



Introduction to Metadata

R. Venkat Shesu

Scientist

Indian National Centre for Ocean Information
Services (INCOIS)

Hyderabad, India

Learning Objectives

- ▶ Metadata
- ▶ Purpose of Metadata
- ▶ Contents of metadata
- ▶ Types of metadata
- ▶ Metadata lifecycle



Data Exchange

- ▶ Provides the data
- ▶ Receives the data



- ▶ Metadata provides the understandable information
-



Data Exchange

- ▶ **While providing data**
 - ▶ Objective of data creation
 - ▶ Limitations
 - ▶ What is available in data
 - ▶ How to cite
- ▶ **While receiving data**
 - ▶ How data is created
 - ▶ Data charges
 - ▶ What data is available
 - ▶ Software to process data
 - ▶ Is data transferable



What is Metadata?

- ▶ Metadata is **data about data**. Metadata describes how and when and by whom a particular set of data or a service was collected or prepared, and how the data is formatted or the service is available. Metadata is essential for understanding information stored in and has become increasingly important.
- ▶ Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information.

http://www.marinespecies.org/introduced/wiki/Metadata_and_metadata_catalogues



Curriculum Vitae

Personal Information

R. Venkat Shesu
Scientist – D, Data and Information Management Group (DMS)
Indian National Centre for Ocean Information Services (INCOIS)
'Ocean Valley', Pragathi Nagar (BO), Nizampet (SO)
HYDERABAD – 500 090, India

Phone: 91-40-2388 6069 / 91-90102 64931

Fax: 91-40-2389 5001

E-mail: venkat@incois.gov.in

URL: www.incois.gov.in

Date of Birth

30-10-1977 (Thirtieth October Nineteen Seventy Seven)

Education

- M.Tech.(Computer Science & Engineering), Acharya Nagarjuna University, 2005
- M.Sc.(Computer Science), Acharya Nagarjuna University, 2000
- B.Sc.(Maths, Statistics, Computer Science), Acharya Nagarjuna University, 1998

Research Interest

- Ocean data and information management
- Application of data mining algorithm for oceanographic databases.
- Quality control procedures.
- Data interoperability experiments.

Professional Experience

- **Scientist, Data and Information Management Group at INCOIS (May 2007 – till date)**
 - *In-situ* Data acquisition, processing, quality control, database management and data services.
- **Assistant Professor (June 2006 – May 2007)** at Nalanda Institute of Engineering & Technology, Sattenapalli.

- **Lecturer (September 2001 – September 2003)** at V.S.R. & N.V.R. College, Tenali.
- **Lecturer (June 2000 – September 2001)** at R.V.V.N. College, Dharanikota.

Awards / Certifications

- **Certificate of Merit** for outstanding contribution in the field of Ocean Science & Technology by Ministry of Earth Sciences, July 2012.
- Qualified in "**National Eligibility Test(NET)**" in the professional subject of Computer Application in Agriculture conducted by Indian Council of Agricultural Research in October 2001.

Professional Memberships

- **Secretary**, Working Group (WG-IV/4: Spatial Data Infrastructure) of Technical Commission-IV (Geo Databases and Location-based Services), International Society for Photogrammetry and Remote Sensing (ISPRS), 2012-2016.

Major Achievements

- **Ocean Data and Information System (ODIS): An integrated geospatial solution for ocean data**

