

## A novel metadata standard for *in situ* marine spectroscopy campaigns

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 Land and Water  
Canberra, Australia  
tim.malthus@csiro.au

### ABSTRACT

Metadata are an important component in the cataloguing and analysis of *in situ* spectroscopy datasets because of their central role in identifying and quantifying the quality and reliability of spectral data and the products derived from them. This paper presents approaches to constructing a novel metadata standard for marine spectroscopy that serves to ensure a high level of reliability, integrity, and longevity for a spectroscopy dataset. Examined are the challenges presented by designing a standard that meets the unique requirements of *in situ* marine spectroscopy datasets, including the special case of measuring reflectance for underwater coral targets. Issues such as field measurement methods, instrument calibration, and data representativeness are investigated. The proposed metadata model incorporates expert panel recommendations that include metadata protocols critical to all campaigns, and those that are restricted to campaigns for specific marine environments. The implication of semantics and syntax for a robust and flexible metadata standard are also considered. Approaches towards an operational and logistically viable implementation of a schema are discussed. This paper also proposes a way forward for adapting and enhancing current geospatial metadata standards to the unique requirements of field spectroscopy.

*Keywords: Remote Sensing, Databases, in situ Observations, Metadata, Field Spectroscopy*

### 1 INTRODUCTION

Data collection protocols, encompassing both field spectral measurement methods and the metadata associated with them vary widely across the breadth of scientific inquiry applied to *in situ* spectroscopy. Metadata is a central component to the reliability, integrity, and legacy of a spectroscopy dataset because it serves to mitigate systematic and random errors on recorded radiance, target discriminability and contrast (Duggin, 1985) and reduce system bias and variability (Pfitzner *et al.*, 2006). On international and national scales, *ad hoc* data collection protocols are the norm as no formal standards exist within the remote sensing community for *in situ* metadata collection and rather arise from the expertise and knowledge of the scientists carrying out the campaign. Metadata recorded during a campaign may vary in format (hardcopy log sheets, excel forms, rudimentary database) and in volume (inclusive of documentation of all relevant campaign protocols to a minimum of metadata describing only the target being sampled). Metadata collection protocols diverge along the lines of the purpose of the campaign (calibration and validation, creation of a spectral library) and the target (tree crown, soil, seagrass, etc). Logistics, environment, instruments and target type all affect the design and implementation of a practical metadata standard.

Here the special case of a metadata standard for a marine campaign for underwater coral reflectance is presented. Marine campaigns are unique from terrestrial campaigns in terms of the instrumentation, specialized requirements for *in situ* data collection and environmental variables. Targets can include seagrass, macro-algae, corals and sponges, spectral measurements may be taken above surface or below surface and opinions differ on how inclusive a metadataset must be to document environmental and target properties (Bhatti *et al.*, 2009 and Dekker *et al.*, 2010). Instrument housing is often necessary to permit submersion and in some instances the instrument must be specially adapted to the underwater light field. Spectral measurements are recorded in a potentially unsafe environment with often continuously variable viewing conditions (illumination, viewing geometry, turbidity, etc.). At the University of Queensland, a customized underwater spectrometer system was developed and tailored specifically to coral reef ecology, and the ecology and physiology of animal colour vision. The accompanying protocols

for recording metadata *in situ* are interdependent with the challenges of radiometric data collection underwater as they are designed to simultaneously ensure the requisite operator safety (Roelfsema *et al.*, 2006).

## 2 A SPECIALIZED MARINE SPECTROSCOPY METADATA STANDARD

To ensure a high quality and practical metadataset, a metadata standard for underwater coral reflectance should have the following properties: 1) the metadata fields are sufficient to comprehensively and explicitly document the activities that took place and quantify and qualify influencing factors to the spectral measurement 2) allow replication of the campaign if required 3) and be flexible and broad enough in the scope of data capture to permit interoperability with other datasets. Granularity (the degree of specificity of the variable being recorded), syntax of the fields, and their data format (numeric/text/timestamp) affects the potential for data export, mining, and sharing.

