ELSEVIER

ISPRS Journal of Photogrammetry & Remote Sensing 57 (2002) 204–211

PHOTOGRAMMETRY & REMOTE SENSING

www.elsevier.com/locate/isprsjprs

SPECCHIO: a Web-accessible database for the administration and storage of heterogeneous spectral data

S. Bojinski*, M. Schaepman, D. Schläpfer, K. Itten

Remote Sensing Laboratories (RSL), Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Received 20 July 2000; accepted 30 July 2002

Abstract

The administration of spectral remote sensing data is a key factor for thorough, comparative analyses which take into account the entire information available to the investigator. Today, a great variety of spectrometers operate in the field, giving rise to a strong heterogeneity in data and meta-data formats. For a holistic view on these data, a spectrum database must adhere to certain principles: independence of file format, flexibility to attribute changes, establishment of relations between data, content-based search capability, common interfaces and scalability. SPECCHIO as a Web-accessible spectrum database represents a new approach that incorporates the mentioned principles and overcomes the drawbacks of file-based solutions. It is based on a relational data model and provides a prototype that accounts for the complexity of heterogeneous spectral data. SPECCHIO can be used for administration and storage purposes, as well as for online data inspection. It may also serve as a reference database for the characterization of Earth surface targets.

Keywords: spectrum database; web access; reference spectral collection

1. Introduction

The high diversity of spectral measurements in remote sensing gives rise to a great variety of spectral data as well as meta-data formats, where the latter describe the former by a set of attributes. Meta-data, e.g. date and location of a measurement, are frequently being stored in a file header, which is defined by some standard. There are several examples of format standards for spectral data other than columnar ASCII, such as HDF-EOS (NASA1, 2002), JCAMP (McDonald and Wilks, 1988), or ENVI spectral library (RSINC, 2002).

Along with a file-based organization of the data comes a file system tree that provides a logical organizational structure. However, experience shows many limitations of this approach, especially for large amounts of data, as it poorly represents the whole complexity of information prevailing in remote sensing applications. As an example, given spectral measurements from many field campaigns, users can easily lose sight of the entirety of information available to them. For a comprehensive analysis of spectra, an intercomparison of data from different sensors is desired, but hard to achieve in practice. Even if spectra and meta-data are available online, effective

* Corresponding author. Tel.: +41-1-6355149; fax: +41-1-6356846.

E-mail address: sbojinsk@geo.unizh.ch (S. Bojinski).

0924-2716/02/$ - see front matter
PII: S0924-2716(02)00119-3

queries and compact downloads remain a challenge in file-based solutions. Most of the references given in Section 2 of this work, however, use the file-based method of data administration.

Six principles can be identified that are to be fulfilled by a spectrum database that account for the heterogeneity of spectral remote sensing data: independence of file format, flexibility to attribute changes, establishment of relations between data, content-based search capability, common interfaces and scalability.

In this paper, we present the spectrum database SPECCHIO (Spectral Input Output) as a new prototype that matches the abovementioned criteria. It consists of a relational database system and appropriate user interfaces, script template-based and Web-based. Section 3 presents the technology, Section 4 the conceptual side of the system. Section 5 provides an insight to the interfaces to the database, followed by an outlook that outlines applications for SPECCHIO in remote sensing.

2. Overview of spectral remote sensing data resources

There exists a large number of online remote sensing data archives that can be accessed by the scientific or commercial user to query and order the data needed. For the most part, these sites are set up by the data distributor for a particular sensor or mission (e.g. USGS1, 2002; DLR, 2002) and provide very large collections of airborne or spaceborne spectral scenes. The granularity of data is usually at scene level and does not break down spectral information to the single spectrum.

Only two publicly accessible sites that contain single spectra retrievable in ASCII format could be identified: the USGS Denver Spectroscopy Lab (Clark et al., 1993; USGS2, 2002), which focuses on mineral spectra and the ASTER Spectral Library (NASA2, 2002) which includes about 2000 spectral entries from natural and man-made materials. Price (1995) created an offline collection of over 3400 spectra with reference character from various sources in ASCII format. These spectral collections are all file system based.