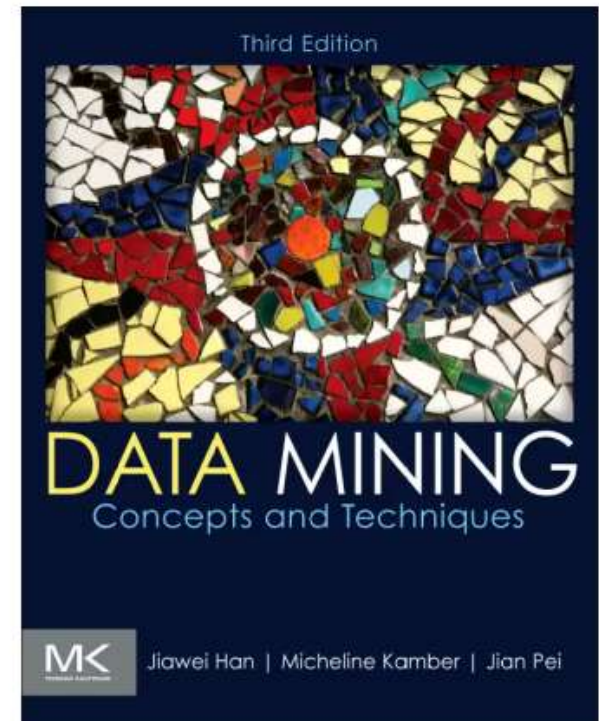
The three-tier open source architecture was adopted in developing the Ocean Data and Information System (ODIS). The ODIS is an end-to-end system comprising of acquisition (satellite and in-situ) means), processing, quality control and data dissemination to the users for operational and research activities. The ODIS architecture is supported by Open Source GIS Server (PostGIS on PostgreSQL Server), OpenLayers as Frontend, MySQL database layers, Apache as the Web server and Joomla! for user interface and Joomla as Content Management System. The implementation of various components of the system was achieved in two phases i.e., (i) development of integrated in-situ database management system and (ii) the Web-GIS interface that extracts with heterogeneous and voluminous data from variety of observing systems and provide online data services including on-the-fly visualization. The ODIS (<http://www.incois.gov.in>), presently hosts data (real-time as well as past data) from Drifting Buoy, Moored Buoys, XBT, Ship mounted Automatic Weather Stations, Wave Rider Buoys, Water Level Sensors, Coastal HF Radars, Tide Gauges, etc. The ODIS is scalable and designed to host incoming data sets from CTD, Radiometers etc. The web-interface enables the user to (i) display the geographic locations of various ocean observing systems (ii) query real-time data on user defined spatial and temporal selections (iii) on-the-fly visualization of time graphs (iv) downloading the selected and (v) provides bilingual web interface to the users.

The entire system provides a unique opportunity to access in-situ data through a single window solution that facilitates users to compare, validate, visualize and analyze the data on-the-fly received from various oceanographic systems.
- **Software development** for acquisition of real-time data through INSAT communication

