GeoInformation in Europe, M.A. Gomarsca (ed.)

Data and metadata coding and network services for the interoperability of satellite geophysical products

P.F. Pellegrini, M. Tommasini
DET – Telecommunication and Electronic Department, University of Florence, Florence, Italy

M. Gherardelli, G. Poli & G. Adembri
Satellite Receiving Station of Polo Universitario di Prato (PIN), University of Florence, Prato, Italy

Keywords: INSPIRE, interoperability, metadata, ISO, XML, satellite, remote sensing

ABSTRACT: The recent European Community policy on the environment is working in defining a Directive for the Creation of an Infrastructure for Spatial Information in the Community (INSPIRE); this Directive aims at promoting geographic data interoperability between governments or other public administrations, ordering the utilization of metadata and proper data coding formats and the establishment of network services for geographic information distribution.

Direct satellite data reception and monitoring of sea and land geophysical quantities on a regional scale are regularly carried out at the Satellite Receiving Station of Polo Universitario di Prato (PIN), Italy. Within those activities, issues were considered that deal with geographic data coding, metadata organization and network services management to search and share all the obtained products. A monitoring activity specially addressed to the Tuscany Region, was the basis of a collaboration with the Tuscany Region Environmental Protection Agency (ARPAT) since 1999.

Within the investigation of aspects concerning the interoperability of satellite geographic products, a review of the current reference standards was performed; the ISO 19100 series gives the rules for metadata organization and coding with XML (*eXtensible Mark-up Language*). Because of the lack of precise standards about data coding, at the Satellite Receiving Station, the HDF (*Hierarchical Data Format*) and GeoTIFF (*Geographic Tagged Image File Format*) file formats were considered, which are widely used by the scientific community.

1 INTRODUCTION

The Satellite Receiving Station of Polo Universitario di Prato (PIN) is equipped with the instrumentation for direct and real time the reception from NOAA and METEOSAT Second Generation (MSG) satellites. Using original software (developed in the last 10 years at the Station and at the Telecommunications Laboratory–University of Florence), the Station's activities involve environmental monitoring on a sub-regional, regional and national scale. Such an activity consists in obtaining various geophysical sea

and land products from satellite data and in particular: SST (Sea Surface Temperature), LST (Land Surface Temperature), NDVI (Normalized Difference Vegetation Index), primarily extracted from AVHRR (NOAA) e SEVIRI (MSG) sensors. Since 1997 a consistent archive (data repository) with raw data and processed images has been built that consists in about 40′000 AVHRR/NOAA, 80′000 METEOSAT-7 and 50′000 MSG passages and more than 2′000 processed images. These geophysical products are of noticeable interest in many activities regarding environmental monitoring, activities which involve public offices or authorities and private companies.

Since 1999, the Satellite Receiving Station's activity has included collaboration with ARPAT, the Tuscany Region Environmental Protection Agency, aimed at monitoring in real time the Tuscany Archipelago Sea (1999–2004) and participation in the SEDEMED II (Sécheresse et Désertification dans le bassin Méditerranée / Drought and Desertification in the Mediterranean Basin) Project, aimed at revealing desertification tendencies in the Mediterranean Sea Basin (2004–2006). The experience acquired during this research brought about a development in the satellite data processing system at the Receiving Station, particularly focusing on aspects related to integrating products obtained from other units of research and to the exchange of geographical information.

Considering current international standards, especially from International Organization for Standardization (ISO) and Open Geospatial Consortium (OGC), about the treatment of geographical information (still being developed), the Station's recent ongoing research pertains to the adaptation of data processing and product archiving procedures. In Europe, such standards inspired the INSPIRE (EU 2007) directive, approved by the European Union and entered into force for Member States in May 2007. This directive furnishes guidelines for creating an infrastructure aimed at exchanging geographical information through the use of metadata and appropriate network services; it also sets 2012 as a deadline for Member States to adopt the provided rules, regarding several categories of geographical information indicated in the same directive. Activities in the framework of the INSPIRE initiative (which is still in progress) have produced implementing rules for organising metadata (INSPIRE 2007).

