Spectral Information System Development for Australia

Andreas Hueni[1], Laurie Chisholm[2], Lola Suarez[3], Cindy Ong[4], Matthew Wyatt[5]

[1] Remote Sensing Laboratories, University of Zurich, Switzerland

[2] Institute of Conservation Biology and Environmental Management, School of Earth and Environmental Sciences, University of Wollongong, NSW, Australia

[3] School of Mathematical and Geospatial Sciences. RMIT University, Melbourne, Australia

[4] CSIRO Earth Science and Resource Engineering, Australia

[5] iVEC/Australian National Data Service

ahueni@geo.uzh.ch

Abstract

Scientific efforts to observe the state of natural systems over time, allowing the prediction of future states, have led to a burgeoning interest for organised storage of spectral field data and associated metadata, seen as being key to the successful and efficient modeling of such systems. A centralised system for such data established for the Australian remote sensing community aims to standardise storage parameters and metadata thus fostering best practice protocols and collaborative research. Supported by the Australian National Data Service (ANDS), whose aim is to promote connections between data, projects, researchers and institutions, a spectral information system based on the already operational SPECCHIO spectral database system is being augmented to specifically meet the needs of the Australian remote sensing community, and is aligned with the Terrestrial Ecosystem Research Network (TERN) Auscover facility. In this paper we outline the envisaged dataflow and usage of the system as a case study within the context of TERN Auscover. The development of a national spectral information system will not only ensure the long-term storage of data but support scientists in data analysis activities, essentially leading to improved repeatability of results, superior reprocessing capabilities, and promotion of best practice.

Keywords: Spectral Database, Spectral Information System, Field Spectroscopy, Data Sharing

Andreas Hueni's main research interests are the development of spectral data processing systems and databases, and the calibration of imaging spectrometers. He is the author of the SPECCHIO spectral database and APEX (Airborne Prism Experiment) calibration and processing scientist.

Laurie Chisholm's research interests focus on the application of advanced spatial science techniques and the development of spectral indices to investigate the impacts of disturbances on ecological systems at a range of scales. She is involved in the compilation of spectral libraries in support of research and validation of image products and manager of the ANDS Project DC-10, development of a national spectral database system.

Lola Suárez holds a PhD in remote sensing of vegetation stress. Her research interests include vegetation biophysical properties estimation and up-scaling methodologies to different spatial resolutions using hyper/multispectral remote sensing, and radiative transfer modeling.

Cindy Ong's research interests are focused on deriving quantitative measurements from hyperspectral data. She has extensive experience on the application of hyperspectral technology for environmental measurements for the mineral industry and regulatory requirements.

Mathew is a technical manager and analyst on a number of eResearch projects around Australia, and avid supporter of open and collaborative data projects such as the DC-10 project.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

Introduction

Field spectroscopy, particularly the measurement of the spectral reflectance of terrestrial or aquatic features, is widely used as a tool for a variety of research operations and applications and plays a critical role in the calibration and validation of image data from remote sensing campaigns (Milton et al., 2009, Curtiss and Goetz, 1994).

In Australia there is a wealth of field spectroscopy data collected from a variety of instruments providing measurements at multiple scales and for wide-ranging purposes:

- as ground truth collected in support of a remote sensing validation campaign to relate image data to real features and materials on the ground;
- to calibrate remote sensing data and sensors;
- to provide a spectral reference for materials on the Earth surface;
- to determine mixtures of materials or investigate discriminatory capacity to discern a particular material or examine the behaviour of spectra from influencing factors; and
- to answer research questions across a multitude of disciplines.

Data from these wide-ranging capture events are often stored in silos, for example, on ftp stores, local machines, in data disks. Historically such data are not well documented. These data, especially if independent validation data were collected concurrently, are extremely valuable yet are not discoverable by others doing experiments in similar areas, on similar sample types or requiring references for certain materials. There are also generally no linkages between the remotely-sensed data, the field validation spectra and the independent validation data for example XRD (X-ray diffraction) data for mineralogy; vegetation moisture content for bushfire fuel assessment.

Metadata are paramount to broad and long-term use and interpretation of scientific data and must thus be acquired and stored in a rigorous way (Michener, 2000, Michener et al., 1997, Curtiss and Goetz, 1994, Latham et al., 2009, Lawrence et al., 2009). This is particularly true for field spectroradiometer measurements where only detailed metadata permits assessing the environmental conditions and applied sampling setup.