MedSpec (Preissler et al., 1998) is a stand-alone spectrum database solution that features a large number of meta-data attributes for campaign data. It allows input and query of spectral data in ASCII format only, as well as of ancillary data.

3. SPECCHIO: technology

Web access to databases is generally being established by (1) an application programming interface that enables communication between a programming language and a database and (2) a method to call executable programs on a Web server. As for (1), the scripting language TCL (TCL, 2002) with a database interface extension is used to process structured query language (SQL) database queries. (2) has been implemented using the Common Gateway Interface (CGI) (Gundavaram, 1996) for Web servers as provided by the Webshell (WEBSH, 2002) Web application interface. All of these programming tools except the database software are developed and distributed as Open Source software (OSS, 2002).

We use this approach for reasons of implementation speed and ease of maintenance. It uses the standard HTTP protocol for all communication between client and server. Users supply query or update parameters by means of HTML forms, and these parameters are sent in HTTP to the Web server, which calls a CGI program to interpret the request and establish a connection to the SPECCHIO database. The query results provided by the database are encoded in HTML documents and returned to the client. Afterwards, it is possible to select individual data sets for run-time-generated graphics display in the browser window, technically realized by a background process that generates GIF files, and for download to the local file system.

In many scientific applications, automated processing tasks are called for, performed by executable scripts. This gives rise to a script-based feed and query option to access the database. These scripts directly connect to the database and execute the requested transactions. Output is written into files on the local file system.

The relational database system Oracle 8i (ORA, 2002), accessed by transactions coded in embedded SQL, is particularly suited for the consistent and reliable storage and backup of large amounts of data. Processing and display of spectral data has been

implemented with IDL (RSINC, 2002). Currently, SPECCHIO runs on a UNIX/Sun Solaris system with an Apache Web server.

4. SPECCHIO: data model

The major goal of SPECCHIO in terms of data description is breaking down the 'spectral remote sensing data' aspect of reality to a data model in order to manage all spectral remote sensing data and metadata within one single relational database instance. In doing so, data security issues can be delegated to the database management system, such as the integrity of data, access control, consistency and the recovery of data in case of hardware failure. Furthermore, the stored information can be efficiently retrieved by the SQL statements.

