CVRD Metadata Standard

Marina Hitomi Ando
Marcelo Roberto Barbosa
Alessandro Frizzera

Abstract

Companhia Vale do Rio Doce (CVRD) is a leader diversified mining company of Americas and the world biggest iron ore producer and supplier. Our business also includes logistics and eletric power generation. During 2004 the GIS team has implemented the first module of its Enterprise GIS (Metadata Server) and has surveyed and organized geographic data among all its business units. This presentation shows CVRD's Metadata Standard which was established in the company to provide identification, spatial reference, data quality and metadata reference of common use geographic data, such as digital satellite image, scanmaps, aerial photography and vector data. The standard created was based on a survey that identified alignments with FGDC and adjustments according to CVRD needs and generates several benefits for the company's GIS community. The technology is based on ESRI products (ArcSDE, ArcIMS, ArcGIS).

Why Use Metadata

Metadata consists of properties and documentation of data. Properties are derived from the item itself, documentation is descriptive information supplied by a person.

Documenting your data protects your organization's investment in that data. When detailed metadata has been created, it can answer your questions and help you make decisions. For example: it can help you to determine when your data is out of date, what map scale is appropriate for presenting your data, or how accurate your data is.

Usually, the existing documentation about the data is made in a few lines in a text file. This kind of document is useful only for the person who create it. Some historical processes may be omitted or lost because it may seem obvious at that moment. It is important that other people have access to it and be able to better comprehend this data information, particularly information describing how data should be used.

To unify all information registration, and even prevent errors of omission and misinterpretation, a standard to document CVRD spatial data was created, and a sytem based on GIS was implemented to provide a set of tools for data catalogue accessible for all CVRD units throughout Brazil.

In most cases, the data was acquired according to specific needs (just enough for the project), and that may turn the data not appropriate for other projects.

And with no defined procedure to distribute the data, these information used to be transmitted from one person to another and sometimes it could cause some distortion or error about data information.

Also the data might not be transmitted with enough information causing the loss of important historic data, so that workers have little understanding of the contents and ways to use this data and because of that the workers won't find the results, generated from these data, reliable .

Other examples of problems we had in the past:

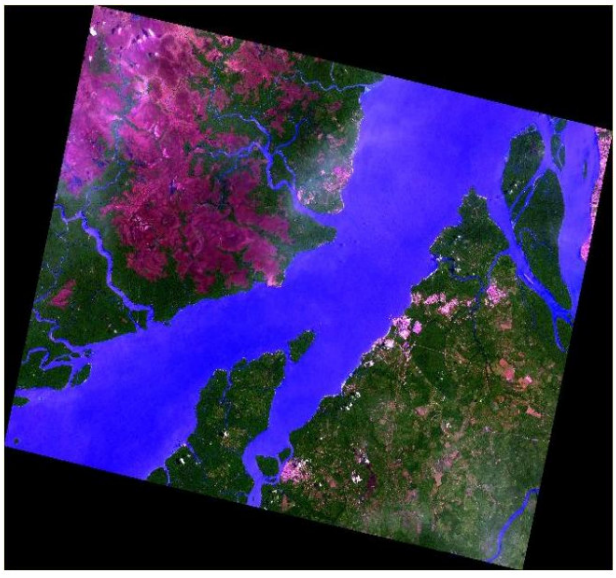
- data without spatial reference
- data without processing steps, original files and their data acquisition or RMS obtained from georeferencing
- package of data without purpose or project history
- Unusual image formats

Generating Metadata

With metadata, these attributes and history described above is associated with geographic data.
All of these issues now can help workers to decide if the data can be used on other projects.

With these informations the worker can better understand, evaluate and decide if it's necessary to acquire a new data or to request it from the CVRD Unit responsible for the data.

Other actions like file naming, storage conventions, use of properties domains for most of the attributes and providing CVRD Unit contacts, also was implemented to speed up data search.



File Name:
img_spo_040907_701353.tif

Presentation Form:
Raster digital data

Acquisition Technology:
Spot5

Scene: 701/353

Bands: 3,2,1,8

Acquisition Date:
07/09/2004 (dd/mm/yyyy)

Project: ABC Refinary

Processes:
Equalization; Georeferencing

CVRD Unit: DIAT

Figure 1 – Example of geographic data and metadata.