Spectral databases are systems for the storage of spectral data acquired by spectroradiometers under both field and laboratory conditions, augmented with associated auxiliary data (Hueni et al., 2011). By definition, data are being transformed into information when they are being structured, edited or processed. Information may be processed to form new information and hence the data and information tiers as depicted in the DIKW (Data, Information, Knowledge, Wisdom) (Ackoff, 1989, Rowley, 2007) pyramid in Figure 1 may be regarded as a continuum of both spectral data and associated metadata to higher-level spectral information and associated information (Herold, 2003). A spectral database requires data input mechanisms that store data acquired by sensors in the database, usually also entailing the automatic ingestion of metadata. Hence, spectral databases build a certain amount of information upon data load.

Spectral information systems take spectral databases a step further by making data held by the databases retrievable and usable by other users or systems and by adding processing functionalities that further transform the data or information held by the system, in turn generating more information. This could e.g. involve the generation of higher-level products or spectral data corrected for sampling equipment or sensor artefacts.

Most essentially, spectral information systems are no longer mere data repositories but systems that support the scientists in analyzing their data utilizing the full potential of combined metadata spaces (Wason and Wiley, 2000) and spectral spaces. It this point, such systems will no longer be perceived as an encumbrance but as a tool to improve the research process. We demonstrate this new paradigm in the case study provided in this paper. Furthermore, such spectral information systems allow data exchange and support the reuse of existing spectral data for different applications.

Funded by the Australian National Data Service (ANDS), and stimulated by the TERN AusCover and ACEAS (Australian Centre for Ecological Analysis and Synthesis) facilities, and members of the wider Australian remote sensing community, Australian efforts to develop international standards for the exchange of spectral data and related metadata are reaching fruition in the form of a comprehensive national spectral database system.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

The aim of this project is to develop a centralised system to house spectral libraries and associated metadata. The repository will allow the Australian remote sensing community to collate, share and discover existing spectral libraries and facilitate the capture of new datasets as they are formed.

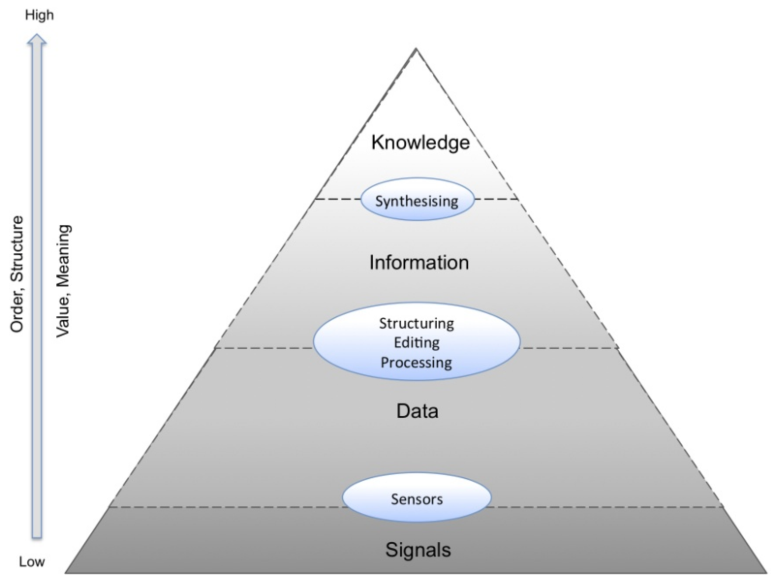

Figure 1: DIKW hierarchy adapted to spectroradiometry, encompassing both spectral data and metadata evolution

Spectral Information System Requirements

The system specifications presented hereafter are based on a requirements and gap analysis carried out during a TERN ACEAS workshop on best practices associated with bio-optical data in mid-2012. Involving stakeholders from multiple disciplines within the Australian and international remote sensing community, the gap analysis has been expanded with an aim to consolidate and prioritise system requirements to ensure the final product meets the needs of end users. Many required features are already provided by the current SPECCHIO system (Hueni et al., 2009) that was re-engineered in 2006 and has been constantly upgraded and improved since then; these supported features are listed in Table 1. SPECCHIO uses a relational database (DB) as a data repository, storing both spectral data and metadata in the DB. The system is preconfigured with over 40 metadata parameters (Hueni et al., 2009), including categorical variables, and further parameters can be defined dynamically, thus supporting the adaptation to specific use cases. Spectral data are stored as vectors, referencing sensor definitions that define the centre wavelengths per spectral band and new sensors can be easily added to the system. Details on the current functionality of the system can be found in the SPECCHIO user guide (Hueni, 2012), available online[1].