Presented here (Table 2.1) is a metadata standard for underwater coral reflectance spectroscopy. It is derived from input from an expert panel of marine remote sensing scientists at the ACEAS (Australian Centre for Ecological Analysis and Synthesis) Bio-optical workshop held in Australia in 2012. While not inclusive of all metadata (instrument, calibration activities, reference standards, etc.) that should be recorded for an *in situ* campaign, it documents those metadata that describe field methods and variables unique to underwater coral reflectance measurements. The metadataset is divided into four main categories: 'Location and Environment Information', 'Illumination Information', 'Viewing Geometry', 'Coral Target Properties'. A description and reasons for inclusion of each field is provided, as well an example of each. An optionality designation of either 'Critical' or 'Useful' has been assigned to each field. Assuming that campaign logistics are not always favourable to documenting all necessary metadata, a prioritization model for criticality can form the basis of a standard that is both practical and fits the purpose for which the data is being collected. Critical fields are those that ensure the integrity of the dataset and cannot be excluded; useful fields are those that increase the robustness of the dataset for purposes beyond which it was originally intended. The data type specifies the most suitable format (text/numeric/timestamp/binary/image) for a given metadata parameter. A 'GML Object Type' column is included to indicate those metaparameters that can be expressed as GML 3.3 (Geographic Markup Language) objects. GML 3.3 is an implementation of ISO 19107 (specifying conceptual schemas for geographic features) and is used here simply as an example of a vocabulary that could be used to implement the metaparameters as objects in a metadata schema. Reference to a standard vocabulary, such as that provided by GML, permits translating the standard into a schema with maximum interoperability.

The most populous category (23 fields) is 'Location and Environment Information'. This is due to the high number of variables found within the marine environment that influence spectral measurements (water column properties, subsurface conditions, CDOM, etc.). There are commonalities with terrestrial campaigns (GPS coordinates, location description) but even in these cases special considerations must be made for the feasibility of recording these *in situ*. The 'Illumination Information' metadata category, while again sharing common fields with other non-marine campaigns, must make allowances for wave lensing and artificial light fields. The 'Viewing Geometry' category is identical to metadata requirements for most terrestrial campaigns except for documenting an operator's position relative to the target when they must provide shading over the target with their body to compensate for the fluctuating light field. The 'Coral Target Properties' category, similar to 'Location and Environment Information', contains fields relevant to marine campaigns only and reflects the special requirements of documenting underwater coral reflectance measurements.

**Table 2.1** Metadata standard subset for underwater coral reflectance measurements

Location Information Metadata

METADATA FIELD	REASON FOR INCLUSION / COMMENTS	OPTIONALITY	EXAMPLE	DATA TYPE	GML OBJECT TYPE
Location description	Qualitative description of surrounding environment	Useful	5 km offshore	text	gml:location
GPS coordinates	Permits referencing to aerial/satellite/other campaigns  Difficult to do; done on the dive site  Coordinates, datum + projection can be determined from Google Earth	Critical	x,y,z	numeric	gml:CoordType
Manual coordinate determination with map and compass	Substitutes GPS coordinates in instances of poor positional accuracy	Useful	x,y	numeric	gml:CoordType
Reference to photo of local relevant environment + target	Provides additional visual data where recording additional metadata of target and environment is not possible or feasible	Critical	photo # or name	text	gml:stringOrNull
Date of associated photo	Provides timestamp for photo	Critical	11/28/2012	timestamp	gml:TimePositionUnion
Water type (freshwater, saltwater)	for water column profiles	Useful	Fresh/brackish/salt	text	gml:CodeType
Depth	From lowest astronomical tide	Critical	18 m	numeric	gml:doubleOrNull
Above surface conditions	AOT/ atmospheric visibility/ clouds	Useful	high ceiling	text	gml:stringOrNull
Subsurface conditions	qualitative description of visibility	Useful	2m vis	text	gml:stringOrNull
Wave height and period (for reflectance measures)	Input for determining true depth relative to datum and wave lensing effects	Critical	0.25 m	numeric	gml:doubleOrNull