Nowadays the procedure for metadata generation occurs when the data is produced or acquired.

Data is loaded into the GIS software and overlayed with other cartographics data.

At this moment:

- a visual check should be made to certify that the image is geographically correct
- an extent polygon of the data is created
- the CVRD Metadata Standard is associated with polygon shapefile and xml file.

The automation of metadata generation has been applied during evaluation of existent dataset in a CVRD unit. So, all medias (CD-ROM, DVD, Servers, Hard Disk) have been scanned to obtain data extents and physical properties.

On the other hand, when a package of data (like mosaics) is acquired, the source is asked to deliver the metadata, already under CVRD's metadata standard.

Mostly because of time consuming loading raster data, the use of extent polygons was the best solution to speed up data cataloguing and metadata publication – making the data extent (polygon) an important metadata too.

Enterprise GIS

The Territorial Management System (Territorial GIS) was identified as important business tool during 2002 through a directory decision. Its definition was made by a committee formed by mining, exploring and environmental departments of the company.

The system has two modules **Metadata Module** and **Management Module.**

The first module has been implemented at CVRD since September, 2004.

The evaluation of existent geographic data has been made through CVRD business units. This generated an initial amount of more than 2.000 metadata of common used geographic data.

The second module Management Module is already at CVRD under test phase. Its implementation must occur until the end of this year.

With these two modules, CVRD has been creating an important vehicle for data sharing and GIS dissemination to most of their business units carrying answers to corporative needs:

- What data are available
- Whether the data meets specific needs
- Where to find the data
- How to access the data
- Where to manipulate data (Map Production)

Many CVRD business units still have their own geographic database and GIS applications that runs isolated or shared within a few departments.

Territorial GIS has been allowing common datasets sharing, to be an important tool to eliminate redundancies, to increase spatial and information accuracy and to build applications for data dissemination throughout the organization.

Its integration with mainstream IT turns it more powerful and pervasive in the company.

Integration between Departmental and Enterprise GIS's datasets (models) still permits development of specific applications, databases and it guarantees the elimination of redundant data in the company, maintaining foundation datasets available for department applications and vice-versa.

This integration has been creating a common acquisition process, a database which is useful for most of CVRD projects.

System Architecture

The CVRD system architecture assists to at least 3 CVRD GIS projects:

- Territorial GIS (enterprise GIS)
- Mineral GIS (departmental GIS)
- RailRoad GIS (departmental GIS)