2 DATA PROCESSING AND METADATA MANAGEMENT

The Satellite Receiving Station of Polo Universitario di Prato (PIN) directly receives data from NOAA and MSG satellites (Figure 1) and from other satellites such as SSM/I, MODIS, ENVISAT via internet. Data processing, executed in almost real-time, makes it possible to obtain different geophysical products (from single satellite passages or monthly mean observations), for example Sea Surface Temperature (SST), Land Surface Temperature (LST), Vegetation Index (NDVI); the products are obtained with different typologies: geolocated or non-geolocated, with or without land/sea or cloud pixel masking, with or without the presence of coastlines, geographical grids, palettes, and other artefacts. According to the typology, the geophysical products are stored on files

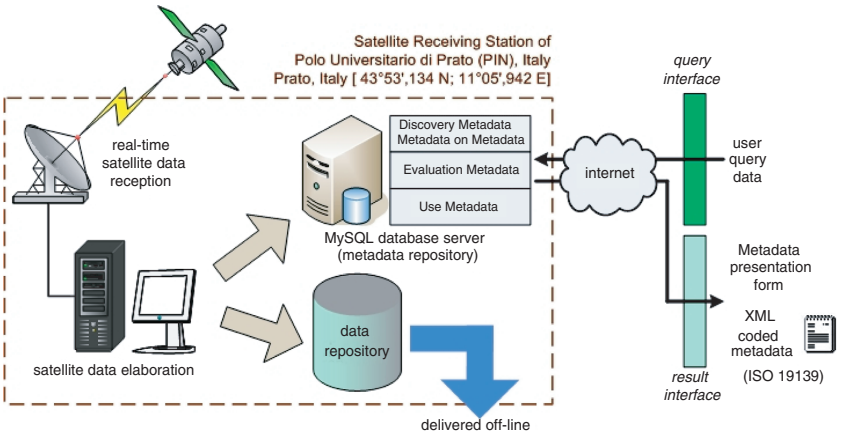

Figure 1. Data processing and metadata management at the Satellite Receiving Station of Polo Universitario di Prato (PIN).

Table 1. Characteristics of the geophysical product types obtained at the Satellite Receiving Station of Polo Universitario di Prato (PIN): products can be geolocated or non-geolocated, with or without land/sea or cloud pixel masking and with or without the presence of coastlines, geographical grids, palettes, and other artefacts.

Product type	Land/sea mask	Cloud mask	Coastlines, reticulate, palette, other artefacts	Storage file format
Non-gelocated (only single-pass products)	no	no	no	HDF
Geolocated (single-pass and monthly products)	yes	optional (single-pass products); yes (montly products)	no	HDF, GeoTIFF
TIFF/JPEG image (single-pass and monthly products)	yes	yes	yes	TIFF, JPEG

using some standard formats widely used in the international science community, like HDF (HDF 2005), GeoTIFF (Ritter & Ruth 2000), JPEG, and TIFF (Table 1).

During the processing phase, metadata related to every obtained product is extracted and archived on a MySQL database server (metadata repository). As indicated in the INSPIRE implementing rules (INSPIRE 2007), the extracted metadata is divisible into four categories, whose contents are shown in Table 2: discovery metadata (level 1 & 2), metadata on metadata, evaluation metadata and use metadata. The contents of discovery metadata and metadata on metadata are set in (INSPIRE 2007), whereas the contents of evaluation metadata and use metadata are left undefined, being strictly dependent on the geographical data typology provided.

Table 2. Contents of metadata categories as indicated in the INSPIRE implementing rules (INSPIRE 2007).