Table 1: List of functionalities supported by the current SPECCHIO version

Feature	Description / Implementation details
Public read access for automatically created users	Standard operational mode of the current system
Electronic insert of new data	Supported via various file format specific data loaders
Definition of core metadata set	Via SPECCHIO policy concept, essentially based on user sampling protocols
Organised storage of data	Data are stored in campaigns, each spectrum has a unique identifier
Automated output formatting when exporting data	SPECCHIO file output routines

[1] www.specchio.ch

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

Insert of new data records into existing collections	SPECCHIO delta loading capability
Delete data from database	SPECCHIO data remover function
Granularity of metadata description	Metadata are stored at single spectrum level, but group updates are possible to streamline input processes
Maintaining quality (number of metadata entered)	Via SPECCHIO policy concept and SPECCHIO metadata space density measure (Hueni et al., 2011)
No anonymous data access	Guaranteed by database access rights
Backup of data owned by a group	Storage as XML files (Hueni et al., 2011)
Data access options: file export, direct DB access	SPECCHIO file output routines and direct DB access using Java bridging
Download data in a specific format	CSV and ENVI SLB are currently supported
Download of metadata	Full metadata are included in the CSV file output
Interaction with other systems	Via Java bridging (Matlab, Python, R, etc)
Import of data into statistics packages	Via CSV files
Data analysis in open source software	Via CSV files or via direct database access (Python, R and similar open source languages supporting Java bridging)
Online and offline work possible	Local DB installation is possible. This allows the use under field conditions.
Selecting data for download	Via metadata space restrictions defined in the SPECCHIO query builder (Hueni et al., 2009) or queries written by user for specific cases
Support for distributed systems	Data exchange via XML format (Hueni et al., 2011)

The features listed in Table 22 represent a compilation of mandatory as well as desirable functionality that will be prioritised within the project.

Table 2: List of new functionalities

Feature	Description
Research group concept	Data can be owned by a group instead of only a single user
User data update via web interface	-
Concept of PI's	A data set is also owned by a PI (principal investigator)
Loading of ancillary files	Augmenting spectral collections with metadata held by files
Data import from another system	Add new file reading routines as interfaces to such systems
Metadata updates overwriting existing entries should be logged	Metadata editing action logs with reference to involved user (provenance information)
Assessing data quality	Use concepts developed at RMIT, Melbourne
Support of usage profiles	User group specific data editing masks, based on sampling protocols
Archive data	Freeze data at a given state and disable editing
Intended context indicator	Educational, training, test, research, etc. Defined on campaign level.
Invite 3rd parties to access data	Adding new persons to an existing group for a limited time, e.g. for reviewing
Community building	Facebook similar features
Web access of DB	Data can be accessed without local installation of SPECCHIO desktop application
Public/private visibility of campaigns	Some data can be embargoed by the PI for a certain time while work or publishing is ongoing
Download of original data files	Store a copy of the files on a fileserver during data

ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

	ingestion and make them available for download
Export of data to e.g. GIS formats	Add exporting routines for identified software required by end users

Methods

An initial gap analysis identified the potential of the SPECCHIO system to serve as a basis for the development of a national spectral database for the Australian remote sensing community. The following methods thus represent a mix of the current capabilities of the SPECCHIO system and of features that will be developed or enhanced in the scope of the spectral database development.

The general dataflow (Figure 33) starts with a data ingestion process, loading spectroradiometer generated data files to the database. This step involves extracting as many metaparameters as possible from the files to reduce the manual metadata input. Input files supported at the time of writing are given in the appendix.

Metadata of data contained in the database can be augmented by using a metadata editing graphical user interface (GUI). This process is streamlined by the concept of group updates, where a spectral collection can be updated in one operation. Data held by the system can be processed and visualised within the SPECCHIO Java application, but these capabilities are rather generic and rudimentary as it is virtually impossible to cater for all possible user needs within the Java application. Data may be exported to regular files with CSV (comma separated values) and ENVI spectral library being the current output options.

More complex or user specific processing, analysis and visualization can be implemented in programming languages that support scientific, exploratory work at a much greater extent than Java. Such higher-level algorithms are easily implemented in any language that supports Java bridging in some fashion. Generally, SPECCHIO Java Objects are used to connect to the database, making spectral data and metadata available to the runtime environment of choice. Matlab and Python have been proven to seamlessly interact with SPECCHIO databases and the same is expected to hold true for further languages such as R and IDL. APIs (Application Programmer Interfaces) provided by the SPECCHIO system software support all interactions with the relational database management system (RDBMS), i.e. querying, inserting, updating and deleting (Figure 2).