**Table 2.1** (continued) Metadata standard subset for underwater coral reflectance measurements

### Location Information Metadata

Wave height and period (for radiance measures)	Input for determining true depth relative to datum and wave lensing effects	Useful	0.25 m	numeric	gml:doubleOrNull
Tide conditions H or L	Input for determining true depth relative to datum and wave lensing effects	Critical	6:36 PM	time	gml:TimePositionUnion
Swell, wave height, long period waves	Input for determining water column depth	Useful	1 m	numeric	gml:doubleOrNull
Wind speed	optionality ranking dependent on severity of conditions	Critical/Useful	5 kn	numeric	gml:Quantity
Wind direction	optionality ranking dependent on severity of conditions	Critical/Useful	Ssw	text	gml:Direction
Height of sensor from surface (if characterizing water column properties)	for water column profiles	Critical	1.75 m	numeric	gml:doubleOrNull
Depth of sensor from surface (if profiling water column)	for water column profiles	Critical	7 m	numeric	gml:doubleOrNull
Natural canopy structure	Reference to photo illustrating canopy structure surrounding target	Useful	photo filename	text	gml:stringOrNull
Suspended sediment concentration (for water column studies)	Not useful for habitat spectral library	Critical	#mg l <sup>-1</sup>	numeric	gml:Quantity
Chlorophyll concentration	for water column profiles	Useful	#mg l <sup>-1</sup>	numeric	gml:Quantity
Secchi disk transparency/turbidity measure	for water column profiles	Useful	M (?)	numeric	gml:Quantity
CDOM spectral slope	Coloured dissolved organic matter for water column profiles	Critical	-S value	numeric	gml:Quantity
CDOM concentration	Coloured dissolved organic matter for water column profiles	Critical	A 440 nm	numeric	gml:Quantity
Detritus concentration	for water column profiles	Critical	1200 µg C•l <sup>-1</sup>	numeric	gml:Quantity
Phytoplankton species/classes	for water column profiles	Critical	Gymnodinium spp.	text	gml:stringOrNull

**Table 2.1** (continued) Metadata standard subset for underwater coral reflectance measurements

### Illumination Information Metadata

METADATA FIELD	REASON FOR INCLUSION / COMMENTS	OPTIONALIT Y	EXAMPLE	DATA TYPE	GML OBJECT TYPE
Optical measure of ambient conditions (direct, diffuse)	Description of general illumination conditions; useful for water column profiles	Useful	diffuse light field	text	gml:stringOrNull
Source of illumination (e.g. sun, lamp)	Type of illumination	Critical	halogen lamp	text	gml:CodeType
Bulb intensity	Input parameter for downwelling radiance calculation	Useful	100 W	numeric	gml:Quantity
Light spectrum	Range of irradiance spectrum	Useful	VIS/NIR	text	gml:stringOrNull
Single beam/multi beam	Input parameter for downwelling radiance calculation	Useful	single	boolean	gml:boolean
Beam coverage (as a degree measure)	Target surface area exposed to bulb radiance varies with beam spread	Useful	25°	numeric	gml:degrees
Time interval for weather station data logging	Used for cross-referencing weather station data with time of spectral measurement	Useful	15 min	numeric	gml:Quantity
Optical thickness of atmosphere	Qualitative description of visibility	Useful	good visibility	text	gml:stringOrNull
Visibility estimate	Estimated quantitative visibility	Useful	100 km	numeric	gml:Quantity
Cloud cover %	Estimated percentage of sky covered by clouds	Useful	25%	numeric	gml:Quantity
Cloud cover model	Model used to describe cloud cover	Useful	octave / quadrant / other	text	gml:CodeType
Cloud cover threshold for this project	Only useful if overcast	Useful	50%	text	gml:Quantity
Photo of sky (zenith to horizon)	Qualitative visibility data	Useful		image	
Wave lensing	Can't be measured in situ; Will know this from wave height data	Useful	yes/no	boolean	gml:boolean
Natural canopy shading	Only in seagrass, branching corals	Useful	seagrass shading	text	gml:stringOrNull
Artificial light canopy effect	Shadowing with diver's body to eliminate influences (eg. Wave lensing) If measurement is from a boat, then boat may shade	Useful	shadowing of target from diver	text	gml:stringOrNull