Category of metadata	Contents
Discovery metadata (level 1 & 2)	The minimum amount of information that needs to be provided to convey to the user the nature and content of the geographic resource
Metadata on metadata	Information about metadata management, such as metadata point of contact, creation date and language
Evaluation metadata	The amount of information sufficient to enable a user to ascertain that a spatial resource fit for a given purpose exists, to evaluate its properties, and to reference some point of contact for more information
Use metadata	Information required to access, transfer, load, interpret, and apply a spatial resource in the end application where it is exploited (including projection and geometric characteristics, and other parameters that are useful to human and machine in the proper use of spatial resources)

Particular attention has been given to information, within the evaluation metadata category, on the furnished product's quality, meaning the combination of information that describes the product's characteristics (dataset) and circumscribes its use. The organisation of this information is described in (ISO 2003a) and (ISO 2003b). First of all, descriptions of the processing phases are provided which lead to the obtained products (lineage); following is the quantitative quality information concerning the product: the commission/omission (excess or lack of information), the logical consistency in respect to the scope of the dataset (e.g. an SST image scope is to furnish the sea surface temperature for every image pixel), format and domain consistency of numerical values associated with image pixels (presence/absence of out-of-domain values or incorrect formats), the positional accuracy of image pixels (in the case of geolocated products), the accuracy of dataset temporal references (e.g. the time acquisition estimate of satellite scan lines), thematic accuracy (e.g. correctness in attributing a pixel image to a certain class, like land, sea or cloud, in case the images include classification information), quantitative accuracy in estimating the considered geophysical product (e.g. the precision with which the acquisition system and the algorithm used allow the evaluation of the physical quantity investigated). As recommended in ISO 19114 and 19115 (ISO 2003a, ISO 2003b), for each of the classes providing quantitative quality information, the quality measure description, the evaluation method description, the measure numerical result and its units are provided. Table 3 shows the adopted quality classes and the related evaluation method descriptions.

3 SEARCH NETWORK SERVICES

The ongoing activity at the Satellite Receiving Station also includes developing and managing network services for consulting the database server (metadata repository)

Table 3. Evaluation method description (evaluationMethodDescription, attribute nr. 104 in ISO 19115) for the adopted quantitative quality information classes defined by ISO 19114 and 19115 (ISO 2003a) e (ISO 2003b); for each class also the measure description and the numerical result of quality evaluation are provided.

ISO 19115 data quality class	evaluationmethoddescription (attr n. 104)
DQ_CompletenessCommission	Divide number of pixels not included in the image by count of image pixels
DQ_CompletenessOmission	Divide number of lacking pixels by count of image pixels
DQ_ConceptualConsistency	Divide number of pixels not representing SST (e.g. land pixels) by count of image pixels
DQ_DomainConsistency	Divide number of pixels whose value is out of the range of acceptable values by count of image pixels
DQ_FormatConsistency	Divide number of pixels whose data type is incorrect by count of image pixels
DQ_AbsoluteExternalPositionalAccuracy	Evaluate RMSE (root mean square error) of image pixel geolocation (in meters)
DQ_AccuracyOfATimeMeasurement	Evaluate absolute error in the instant of the first satellite scan line processed to obtain the image (in seconds)
DQ_ThematicClassificationCorrectness	Divide number of badly classified pixels (e.g. land instead of sea and cloudy instead of free) by count of image pixels
DQ_QuantitativeAttributeAccuracy	Evaluate RMSE (root mean square error) of image pixel SST values (in Kelvin) with respect to true values (e.g. derived from *in situ* measurements, if available)

that keeps metadata related to elaborations which have taken place (Figure 1). To perform the database search an interface query is adopted with a reduced number of fields whose content is chosen among multiple pre-set choices. The data to be inserted in the query concern the name and type of product, the satellite data's origin, the temporal interval and the geographical extent of the region of interest. The result of the user query is a list of products that are likely to satisfy the user request; for each of them it is possible to visualize all the metadata, belonging to the categories described in Table 2.

