……..……..……..……..……. ………..…………..……..

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DISSERTATION

Remote Sensing Data Handling to Improve the System Integration of Indonesian National Spatial Data Infrastructure

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften unter der Leitung von

Ao.Univ.Prof.Dipl.-Ing. Dr.techn. Josef Jansa
Institut für Photogrammetrie und Fernerkundung (E122)

Eingerichtet an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Raden Venantius Hari Ginardi
Matr. Nr. 0027853
Dominikanergasse 4/3, 1060 Wien

Wien, am 5. August 2010 ………………..

Kurzfassung

In den letzten Jahren hat die Verwendung von Fernerkundungsdaten in Indonesien laufend zugenommen. Bakosurtanal, als nationale indonesische Organisation für die Koordination von Vermessung und Kartierung hat begonnen, die digitale Kartierung hauptsächlich auf Fernerkundungsdaten zu stützen. Im Jahr 2005 konnte Bakosurtanal für das gesamte Land die Karten in den Maßstäben 1:1.000.000 und 1:250.000 fertigstellen und im Augenblick liegt der Schwerpunkt bei den Karten 1:25.000 für Java, Bali und die Nusa Tenggara Inseln. Bakosurtanal und die anderen teilnehmenden Institutionen treffen einander regelmäßig, um ein indonesisches nationales räumliches Informationssystem zu entwickeln, welches durch die indonesische Regierung mit dem Präsidialerlass 85/2007, der sich mit dem nationalen Netzwerk räumlicher Daten befasst, gestützt wird. Zur Implementierung dieses nationalen räumlichen Informationssystems gehören neben anderem die Entwicklung der Infrastruktur der indonesischen nationalen räumlichen Daten und die Entwicklung eines entsprechenden Vertriebssystems.

Neben dem Nutzen, welchen Fernerkundungsdaten bringen, gibt es auch einige Probleme, die mit der Datenhandhabung der rasch anwachsenden Datensammlungen verbunden sind. Forscher haben verschiedene Methoden entwickelt, Rasterdaten nach ihrem Inhalt abzufragen und durch Data Mining Information über sie herauszufinden. Es gibt auch Information, welche nicht aus den Daten herausgezogen werden kann, zum Beispiel externe Information über einen Datensatz, welche in Metadaten abgespeichert ist. Der Metadaten-Standard, wie er durch United States Federal Geographic Data Committee (FGDC) erstellt und in ISO-19115 übernommen und ergänzt wurde, dient gegenwärtig als Standard für Fernerkundungsdaten und findet in vielen Katalog-Systemen für räumliche Daten Anwendung. Der Metadaten-Standard nutzt XML als Plattform.

Durch die Verwendung von Metadaten als Referenz für räumliche Abfragen werden Fernerkundungsdaten und andere räumliche Datensätze mit ihrer semantischen Information verbunden. In den gegenwärtigen Katalog-Systemen, wie etwa jenen von Satellitendaten-Anbietern und in den Daten-Vertriebssystemen wird jedes Fernerkundungsbild als unabhängige Einheit gehalten. Es gibt sehr eingeschränkte Möglichkeiten zu erfahren, wie ein Bild zu einem anderen verbunden ist, auch wenn ein Bild von einem anderen abgeleitet wurde. Für viele Zwecke ist vorteilhaft, wenn eine Verbindung zwischen Fernerkundungsbildern oder anderen räumlichen Daten erhalten oder rekonstruiert werden kann. Diese Forschungsarbeit soll erkunden, wie ein Bild zu dessen Information verknüpft ist und wie es zu anderen Bildern verknüpft werden kann.

Indem die Verbindungen zwischen den Fernerkundungsbildern untersucht werden, kann eine Abfrage in einer Sammlung von Fernerkundungsdaten erweitert werden, zum Beispiel, um die Antwort auf die Frage zu finden: „Welche Bilder wurden verwendet, um einen bestimmten Datensatz zu erzeugen?“, oder „Welche Bilder wurden unter Verwendung eines konkreten Datensatzes generiert?“, oder „Gibt es eine Beziehung zwischen Bild A und Bild B in Bezug auf die durchlaufenen Prozessierungsschritte?“ Indem Verbindungen zwischen den räumlichen Datensätzen innerhalb einer Datensammlung aufgebaut werden, welche die Prozessierungskette zur Grundlage nehmen, entsteht eine weitere Möglichkeit die Organisation räumlicher Daten zu unterstützen.

Der erste Teil der Dissertation untersucht die indonesische nationale Infrastruktur für räumliche Daten und dient als Ausgangspunkt für weitere Forschung. Das Konzept der Infrastruktur wird erklärt, einschließlich der verwendeten Standards und der Protokolle. Auch wird gezeigt, was aus über ein Vertriebssystem erhalten werden kann. Dieser Teil beschreibt, wie Fernerkundungsdaten heute in Indonesien verwaltet werden. Die Einteilung der Fernerkundungsdaten in Kategorien nach ihrem Processing Level, wie sie jetzt schon existiert, wird in der Dissertation berücksichtigt, und daher wird die Level-Kodierung hier weiter verwendet (Level 0 für Originalbild, Level 1 für geometrisch rektifiziertes Bild oder Geländemodell, Level 2 für klassifiziertes und weiterverarbeitetes Bild, Level 3 für Bilder, die zur Präsentation aufbereitet wurden).

Der zweite Teil der Arbeit ist einem Vorschlag gewidmet, der beschreibt, wie das Ergebnis der Handhabung räumlicher Daten aussehen könnte. Die derzeitig existierende Methode und Implementierung des Fernerkundungskatalogsystems und sein Metadaten-Standard wird weiter verwendet als Ausgangssituation. Außerdem wird angenommen, dass jeder Datensatz (d.h. jedes Fernerkundungsdatenfile) eine eindeutige ID besitzt und dass die Prozessierungsschritte in den Metadaten aufgezeichnet werden können, einschließlich einer Liste jener Datensätze, die als Ursprungsdaten Verwendung fanden.

Innerhalb einer Sammlung von Datensätzen werden die Prozessierungsschritte dann aus den Metadaten extrahiert und der Ablauf rekonstruiert in Form einer gerichteten azyklischen Graphen-Struktur (DAG). Durch die Verwendung einer Nachbarschaftsliste, um alle Nachbarschaften in einem Graphen zu organisieren, kann der komplette Verlauf der Prozessierung für jeden Datensatz nachvollzogen werden, und zwar in zwei Richtungen: in Richtung zum Originaldatensatz und in Richtung zum (erzeugten) Zieldatensatz. Mithilfe dieser Nachbarschaftsliste wird ein Verfahren eingeführt, durch welches eine Beziehungsgröße (relatedness) innerhalb eines Datensatzpaares ermittelt werden kann. Da dieses Konzept der räumlichen Metadaten sowohl für Vektor- als auch für Rasterdaten verwendet werden kann, ist die vorgeschlagene Datenverwaltung auch für Datenvarianten nutzbar.

Die Anwendbarkeit der vorgeschlagenen Methode für und ihre Kompatibilität mit der momentanen Plattform wird berücksichtigt. Sie wurde im selben Standard und mit demselben Protokoll implementiert, wie durch das vorhandene System vorgegeben ist. Diese Vorgehensweise würde auch die Implementierung in vielen anderen Ländern erlauben, falls dieselbe Infrastruktur verwendet wird. Um die Zweckerfüllung zu beweisen, wurde ein Prototyp entwickelt, der auf einer offenen Plattform aufbaut mit PostgreSQL, Apache Webserver, Mapserver WebGIS und PHP Programmierumgebung.

Das Ergebnis der Forschungsarbeiten führt zu einer Verbesserung der Handhabung räumlicher Daten, wobei eine Nachbarschaftsliste geführt wird, um die Verbindung zu existierenden räumlichen Daten aufzubauen. Die Verbesserung steigert das Potenzial der Abfragen in räumlichen Daten innerhalb eines Katalog-Systems und ermöglicht das Gruppieren oder auch die Klassifizierung der Datensätze, abhängig von der Nutzung der Datensätze. In einem System, welches Entscheidungen unterstützen soll, kann die Aussage über die Nutzung jedes Datensatzes zur Qualitätsbeurteilung herangezogen werden, oder aber sie kann als ein Parameter dienen, mit welchem die Produktivität eines Datensatzes innerhalb einer Datensammlung bewertet werden kann.

