GSR_3
Geospatial Science Research 3.
School of Mathematical and Geospatial Science, RMIT University
December 2014

Sharing field spectroscopy data within large data sharing systems

Barbara Rasaiah, Simon Jones, Chris Bellman
RMIT University
Melbourne, Australia
barbara.rasaiah@rmit.edu.au, simon.jones@rmit.edu.au, chris.bellman@rmit.edu.au

Tim Malthus
CSIRO
Canberra, Australia
tim.malthus@csiro.au

Authors

Barbara Rasaiah

Barbara Rasaiah is completing her PhD in geospatial science at RMIT University. She has industry experience as an IT operations analyst and programmer. Barbara's research interests include geospatial data and metadata exchange and distribution, data analytics, datamining, geospatial databases, and large information sharing systems.

Simon Jones

Simon Jones is Professor of Remote Sensing and director of the Remote Sensing and Photogrammetry Research Centre at RMIT University. His research activities include biophysical remote sensing of terrestrial environments, in situ observations (including spectral-radiometry), scaling ground observations to the image and landscape level, and spatial data uncertainty. He was director of the Spatial Sciences Institute of Australia from 2006-08 and co-Chair of the IGARSS, IEEE International Geoscience and Remote Sensing Symposium, Melbourne 2013. Dr. Jones is the lead researcher/chief investigator on numerous concurrent remote sensing projects including the Terrestrial Ecosystem Research Network (AusCover team), the Australian Woody Vegetation Landscape Feature Generation from Multi-Source Airborne and Space-Borne Imaging and Ranging Data project, and the Data Integration, Scale and Classification of Remotely Sensed Imagery project.

Chris Bellman

Chris Bellman is associate professor in the School of Mathematical and Geospatial Sciences at RMIT University in Melbourne, Australia. Chris has a strong interest in teaching and learning and research interests in the fields of photogrammetry, GIS and spatial analysis. He was Discipline Head (Geospatial Science) from 2006 – 2012.

Tim Malthus

Dr Tim Malthus is Research Group Leader of the Coastal Monitoring, Modelling and Informatics Group in the Coastal Management and Development Program of CSIRO's Oceans and Atmosphere Flagship. He previously led the Environmental Earth Observation Program in the CSIRO Division of Land and Water. He combines skills in calibration, validation and field spectroscopy with analysis of airborne and satellite Earth observation data, to develop improved monitoring tools for the management of land and water resources informing wider environmental policies. Prior to joining CSIRO, Dr Malthus was Senior Lecturer in Remote Sensing, University of Edinburgh and Director of the Natural Environment Research Council (NERC) Field Spectroscopy Facility based at the University. He oversaw the expansion and diversification of the Facility to maintain its world leading status in field spectroscopy research and related instrumentation. Tim has a BSc and PhD from the University of Otago, New Zealand.

Abstract

There is urgency in acquiring continuous high quality spectroscopy data to solve problems in Earth systems science (Milton *et al*., 2009). Informing users and stakeholders of field spectroscopy datasets of the impact of high-quality data and metadata in the context of Earth observing data systems is an additional challenge facing the remote sensing community. Quality assurance of field spectroscopy datasets necessitates oversight and standardization, both at local, national, and international scales and is a way of ensuring robust metadata protocols for field spectroscopy. The need for a standardized methodology for collecting field spectroscopy metadata has increased with the emergence of data sharing initiatives such as NASA's EOSDIS (Earth Science Data and Information System) LTER (Long Term Ecological Research) network, Australian Terrestrial Ecosystem Research Network (TERN), SpecNet, and some of the smaller *ad hoc* spectral libraries and databases created by remote sensing communities internationally. This paper presents the central considerations for large-scale distribution and discoverability of field spectroscopy datasets and their metadata.

Keywords: metadata, field spectroscopy, database, quality assurance, information

1. Introduction

The volume of information derived from *in situ* field spectroradiometers, across a broad variety of, often costly, applications and instrumentation, grows each year. There is a recognized need within the international remote sensing community to document, store, and share field spectroscopy data and metadata in consistent formats within dedicated data sharing and other intelligent archiving systems (CEOS 2013; GEO 2014). Establishing and maintaining optimal integrity of the data is a key priority to ensure effective re-use of the data, and to enable more efficient and higher impact research.

Metadata is an important component in the cataloguing and analysis of field spectroscopy datasets because of their central role in identifying and quantifying the quality and reliability of spectral data and the products derived from them. There is currently no international standard methodology for collecting field spectroscopy metadata (Rasaiah *et al*. 2014). This makes rich and flexible metadata capabilities a critical factor in the interoperability and quality assurance of datasets. The largest publicly available spectral databases (including SPECCHIO, DLR Spectral Archive, USGS Spectral Library) do not have a full suite of standardized metadata definitions, nor do they provide quality assurance for the data or metadata. A pervasive lack of quality assurance for these data is a barrier to integration with existing larger-scale data sharing systems that adhere to data quality assurance protocols.