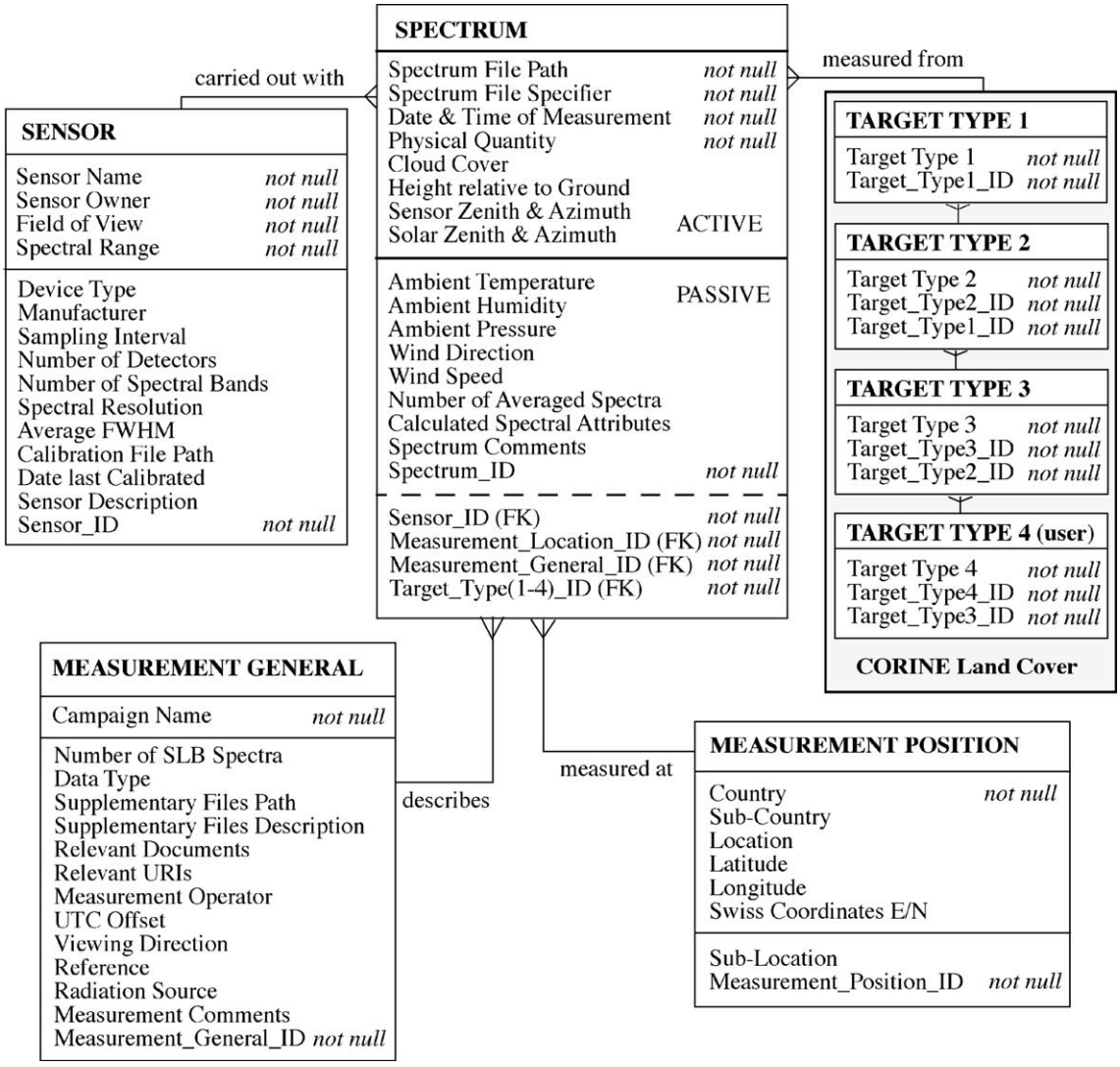

Fig. 1. Entity–relationship diagram reflecting the data model, which describes the 'spectral remote sensing data' aspect of reality. The principal 'Spectrum' table is related to the surrounding tables by foreign keys (FK) defined on the respective IDs. 'Active' attributes can be queried as opposed to 'Passive' attributes; 'not null' tags denote compulsory attributes. The hierarchical target type structure (Target Type 1–3) represents the CORINE Land Cover Schema (European Commission DG XI, 1993).

In every relational database, all information is stored in two-dimensional tables that consist of columns and rows. Tables represent semantic entities, columns semantic attributes of an entity and rows the actual data sets. Data cells that are defined by column name and row number contain atomic values. The data model consists of entities that are specified by attributes and logically related by primary and foreign keys, all of which can be displayed in an entity–relationship diagram (Fig. 1). The relations are in the third normal form to preclude update anomalies (Date, 1995).

Grouped around the main entity 'Spectrum' are four other entities that contain meta-data, 'Sensor', 'Measurement General', 'Measurement Position' and 'Target Type' that are related to 'Spectrum' by 1:n relations. This means that each 'Spectrum' data set is uniquely related to exactly one entry in the 'Sensor' table, 'Measurement General' table, etc. Vice versa, an entry in the 'Sensor' table can be related to one or many entries in the 'Spectrum' table, and accordingly. As a result, a new entry into SPECCHIO is equivalent to $m \geqslant 1$ entries or updates in the 'Spectrum' table, associated with $m \leqslant n$ entries or updates in each meta-data table.

Within each entity, a distinction is made between active and passive attributes. Active attributes can be queried in the query interfaces; passive ones come along as the result of a query transaction, but cannot be queried in their own right. As a quality control measure for the database, a number of attributes has been defined as not null, i.e. they must be stated at all times.

SPECCHIO stores data independent of sensor types and file formats. The actual data defining the physical quantity of the spectrum physically remain outside the database, addressed online by a file location pointer, or offline by an external medium identifier. This concept makes the database server a 'thin' server. Access to the database is provided by scripts as well as by standard Web browsers, which are shown in Section 5.

