DATA EXCHANGE BETWEEN SPECTRAL DATABASES

Andreas Hueni, Mathias Kneubuehler & Michael Schaepman

Remote Sensing Laboratories, Institute of Geography,
University of Zurich, Switzerland
ahueni@geo.uzh.ch

1. EXTENDED ABSTRACT

The vision of data sharing and long-term usability of spectral ground data is enticing, as it will provide a wealth of information at a low cost when compared to project specific data acquisitions [1]. However, this vision is dependant on a number of factors including rigorous documentation of the sampling process by standardised protocols, organised storage of spectral signatures and associated metadata and development of data quality indicators. All these factors are co-related and must be considered when designing spectral databases. Only when sampling and documenting procedures have been established, storage systems coping with the accumulated data have been implemented and the quality may be assessed will data sharing move beyond the research realm and into operational use [2].

Spectral databases are essentially repositories for spectral data and metadata and have attracted increased interest of the remote sensing community over the past few years. With consequent installations of separate instances of spectral databases within and across institutes, the issue of data exchange has become more pressing. Typical use cases include: (a) operation of a central spectral data repository with constant intra- or internet accessibility, (b) parallel operation of several server based spectral database instances for e.g. operational, educational and testing purposes, (c) instances installed on local workstations for individual use by researchers and (d) instances run on field laptops for documentation of ongoing field campaigns with little or no network access. All these use cases may exist in parallel and lead to the need for data exchange, such as: transfer from distributed systems to a central server, replications of existing datasets for test and educational purposes or specific copy of a dataset between research partners with no common database infrastructure. In all these case, one may assume the existence of a full metadata set describing the primary resource, i.e. the spectral signature. The process of data exchange must therefore retain the full metadata context of the spectral information, as therein lies the key to successful data sharing and long-term use [3].

In this paper we present the concepts for data exchange between distributed spectral databases of the same schema in normalised form while conserving the full metadata context. Metadata of spectral information are essentially redundant for many parameters, e.g. the target descriptions are identical for all spectra obtained by multiple scans of the same object. The consequent use of relational databases necessitates the normalisation of the data model during the design step; in this process, redundancies are removed and data stored on their normalised form. The normalised form not only reduces the needed storage space but also aids data consistency and speeds up data entry if supported by intelligent graphical user interfaces. As a consequence, data transfer concepts should aim at preserving the normalised form to minimise storage size and enable the generation of an exact copy in the target system. Any denormalisation step, such as would be needed for export to a table based file format would substantially increase the amount of data and the exact restoration of the original normalised form would be exceedingly difficult [4-6].

The concepts for automated extraction of the full metadata context, storage of spectral information and metadata in XML format and non-conflicting import into target systems have been implemented in the SPECCHIO [7] spectral database and offer users the exchange of spectral campaign data with other SPECCHIO instances using an XML style data format, thus allowing data transfer between machines with no live database connection.

The presented concepts hold essentially also true for the data exchange between systems of differing schemata. However, an additional abstraction layer providing mappings of parameters from one schema into another would be needed. According efforts for the generation of such a layer should be combined with standardisation processes regarding the recorded metadata and the utilised sampling approaches.

11. REFERENCES

[1] K. Pfitzner, A. Bollhöfer, and G. Carr, "Standards for reflectance spectral measurement," Australian Government, Department of the Environment, Water, Heritage and Arts, Supervising Scientist, Darwin, NT, Australia, 2009 (in preparation).

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

[3] A. Hüni, J. Nieke, J. Schopfer, M. Kneubühler, and K. Itten, "Metadata of Spectral Data Collections," in *5th EARSeL Workshop on Imaging Spectroscopy*, Bruges, Belgium, 2007, p. 14.

[4] J. Fong, "Methodology for schema translation from hierarchical or network into relational," *Information and Software Technology,* vol. 34, pp. 159-174, 1992.

[5] J. Fong, H. K. Wong, and Z. Cheng, "Converting relational database into XML documents with DOM," *Information and Software Technology,* vol. 45, pp. 335-355, 2003.

[6] J.-K. Min, C.-H. Lee, and C.-W. Chung, "XTRON: An XML data management system using relational databases," *Information and Software Technology,* vol. 50, pp. 462-479, 2008.

[7] A. Hueni, J. Nieke, J. Schopfer, M. Kneubühler, and K. Itten, "The spectral database SPECCHIO for improved long term usability and data sharing," *Computers & Geosciences,* vol. 35, pp. 557-565, 2009.