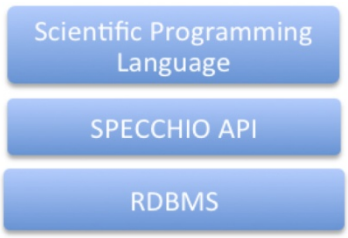

Figure 2: SPECCHIO system layers

This capability is extremely powerful, as it gives scientists the freedom to develop their own algorithms and data analysis/processing flows while basing on a stable and abstracted foundation that handles all the querying, data loading, inserting as well as housekeeping regarding metadata and consistency of spectral spaces.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

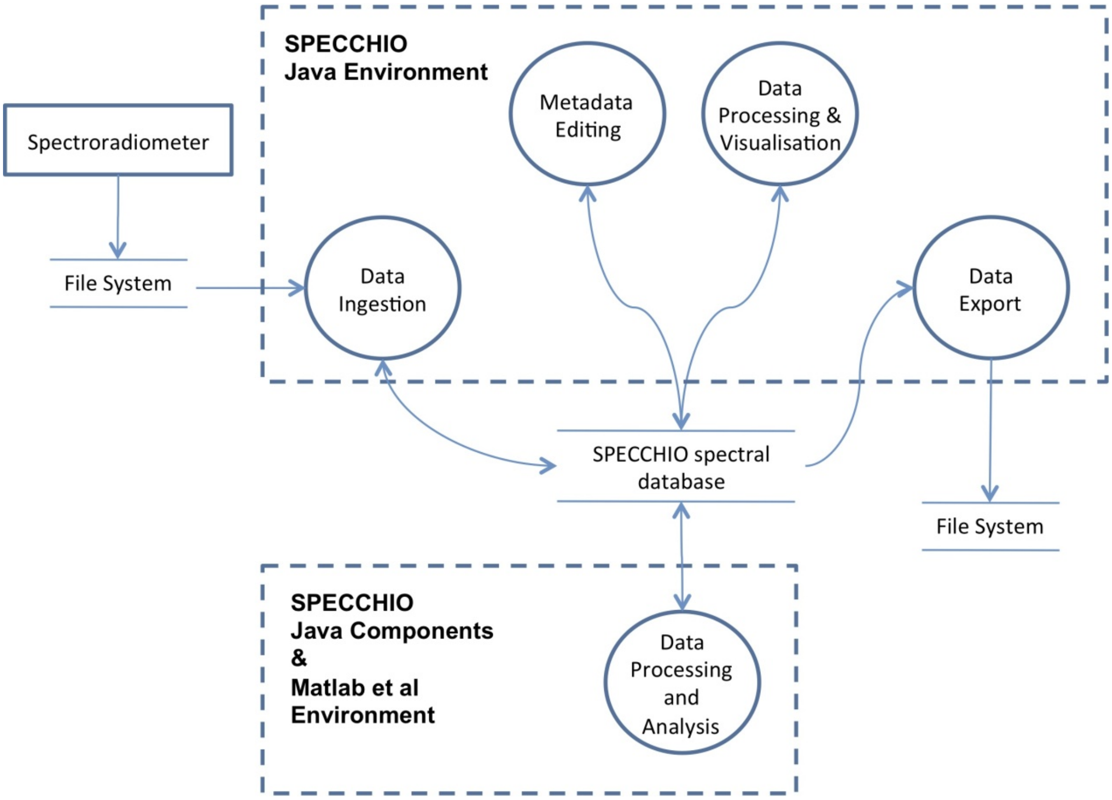

Figure 3: Main dataflows of the SPECCHIO system

The SPECCHIO system can be easily deployed on database servers with or without Internet access or on local workstations or even on field laptops. A possible ontology is shown in Figure 44. This allows a flexible and user-need tailored use of the system, e.g. to limit the data access to in-house researchers, or to facilitate data storage and processing during field campaigns where Internet access may be impossible. SPECCHIO offers a file-based option to exchange data between distributed databases and thus allows consolidating the data in a central server once deemed appropriate to do so by the data producer (Hueni et al., 2011).