**Table 2.1** (continued) Metadata standard subset for underwater coral reflectance measurements

### Viewing Geometry Metadata

METADATA FIELD	REASON FOR INCLUSION / COMMENTS	OPTIONALITY	EXAMPLE	DATA TYPE	GML OBJECT TYPE
Distance from target	Measure of distance of sensor from the target	Critical	0.75m	numeric	gml:doubleOrNull
Distance from bottom/substrate	Yes, if 3D structure (seagrass, branching coral)	Critical	3m	numeric	gml:doubleOrNull
Area of target in field of view	Calculated if FOV specified	Useful	100%	numeric	gml:Quantity
Illumination zenith angle	Declination of illumination source from the zenith	Useful	15°	numeric	gml:degrees
Illumination azimuth angle	Horizontal angle of illumination source measured clockwise from a north base line	Useful	205°	numeric	gml:degrees
Sensor zenith angle	Declination of sensor from the zenith	Useful	5°	numeric	gml:degrees
Sensor azimuth angle	Horizontal angle of sensor measured clockwise from a north base line	Useful	75°	numeric	gml:degrees
Foreoptic	Degree measure of adjusted field-of-view of bareoptic fibre (due to attached foreoptic)	Critical	8°	numeric	gml:degrees
Distance of operator from sensor	Only applies if there is presence of shading from operator's body	Critical	0.25 m	numeric	gml:doubleOrNull

### Coral Target Properties Metadata

METADATA FIELD	REASON FOR INCLUSION / COMMENTS	OPTIONALITY	EXAMPLE	DATA TYPE	GML OBJECT TYPE
Target ID	Code identifier/tag for sample	Critical	Name code	text	gml:stringOrNull
Type	Qualitative descriptor of target type	Critical	Coral algae etc.	text	gml:CodeType
Species or name	Coral species	Critical	Diploria strigosa	text	gml:stringOrNull

**Table 2.1** (continued) Metadata standard subset for underwater coral reflectance measurements

Coral Target Properties Metadata (continued)

Size (diameter)	Size of target	Useful	30 cm	numeric	gml:Quantity
Location description (in situ/on boat/in lab)	Critical to quantifying environmental factors to spectral measurement	Critical	Lab/boat/in situ	text	gml:CodeType
Density of growth	Quantitative measure of density of target	Critical	2.94 g cm <sup>-3</sup>	text	gml:Quantity
Homogeneity/heterogeneity	Qualitative description of degree of homogeneity of target being sampled	Useful	homogeneous	text	gml:stringOrNull
Homogeneity/heterogeneity (photo)	Attached photo can be used as a reference	Useful		image	
Presence of epiphytes	Useful for endmember analysis of spectral measurements	Useful	Numerous epiphytes	text	gml:stringOrNull
Presence of epiphytes(photo)	Attached photo can be used as a reference	Useful		image	
Benthic microalgae (absence/presence)	Useful for endmember analysis of spectral measurements	Useful	Chla sampling	text	gml:stringOrNull
Distance from bottom	Input parameter for determining upwelling radiance/background reflectance affecting spectral measurements	Critical	20 m	numeric	gml:doubleOrNull
Substratum height	Input parameter for determining upwelling radiance/background reflectance affecting spectral measurements	Critical	4 m	numeric	gml:Quantity
Slope	Input parameter for determining upwelling radiance/background reflectance affecting spectral measurements	Useful	5%	numeric	gml:Quantity
Strike	Input parameter for determining upwelling radiance/background reflectance affecting spectral measurements	Useful	25°	numeric	gml:degrees



### 3 IMPLICATIONS FOR METADATA SHARING AND INTEROPERABILITY

A viable and practical metadata standard for underwater coral reflectance measurements must provide flexibility for data sharing in a common exchange format, while being suitably comprehensive in documenting the data relevant to the campaign. In the context of international data sharing of substratum and benthic spectral data, the establishment of standards for the capture, storage, and use of spectral signature files with associated metadata is required due to the effect of environmental factors in shallow water environments on the derived data (Dekker *et al.*, 2010). The standard proposed in Section 2 can be easily implemented as a schema in a common exchange format such as GML and XML (Extensible Markup Language). XML is self-descriptive with extensibility features (Mahboubi and Darmont, 2010) and can facilitate progress towards integration of *in situ* coral reflectance data with multi-dimensional remote sensing data sets, both within the marine context and near-shore terrestrial campaigns. One of its greatest strengths is platform independence, and a framework for XML-based data interchange is espoused in the Common Warehouse Metamodel, which includes XML Metadata Interchange (XMI) standards for datawarehouses (Mangisengi *et al.*, 2001 and Torlone, 2009). XML also facilitates searching and selection, it is human and machine readable, platform independent, convertible to other formats and allows quick assessment of suitability for other research products (Malthus and Shironola, 2009); it provides the greatest potential for data discoverability compared to the spectral archiving structures currently used by marine scientists in coral spectroscopy campaigns (including excel sheets and text files). The XML format can be easily accommodated in a variety of data archiving schema and software, including spectral libraries, databases, and datawarehouses.