Hardware Specifications	
Product Name	HP ProBook 450 G1 Notebook PC
Operating systems	Preinstalled: Windows 8 Pro 64 ¹ Windows 7 Professional (available through downgrade rights from Windows 8 Pro) ² Windows 8 64 ³ Windows 7 Professional 64 ³ Windows 7 Home Premium 64 ³ Windows 7 Home Basic 64 ³ SUSE Linux Enterprise Desktop 11 FreeDOS
Processor	Intel®Quad-Core™ i7, Dual-Core™ i5, and i3, Celeron and Pentium
Chipset	Mobile Intel®HM77 Express
Memory	DDR3L SDRAM, 1600 MHz, two slots supporting dual-channel memory, 2 / 4 / 8 GB 5001MM, up to 16 GB total ⁴
Internal Storage	320/500/750 GB 5400 rpm SMART SATA II HDD, 500 GB 7200 rpm SMART SATA II HDD ⁵ 500 GB 5400 rpm SMART SATA II Hybrid HDD ⁶ /SSHD, 1TB 5400 rpm SMART SATA II HDD ⁵ 128 GB 2.5" Solid State Drive ⁷
Removable Storage	Fixed 9.5 mm SATA optical drive: DVD+/-RW SuperMulti DL Drive; Blu-ray ROM DVD+/-RW SuperMulti DL Drive; DVD-ROM and option of no optical drive
Display	15.6" diagonal LED-backlit HD ⁸ anti-glare (1366 x 768) 15.6" diagonal LED-backlit HD ⁸ with Touch Display (1366 x 768) Shatter, scratch, reflection and smudge resistant.
Graphics	Integrated: Intel®HD ⁹ Graphics 4000 (Intel®Quad-Core i7 and Dual-Core i5 and i3); Intel®HD ⁹ Graphics (Celeron and Pentium configurations) Switchable Discrete: AMD Radeon™ HD 8750M, with 1 or 2 GB dedicated DDR3 video memory ¹⁰
Audio/Visual	DTS Sound+, stereo speakers, single integrated digital microphone or dual array microphone on models with optional ¹¹ 720p HD ⁸ webcam
Wireless Support	WLAN: Realtek 802.11 b/g/n 1x1 ^{12,13} Mediatek 802.11 b/g/n 1x1 with Bluetooth®4.0 Combo ^{14,15} Intel®802.11 a/b/g/n/ac 1x1 with Bluetooth®4.0 Combo ^{16,17} Intel®802.11 a/b/g/n 2x2 with Bluetooth®4.0 Combo ^{18,19} Intel®802.11 a/b/g/n 2x2 with Bluetooth®2.1 (on Linux operating system) ²⁰ WWAN: HP h33 110 HSPA+ Mobile Broadband Module ^{21,22} HP h4112 LTE/HSPA+ Gobi 4G Module ^{23,24}
Communications	Realtek 10/100/1000 Ethernet Controller
Ports and Connectors	2 USB 3.0 ports, 2 USB 2.0 ports, 1 HDMI 1.4a, 12.1 VGA, 1 stereo microphone input, 1 headphone/line out, 1 RJ-45 (Ethernet), 1 power connector
Expansion Slots	Multi Media Reader Slot supporting SD, SDHC, SDXC, Memory Stick, MSXC
Input Device	Full-sized, spill-resistant keyboard, touchpad with scroll zone and gestures support, power button, 2 launch buttons (Wireless on/off and speaker mute), Function Keys, full separate numeric keypad
Software (Windows OS only)	Buy Office, HP 3D DriveGuard, ²⁵ HP Connection Manager (Win 7), HP Wireless HotSpot, ²⁶ HP Mobile Connect (EMEA only, requires Windows 7), HP PageIt (Win 8), HP Recovery Manager, HP Support Assistant ²⁷ , HP ePrint ²⁸
Security	HP Client Security Suite includes: HP Password Manager, ²⁹ HP File Sanitizer ³⁰ and HP Device Access Manager with Just in Time Authentication, HP Drive Encryption, ³¹ HP Secure Erase, ³² HP Trust Circles, ³³ Microsoft Security Essentials (Win 7) ³⁴ & Microsoft Defender (Win 8) ³⁵ . HP SmartKey ³⁶ requires initial user setup. Absolute Data Protect ³⁷ , security lock slot
Dimensions (w x d x h)	Non-Touch 14.75 x 10.09 x 0.9 in (at front) 37.5 x 25.62 x 2.28 cm (at front) Touch 14.76 x 10.09 x 0.95 in (at front) 37.5 x 25.62 x 2.42 cm (at front)
Weight	Starting at 5.22 lb (2.37 kg) with optical drive Weight will vary by configuration.
Power	9-cell (93 Whr) lithium-ion battery; 6-cell (47 Whr) lithium-ion battery; Fast Charge ³⁸ 90W Smart AC Adapter; 65W Smart AC Adapter
Expansion Solutions	Optional: ³⁹ HP 300SPr USB 3.0 Port Replicator, HP 200SPr USB 2.0 Port Replicator
Warranty	Limited 3 year, 1 year and 90 day warranty options available, depending on country, 1 year limited warranty on primary battery. Optional ⁴⁰ HP Care Pack Services ⁴¹ are extended service contracts which go beyond your standard warranties. For more details visit: http://www.hp.com/go/pcp



Author(s)	: Jiawei Han, Micheline Kamber, Jian Pei
Title(s)	: Data Mining : Concepts and Techniques 3rd Edition
Publisher	: Elsevier
Edition	: 3rd Edition, 2011
ISBN	: 9789380931913, 9380931913
Language	: English
Pages	: 702



Purpose of Metadata

- ▶ Support discovery of scientific data
- ▶ Facilitate acquisition, comprehension, and use of data by HUMANS
- ▶ Enable automated discovery, ingestion, processing and analysis by MACHINES
- ▶ Organize and maintain the organization's data investment
- ▶ Share data set information through data catalogs
- ▶ Provide access for download or display and analysis through web services

Good metadata means good data management



Purpose of Metadata

▶ For data producers

- ▶ Avoid data duplication
- ▶ Share reliable information
- ▶ Publicize efforts – promote the work of a scientist and his/her contributions to a field of study

▶ For Data users

- ▶ Search, retrieve, and evaluate data set information from both inside and outside an organization
- ▶ Find data: Determine what data exists for a geographic location and/or topic
- ▶ Determine applicability: Decide if a data set meets a particular need
- ▶ Discover how to acquire the dataset you identified

- ▶ Process and use the dataset

Purpose of Metadata

▶ For Organization

- ▶ Metadata helps ensure an organization's investment in data
 - ▶ Documentation of data processing steps, quality control, definitions, data uses, and restrictions
 - ▶ Ability to use data after initial intended purpose
- ▶ Transcends people and time
 - ▶ Offers data permanence
 - ▶ Creates institutional memory
- ▶ Advertises an organization's research
 - ▶ Creates possible new partnerships and collaborations through data sharing