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

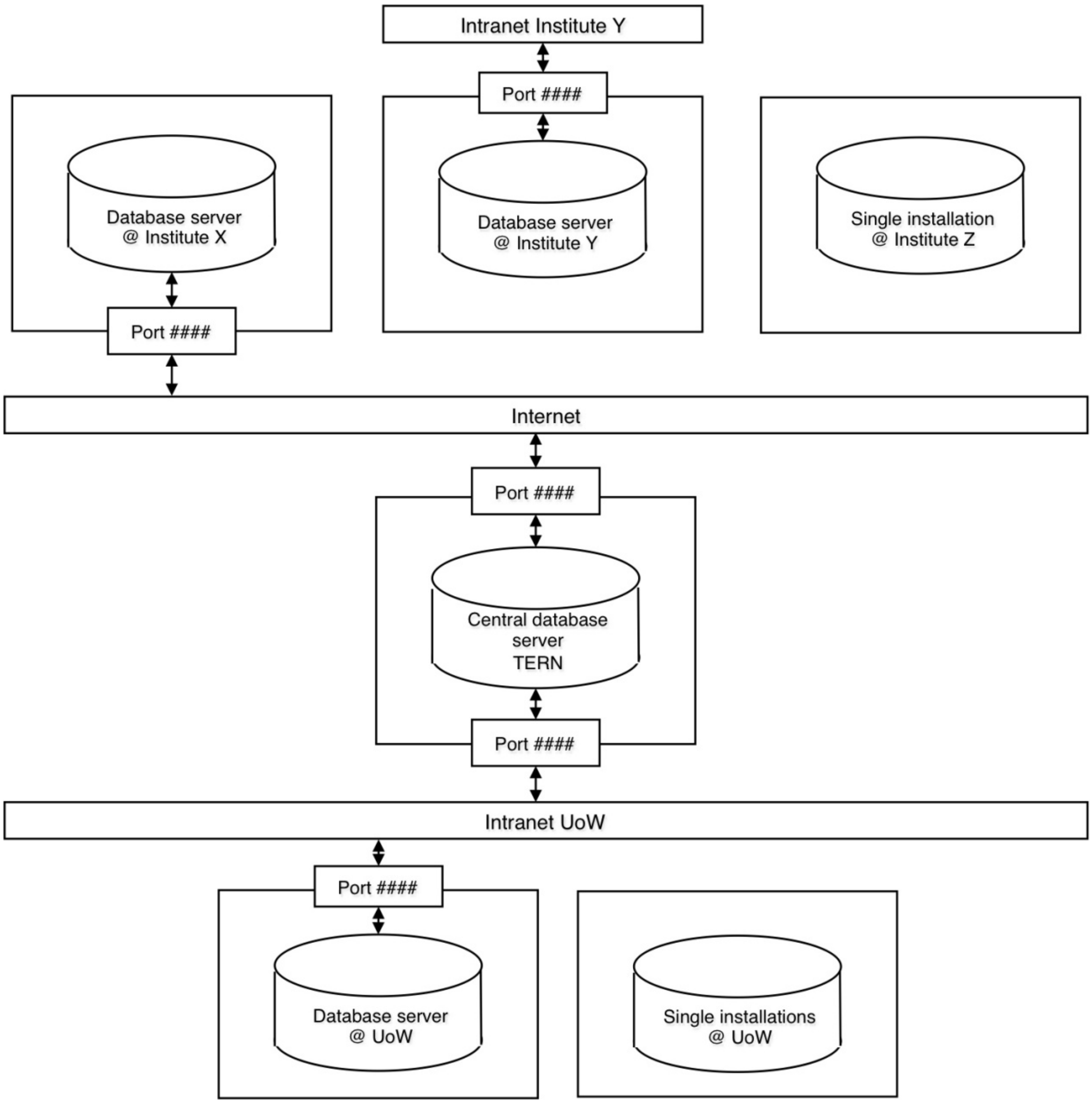

Figure 4: Possible ontology of database instances, showing the connections to various networks and accordingly varying degrees of accessibility

TERN Auscover Case Study

The Terrestrial Ecosystem Research Network (TERN) AusCover facility was created to facilitate the production of validated satellite-derived biophysical map products for the ecosystem research community and natural resource managers (TERN 2010). The Auscover network strives to create these products within the framework of established Australian validation programs following established international earth observation protocols.

Using a TERN Auscover field campaign as a case study illustrates the use of the current SPECCHIO system as data repository and as a platform for post-processing and storage of according results in the database. This study was chosen to demonstrate the generic nature of the SPECCHIO system for the handling of in-situ spectral data and metadata and thus provides a preview of the future system functionality.