5. Interfaces to SPECCHIO

Interaction with the SPECCHIO database instance is intended to be easy to implement and widely usable. For remote and interactive access, dynamically created Web forms provide a universal way to communicate with the database via a standard browser. Alternatively, for automated tasks, predefined script templates can be employed.

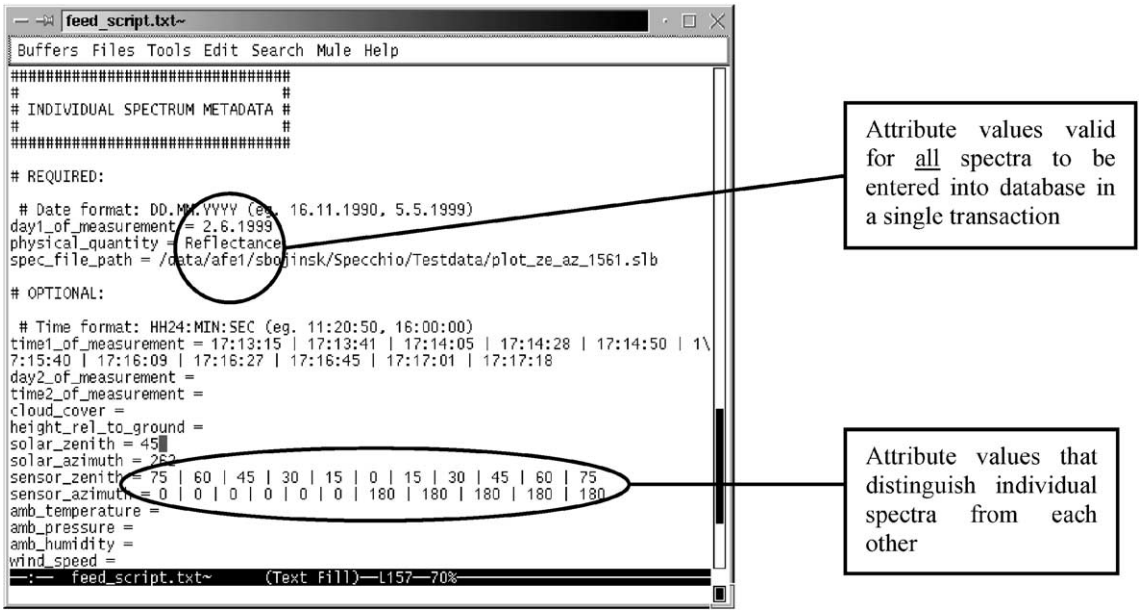

Fig. 2. Feed input file syntax.

5.1. Feed data

The Web-based feed site grants a user-friendly possibility to enter information into the database. Returned dynamically by the server, the feed site shows the current range of values already present in the database for attributes with relatively small semantic variation. This feature reduces semantic redundancies, i.e. different concepts with the same meaning are prevented from being entered and conceived as different by the database. For data quality reasons, it is advisable to give as many attributes as possible, though some are compulsory and denoted as such. Currently, the implementation is mainly intended for small numbers of spectra to be entered. The issue of large spectral sets in which individual spectra differ only in one or two attributes (e.g. sensor angles) is still to be considered closely.

The script-based feed is mainly intended for automated or large-scale input to the database (Fig. 2). It is currently optimized for spectra in the ENVI spectral library format. In principle though, any set of spectra stored in a number-distinguishable way can be entered, as they are automatically allocated a spectral specifier.

The feed Web site may also serve as a database view to support the script-based alternative in terms of semantic redundancy. This means that users can look at existing attribute values in the database (e.g. 'Switzerland' as value for the attribute 'Country' in the 'Measurement Position' entity) and align their feed script terminology accordingly.