Abstract

In recent years, the usage of remotely sensed data in Indonesia has continually increased. Bakosurtanal, as the Indonesian national institution for survey and mapping coordination has started the digital mapping process mainly based on remote sensing data. In 2005, Bakosurtanal has finished the 1:1,000,000 and 1:250,000 scale maps of the whole country and the current production focuses on 1:25,000 maps for Java, Bali, and Nusa Tenggara islands. Bakosurtanal and participated institutions meet regularly in developing the Indonesian National Spatial Information System, which is endorsed by the Government of Indonesia with Presidential Decree number 85/2007 about National Spatial Data Network. Implementations of this national spatial information system among others are the development of Indonesian National Spatial Data Infrastructure, and the development of National Spatial Data Clearinghouse.

Besides the benefit that comes from remote sensing data, there are also several problems with the handling of remote sensing data collections which grow rapidly. Researchers have developed various methods to query raster data based on its contents and data mining approaches to disclose information from raster data. There exists also information which can not be extracted directly from raster data, for example the exterior information about the dataset, which is stored in image metadata. The metadata standard as introduced by The United States Federal Geographic Data Committee (FGDC) and adopted and augmented by ISO-19115 currently serves as standard for remote sensing metadata, and it is broadly used in many spatial data catalogue systems. This metadata standard is based on the XML scheme.

With the usage of metadata as a reference for spatial data query, remote sensing images and other spatial datasets have been linked to their related semantic information. In the current catalogue systems, like those or satellite data provides, or clearinghouses, each remote sensing image is maintained as an independent entity. There is a very limited possibility to know the linkage of one image to another, even if one image has actually been derived from the other. It is an advantage for many purposes if the linkage among remote sensing image or other spatial data can be maintained or at least reconstructed. This research will explore how an image is linked to its related information, and how an image can be linked to another images.

By exploring links among remote sensing images, a query of remote sensing data collection can be extended, for example, to find the answer of the query: “which images are used to create certain dataset?”, or “which images have been created from a concrete

dataset?”, or “is there a relationship between image A and image B based on their processing steps?”. By building links among spatial datasets in a collection based on their creation process, a further possibility of spatial data organization can be supported.

The first part of this dissertation will explore the Indonesian National Spatial Data Infrastructure as our research starting point. The concept of this infrastructure will be explained, including which standards and protocols are used, and what output can be obtained from the clearinghouse. This part will also discuss how remote sensing data are currently managed in Indonesia. The categorization of remote sensing data based on its processing level, as it already exists, is considered in this dissertation, and its levelling code (level-0 for raw image, level-1 for geometrically rectified image or Digital Terrain Model, level-2 for classified or further processed image, and level-3 for presentation quality image) is used in our research.

The second part of this research is our proposed achievement in spatial data handling. The current method and implementation of remote sensing catalogue system and its metadata standard is used as our reference system, and we assume that each dataset (remote sensing data file) has a unique dataset ID, and the processing steps is recorded in its metadata including the list of datasets which are used as its source.

Within a collection of datasets, processing steps are extracted from dataset metadata, and reconstructed using a directed acyclic graph (DAG) structure. By using an adjacency list to organize all adjacencies in the graph, a complete processing step for each dataset can be tracked in two directions: source datasets and destination (created) datasets. With the help of this adjacency list, an approach to evaluate the relatedness of a pair of dataset is introduced. Since the concept of spatial metadata can be applied to both vector data and raster data, the proposed data handling can be used for both sorts of data.

The applicability and compatibility of the proposed method with the current platform is also considered. The proposed method can be implemented using the same standard and protocol and using the same metadata file as used by the existing system. This approach makes it also possible to be implemented in many countries which use the same infrastructure. To prove this purpose, we develop a prototype based on open source platform, including PostgreSQL, Apache Webserver, Mapserver WebGIS, and PHP programming environment.

The output of this research leads to an improvement of spatial data handling, where an adjacency list is used to maintain spatial dataset history link. This improvement can enhance the query of spatial data in a catalogue system, and enables dataset grouping or classification based on the usage of the dataset. For a decision support system, the list of the usage of each dataset can be applied for quality assessment or may serve as one parameter to evaluate the productivity of each dataset in a collection.

Acknowledgement

I would like to thank Prof. Josef Jansa, my supervisor, for his continuous help, encouragement, and support. Prof. Jansa always find time in his busy working schedule for frequent discussions with me, and his insight and careful make our discussion interesting and fruitful.

This dissertation has been supported by the Neumaier Award, provided by the Institute of Photogrammetry and Remote Sensing, Vienna University of Technology. I am very grateful to this scholarship, and I always remember the late Prof. Karl Kraus for his support, help, and care.

My thanks also go to ITS Surabaya which has given me supports and many people in the Institute of Photogrammetry and Remote Sensing for their supporting ideas, administrative supports, and technical supports.

The completion of this work has never been out of support from my beloved family and friends. I want to thank all of them. My sincerest thanks go to my mother Anneke Soebardi, and my wife Dian Sulistiani. Their support, understanding, care, and love are really of my indebtness.

Contents

List of Figures

List of Tables

1 Introduction

A huge earthquake and tsunami hit Northwest part of Sumatra Island in December 2004, destroying Banda Aceh city, Meulaboh, and much coastal area in Aceh Province and also Nias Island in North Sumatra Province. To rebuild the damaged area and recover the economy, a coordination agency for rehabilitation and recovery (named *Badan Rehabilitasi dan Rekonstruksi* or *BRR*) of Aceh and Nias area has been established[1]. This agency uses a multidiscipline documentation including remote sensing images and digital maps to support its rehabilitation and recovery planning.

The remote sensing image collection for Aceh and Nias area is well organized, consists of images before tsunami and images from later acquisitions (after tsunami) using airphotos or satellites. This collection is then linked into the Indonesian National Spatial Data Infrastructure, operated by Indonesian National Agency for Survey and Mapping (BAKOSURTANAL[2]) in a digital catalogue system which is also known as Clearinghouse[3]. By implementing clearinghouse, Bakosurtanal can share the information of its spatial data collection to broader users.

A query system in Indonesian spatial data clearinghouse supports search keys and several popular filters such as spatial data types or sensor types, which makes it user friendly in terms of user interface and query filter facility. This type of query facilities is commonly used in many spatial data clearinghouse. The implementation of metadata as a source for data search enhances the speed of data search, and optimizes the storage capacity of the catalogue server.

The main purpose of the above mentioned clearinghouse is to find whether the needed spatial data exists in the collection, and how to access the data. In practice, it is also of interest to know the creation history of the dataset in the collection. Users may also want to know which datasets used a respective dataset for their creation and which datasets have been used to create the respective dataset, i.e.: “which datasets use the current dataset?” or “which datasets have been used to create this dataset?”

---

[1] The recovery programme of BRR: http://know.brr.go.id/Modules/Home-Accordion/recovery-programme.html, last accessed Jan 2010
[2] www.bakosurtanal.go.id

1.1 Motivation

This research started with the above mentioned questions, and the result is expected to be used inside or at least compatible with the existing Indonesian spatial data clearinghouse. Since this clearinghouse is compatible with many other clearinghouses, the output can also be used by clearinghouse in common, not necessarily limited to Indonesia. There are many types of spatial data stored and distributed in this clearinghouse, but we focus on the handling of remote sensing data and all raster data which are derived from remote sensing data.

By the term *remote sensing data* we understand in the following images or raster data sets,

- taken by remote sensing acquisition techniques (i.e. remote sensing images), such as aerial photographs, various satellite imagery comprising multispectral images, thermal images, Radar images, etc; and
- images derived from them, such as mosaicking products, thematic images, etc.

Remote sensing takes an important role in digital mapping and Geographic Information System. It enables wide possibilities for data acquisition from the real world using many types of passive or active sensors. To store and distribute this type of geospatial data, Bakosurtanal adopts the Federal Geographic Data Committee (FGDC)-Standard metadata for its clearinghouse and develop an inter-operability of geospatial data in Indonesian National Spatial Data Infrastructure named IDSN (Puntodewo and Nataprawira, 2004).