Large-scale implementation of standards for encoding and sharing coral reflectance metadata is best facilitated by national and international agencies responsible for safeguarding and distributing these datasets. OGC (Open Geospatial Consortium) launched the Marine Metadata Interoperability Project to make data available from various ocean observing systems (OGC, 2012); however there are no specific metadata standards for *in situ* marine spectroscopy. IMOS (Integrated Marine Observing System, Australia) provides NetCDF specifications for *in situ* marine observations but are biased towards biochemical sensors and recording environmental variables, with no reference to spectroscopy measurements (IMOS, 2012). The ISO19115 sets of standards for geospatial metadata provide general guidelines, but do not explicitly address the metadata requirements of marine field spectroscopy collection techniques, or the ontologies and data dependences required to model the complex interrelationships among the observed phenomena as data and metadata entities (ISO, 2012). The lack of international standards impedes wide-scale mining and sharing of *in situ* marine spectroscopy datasets generated by remote scientists around the world. Adopting an XML-based metadata model for coral reflectance measurements is an initial step in establishing the foundations for a standard.

### 4 CONCLUSION

A practical and viable metadata standard for *in situ* coral reflectance can be used to inform a common data exchange standard for spectroscopy datasets in general. The model presented in this paper meets the requirements for a metadataset that is comprehensive, explicit, allows replication of the campaign if required, and is suitably broad in the scope of data capture to permit interoperability with other datasets. The standard is flexible by specifying both critical and useful metadata fields that are populated dependent upon the logistics of the campaign and the purposes for which the data will be used. *In situ* spectroscopy metadatasets are currently generated based on *ad hoc* data collection protocols that impede wide-scale data mining, sharing, intercomparison and interoperability of datasets. A metadata model based on the standard proposed here, in a common exchange format such as XML would facilitate convenient and practical data exchange among the remote sensing community.

### ACKNOWLEDGEMENTS

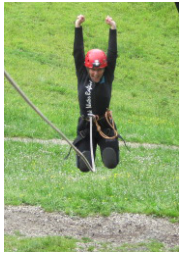
Marine remote sensing scientists at the ACEAS (Australian Centre for Ecological Analysis and Synthesis) Bio-optical workshop held in Australia in 2012 who generously provided input to the coral reflectance metadata schema proposed here.