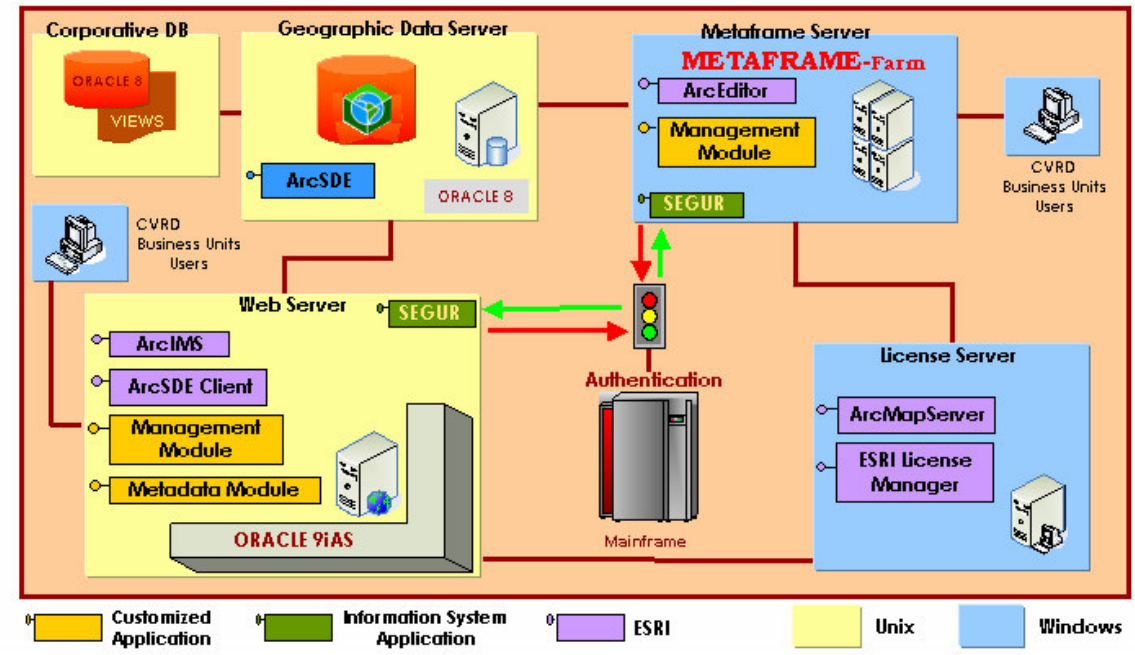


Figure 2 – CVRD GIS Architecture

These servers are connected by an optical fiber network and they provide:

- Geographic data with ArcSDE 8.3 and Oracle8i
- Intranet applications with ArcIMS, powered by ArcMapServer extension
- ArcMap customized applications on metaframe
- ESRI license manager

The Territorial GIS can be accessed via:

- intranet for metadata catalogue and map visualization
- citrix metaframe for data editing and map production

The user is authenticated through CICS mainframe and the user permissions are evaluated by SEGUR which is used by most of CVRD System Applications.

Functionalities

This is the UML diagram showing detailed tools of Metadata Module for two user profiles.

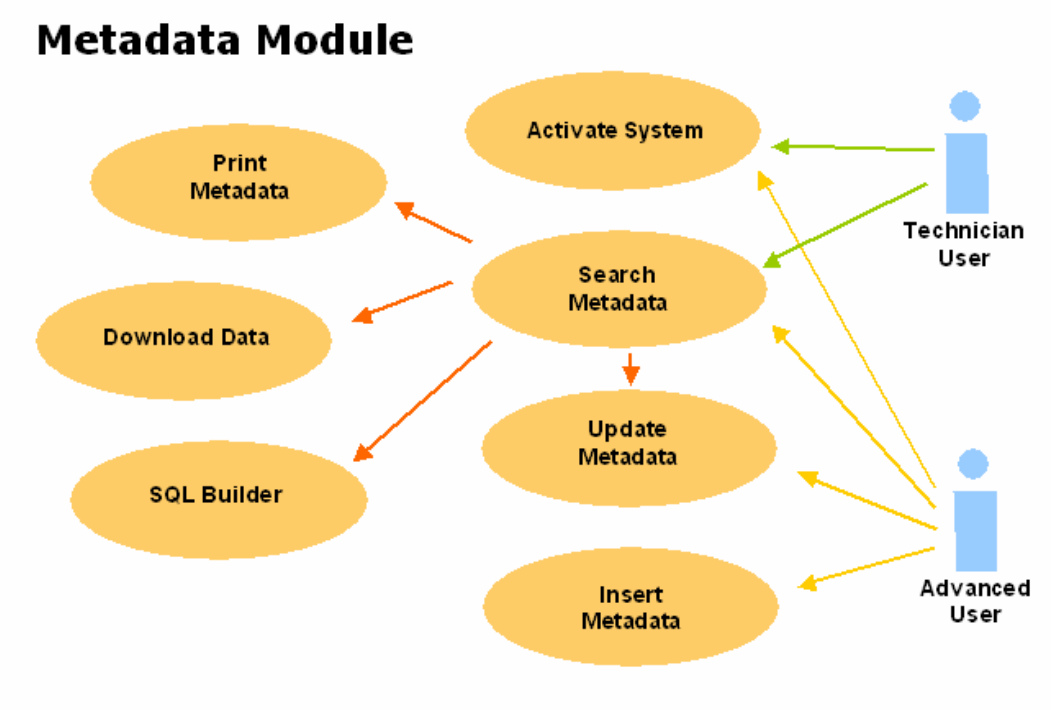


Figure 3: Functionalities and User Profiles

The **Technician** level user is the professional who uses geographic data in different ways: to acquire data, to provide data to external companies, to use maps, to prepare reports containing maps and researches.

The **Advanced** level user is the one who already has a good GIS experience and can make database maintenance like generating, processing geographic data for any project.

Searching Metadata

The Metadata Module provide data searching through region selection (a state, county or coordinates) and alpha-numeric (metadata attributes) filter application.

The result of the searching brings all of polygons that intersect the selected region. Each enumerated polygon has a link that provide the metadata editor interface with attributes of the metadata polygon.