In recent years, the usage of remotely sensed data in Indonesia has continually increased. Multispectral images from the satellites such as Landsat-TM, SPOT, IKONOS and ALOS are widely used for mapping purposes and geo-related analysis. Bakosurtanal has started the digital mapping process mainly based on remote sensing data since 1993 (Ikawati and Setiawati, 2009; pp. 69-86). In 2005, Bakosurtanal has finished the 1:1,000,000 and 1:250,000 scale maps of the whole country and currently Bakosurtanal is producing the 1:25,000 maps for Java, Bali, and Nusa Tenggara islands. Each department in Bakosurtanal has a specific field of interest and plays a role also as a center of survey and mapping and producing thematic maps in that field, i.e. Centre for

---

[3] This clearinghouse can be accessed online through Bakosurtanal web page (http://www.bakosurtanal.go.id) or www.idsn.or.id

Natural Resources Survey, Centre for Aeronautic Survey, Centre for Topography and Base Mapping[4].

Besides the benefit that comes from the remote sensing data, there are also several difficulties or limitations in the handling of their collection. The integration of vector data with attribute database is better studied than the integration of raster data or images from remote sensing. To increase the productivity of raster data, many researches have achieved the automation of image processing and analysis with more sophisticated features. In the literature, several groups have developed concepts of managing raster or image datasets. Researchers in the fields of image database and image information mining have developed models for Information extraction and retrieval from image content in large archives (Vinhas et al., 2003). Researchers of spatial and geographic information system have developed concepts of spatial query and efficient spatial archive (Vatsavai et al., 2001). Researchers of image processing have developed many types of feature extraction and pattern analysis (Schroeder et al., 1997; Harvey et al., 2002). Researchers of knowledge-based and artificial Iintelligence have developed concepts of mining knowledge in Geographical Data (Guo et al., 2003).

Those developments mostly apply to the homogeneous image collections, for example the collections of NOAA datasets, Landsat-TM datasets, SAR datasets or IKONOS datasets. In practice, there are many new images or datasets that are derived from other datasets. Those derived datasets usually have a higher level of information compared to the source datasets. The current condition in Bakosurtanal and also in other centers for remote sensing data shows that the collection of remote sensing data grows and produces several level of processed dataset in form of images for many purposes of analysis. Since the collection of images becomes multi-level, it requires a better data handling and management compared to homogeneous collection. The problem that is also important to be solved is ***'how to link the images with their related information'*** and the next question is ***'how to link the images which are related to each other'***

This research focuses mainly on the above subject. Considering the possibility to extract processing steps from image metadata, the history of creation of each remote sensing data in a collection can be constructed. This research proposes a method to organize spatial data which extend the possibilities of browsing remote sensing data. It will provide links among datasets within strict FGDC standard which is used as a standard for geospatial metadata in Indonesian national spatial data infrastructure. This FGDC metadata standard is compatible with ISO 19115 standard.

---

[4] Organization scheme of Bakosurtanal: http://www.bakosurtanal.go.id/bakosurtanal/organization/

The development of internet and web technology and open source spatial database open the opportunity to develop a portable web-based application which enable the sophisticated database system to be accessed by public using a thin client system (Bojinski et al., 2002). Other part of this dissertation discusses the programming environment to implement the proposed remote sensing data handling. This programming environment includes the employment of web-server technology and spatial database as used in many spatial catalogue systems.

1.2 Problem Statement and Description of Approach

Considerations outlined above are the premises for our research. Amhar (2001) has introduced a categorization of remote sensing data based on processing status. The data handling that we develop in this thesis should cover raster datasets from any sources / sensors (air photo to meteorological satellite), it covers raster data in form of raw (assigned as level 0), pre-processed (assigned as level 1), analyzed / thematic (assigned as level 2), or presentation data (assigned as level 3). It should also maintain the link between each dataset and its information, and also maintains the link among datasets which have a relationship to each other. With this enhancement, spatial database and information system can be enhanced to support better features.

With the above proposed enhancement, the information which is extracted from datasets or linked with datasets should support the following user queries:

- What is provided by the dataset? This includes extraction of attributes of dataset
- Where is it stored? This query tracks the data provider information
- What feature can be extracted? This query extends the type of output based on commonly known algorithms. For example, NDVI can be extracted if both the near-infrared-band layer and the red-colour-band layer exist in remote sensing data
- Which datasets are related with this dataset? This information is derived from the processing history.

The current working scheme of spatial data organization in Bakosurtanal as shown in Figure 2.2 is used as our starting point. This scheme shows spatial data sharing within many institutions in Indonesia using a clearinghouse. This clearinghouse has been developed using GeoNetwork Opensource catalogue system which enables the interoperability with many spatial data provider in Indonesia and the regional or global internet community. The functionality of this clearinghouse can be extended using the proposed data handling which maintain the history of creation process and enhance the possibility to maintain the datasets relationship.

The clearinghouse as introduced above uses metadata as its primary source to organize spatial data collection. Bakosurtanal implements the FGDC standard metadata and uses several publicly available tools to create or edit metadata in XML format. The content of FGDC and ISO standard for spatial data metadata is taken into consideration, and the proposed data handling should conform these standards.

A feasibility to implement the proposed method of this research is also taken into our consideration. The prototype will be developed using several levels of tools or applications. The core level of this prototype is a Database Management System (DBMS). The second level of this prototype system is a spatial engine which enhances the DBMS ability to support spatial data handling. The third level of this prototype is a programming environment and user interface. All of these tools are available as open sources.

The study area for this thesis is based on the availability of spatial data in Indonesian spatial catalogue system, and the proposed method is also considered to be applicable in the Indonesian regional institution. There are limitations to this design that are: (i) with more than 17.500 islands, there are too many possibilities for object selection (e.g: island selection) based on shape or topology matching or other query based on feature extraction. Spatial query is based on the geo-referenced / coordinate system. (ii) With the existing national digital map collection, only datasets which have metadata in XML format and FGDC standard metadata are used. (iii) High cost for software license or maintenance to be avoided.

1.3 Hypothesis

Currently, most of spatial data, which are stored in the spatial database of the Indonesian National Spatial Data Infrastructure, are vector data. Therefore, it would make sense to build up an organisation of the vector data sets. Since many vector data have been derived from raster data, both data types concurrently exist in a collection. As a conclusion, the organisation of spatial data must not be limited neither to pure raster nor to pure vector data. In the following the concentration will be put on the organisation of raster data, keeping in mind that eventually both types of data shall be supported.

Extracting stored information or attributes data from raster datasets and maintaining their values within a database will improve the integration of remote sensing data collection into spatial database system or catalogue system. By maintaining the history link among raster data based on their processing steps, we expect an improvement of spatial data infrastructure, for example a wider query possibility in spatial data catalogue systems.

A specific data structure will be required to handle the history link among datasets. This data structure may also provide a possibility to evaluate the relationship or relatedness between two datasets. With this evaluation, a further raster data organization can be considered, for example a partitioning of raster data collection into smaller groups.

The selected data structure, proposed algorithms and methods, which are discussed in this research, are not related with a certain software product. The implementation of this research can be done in many platforms including the open source environment. The development of open source GIS, including the free web mapping tools, opens the possibility of implementing the proposed method without employing high cost hardware and software.

1.4 Related Work

The idea of enhancing the spatial data handling in this research covers many scopes of research fields. Since raster data from remote sensing is used as our study domain, the state of the art of remote sensing data organization will be explored. The development of geospatial metadata is also explored. The currently available standard and the implementation of metadata in spatial clearinghouse are documented. The progress of the Indonesian national geospatial data infrastructure is also documented. The following part is dedicated to the state of the art of current research in these fields.

In the literature many proposals have been presented to organize and maintain remote sensing data. Topographic Mars Information System (Dorninger, 2003), for example, is a real-time distributed remote sensing catalogue system with the 3D visualization. An automatic extraction of linear spatial feature from remote sensing data has been introduced (Doucette et al., 1999), and the proposal of content based image classification for satellite image database has been introduced (Holowczak et al., 2002). Hsu (Hsu et al., 2002) discusses the trend in image mining. The image mining framework as discussed here is not limited to remote sensing data. However, most of the literatures give the concepts or models for organizing remote sensing data which are homogeneous in term of sensor or level of processing. There are several examples of spatial catalogue systems from the remote sensing data providers like GeoEye or NOAA which offer certain processing level but they do not provide relationship of images which shows how the higher level datasets are created from other datasets. In the best of our knowledge, no comprehensive approach has been proposed for heterogeneous datasets (various sensors, various levels of processing) which is suitable for a heterogeneous datasets like the Indonesian spatial data clearinghouse.