In April 2012 hyperspectral and Lidar imagery were acquired over Rushworth State Forest in Victoria. The study site consists of a reference area located within the main box-ironbark forest 19 km NW from Nagambie. Simultaneously, field

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

data collection was conducted consisting of vegetation structure measurements and leaf sampling. The aim of this field campaign was to characterise leaf chemistry and spectroscopy for a number of trees in the study site. A total of 96 stands representative of the five most abundant species were selected. Only trees which crown spectra could be extracted from the imagery were considered for selection. As a result, none of the selected trees were suppressed or shadowed by the surrounding crowns. From each selected tree, one branch from the upper-most third of the crown was shot down and a set of mature leaves was taken for further analysis. Additional metadata collected for every tree consisted of stand coordinates, tree species, height, trunk diameter at breast height (DBH), crown diameter, crown percentage cover and position of the crown in relation to the surrounding trees (i.e. dominant, co-dominant or isolated). In the following hours, an integrating sphere (Analytical Spectral Devices, Boulder, CO) was used to measure reflectance and transmittance spectra of three of the leaves collected per tree. Another group of leaves were scanned and weighed to estimate specific leaf area, and a third set was sent to a laboratory for chemical analysis.

The spectral measurements were acquired using a FieldSpec Pro spectrometer (Analytical Spectral Devices, Boulder, CO) attached to the integrating sphere through a bare fiber. Measurements to correct for the misalignment of the light source were taken for every reflectance and transmittance measurement.

During data ingestion the SPECCHIO ASD file format reader loads data from the file system into the database. This is a standard process and data are immediately available for sphere post-processing. To give users a better control of the processing applied, an interactive graphical user interface written in Matlab using SPECCHIO Java components is used (Figure 66). The left hand side of the GUI holds the Spectral Data Browser component, displaying the content of the database in hierarchical form and allowing the selection of spectral data. Such a selection is the starting point for the dataflow ensuing as illustrated in Figure 5. Once data are selected, their database identifiers will be loaded. These are then iterated over during the spectra loading, selecting spectrum records from the database and making the spectral data available as Matlab matrices. Two modes are supported by the GUI: (a) group by group interactive processing and (b) automatic processing. The former option processes one group consisting of the four raw measurements, displays the raw spectra and the resulting corrected reflectance and transmission spectra in the six dedicated display panels of the GUI. The automated mode processes all groups in one operation and displays all corrected reflectance and transmission spectra; no raw spectra are displayed in that case.

These corrected spectra are then augmented with metadata and inserted into the database, essentially copying the metadata of the input spectra and changing them to reflect the processed state, i.e. setting the correct measurement type of either reflectance or transmittance.

Implementing such a tool can be accomplished in a matter of hours by basing on SPECCHIO API's, resulting in most of the development effort being concerned with data handling and GUI representations within Matlab.

The case study proved that the current system is very well capable of storing and processing data for a specific use case in an operational manner by relying on generic system functions. In particular, it could be shown that (a) data can be automatically ingested, (b) data may be retrieved from the system using metadata queries in a sequence dictated by the scientific experimental setup, (c) a use case specific higher-level interactive component could be easily developed, and (d) extracted information could be re-inserted into the database, automatically linking with the existing metadata of the input spectral information. The system thus greatly increases the speed of data processing while reducing operator errors and linking resulting information to existing metadata, thus ensuring retrievability.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