Figs. 2–4 show an example of metadata repository interrogation, in the search for a geolocated SST image from AVHRR data of a single NOAA pass, between 11.07.2002 and 12.07.2002 in the geographic area approximately corresponding to the Tuscany Region. The interrogation (Figure 2) produces the list of results seen in Figure 3, where, for the data file names, a naming convention described in (WMO 2004) was followed. For every data file listed as query results, it is possible to visualize the archived metadata, in order to verify whether or not the available products are actually suitable for the user's requests. First metadata from the discovery (level 1 & 2) and metadata on metadata categories are visualized, followed by metadata from evaluation and use categories. For the example, the metadata is visualized of an SST image of the Tuscan

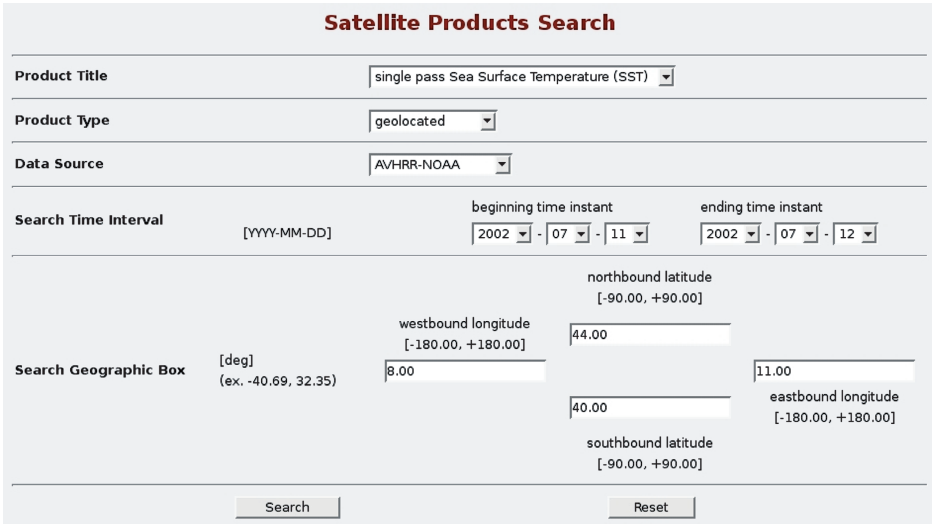

Figure 2. Query interface: satellite products are searched in the metadata repository of the Satellite Receiving Station of Polo Universitario di Prato (PIN) using a few attributes whose values are selectable among a group of fixed entries.

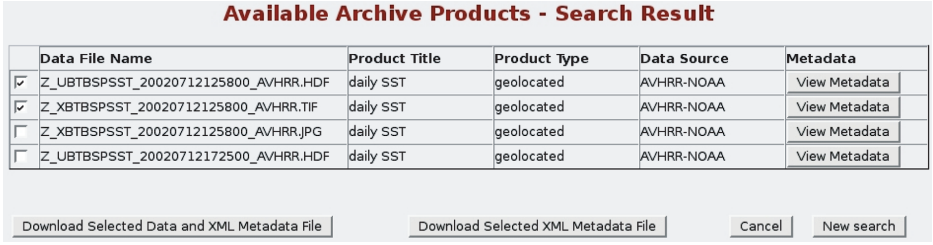

Figure 3. Result interface: the list of available archive products is displayed; for each product the whole metadata set can be viewed or downloaded as an XML coded text file (ISO 2007).

Archipelago taken from of a single NOAA pass (AVHRR sensor), registered with bilinear interpolation over a WGS84 Mercator projection grid, with land/sea and cloud pixel mask, without artefacts and using the HDF file format to store image data. From the page where the query results are visualized (Figure 3), a link is present for each of the available products for viewing the entire metadata set; Figure 4 shows some of the metadata (only those from the discovery level 1 category) associated to the first item of the list in Figure 3.

Metadata can also be downloaded as XML coded text files; XML coding is carried out following the schema provided by ISO 19139 (ISO 2007) and is validated using the Xerces C++ software (Apache 2005).