The Federal Geographic Data Committee Standard Metadata (FGDC, 2000) and ISO-19115 are designed to be used in spatial data metadata based on XML schema. Di (2003) from the Laboratory for Advanced Information Technology and Standards (LAITS), Georg Mason University, describes the development of remote sensing related standards by the FGDC[5], the Open GIS Consortium (OGC)[6], and the International Organization for Standardization (ISO) Technical Committee 211 (TC 211)[7]. Di introduces the individual remote sensing standards developed or being developed by those organizations, and discusses the relationship among those standards and how to use the standards in the development of interoperable, standard-compliant geospatial data and information systems for remote sensing. Di also develops a prototype of geospatial web service system for remote sensing data which implements several interoperability possibilities. The extension version of standard metadata for remote sensing data is currently available by ISO and FGDC. This extension adds more detailed information within XML schema which support more features for remote sensing data organization while keeping the current standard XML schema.

Organizing the remote sensing metadata files, which is stored in XML schema, requires a spatially enabled database with XML support. With the integration of a spatially enabled database system and World Wide Web, the spatial database can be accessed globally through the internet. Garnett and Owens from Refractions Research Inc describe the requirements for Web Feature Service (WFS) as a part of Open GIS Consortium (Garnett, 2003). Along with the XML requirements, ArcSDE, Postgis and Oracle Spatial are analyzed to evaluate their suitability and capability in relation to Open GIS WFS requirements. The guideline of the implementation of current spatial metadata standard for imagery has also been introduced (Schlessinger, 2005).

Bakosurtanal and participated institutions meet regularly in developing the Indonesian National Spatial Information System (SISN). The progress shows that SISN, which started in 2004 implementing MS Access for maintaining database of poverty, grows to a wider application which employs SQL server as an engine, a web mapping service, and FGDC metadata standard (Matindas et al., 2004). Indonesian Government endorses this

[5] http://www.fgdc.gov
[6] http://www.opengeospatial.org/
[7] http://www.isotc211.org/

SISN by Presidential Decree number 85/2007 about National Spatial Data Network. With this strong infrastructure, the development and the usage of spatial data are increased. This progress requires a better data handling for the increasing number of spatial datasets, mainly in the clearinghouse.

A guideline of spatial data metadata development, which is used by Indonesian geospatial data clearinghouse, has been published by Clearinghouse Technical Working Group of Bakosurtanal in 2005. This guideline defines FGDC standard metadata as a standard for Indonesian National Spatial Data Infrastructure. The Indonesian Geospatial Data Clearinghouse currently provides spatial data in many forms, includes: aerial photo, images (remote sensing dataset), atlas, map, DTM, toponymy, and control points.

1.5 Contribution

As previously pointed out, this research is carried out to bring an improvement of remote sensing data handling and other type of spatial data. We take into consideration the current situation and development in Indonesia and propose a new approach to manage raster datasets and to support users with further options of spatial data query. Later on, we also aim at opening the way to the development of a decision support system by evaluating the usage of every raster dataset in a creation process within a collection.

By organizing the raster data collection in a database and maintaining their processing steps or creation history, we can keep track the temporal information of certain area, trace back the previous processing applied to datasets, and enhance the analysis of spatial data by listing all possible raster datasets which support the requested area of interest. Evaluating the relationship or relatedness between two datasets based on their processing steps may be used to group raster datasets or used as an additional parameter for dataset query.

The data handling as proposed by this thesis can also be used as a comparison system for a spatial catalogue system with different approach of dataset relationship or different characteristic of geographic area. If the GIS capabilities are not required, this proposed design can also be implemented as a lighter remote sensing catalogue system which supports an enhanced user query.

1.6 Overview

The following Chapter 2 introduces the currently established Indonesian national spatial data infrastructure. This study area is used to define the consideration of data model. This dissertation is proposed to be tested, evaluated, and possibly implemented in this study area. Chapter 3 describes the state of the art concerning metadata extension in the field of remote sensing. It shows the current characteristics of remote sensing data exchange and their standards.

Chapter 4 analyses the current progress in spatial databases and information systems which is related to the integration of raster data (and its metadata) into spatially-enabled DBMS. This chapter also analyses the possible approach of the current web-mapping, especially the open source web mapping solution to extend the usability of the proposed data handling method.

Chapter 5 introduces a new approach for raster data handling with the Indonesian National Spatial Data Infrastructure as the study area. An appropriate data model to represent the relatedness among datasets is evaluated. The characteristic of dataset relatedness is also evaluated to find the suitable data structure representation. This chapter also shows how the proposed method can use the current standard and it is feasible to be implemented.

Chapter 6 describes in more detail the prototype of implementation, validation, and testing of the proposed method. This includes the input-output specification, system configuration, sample data preparation, and a suggested user-interface to explore the remote sensing data collection. The existing catalogue system used by Bakosurtanal serves as a comparison system or benchmarking system. This chapter also discusses the relationship or relatedness between two datasets as a further analysis of dataset history, and a method to measure the relatedness is introduced. Chapter 7 is dedicated to the conclusions to the topic and future research.

2 Indonesian National Spatial Data Infrastructure

This chapter concentrates on current developments of Indonesian spatial data management. It describes the infrastructure of national sharing of spatial data in Indonesia. Among many institutions which deal with spatial data, Indonesian Coordinating Agency for Survey and Mapping called Bakosurtanal is the centre where most of the development processes take place. This chapter discusses also the concept of clearinghouse which is developed by Bakosurtanal with other institutions to provide a platform for spatial data sharing among spatial data stakeholder and user in Indonesia. The state of the art of remote sensing data handling by Indonesian National Spatial Data Infrastructure is also explained in this chapter.

2.1 Geography of Indonesia

Indonesia is an archipelagic country in Southeast Asia, located between 95° E to 141° E and 6° N to 11° S. It spreads more than 5100 km from east to west and about 1800 km from north to south. There are more than 17500 islands in Indonesia, about 6000 of which are inhabited. The total land area of Indonesia is 1.919.440 square km. The capitol of Indonesia is Jakarta which is located in Java Island. Java is the most populous and dense region in Indonesia.

The administrative of the Republic of Indonesia is divided into 33 provinces as a first level of regional autonomy, and 498 cities (in Indonesian: *kota*) or regencies (in Indonesian: *kabupaten*) as a second level of regional autonomy.

The following Figure 2.1 shows the map of Indonesia. The projection which is commonly used in Indonesian map is the Universal Transverse Mercator (UTM) with WGS84 as a commonly used Datum. The projected map is required for spatial analysis purpose; therefore many digital maps in Indonesia are stored in UTM projection which spread into several UTM zones.

Fig 2.1: Map of Indonesia

2.2 The Development of Indonesian National Spatial Data Infrastructure

The Government of Indonesia has initialized the national geographic information system (National-GIS) in 1997 (Matindas et.al. 2004) which is called SIGNAS. SIGNAS builds a national platform of implementation of GIS in Indonesia. It establishes the role and responsibility of Indonesian government, national survey and mapping institutions, regional institutions, and private institutions, focuses on the following main issues:

- National Laws which are related to surveying and mapping
- Coordination among spatial data stakeholder
- Problem identification

Following the development of SIGNAS, a national coordination meeting for surveying and mapping was held in 2000 where a concept of Indonesian National Spatial Data Infrastructure (IDSN, *"Infrastruktur Data Spasial Nasional"*) has been worked out. The main target of the development of IDSN includes:

- Integration of spatial data providers
- Standardization of spatial data
- Accessibility of spatial data by users and spatial data providers

As a national coordinating agency for surveying and mapping in Indonesia, Bakosurtanal organizes the development of IDSN and its yearly coordination meeting, which has started in 2001. Both the documentation of the meeting and information about the progress of the development are provided by an internet portal of IDSN[8].