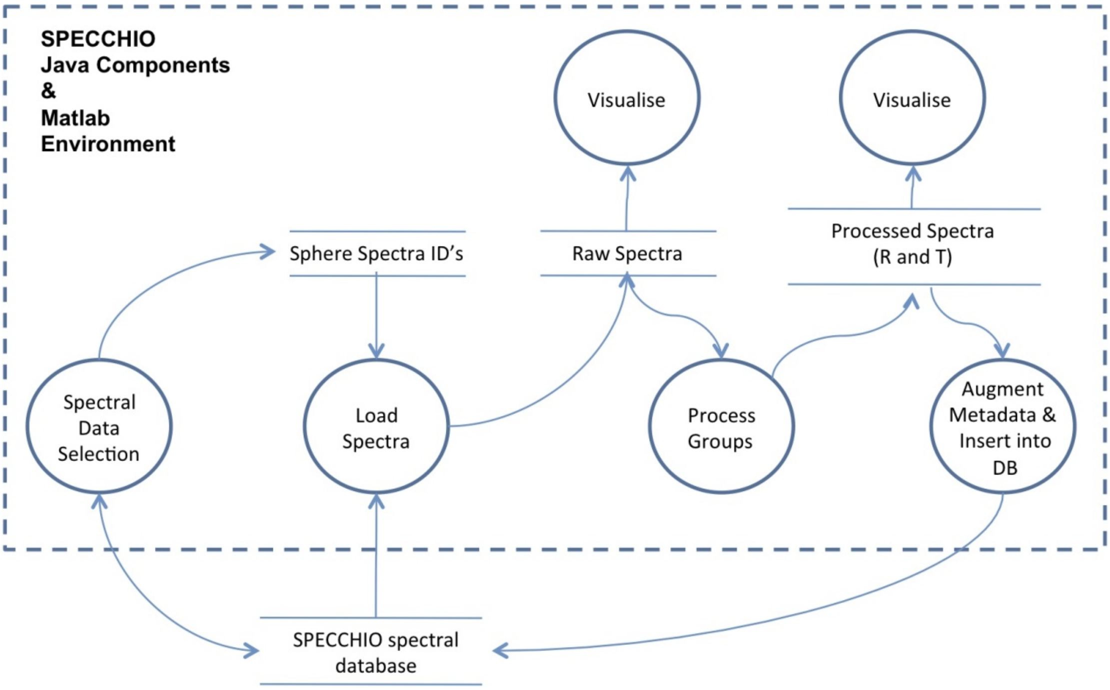

Figure 5: Dataflow of the interactive sphere data processing

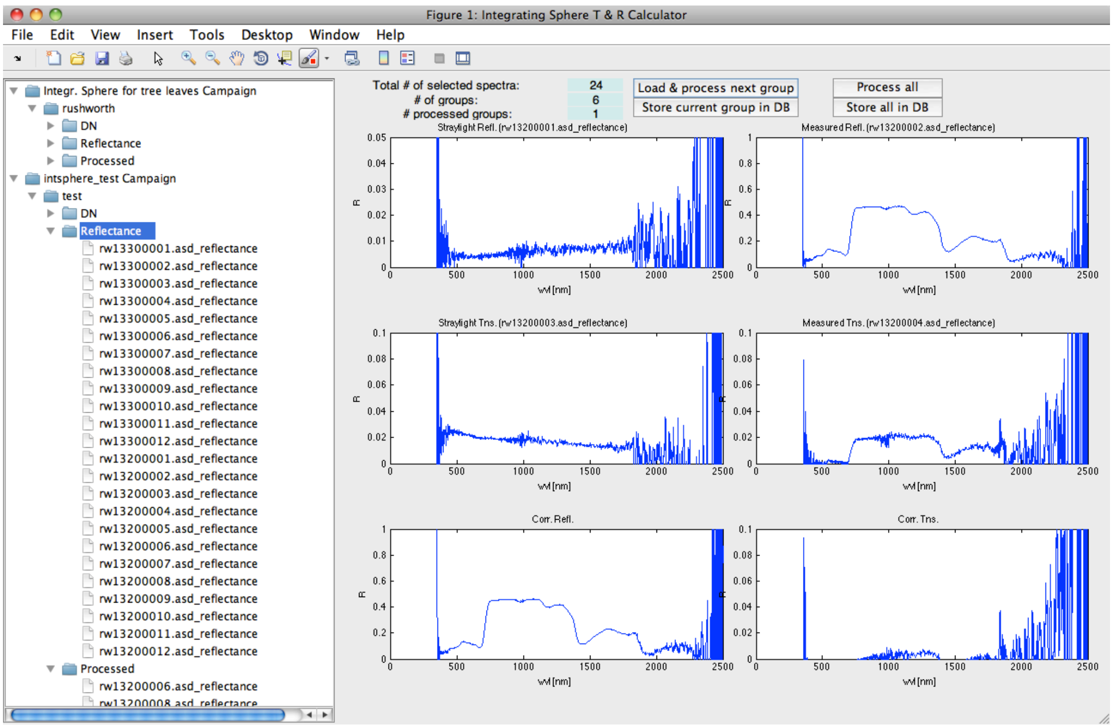

Figure 6: Interactive Matlab GUI with direct database access for integrating sphere data processing

ISBN: 978-0-9872527-1-5

Conclusions

As outlined, the development of a national spectral information system for the Australian remote sensing community, aims to standardize spectral storage parameters and metadata, thus promoting best practice protocols. By enlisting the input of remote sensing stakeholders across wide-ranging disciplines, it is anticipated that the system developed will provide tools that researcher's need and desire, thus enhancing current workflows and facilitating collaborative research.