Z_UBTBSPSST_20020712125800_AVHRR.HDF Metadata

Discovery Metadata - level 1	attribute name (ISO nr., attr. nr.)	value
Resource title	title (19115, 360)	single pass SST 16 bit/pixel, geolocated (bilinear interpolation), land+cloud mask
Temporal reference (begin position)	gml:beginPosition (19108)	2002-07-12T12:58:00
Temporal reference (end position)	gml:endPosition (19108)	2002-07-12T12:58:00
Geographic extent (westBoundLongitude)	westBoundLongitude (19115, 344)	9.40
Geographic extent (eastBoundLongitude)	eastBoundLongitude (19115, 345)	11.40
Geographic extent (southBoundLatitude)	southBoundLatitude (19115, 346)	42.20
Geographic extent (northBoundLatitude)	northBoundLatitude (19115, 347)	43.60
Resource language	language (19115, 39)	ita
Resource topic category	topicCategory (19115, 41)	oceans
Keyword	keyword (19115, 53)	SST
Service type	serviceType (19119)	archive search, data ordering
Resource responsible party (individual name)	individualName (19115, 375)	Tommasini, Maurizio, PhD
Resource responsible party (organisation name)	organisationName (19115, 376)	Stazione Ricezione Satelliti - Polo Universitario di Prato (PIN)
Resource responsible party (role)	role (19115, 379)	originator
Abstract	abstract (19115, 25)	Sea Surface Temperature
Resource locator (linkage)	linkage (19115, 397)	http://maresat.ing.unifi.it/archivio
Resource locator (connect point)	connectPoint (19119)	http://maresat.ing.unifi.it/ricerca

Figure 4. Result interface: for a selected product, discovery metadata (level 1 & 2), metadata on metadata, evaluation metadata and use metadata can be viewed on an internet page.

NOTE

MSG data (Meteosat Second Generation) referred to in this paper originate from EUMETSAT. They were obtained through the license issued by Servizio Meteorologico dell'Aeronautica Militare (Ministero della Difesa conv. N. 4123/62 of 15.11.2005 and Prot. N.M-DGTEL/4123/2508 of 25.01.2007)

REFERENCES

EU (EUROPEAN UNION) 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007, establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, L108/1-14, 25 April 2007.

HDF (Hierarchical Data Format) Group 2005. HDF5 User's Guide (Preliminary Version for Release 1.6.5). National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana–Champaign (UIUC), November 2005 (http://hdf.ncsa.uiuc.edu/HDF5/doc/PSandPDF/HDF5_UG_r165pre.pdf).

INSPIRE (Infrastructure for Spatial Information in Europe) – DT Metadata 2007. Draft Implementing Rules for Metadata (D1.3). DT Metadata, 2007 (http://www.ec-gis.org/inspire/reports/ImplementingRules/draftINSPIREMetadataIRv2_20070202.pdf).

ISO (International Standards Organization) 2003a. ISO 19114:2003, Geographic information – Quality evaluation procedures. TC 211 Technical committee "Geographic information/Geomatics", ISO 2003.

ISO (International Standards Organization) 2003b. ISO 19115:2003, Geographic information – Metadata. TC 211 Technical committee “Geographic information/Geomatics”, ISO 2003.
ISO (International Standards Organization) 2007. ISO/PRF TS 19139:2007, Geographic information – Metadata, XML schema implementation. TC 211 Technical committee “Geographic information/Geomatics”, ISO 2007.
Ritter N. & Ruth M. 2000. GeoTIFF Revision 1.0, Version: 1.8.2. 28 December 2000 (ftp://ftp.remotesensing.org/pub/geotiff/spec/geotiff.rtf.Z).
The Apache Software Foundation 2005. Xerces-C++ Documentation, Version 2.7.0, 19.09.2005. (http://xml.apache.org/xerces-c/pdf/xerces-c.pdf).
WMO (World Meteorological Organization) 2004. Manual On The Global Telecommunication System, Volume I, Global Aspects, WMO–No. 386, Secretariat of the World Meteorological Organization, Geneva, Switzerland, 2004, ISBN 92-63-12386-14.