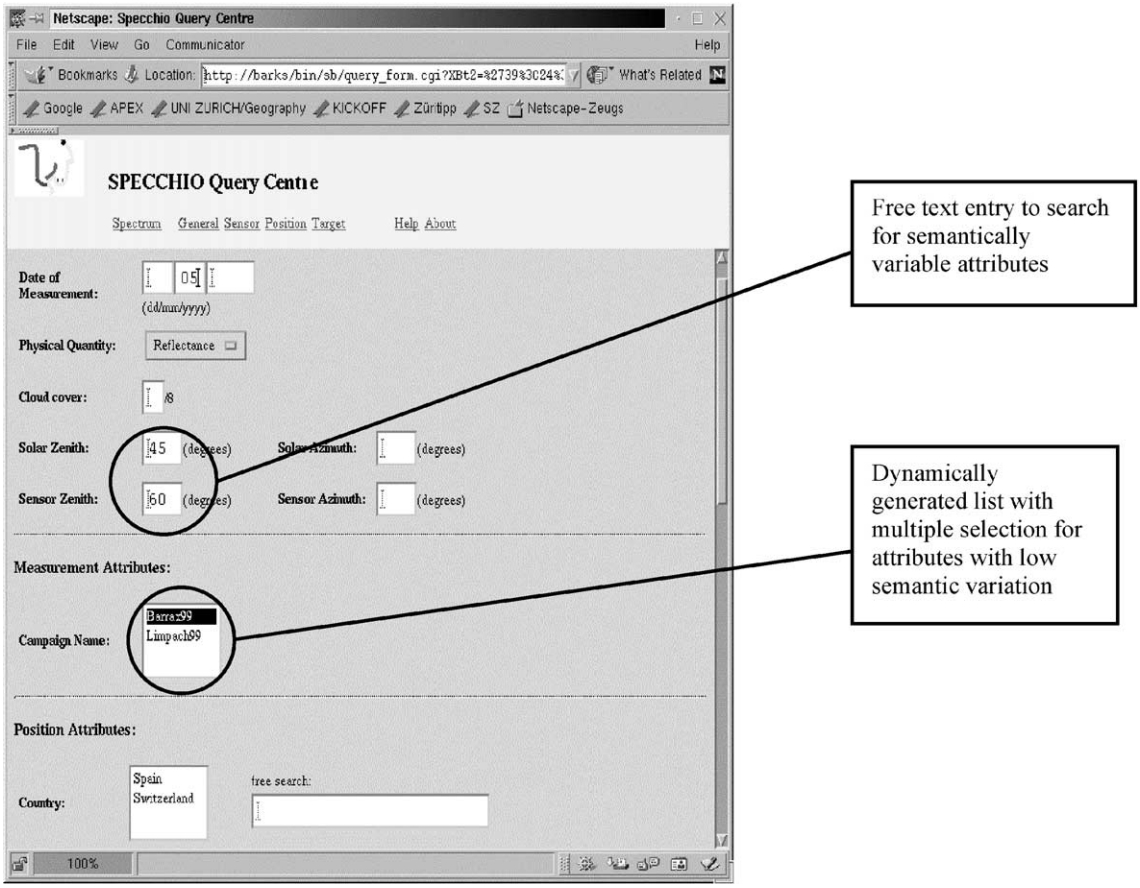

Fig. 3. Query Web site with different attribute input modes.

5.2. Query data

The majority of SPECCHIO users may have questions directed to the database that are translated into a combination of selected attributes. For example,

- Who measured GER 3700 barley spectra in 1997, and how exactly were they measured?
- Show me all the grass spectra measured at sensor angle 45°!
- Compare all bare soil spectra measured with the ASD device during the Barrax campaign!