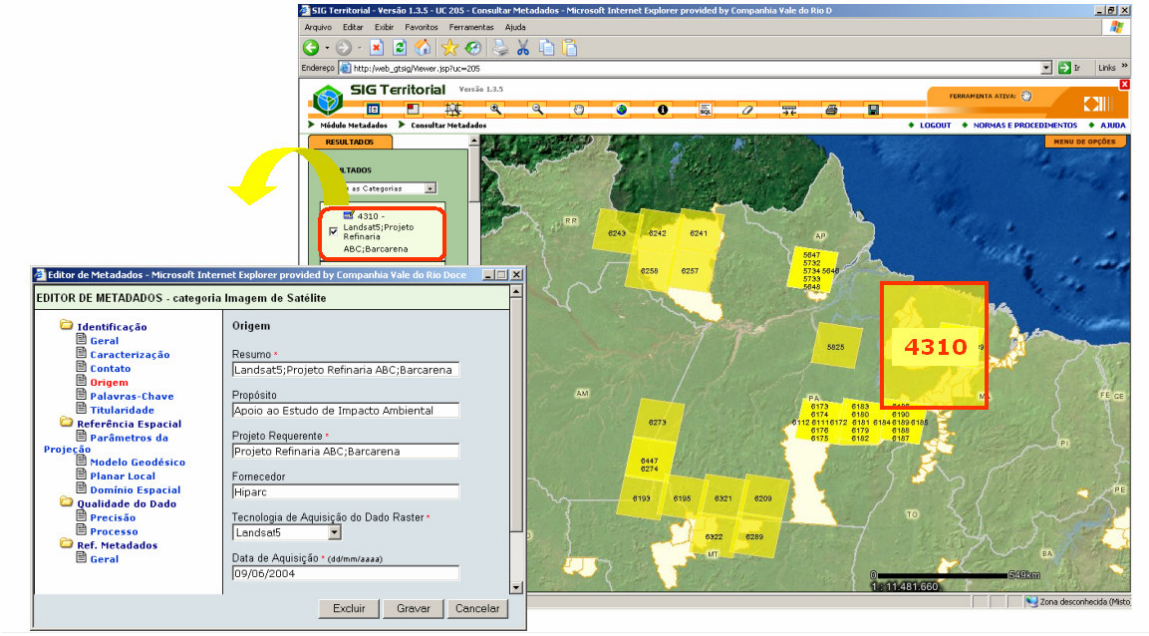


Figure 4: Metadata searching through Territorial GIS.

The first metadata package loaded in database summarized 2163 records representing Satellite Images (Landsat, Ikonos, SPOT) and Scanned Maps.

When an item is located, the file can be requested to the person indicated in the Contact Identification section. Eventually, the file can be downloaded, if its size is lower than 50Mb.

Editing Metadata

To insert a metadata the data extent must be indicated (by coordinates) or selected (the scene) in map, then a category must be selected (Database, Shapefile, CAD, Scanmap, Aerial Photography or Sattelite Image) and finally the data information according to CVRD Metadata Standard.

CVRD Metadata Standard

The CVRD Metadata standard is divided into 4 categories: satellite image, scanmaps, aerial photography and vector data.

This standard is a simplification and an adaptation of FGDC[1] (Federal Geographic Data Committee) according to GIS team experience.