Figure 2.2 shows the working scheme of Indonesian spatial data clearinghouse which is provided by IDSN. This distributed database server system deals with geospatial data which are produced and used in Indonesia. It uses TCP/IP protocol through internet and computer network to connect users to multiple nodes of data providers and provides many useful tools to search, find, evaluate and obtain any kind of geospatial data (Puntodewo and Nataprawira, 2004). With the use of internet protocol, all metadata servers in Indonesia can be connected.

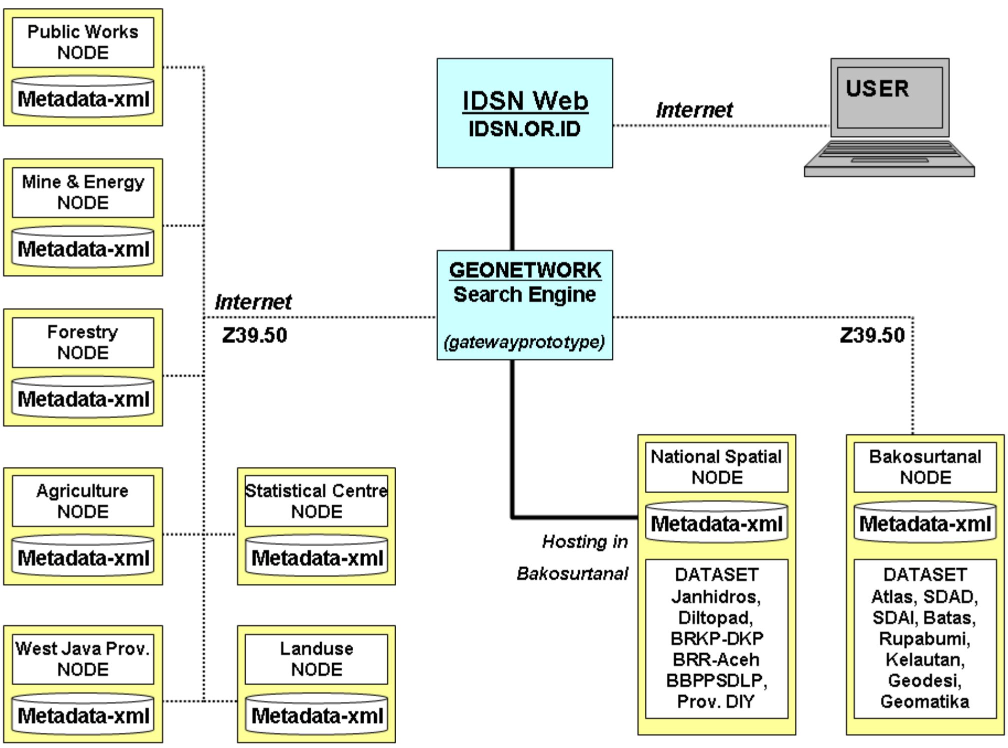


Fig 2.2: Clearinghouse concept from Bakosurtanal[9]

[8] http://www.idsn.or.id, last accessed 1 April 2009
[9] Original diagram by Bakosurtanal, translated into English

The clearinghouse by IDSN uses Geo-profile with Z39.50 (ANSI, 2008) protocol to maintain geospatial metadata database and to support queries from users.

The national coordination meeting in 2000 also directs the development of IDSN into five main aspects or activities which are known as components or pillars of IDSN as follows:

- Institutional aspect
- Legal aspect
- Geospatial data
- Research and development
- Human resources development

This meeting also emphasizes the importance of IDSN to manage Indonesian natural resources as one of the key factors for the planning of sustainable development in Indonesia. A wider coverage of spatial data and the accuracy of information from IDSN can support the decision making in Indonesian development planning. The next section discusses the explanation of the above five-pillars of IDSN.

2.3 Five Pillars of IDSN

The development of IDSN opens the possibility of establishing a new communication platform which easily and rapidly covers nationwide public users. The distributed spatial data server or clearinghouse which is provided by IDSN gives many benefits to the users, data providers, and the government of Indonesia. With the implementation IDSN, each spatial data can be shared nation wide and this reduces the cost of data production. The national coordination meeting for surveying and mapping in 2000 described the development policy of IDSN as “Five Pillars of IDSN”. The detailed of those five pillars of IDSN will be described in the following sections.

2.3.1 Institutional Aspect

The Indonesian National Spatial Data Infrastructure (IDSN), led by Bakosurtanal, is an integration of spatial data providers in Indonesia which helps to distribute existing spatial data produced by many institutions including private companies or non-government organizations (Amhar, 2001; Matindas et al., 2004; Sukmayadi et al., 2004). This IDSN established since 2001 as the successor of National GIS coordination called SIGNas (*Sistem Informasi Geografi Nasional* / National Geographic Information System).

As illustrated in Fig 2.2, IDSN covers many institutions as its nodes. The development of current active nodes of IDSN can be found on the IDSN website. There were 14 active nodes available in February 2009, consisting related national institutions and regional governments. The following Table 2.1 shows the participant nodes in IDSN as of February 2009.

Table 2.1: IDSN nodes status, February 2009[10]

Network Node	**Node Status**			
	Metadata	**Spatial Data**	**Intranet**	**Internet**
BAKOSURTANAL	Ready	Ready	Ready	Ready
BPN (Land Authority)	Ready	On Progress	Ready	Ready
BPS (Statistic Bureau)	Ready	Ready	Ready	Ready
BMKG (Meteorology)	Ready	Ready	Ready	Ready
Dephut (Forestry)	Ready	Ready	Ready	Ready
Deptan (Agriculture)	Ready	Ready	Ready	Ready
Dep ESDM (Energy)	Ready	<unknown>	Ready	Ready
Dephub (Transportation)	Ready	Ready	Ready	Ready
Depdagri (Internal Affairs)	On Progress	On Progress	Ready	Ready
Depkominfo (Information)	Ready	<unknown>	Ready	Ready
Depbudpar (Culture)	Ready	Ready	Ready	Ready
Dep PU (Public Works)	Ready	Ready	Ready	Ready
DKP (Fishery)	Ready	Ready	Ready	Ready
LAPAN (Aeronautical)	Ready	Ready	Ready	Ready
Prop Jabar (Regional)	Ready	Ready	Ready	Ready

As the initiator and main node of IDSN, Bakosurtanal improves the IDSN by supporting the regional nodes of IDSN throughout regional/local governments. From the target of 483 nodes[11], Bakosurtanal provided grants to establish 35 nodes in 2006, and continues supporting this project in cooperation with Intergraph to cover more regional institutions.

Spatial data clearinghouse of IDSN provides interoperability within IDSN nodes and regional or international nodes. To integrate all IDSN nodes with regional and international nodes, internationally recognized standards are used. Data retrieval protocol

[10] From http://www.idsn.or.id, accessed 1 April 2009.
[11] Consist of national institution nodes, 1st level regional autonomy nodes, and 2nd level regional autonomy nodes

Z39.50 is implemented in this spatial data clearinghouse. This protocol is also used in Asia Pacific Spatial Data Infrastructure (Puntodewo and Nataprawira, 2004). With this development, IDSN is not limited to cover the Indonesian spatial data, but also takes a part in building a wider spatial data infrastructure. The following Figure 2.3 (Bakosurtanal, 2005) shows the interoperability of IDSN with national and international users to build a global spatial data infrastructure.