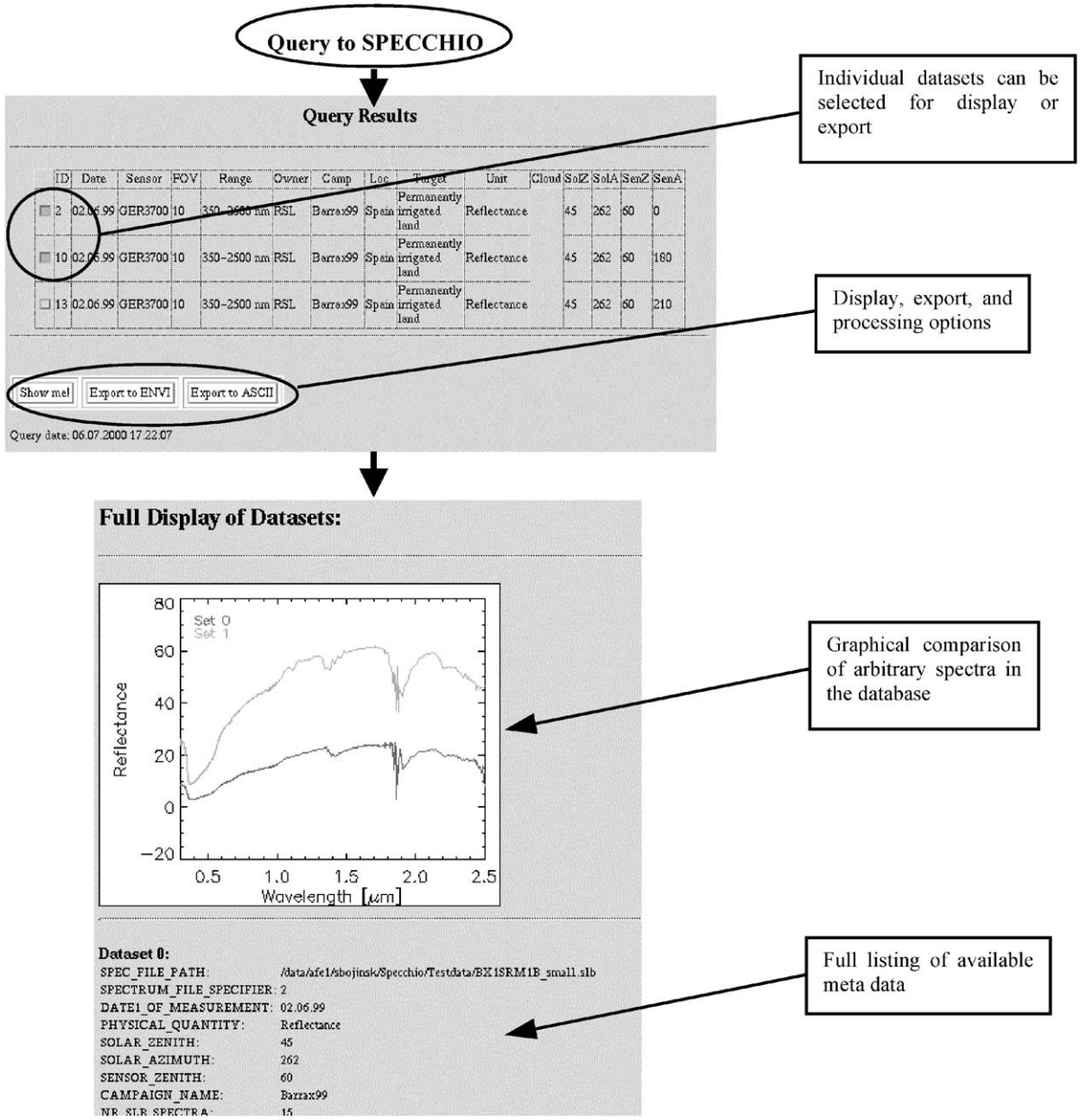

Fig. 4. Query and display of spectral data.

For fast and possibly automated query of the database, a script solution exists, which is syntactically similar to the feed script. The output consists of a text file containing a list of data sets that match the search criteria, including all the associated meta-data.

The Web site allows an interactive one-step query setup for a selection of attributes which are considered to be most important (see Fig. 3). Due to procedural modular internal programming, the appearance of the query site can be altered easily. Some multiple attribute lists are dynamically generated and form an up-to-date and distinct selection of attribute values, enabling more precise searching, as opposed to free text entries.

For both query interfaces, a specification of different attributes is logically equivalent to a Boolean AND operator; a multiple statement within a single attribute corresponds to an OR operator. If the Web site query yields data sets that match the search criteria, a clickable table shows up with columns corresponding to the searchable attributes (cf. Fig. 4). It is then possible to select data sets for graphical display and full statement of all associated meta-data, as well as download to the local system. Spectral plots can be chosen as single curves or overplots of all the selected spectra. The display feature as well as the export of spectral files to a file system is currently enabled for ENVI spectral library and two-column ASCII data formats.