2. Central issues to sharing field spectroscopy data and metadata within large geospatial information sharing systems

Storing and distributing field spectroscopy data and metadata within large data sharing systems requires specific consideration for the data and metadatasets, the data stakeholders, the IT infrastructure, and protocols for data and metadata distribution (Figure 1).

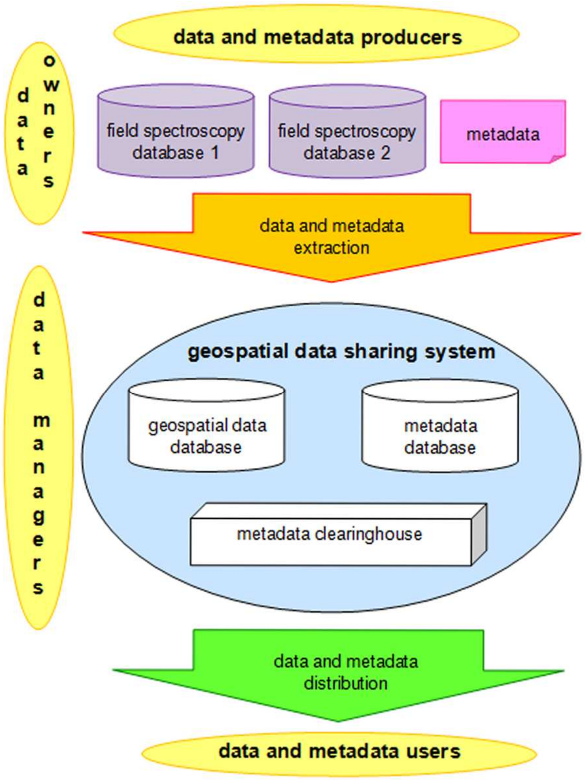

Figure 1 Relationships among field spectroscopy data and metadata, data producers, owners, managers and users relating to a large geospatial information sharing system

2. 1 Data and metadata

Field spectroradiometer data relies on its associated metadataset for discoverability, proof of quality control and assurance. The associated metadata also permits a data user to assess whether a given dataset is suitable for their purpose based on information including the general target and sampling properties, instrument properties, reference standards, calibration, hyperspectral signal properties, and general project details. More specifically, users can use this metadata to identify the impact of their experiments and allow intercomparison of datasets (Duggin 1985; Kerekes 1998; McCoy 2005; Stuckens *et al.* 2009). An effective and reliable information sharing system must incorporate capabilities for the storage and discoverability of both the data and associated metadata.

2.2 Data producers and data users

Identifying the needs of users who will access and use the data, identifying an application profile, and the direct involvement of interested stakeholders are critical to designing and implementing robust metadata standards and data sharing protocols. Engagement with data producers and data users with the requisite expertise in application domains ensures that a metadataset is aligned with a data user's needs. As stakeholders of the data, field spectroscopy scientists have a vested interested in adopting a standard most suitable to their needs as both data and metadata creators and users of these data. It is important that data producers, owners, and managers coordinate their efforts to ensure that a metadataset is as complete and high quality as possible before it is uploaded to databases, datawarehouses, cloud platforms, or otherwise made available for distribution (Orr 1998; Bruce and Hillman 2004; Loshin 2010; da Cruz *et al*. 2011).

2.3 Rules and protocols for data and metadata production and distribution

There is no currently no standard for field spectroscopy data and metadata documentation or exchange. Numerous bodies overseeing and advising the geospatial sciences have adopted standards based on the ISO 191__ standard family relating to storage, encoding, and quality evaluation of geographic data. OGC (Open Geospatial Consortium) and INSPIRE (Infrastructure for Spatial Information in the European Community) have both adopted architecture and data interoperability protocols for geospatial metadata based on EN ISO 19115 and EN ISO 19119 (INSPIRE, 2009; OGC 2012). These standards, however, fail to explicitly address the metadata requirements of field spectroscopy collection techniques, or the ontologies and data dependencies required to model the complex interrelationships among the observed phenomena as data and metadata entities. Critical metadata for field spectroscopy campaigns has been identified (Rasaiah *et al*. 2014), but not yet incorporated into a formal standard.

2.4 IT infrastructure

The absence of a central archiving apparatus for field spectroscopy data either for a specific campaign or on an international scale is a barrier to the efficient archiving of data and metadata by spectroscopic scientists. Recent developments in relational spectral databases include the publicly accessible DLR Spectral Archive (http://cocoon.caf.dlr.de) and SPECCHIO (http://www.specchio.ch/), as well as others designed in-house for organizations engaged in field spectroscopy research; these have allowed a more structured storage for spectral measurements and their associated metadata (Pfitzner *et al*. 2006; Hueni *et al*. 2009).