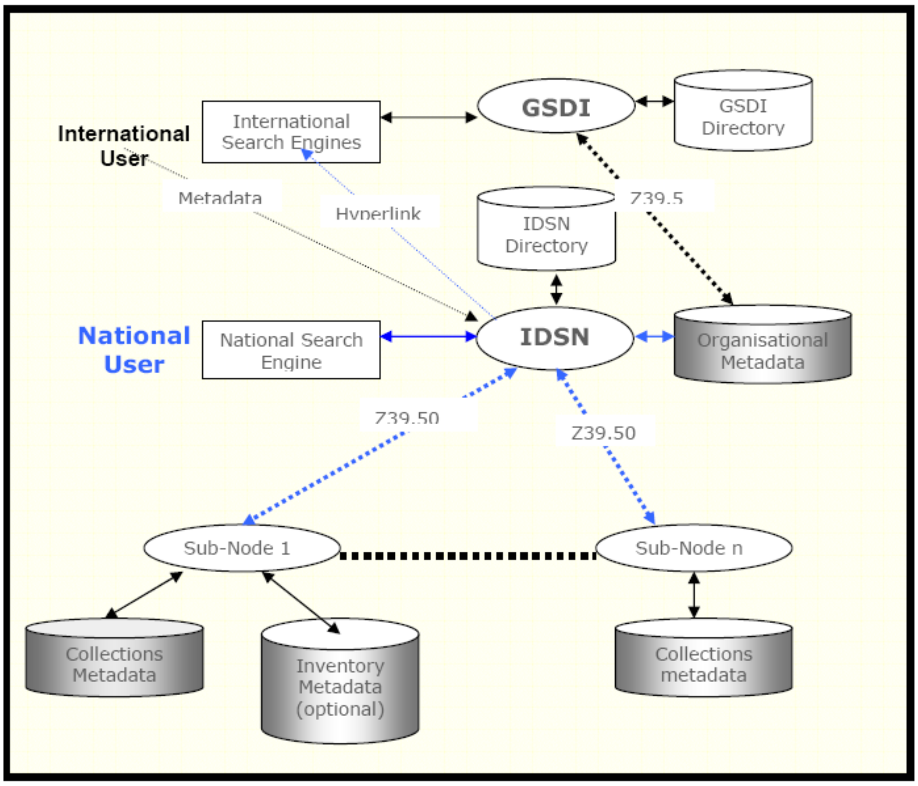


Fig 2.3: Interoperability of Indonesian National Spatial Data Infrastructure (IDSN)

This spatial data infrastructure gives room for spatial data integration and information exchange for the purpose of national development. This integration will also maximize the use of spatial data, provide easier access and distribution, avoid data duplication, and support the decision making.

During the establishment and expansion of this infrastructure, an application that gathers information from IDSN and provides user with spatial analysis is created. This application is based on Spatial Application Service Provider (Spatial-ASP) mainly from current IDSN. It creates reports on many national development aspects, for example: spatial information about unemployed population, poverty map, economic growth map, education development map, and other maps or spatial information. This application, called SISN

(*Sistem Informasi Spasial Nasional* / National Spatial Information System), is currently used by national government and has the following characteristics:

- Based on web and GIS
- Provide spatial and textual information of regional potential resources
- Temporal data covers several years of information; it covers the regional level of province, district and county.
- Data updating can be done online by each IDSN node.
- Very useful to support decision making with an accurate and integrated data
- At the early stage of development, it is operated by Vice President Office, Coordination Ministry of Public Wealth, and Bakosurtanal.

The following Figure 2.4 shows the working diagram of Indonesian National Spatial Information System, developed by Bakosurtanal together with Indonesian spatial data stakeholders.

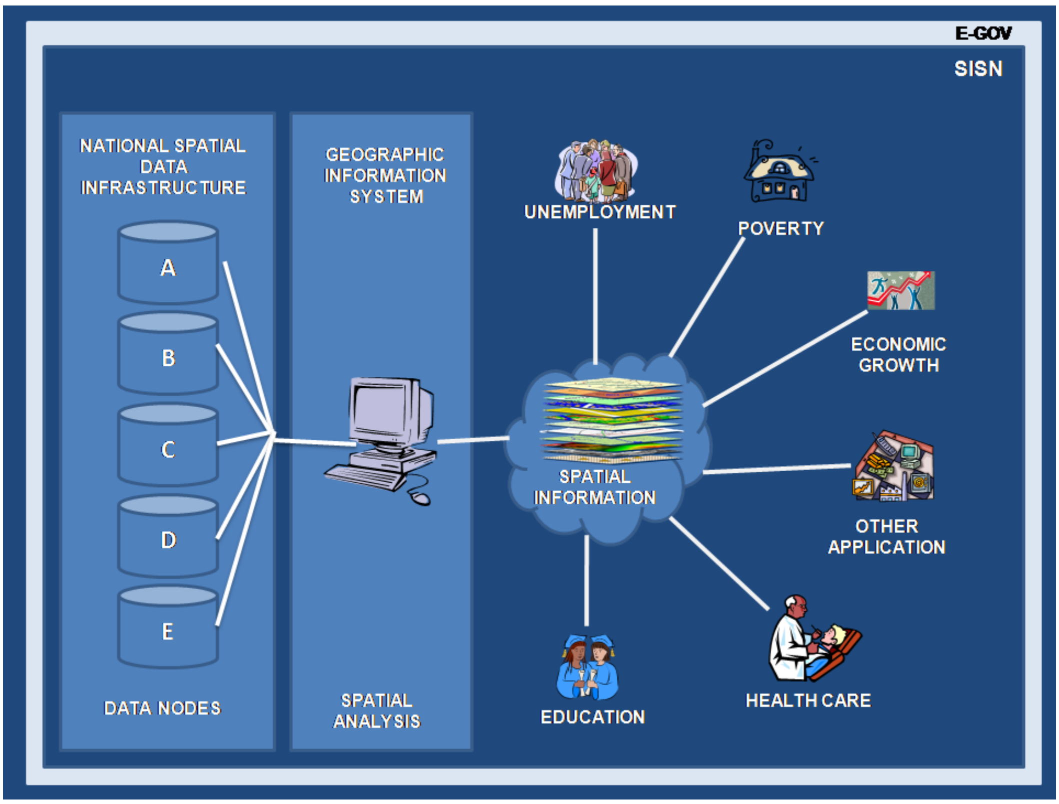


Fig 2.4: Indonesian National Spatial Information System (SISN)[12]

[12] Diagram by Bakosurtanal, internal report, translated into English

2.3.2 Legal Aspect

Several aspects of laws and regulations have been taken into consideration during the development of spatial data infrastructure in Indonesia to ensure mutual cooperation among surveyors, data providers and data users. These laws and regulations cover the protection of copyrighted data, standardization of data, detailed specifications of data, procedures and institutions.

Current developments in establishing national law of spatial data infrastructure include the following coverages[13]:

- National and regional coordination
- Participation and responsibility of government institutions
- Participation and responsibility of non-government institutions
- Defining key institutions
- Data distribution rules
- Rules for pricing, copyright, security, privacy and liability
- Access rules

2.3.3 Geospatial Data

Spatial data and spatial information are important parts of the infrastructure. It is needed by many users for planning and implementation of regional or national development. Spatial datasets from remote sensing data, geodetic base map, thematic maps, and other types of maps are formulated with their type, standard, specification, availability and accessibility to develop national main spatial data. A coordination forum of national main dataset development has been formed with the following targets:

- Formulation of main dataset standard
- Formulation of data acquisition specification
- Formulation of database standard
- Formulation of data distribution protocol
- Formulation of work area network

The main dataset or fundamental spatial dataset is provided by national institutions and intended to be the main source of spatial data. Many types of data and information can be derived through integration and addition.

---

[13] From IDSN Guide, Bakosurtanal internal report

The importance of the main spatial dataset and the consideration for optimizing the usage of this main dataset require an agreement of how far the government should be involved in providing the data; in which form, by which institutions, and how to access the data. Coordination among institutions is needed to ensure that all datasets are collected using applicable standards in accordance with policies determined by IDSN.

2.3.4 Research and Development

Dynamic innovation and achievement in technology should be thought as a chance to find a better solution of the current system, to solve problems, and give a better added value to the spatial data. Several main objectives of the research and development in IDSN are[14]

- Supporting the continuity of research and development activity
- Improvement of knowledge and implementation skill
- Achievement of new invention in the field of spatial data.
- Giving an added value to fundamental spatial data.

2.3.5 Human Resources Management

Human resources development in IDSN is directed to professionalism with a competitive skill compared to international resources. IDSN considers the development of human resources as an important part, and it covers three major directions:

- Developing skills in mapping
- Accreditation and certification of surveying and mapping profession
- Standardization of competence in surveying and mapping

2.4 Clearinghouse

The fundamental goal of the geospatial clearinghouse is to provide access to digital spatial data and related online services for data access, visualization, or order[15]. It is a distributed online system which provides a catalogue system where the public can browse collections of spatial data, and locate the original data and its provider. The

[14] IDSN Guide (Internal report from Bakosurtanal), 2003
[15] Clearinghouse Concepts Q and A: http://www.fgdc.gov/dataandservices/clearinghouse_qanda, last accessed Oct 2009.

original data are usually stored in their original format. Publicly available data can be accessed or downloaded directly, or user can get the information about the provider who owns that spatial data.