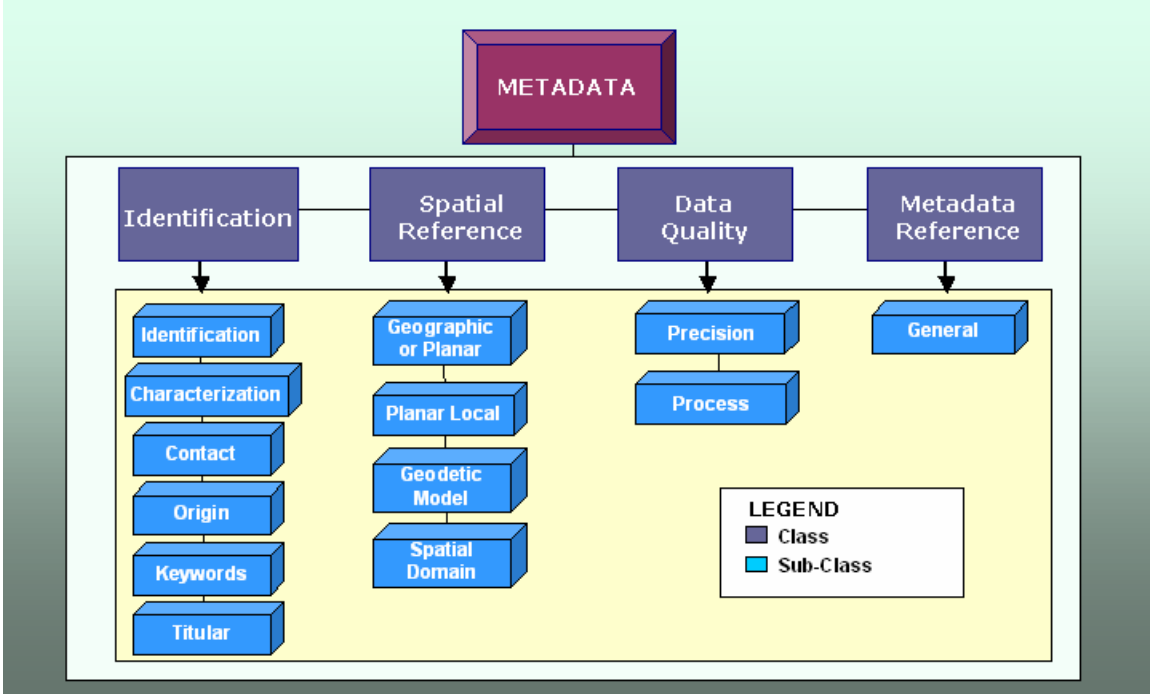


Figure 5: The CVRD Metadata Standard and its classes and subclasses

Basically was identified 4 classes equivalent to the FGDC sections:

- Identification: basic informations like physical characteristics, data localization, origin, keywords and security.
- Spatial Reference: system coordinate and other projection parameters of data.
- Data Quality: properties that indicate data accuracy and how transformed was the data.
- Metadata Reference: information about metadata registration.

The behavior of standard varies according to data type (vector or raster), some attributes can be another one or suppressed.

To understand CVRD Metadata Standard some subclasses above will be explained for better comparison with FGDC.

[1] FGDC Content Standard for Digital Geospatial Metadata (STD-001-1998) – Version 2.0

Identification
The **Origin** subclass provides attributes that registers historic events that determined the data acquisition.