The implications for maintaining integrity of field spectroscopy and metadata are magnified in large information sharing systems and ‘big data’ environments. There are several IT infrastructure models that have been adopted for the sharing and distribution of scientific research in general and for geospatial data specifically. Metadata clearinghouses (NASA's Global Change Master Directory) (NASA 2013) are public metadata inventories of a broad spectrum of Earth science data and more specifically, authoring tools, data discovery, and metadata transformation and conversion tools in accordance with ISO, FGD, ESRI, Dublin Core, ANZLIC standards for geospatial metadata. There exist data exchange networks among the geospatial community that are evolving towards an integrated datawarehousing, cloud-based, big data model (including EOSDIS, GALEON and TERN). Although none of these systems have formally integrated field spectroscopy data and metadatasets, it is incumbent upon the field spectroscopy community to actively participate in the design and implementation of such systems, which includes supporting a field spectroscopy metadata standard for maximizing the discoverability and quality assurance of their datasets.

3. Steps to a solution

Integrating field spectroscopy data and metadatasets in large data sharing systems need not be a challenging task given that the data stakeholders have an understanding of the value of storing and sharing their data on such a platform, and that they have the desire to make their datasets available. Steps towards achieving this require participation of the data and metadata producers, users, managers, owners, and IT systems designers and managers. Collaborative stewardship of data and metadata assigns of responsibility of creating and maintaining data and metadata to multiple individuals and stakeholders (researchers, IT specialists, data managers) according to their domain of expertise. Identifying a purpose for data and metadata collection and use allows data and metadata creators the flexibility to set thresholds for quality and completeness within domain and purpose-specific contexts. Standards-compliant software tools and information systems can comprise data sharing systems and metadata editors that enable and enforce creation and distribution of metadatasets compliant with the field spectroscopy data and metadata standards. Proper oversight of IT infrastructure and management enables data distribution system to provision quality-controlled discoverability and distribution of field spectroscopy data and metadatasets. Education initiatives including workshops and training programs for researchers and field spectroscopy data stakeholders promote community understanding of the benefits of adhering to standards for the data and metadata documentation and discoverability. Much potential exists for adapting and improving current geospatial data exchange environments for the unique requirements of the field spectroscopy community.

References

Bruce, T.R.; Hillmann, D.I. 2004, ‘The Continuum of Metadata Quality: Defining, Expressing, Exploiting’. In Hillmann D. & Westbrooks, E. (Eds.) *Metadata in Practice*. Retrieved from eCommons@Cornell.

Committee on Earth Observing Satellites (CEOS) 2013, ‘CEOS Strategic Guidance Version: November 2013’, Retrieved June 07, 2014 from http://www.ceos.org/images/CSS/CEOS_Strategic_Guidance_Nov_2013.pdf

da Cruz, S. M. S.; Paulino, C. E.; de Oliveira, D.; Campos, M. L. M.; Mattoso, M. 2011, ‘Capturing distributed provenance metadata from cloud-based scientific workflows’, *Journal of Information and Data Management*, 2, 43-50.

Duggin, M. J. 1985, ‘Factors limiting the discrimination and quantification of terrestrial features using remotely sensed radiance’. *International Journal of Remote Sensing*, 6, 3-27.

Group on Earth Observations (GEO) 2013, *What is GEOSS?: The Global Earth Observation System of Systems*, retrieved January 10, 2013 from https://www.earthobservations.org/geoss.shtml

Hueni, A.; Nieke, J.; Schopfer, J. Kneubuehler, J.; Itten, K.I. 2009, ‘The spectral Database SPECCHIO for Improved Long-Term Usability and Data Sharing’, *Computers & Geosciences*, 35, 557-565.

Kerekes, J. P. 1998, ‘Error Analysis of Spectral Reflectance Data From Imaging Spectrometer Data’, *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, July 6-10, in Seattle, USA.

Loshin, D. 2010, *Effecting data quality improvement through data virtualization*. Accessed June 16, 2014 from http://dataqualitybook.com/kii-content/DataQualityDataVirtualization.pdf

McCoy, R.M. 2005, *Field Methods in Remote Sensing*. New York: The Guilford Press.

NASA 2013. *Global Change Master Directory: Metadata Protocols and Standards*, retrieved December 20, 2013 from http://gcmd.nasa.gov/add/standards/index.html

Orr, K. 1998, 'Data quality and systems theory', *Communications of the ACM*, 41(2), 66-71.
Pfitzner, K.; Bartolo, R.; Carr, G.; Esparon, A.; Bollhoefer, A. 2011, *Standards for reflectance spectral measurement of temporal vegetation plots, r*etrieved January 03, 2014 from http://www.environment.gov.au/system/files/resources/bf8002d0-2582-48a1-820f-8e79d056faed/files/ssr195.pdf
Rasaiah, B. A., Jones, S. D., Bellman, C., & Malthus, T. J. 2014, 'Critical Metadata for Spectroscopy Field Campaigns', *Remote Sensing*, *6*(5), 3662-3680.
Stuckens, J.; Somers, B.; Verstraeten, W.W.; Swennen, R.; Coppin, P. 2009, Normalization of Illumination Conditions For Ground Based Hyperspectral Measurements Using Dual Field of View Spectroradiometers and BRDF Corrections', *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, July 12-17, in Cape Town, South Africa.