The project, starting in the last quarter of 2012 with an expected duration of 5-6 months, aligns with ANDS objectives by creating descriptions of the spectral libraries and sharing those descriptions in ARDC (Australian Research Data Commons). Once developed, there will be an automatic feed from the system to ARDC. Data from the libraries will also be available for reuse by other researchers.

The software deliverables will be of use to other research groups who also have spectral databases and its availability will ensure any future libraries are built on spec based on protocols and metadata requirements, thus, facilitating use and growth of the system. Since libraries can be built at multiple scales and levels, the potential use of this development offers high value return on investment, based on a common need across the entire remote sensing community.

In addition, the system will enable reliable transfer of experimental data for calibration and research and promote sharing amongst collaborators and other researchers via ARDC.

Interested users may test the current SPECCHIO system at any time by subscribing to the online SPECCHIO spectral database via the SPECCHIO website www.specchio.ch.

Acknowledgements

The DC-10 project is funded by ANDS (Australian National Data Service).

References

ACKOFF, R. L. 1989. From data to wisdom. *Journal of Applied Systems Analysis,* 16**,** 3–9.
CURTISS, B. & GOETZ, A. F. H. Year. Field Spectrometry: techniques and instrumentation. *In:* International Symposium on Spectral Sensing Research, 1994. 9.
HEROLD, K. 2003. An information continuum conjecture. *Minds and Machines,* 13**,** 553–66.
HUENI, A. 2012. SPECCHIO User Guide. 2.2.0 ed. Switzerland: Remote Sensing Laboratories, Institute of Geography, University of Zurich.
HUENI, A., MALTHUS, T., KNEUBUEHLER, M. & SCHAEPMAN, M. 2011. Data Exchange between distributed Spectral Databases. *Computers & Geosciences,* 37**,** 861–873.
HUENI, A., NIEKE, J., SCHOPFER, J., KNEUBÜHLER, M. & ITTEN, K. 2009. The spectral database SPECCHIO for improved long term usability and data sharing. *Computers & Geosciences,* 35**,** 557-565.
LATHAM, S. E., CRAMER, R., GRANT, M., KERSHAW, P., LAWRENCE, B. N., LOWRY, R., LOWE, D., O'NEILL, K., MILLER, P., PASCOE, S., PRITCHARD, M., SNAITH, H. & WOOLF, A. 2009. The NERC DataGrid services. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 367**,** 1015-1019.
LAWRENCE, B. N., LOWRY, R., MILLER, P., SNAITH, H. & WOOLF, A. 2009. Information in environmental data grids. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 367**,** 1003-1014.
MICHENER, W. K. 2000. Metadata. *In:* MICHENER, W. K. & BRUNT, J. W. (eds.) *Ecological data: Design, management and processing.* Oxford, UK: Blackwell Science.
MICHENER, W. K., BRUNT, J. W., HELLY, J. J., KIRCHNER, T. B. & STAFFORD, S. G. 1997. NONGEOSPATIAL METADATA FOR THE ECOLOGICAL SCIENCES. *Ecological Applications,* 7**,** 330–342.
MILTON, E. J., SCHAEPMAN, M. E., ANDERSON, K., KNEUBÜHLER, M. & FOX, N. 2009. Progress in field spectroscopy. *Remote Sensing of Environment,* 113**,** 92-109.
ROWLEY, J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science,* 33**,** 163–180.
WASON, T. D. & WILEY, D. 2000. Structured Metadata Spaces. *Journal of Internet Cataloging,* 3**,** 263-277.

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012

Appendix

Table 3: File formats support by SPECCHIO version 2.2.2 delta

File Format	Description
ASD binary	Analytical Spectral Devices, old and new file formats
GER	GER instruments text files
MFR	MFR Sun Photometer text files
SVC	SVC HR-1024 files
Apogee	Apogee text files, preliminary support
ENVI SLB	ENVI spectral library files
Ocean Optics	Ocean Optics SpectraSuite text files
UniSpec	UniSpec Single Beam and Dual Beam text files
FGI HDF5	HDF5 files with data structure proprietary to the Finnish Geodetic Institute
Text	Columnar, space separated spectral data
SPECPR	Spectral file format by USGS, also generated by the USGS PRISM software

* Review Paper – accepted after double-blind review.
ISBN: 978-0-9872527-1-5
Melbourne: Geospatial Sciences, RMIT 2012