Spatial data are stored in many forms, which make their discovery using common internet / intranet search engine difficult. Even if a search engine can read all fields in a metadata file, it is required to reconstruct the structure of metadata to get the information as required. The clearinghouse enables a simplification by distributing the description of spatial data through metadata instead of the original dataset.

The main component of a clearinghouse is metadata, where the extracted information from spatial data is stored. The other important component of clearinghouse is the access protocol which allows public access to a catalogue system. The following section will discuss metadata and the protocol in the Indonesian clearinghouse.

2.4.1 Metadata in Clearinghouse

Metadata in a spatial data clearinghouse describes key information of corresponding dataset. In practice, metadata is a separated data file which stores descriptions of a corresponding spatial dataset. Metadata in the Indonesian clearinghouse supports the distribution of spatial data which covers the following benefits:

- support the organization, maintenance and rule of spatial data investment
- inform the data ownership to user and catalogue system
- get information of data processing, or interpretation from external / imported data

The following information is designed to be included in the metadata of the Indonesian clearinghouse as an implementation of above consideration:

- Availability
- Usability
- Accessibility
- Transferability

Center for Network System and Spatial Data Standardization of Bakosurtanal formed a task force for metadata in 2002. Members of this task force are from units in Bakosurtanal and several institutions including Ministry of Agriculture, Ministry of Forestry, Ministry of Infrastructure, Ministry of Energy and Mineral Resources, and Indonesian Navy. As a result, FGDC Metadata standard is adopted to be used as Indonesian national standard for spatial metadata[16].

As a follow-up, a team work is formed to enable the implementation of FGDC Metadata Standard. This team translates the metadata into Indonesian term, introduces a user guide of metadata to data producers and clients, and develops metadata applications. Metadata applications which are developed or collected by Indonesian clearinghouse can be divided into five categories:

- Generator: Application to generate standard metadata file from existing information
- Extractor: Application to extract information from metadata file
- Viewer: Application to view metadata content
- Manager: Application to maintain metadata
- Server: Application to store and integrate distributed metadata

Following the above progresses, Bakosurtanal introduced an application named *Metadata Data Spasial Nasional* (MDSN). This application has two main parts: MDSN Web-based application and MDSN Desktop application. MDSN Desktop application is distributed to every data provider in Indonesia to create metadata which conforms standard. MDSN Web-based application is installed in Bakosurtanal network to maintain metadata in clearinghouse. More details of metadata and its standard will be discussed in chapter 3.

2.4.2 Protocols in Clearinghouse

Bakosurtanal adopts client-server-protocol standard Z.39.50[17] to be used as an open system interconnection in the Indonesian clearinghouse. This protocol is widely accepted worldwide and supports implementation of TCP/IP transmission protocol. This protocol is an application-layer-protocol which mainly used for searching and retrieving in a wide-area-network.

The use of standard protocol is necessary to implement search and retrieval in a cross-platform application which is provided by the Indonesian clearinghouse. To implement this protocol, many open-source applications or modules can be used. A specific use of Z.39.50 protocol for spatial data distribution is implemented using a strict structure

[16] From the conclusion of Indonesian yearly national coordination meeting of IDSN in Yogyakarta, 2004.

description known as profile. GEO-profile[18] is an integration of FGDC standard metadata and Z.39.50 protocol, introduced by US FGDC and USGS to build server nodes in a clearinghouse.

GEO-profile, which is used in the Indonesian clearinghouse, gives the facility to search and retrieve spatial data from a GEO-Server and uses XML data format as its record syntax[19]. The GEO-Profile supports search and retrieval of geospatial metadata entries and related geospatial data sets accessible to GEO-compliant servers in the Internet environment. This includes local applications of isolated TCP/IP networks, local area networks based on ethernet protocol, and wide-area networks using a consistent host addressing scheme. The GEO-Profile will be used by implementations of GEO-server processes as one method of providing standardized access to geospatial metadata and as a reference to access the data they reference. It will also be used by client developers to understand expected behaviors of GEO-servers. A GEO-server accessed using Z39.50 in the Internet environment acts primarily as a direct access point to information stores it controls. These information stores can be accessed through an API, such as is provided with the ZDIST software from the Clearinghouse for Networked Information Discovery and Retrieval.

GEO-servers are required to support three element set names: B (Brief), S (Summary), and F (Full). GEO-servers will interpret the use of the element set names to identify the following elements from the GEO-Schema[20] (marked with their SGML tags in parentheses):

- **B:** includes the Title (title) element.
- **S:** includes the following elements: Title (title), Online Linkage (onlink), Bounding Coordinates (bounding), Extent (extent), Publication Date (pubdate), Beginning Date (begtime), Ending Date (enddate), Browse Graphic (browse), Entity Type Label (enttypl), Attribute Label (attrlabl), and Data Set G-Polygon (dsgpoly). The Browse Graphic (browse) should appear as groups of Browse Graphic File Name (browsen), Browse Graphic File Description (browsed), and Browse Graphic File Type (browset).
- **F:** contains all elements available in the record. The server should include in the retrieved record all of the elements for which there is data available in the database record and which can be encoded in the requested record syntax (e.g.,

---

17 http://www.loc.gov/z3950/agency/
18 http://www.fgdc.gov/standards/projects/GeoProfile
19 Bakosurtanal, „Pedoman Pembangunan Clearinghouse Data Spasial“, Internal report, 2003
20 http://www.fgdc.gov/standards/projects/GeoProfile

some types of locally-defined binary data may not be encodable in a USMARC or SUTRS record).

And, optionally:

- **A:** includes the Title (title) and Abstract (abstract) elements.

2.4.3 Example of Spatial Data Clearinghouses

Bakosurtanal compares or benchmarks its clearinghouse with other spatial clearinghouses, and opens the possibility for integration. This integration is currently on progress in the Asia-Pacific region.

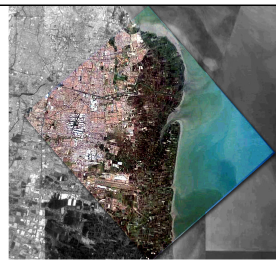
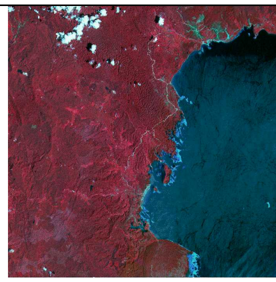
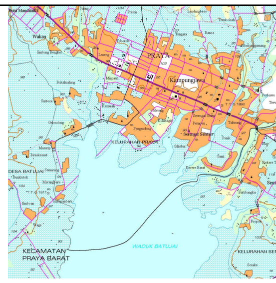
Several of web-enabled clearinghouses:

FAO Geo-Network
    http://www.fao.org/geonetwork/srv/en/main.home
GEO Portal, ESA Earth Observation Community Portal
    http://geoclearinghouse.eoportal.org
Hawaii Data Clearinghouse
    http://hawaii.wr.usgs.gov/
List of active clearinghouse worldwide:
    http://clearinghouse4.fgdc.gov/registry/clearinghouse_sites.html

2.5 Raster Data in Indonesian Spatial National Data Infrastructure

Each node in Indonesian spatial national data infrastructure shares spatial data in many forms of vector data and raster data. Raster data which are stored and distributed within Indonesian national spatial data infrastructure (IDSN) have many levels of processing status, for example raw images from remote sensing data provider, and mosaic images as a result of certain image processing. Amhar (2001) introduces the organization of spatial data as multiple levels based on the processing status of those data. This classification brings the starting point to organize the raster data naming rules, and also possible directory structure to store the dataset. Name and description for each level can be shown in the following Table 2.2.

Table 2.2: Raster data organization as introduced by Amhar

Data Level	Description	Example Image
DB0-R	Raw raster data from aerial photo, Satellite Images or other remotely sensed data.	
DB1-R	Geo-rectified remote sensing data. The raster data contains the original radiometric values / layers	
DB1-ZR	DEM	
DB2-GR	Image interpretation result or remote sensing analysis layer, for example NDVI layer, feature extraction, change detection	
DB3-GR	Presentation data.	