5.3. Advanced features

For the moment, query and feed interfaces to the database are implemented on a prototype level. In many applications of imaging spectrometer data, a demand for information on spectral features or the statistics of spectral ensembles arises. Therefore, a low level of processing functionality is going to be added to SPECCHIO, aimed at the retrieval of spectral values for predefined spectral channels, or the calculation of statistical key values. The latter can as well be stored in the database to save processing time in subsequent applications. Accordingly, biophysical parameters, such as leaf area index (LAI), can be calculated for a selection of spectra and subsequently stored.

SPECCHIO truly serves as a reference database, once the full range of statistical functions exists. Another planned feature is the extension of display options after the selection of a query result, which include scaling and resampling of spectra. Lastly, the semiautomated input and output of selected data formats is realized. The routines called by the database to perform calculations are written in IDL or, if time critical, in C/C++.

6. Outlook

We have presented a prototype for a Web-accessible database for the storage and administration of spectral remote sensing data. After conceptual and technological considerations, a data model has been developed which, to our knowledge, comprehensively represents the field of spectral remote sensing. Large amounts of data with a great variety of meta-data can be stored independently of the data format. User interfaces to the database are shown to demonstrate the state-of-the-art functionality of SPECCHIO. Full implementation, documentation and a users' test phase of the software are the next steps to be taken. The demand for processing features that directly address the physical spectral data gives rise to the development of external routines.

The practical use of SPECCHIO is going to be manifold and beyond mere systematic storage of data: the problem of endmember selection and interpretation in the process of image classification requires reference spectra taken from exemplary targets, which can be automatically drawn from the spectrum database. More specifically, segmentation of imaging spectrometry data and subsequent spectral unmixing of segments can be greatly facilitated. This supports the analysis of bidirectional reflectance properties of the Earth's surface, as well as the characterization of atmospheric variability within imaging spectrometry data.

Acknowledgements

This work was kindly made possible by a Swiss Science Foundation research grant, contract number 2000-050727.97. We would like to thank two anonymous reviewers for their valuable comments.

References

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 μm. USGS Open File Report 93-592, USGS, Denver. 1340 pp.

Date, C., 1995. Introduction to Database Systems, 6th ed. Addison-Wesley, Reading. 839 pp.

DLR, 2002. http://ims.dfd.dlr.de (accessed 24 April 2002).

European Commission DG XI (Environment, Nuclear Safety and Civil Protection) (ed.), 1993. CORINE land cover—Technical guide. EUR 12585 EN, Office for Official Publication's of the European Communities, Luxembourg. 136 pp.

Gundavaram, S., 1996. CGI Programming on the World Wide Web. O'Reilly, Sebastopol. 436 pp.

McDonald, R., Wilks, P., 1988. JCAMP-DX: a standard form for exchange of infrared spectra in computer readable form. Appl. Spectrosc. 42 (1), 151–162.

NASA1, 2002. http://hdfeos.gsfc.nasa.gov (accessed 24 April 2002).

NASA2, 2002. http://speclib.jpl.nasa.gov (accessed 24 April 2002).

ORA, 2002. http://www.oracle.com (accessed 24 April 2002).

OSS, 2002. http://www.opensource.org (accessed 24 April 2002).

Preissler, H., Bohbot, H., Mehl, W., Sommer, S., 1998. MedSpec, a spectral database as a to support the use of imaging spectrometry data for environmental monitoring. Proc. 1st EARSeL Workshop on Imaging Spectrometry, Zurich. European Association of Remote Sensing Laboratories, Paris, pp. 455–462.

Price, J.C., 1995. Examples of high resolution visible to near-infrared reflectance spectra and a standardized collection for remote sensing studies. Int. J. Remote Sens. 16 (6), 993–1000.

RSINC, 2002. http://www.rsinc.com (accessed 24 April 2002).

TCL, 2002. http://dev.scriptics.com (accessed 24 April 2002).

USGS1, 2002. http://edc.usgs.gov/products/satellite.html (accessed 24 April 2002).

USGS2, 2002. http://speclab.cr.usgs.gov (accessed 24 April 2002).

WEBSH, 2002. http://www.websh.com (accessed 24 April 2002).