Contents of metadata

- ▶ **WHO**
- ▶ **WHAT**
- ▶ **WHEN**
- ▶ **WHERE**
- ▶ **WHY**
- ▶ **HOW**

~~Yes/No~~

Content, quality, condition, and other characteristics



Contents of metadata

Who	Collected the data? Have to contact for data?	Processed the data? Owns the data?
What	Are the data about? Format data is available? Are the limitations?	Parameters were available? Is the quality of data? Is the use of data?
When	Collected? Processed?	
Where	Data is collected?	Data is available?
Why	Data is collected	
How	Data is collected and processed? Do i access the data?	Quality control has been done?



Contents of metadata

- ▶ Spatial and temporal extents
- ▶ Source
- ▶ Lineage
- ▶ Responsible parties
- ▶ Descriptive attributes
- ▶ Quality
- ▶ Accuracy
- ▶ Maturity
- ▶ Known limitations
- ▶ Logical organization of data



Types of metadata

▶ Descriptive Metadata

- ▶ Describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author and keywords

▶ Structural Metadata

- ▶ Indicates how compound objects are put together, for example, how pages are ordered to form chapters.

▶ Administrative Metadata

- ▶ **Technical metadata** Provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.
- ▶ **Rights management metadata:** intellectual property rights
- ▶ **Preservation metadata:** to archive and preserve a resource

Types of metadata

Descriptive metadata	For finding or understanding a resource
Administrative metadata <ul style="list-style-type: none">- Technical metadata- Preservation metadata- Rights metadata	For decoding and rendering files Long-term management of files Intellectual property rights attached to content
Structural metadata	Relationships of parts of resources to one another

Markup languages integrates metadata and flags for other structural or semantic features within content.



Metadata lifecycle

- ▶ Creation of metadata
- ▶ Maintenance of metadata
- ▶ Updation of metadata
- ▶ Storing metadata
- ▶ Deletion of data
- ▶ Publishing metadata



Creation of metadata

- ▶ **Automatic or semi automatic procedure**
 - ▶ Document properties generated in metadata tools
 - ▶ Creation date
 - ▶ Spatial and temporal informations
 - ▶ File location or URL
 - ▶ **Manually provided information's**
 - ▶ Linking to a subject of vocabularies
 - ▶ Licences to use the data
 - ▶ Where to find more information, such as web links
 - ▶ Quality information
 - ▶ **Formally put into existence for day-to-day use by appropriate stakeholders.**
-

Metadata Maintenance

- ▶ Metadata is in a mode of constant use and is updated and modified, as needed, to meet the needs of use by various stakeholders.
- ▶ Metadata changes along with the properties of data change
- ▶ Metadata can be relatively stable for static datasets
- ▶ For changing datasets like realtime insitu measurements, metadata have to be coupled with data workflow
- ▶ Changing metadata automatically



Updation of metadata

- ▶ Metadata changes along with
 - ▶ Organisation
 - ▶ PI changes
 - ▶ Project transfers to other organisation
 - ▶ Or merge with other organisation
 - ▶ Usage of data
 - ▶ When new applications emerge around data
 - ▶ Inclusion of new sensors
 - ▶ Controlled vocabularies evolve and get linked
 - ▶ Data standards changes
 - ▶ Evolution of new tools and systems



Storing metadata

- ▶ **Embedded along with data**
 - ▶ Metadata along with data makes data exchange easier
- ▶ **Separately from data**
 - ▶ In separate database / separate directories
 - ▶ With link to corresponding data files
- ▶ **Availability of tools**
- ▶ **Databases**
- ▶ **Markup Languages**



Deletion of data

- ▶ Metadata should be live even if data is deleted completely
- ▶ When data is no longer necessary
- ▶ Data is no longer valid
- ▶ Incorrect data
- ▶ Data withdrawn by owner
- ▶ Still metadata contains information
 - ▶ data was deleted
 - ▶ if archived, how to request archival copy



Publishing metadata

- ▶ Process of making metadata available to external users
- ▶ Open publication
 - ▶ Public access to your metadata, even if data is restricted
- ▶ Options
 - ▶ Build your own HTML pages
 - ▶ Use available tools such as GeoNetwork, ESRI GeoPortal Server
 - ▶ Make your metadata available using SPARQL endpoint, which allows external systems to send queries to Resource Description Framework (RDF)



Questions ?

