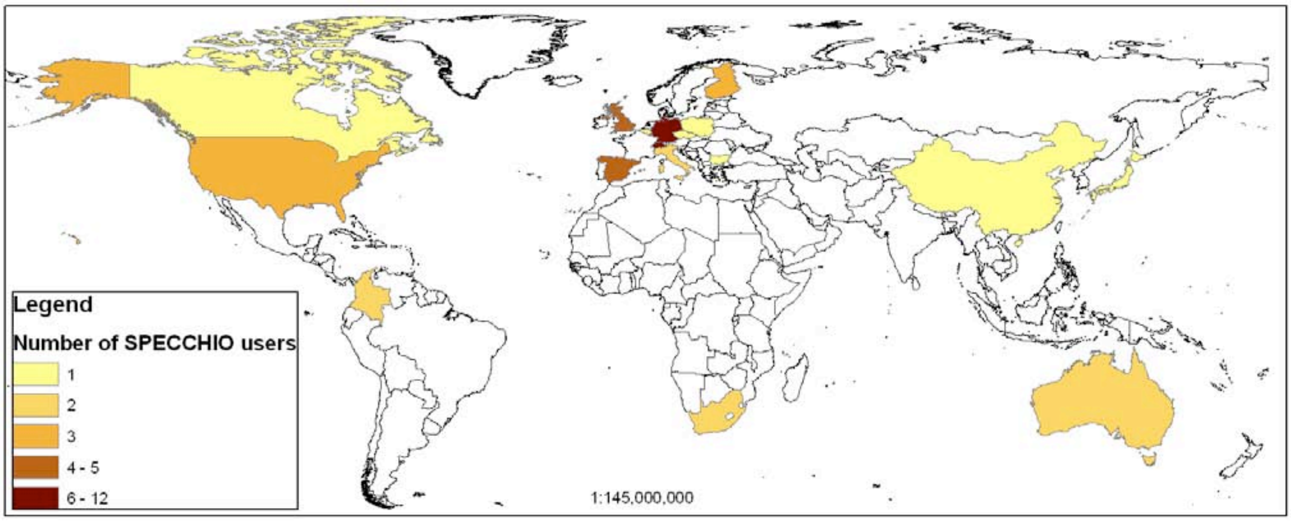

Figure 11: Number of SPECCHIO online users per country (Date: 5th July 2009)

7. CONCLUSIONS

SPECCHIO is a spectral database system, designed to store spectral data and associated metadata. The underlying technologies of MySQL and Java allow its use in heterogeneous computing environments. The major features include the automatic extraction of spectral and metadata from spectral input files, efficient editing of metadata, flexible search queries, user configurable processing networks and various data exploration and export functions. The storage of comprehensive instrumentation information provides the basis for data calibration and can potentially be used for data uncertainty estimations in future versions.

SPECCHIO remains under constant development and users are encouraged to test the system and provide their valuable input regarding further functionalities.

REFERENCES

[1] A. Hueni, J. Nieke, J. Schopfer *et al.*, "The spectral database SPECCHIO for improved long term usability and data sharing," Computers & Geosciences, 35(3), 557-565 (2009).

[2] S. Bojinski, M. Schaepman, D. Schlaepfer *et al.*, "SPECCHIO: a spectrum database for remote sensing applications," Computers & Geosciences, 29, 27-38 (2003).

[3] MySQL AB, [MySQL], (2005).

[4] Sun Microsystems Inc., [JavaTM 2 Platform Standard Edition], Santa Clara, CA(2006).

[5] A. Hueni, T. Malthus, M. Kneubuehler *et al.*, "Data Exchange between distributed Spectral Databases," Computers & Geosciences, (In preparation).

[6] Analytical Spectral Devices Inc., [Technical Guide] Analytical Spectral Devices Inc., Boulder, CO(2007).

[7] D. Landgrebe, [On Information Extraction Principles for Hyperspectral Data] Purdue University, West Lafayette, IN(1997).
[8] A. Hueni, M. Kneubuehler, J. Nieke *et al.*, "PROCESSING EXTENSION FOR THE SPECTRAL DATABASE SPECCHIO."
[9] J. Schopfer, S. Dangel, M. Kneubühler *et al.*, “The Improved Dual-view Field Goniometer System FIGOS,” Sensors*,* 8(8), 5120-5140 (2008).
[10] Research Systems Inc., [ENVI], Boulder, CO, USA(2005).
[11] ESRI, [ArcGIS], Redlands, CA, USA(2006).