## REFERENCES

- A.M. Bhatti, D. Rudnquist, J. Schalles, L. Ramirez and Seigo Nasu, "A comparison between above-water surface and subsurface spectral reflectances collected over inland waters", *Geocarto International Vol. 24. No. 2, 133-141, 2009*.
- A. G. Dekker, V. E. Brando, J. M. Anstee, A. J. Botha, Y. J. Park, P. Daniel., T.J.M. Malthus, S. R. Phinn., C. M. Roelfsema, I.A. Leiper, S. Fyfe, "A Comparison of Spectral Measurement Methods for Substratum And Benthic Features in Seagrass and Coral Reef Environments" *Proceedings of ASD and IEEE GRS; Art, Science and Applications of Reflectance Spectroscopy Symposium, Vol. II, 15pp, Boulder, CO, 2010*.
- M.J. Duggin, "Factors limiting the discrimination and quantification of terrestrial features using remotely sensed radiance", *International Journal of Remote Sensing, 6: 1, 3-27, 1985*.
- IMOS, "IMOS NETCDF FILE NAMING CONVENTION", Version 1.4, February 22 2012, viewed October 14, 2012  
[http://imos.org.au/fileadmin/user\\_upload/shared/IMOS%20General/documents/Facility\\_manuals/IMOS\\_netCDF\\_filenames\\_convention\\_v1.4.pdf](http://imos.org.au/fileadmin/user_upload/shared/IMOS%20General/documents/Facility_manuals/IMOS_netCDF_filenames_convention_v1.4.pdf)
- ISO, "ISO 19115:2003 Geographic information -- Metadata", 2003, viewed October 14 2012,  
[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=26020](http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020)
- H. Mahboubi and J. Darmont, "Optimization in XML Data Warehouses", pp 232-253 *E-Strategies for Resource Management Systems: Planning and Implementation*, E. Alkhalifa (Ed.), University of Bahrain, 2010.
- T. Malthus and A. Shirinola, "An XML-based format of exchange of spectroradiometry data", EARSel Imaging Spectroscopy SIG, Tel Aviv, March 2009.
- O. Mangisengi, J. Huber, C. Hawel, W. Essmayr, "A Framework for Supporting Interoperability of Data warehouse Islands Using XML", *Lecture Notes in Computer Science, 2001 Data Warehousing and Knowledge Discovery, Volume 2114, 328-338, 2001*.
- OGC, "Marine Metadata Interoperability Project", viewed October 14 2012, <http://www.ogcnetwork.net/node/345>
- K. Pfitzner, R Bartolo, G Carr, A Esparon & A Bollhöfer, "Standards for reflectance spectral measurement of temporal vegetation plots", *Supervising Scientist Report 195, Uniprint NT, Darwin, 2011*.
- C. Roelfsema, J. Marshall, E. Hochberg, S. Phinn, A. Goldizen, and K. Joyce, "Underwater Spectrometer System 2006 (UWSS04)", University of Southern Queensland, 2006, viewed August 01 2011,  
<http://ww2.gpem.uq.edu.au/CRSSIS/publications/UW%20Spec%20Manual%2029August06.pdf>.
- R. Torlone, "Encyclopedia of Database Systems", Part 9, p. 1560-1564, Springer Science + Business Media, LLC, 2009.

## Author Biographies



Barbara Rasaiah is a PhD candidate at RMIT University in Melbourne, Australia, investigating approaches to a coordinated evolution of hyperspectral metadata protocols, field spectroscopy methods and data exchange standards within the hyperspectral remote sensing community. Barbara's work has been presented at the ISRSE 34 conference, 7<sup>th</sup> EARSeL workshop, and ISPRS 2012. Barbara has an educational background in computer science and mathematics and has worked in industry as a computer programmer, web designer, and computer operations analyst. She was awarded the 2012 Goetz Instrument Award from ASD Inc., for novel and innovative research in field spectroscopy.



Simon Jones is professor of remote sensing and director of the Remote Sensing and Photogrammetry Research Centre at RMIT University in Melbourne, Australia. His current projects include leading research at TERN (Terrestrial Ecosystem Research Network), Commonwealth Environment Research Fund Hub "Landscape Logic", and organising the 2012 ISPRS International Congress on Photogrammetry and Remote Sensing. Simon's specializes in remote sensing, ground verification (in situ observations), spatial analysis, spatial data uncertainty, land-cover mapping, monitoring & modelling and vegetation. He is a foundation member and former director of the (Surveying and) Spatial Sciences Institute, Australia and has previously worked at the Joint Research Centre of the European Commission (Global vegetation Monitoring Unit).



Chris Bellman is associate professor and discipline head of geospatial science at RMIT University in Melbourne, Australia. His current projects include organising the 2012 ISPRS International Congress on Photogrammetry and Remote Sensing. Chris specializes in photogrammetry, GIS and spatial analysis and computer-aided mapping. He is a previous president of the Surveying and Spatial Sciences Institute of Australia. Chris is winner of the 2008 Victorian Spatial Excellence Award for Education and Professional Development.



Tim Malthus is leader of the Environmental Earth Observation program in CSIRO Land and Water in Canberra, Australia. His current projects include TERN (Terrestrial Ecosystem Research Network), IMOS (Integrated Marine Observing System) and the investigation of land use and land cover classification at high resolution. Tim's specialization in calibration/validation activities, and field spectroscopy with analysis of airborne and satellite Earth observation data, is applied in the development of improved monitoring tools for informing wider environmental policies. He has held positions as Senior Lecturer in Remote Sensing, University of Edinburgh, 1994–2009 and Director of the NERC Field Spectroscopy Facility, UK, 2004–09.