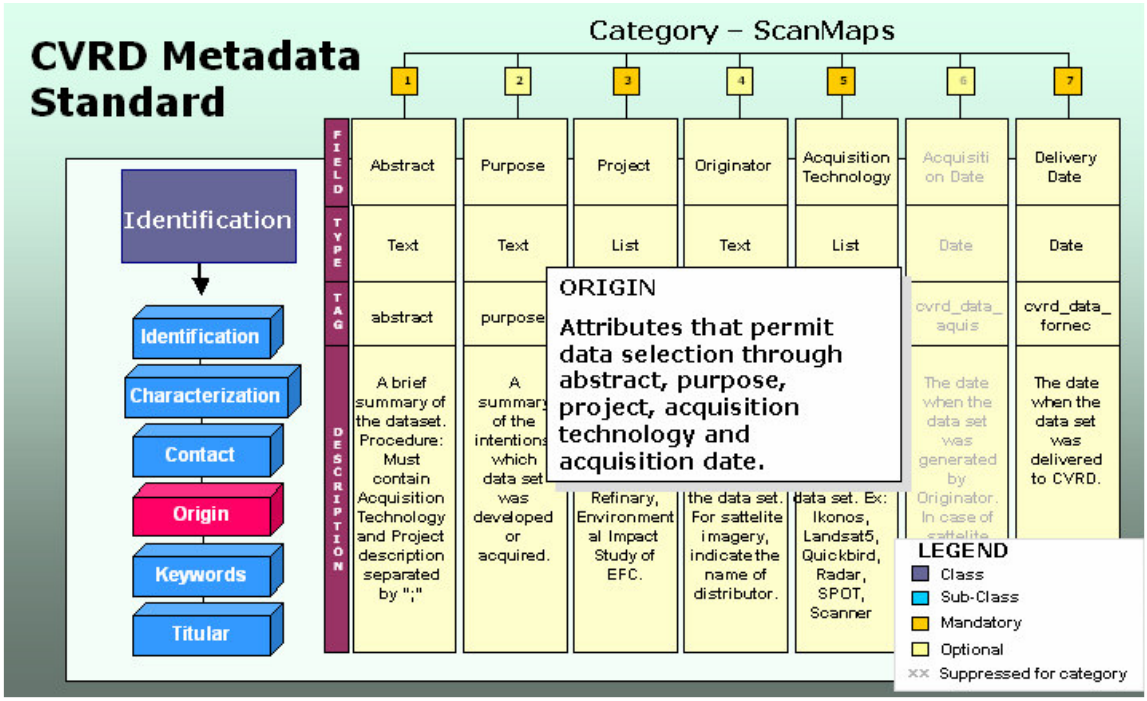


Figure 6: Origin subclass for Scanmap category

Alignments with FGDC:
- Abstract
- Purpose
- Project
- Originator

… and new attributes created:
- Acquisition Technology
- Acquisition Data
- Delivery Date

The legend shows attributes that are mandatory (orange) and attributes that are optional (yellow) for Scanmap data.

The Acquisition Date is not considered mandatory for Scanmap category duo to existent of Publication Date in Characterization subclass. For Satellite Image category Acquisition Date is mandatory.

The **Identification** subclass provides physical identification and data localization through CVRD units.

Filename, Path and Filename and Presentation Form is considered mandatory for all data categories.

Alignments with FGDC:

- Filename
- Path and Filename
- Presentation Form
- File Format

... and new attributes created:

- Unit Identification
- System Identification

As the File Format for Vector category are different from raster categories, another field was created to storage this information for vector data.

To storage these new attributes, a specific set of **tags** was created and implemented in XSL file, like CVRD\IDENTIFICATION\ORIGIN\cvrd_data_fornec.

Data Quality

The **Process** subclass provides historic information about data transformation and the reference for the original data used to generate it.
The attributes of this subclass for Sattelite Imagery category are optional.

Alignments with FGDC:

- Processing Software
- Processing Date

and new attributes created:

- Processing Company
- Processes
- Applied Filter
- Original Files

The field Processes for Vector Category are different from raster categories and another field was created to storage this information.

Spatial Reference

The **Geographic or Planar** subclass identificates dataset coordinate system or projection's parameters.

All attributes are optional and come from FGDC.

Geographic System Name
Projection Name
UTM Zone Number
Latitude of Projection Origin
Longitude of Central Meridian
Scale Factor
Geographic Coordinate System
Planar Coordinate System

Conclusion

The metadata creation takes a principle that the originator must be responsible for data events and data properties registration.

When the workers receives the data they must demand for its metadata too.
In case of data produced or data transformed by the business unit workers the metadata must be created by them.

There are several advantages using metadata, the most importants are:
For the Data

- Metadata publication brings liability through standardized and qualified data
- Optimize geographic data localization
- Provide correct utilization of data
- High level of confidence in existent data
- Minimize duplicated data acquisition

For the Company

- Protect CVRD investment in data
- High level of confidence in decisions based on data
- Better communication between GIS staff
- Provide information about organization's data holdings

Reference

Vienneau, Aleta and Bailey, Jonathan, 2003 - Using ArcCatalog, ESRI Press 2001

Federal Geographic Data Committee, 2000 - FGDC Metadata Workbook, Version 2.0, FGDC-STD-001-1998

Bowman, Scott 2005 - Building na Enterprise GIS in a Limited Fiscal Environment, ArcUser Jan Mar

ESRI, 2005 - Enterprising GIS Management, ArcUser Jan Mar

Authors

Marina Hitomi Ando
GIS Analyst
Cia Vale do Rio Doce
Av. Graça Aranha, 26
Rio de Janeiro – RJ 20030-900
ift002@cvrd.com.br

Marcelo Roberto Barbosa
GIS Coordinator
Cia Vale do Rio Doce
Av. Graça Aranha, 26
Rio de Janeiro – RJ 20030-900
marcelo.barbosa@cvrd.com.br

Alessandro Frizzera
System Analyst
Cia Vale do Rio Doce
Av. Dante Michelini, 5500
Vitória – ES 29090-900
alessandro.frizzera@cvrd.com.br