This raster data organization based on level of processing status has different approach compared to organizing raster data based on spatial coverage or sensor type. This raster data organization opens the possibility to organize the processing levels of remote sensing data collections. In practice, above mentioned processing level codes (DB0, DB1, DB2, and DB3) can be given to each dataset as an additional attribute inside their metadata, or explicitly included in filename path.

There exists also a levelling code for remote sensing data in many remote sensing data providers. The idea of this levelling is similar where the lowest level (level-0) represents the most untouched or original remote sensing data, and the highest level represents the most justified or corrected data. The following list shows the levelling code which is used by NASA[21] and many other data providers[22]:

Level	Description
Level-0	Original resolution with any and all communication artefacts removed.
Level-1a	Original data as of Level-0, with time referenced, and many geometric calibration coefficients and geo-referencing parameters appended.
Level-1b	Several correction, mainly on spectral level have been applied
Level-2	Ground truth is applied, deriving a new variables for example soil moisture, calculated in the original level-1 resolution
Level-3	Interpolated (or extrapolated) to form a uniform spatial coverage, for example a complete region mosaicking
Level-4	Result from analysis of lower data

:

:

[21] Earth Science Reference Handbook by NASA: http://eospso.gsfc.nasa.gov/ftp_docs/2006ReferenceHandbook.pdf, last accessed 1 Oct 2009
[22] As listed in Wikipedia: http://en.wikipedia.org/wiki/Remote_sensing

3 Remote Sensing Metadata

Remote sensing data are usually stored in digital format, or converted into digital format. There exist many different formats and different media. One remote sensing data file can contain more than one image which is called layers or bands. Multispectral and hyperspectral data may consist of several to several hundreds of bands. Besides image data, auxiliary data exists which inform about certain specific characteristics such as geometric and radiometric calibration, sensor type, time stamps, and other exterior information.

Metadata is a data about data. It is a descriptive data which is used by user or application to know what is stored in the data file, and how it is stored. In general, remote sensing metadata contains descriptive information about the image content, and technical information about the file structure. The following discussion emphasizes the use and organization of this descriptive / semantic data which is stored together with geospatial remote sensing data.

3.1 Remote Sensing Data Format

Computer applications for remote sensing usually store image data in indigenous format defined by software provider, rather than commonly used image format. The following Table 3.1 shows examples of various remote sensing file formats with their specific characteristic.

Table 3.1: Popular raster data format used in remote sensing applications

GeoTIFF (.TIFF)	GeoTIFF format fully complies with the TIFF 6.0 specifications. It uses a "MetaTag" (GeoKey) approach to encode information elements into tags, taking advantage of TIFF platform-independent data format representation to avoid cross-platform interchange difficulties. BigTIFF extends GeoTIFF beyond 4GB file size limit. ESRI has a history of using TIFF format for geographic data by introducing a separate metadata file known as a world file (TFW). GeoTIFF covers all features of ESRI TFW world file and organize the metadata within the TIFF data file.
ERDAS Imagine (.IMG)	ERDAS IMAGINE uses .img files to store raster data, extending their hierarchical file format (HFA). This format supports multi-layer bitmap from multispectral sensors, and stores many descriptions for geo-orientation and image processing purpose.
ERMapper (.ERS)	ER Mapper Raster dataset is made up of two files: a header

	file, with an .ERS extension and a data file of the same base name with no extension.
Open Geospatial Consortium GML in JPEG2000	This GML is an XML grammar for the encoding of geographic information including geographic features, coverage, observations, topology, geometry, coordinate reference systems, units of measure, time, and value objects. JPEG2000 is one of the portable format and widely used in current remote sensing file compression.
Intergraph Raster File Format (INGR)	The INGR file format was proprietary for a long time and was not well known outside of INGR customer sites. It has now been released to the public, and its use is encouraged by Intergraph especially for use with Intergraph software.
NOAA Level-1b Data (.LAC)	NOAA Local Area Coverage (LAC) Data consist of binary pattern from sensors which is called “HRPT video data”, and series of spatial-related information stored in its header.

The basic structure of raster or image data consists of header data and bitmap data, either stored as one file or as separate files. The header part is used by computer application to interpret what data are stored, how data are stored, what algorithm should be used to decode, and exterior information (e.g. sensor type, acquisition time). GeoTIFF[23] is one of the commonly used open source format for raster data where header data and raster data are stored together in one file. This format is used in many applications for raster data storage and exchange.

3.2. Organizing Remote Sensing Data

In many data centre, number of remote sensing data is increasing which requires a database system to organize it and support certain user query. In a database system for image (image database) remote sensing data can be presented as a binary large object (Jianting et al., 2009; Cha et at. 1997). This image database is commonly used to support interactive image visualization, image analysis, image interpretation and other image processing. The NASA Magellan Mission to Venus Project is an example how to organize the huge collections of remote sensing data within a database, and implements a combination of Java+HTML for the user interface (Walcher, 1997). Google Maps integrates Tele Atlas GIS data with vector and raster data from around the globe and allow public to use this web-based map. This web-based map can be used as an embedded feature in user website, or a user-defined map on top of Google Maps using Google Maps API.

[23] Detail of GeoTIFF: http://trac.osgeo.org/geotiff/

Query based on image content (binary pattern) has shown many possible applications, for example: Image Content Characterization by Schroeder (Schroeder, et al., 1997), Automated Extraction of Linear Features from Aerial Imagery Using Kohonen Learning and GIS Data by Doucette (Doucette et al., 1999), Content-based Image Classification for Satellite Image Database (Holowczak et al., 2002) and Comparison of GENIE and Conventional Supervised Classifiers for Multispectral Image Feature Extraction by Harvey (Harvey, et al., 2002). Although many researchers have shown the possibility of automatic feature extractions from image data, having the descriptive information together with the remote sensing data has an advantage if this data will be organized within a database, or a description of dataset is required for data exchanging (Goodchild and Zhou, 2003).

Organizing image data for a catalogue system requires appropriate descriptive information about the image to minimize the access to the rather big images. Metadata consist of information that characterizes data. They are used basically to provide documentation for data products. Metadata information can be stored internally within the header of data file like in JPEG or GeoTIFF, or stored externally in a separate file. Metadata file can be in any forms, from a simple text file to a structured record. Figure 3.1 shows an example of a raster dataset with its metadata stored as an independence file. A separate data file, usually in a standard text format, is used to store descriptive information. This descriptive part does not replace the header of the raster data, which contains the technical description. This header is needed to understand the physical structure and to advice an application to read the data in a correct way.

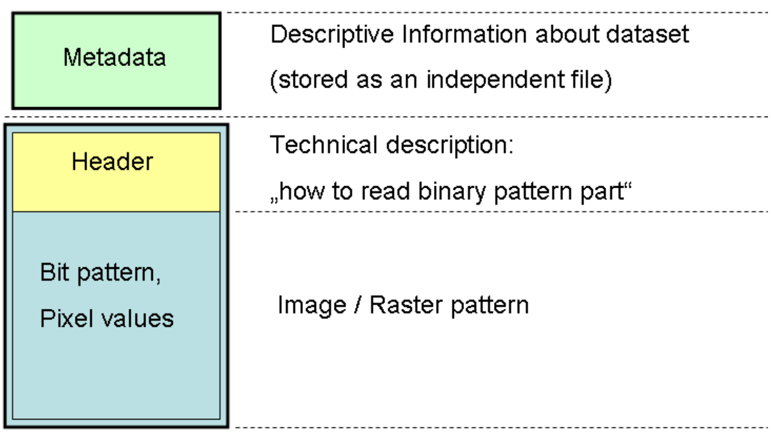


Fig 3.1: Metadata in raster data structure

Within the spatial data user community, consensus was sought for what metadata should be managed with digital spatial data sets for a great variety of uses. The anticipated primary uses of geospatial metadata defined within the preamble to the Federal

Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata are[24] :

- to help organize and maintain an organization's internal investment in spatial data,
- to provide information about an organization's data holdings to data catalogues, clearinghouses, and brokerages, and
- to provide information to process and interpret data received through a transfer from an external source.

The more standardized the structure and content of information, the more effectively it can be used by both humans and machines. A metadata standard is simply a common set of terms and definitions that are presented in a structured format or also known as schema. ISO uses the following term as a design philosophy of ISO 19115 (ISO TC211, 2008): "Data describing a data set is becoming ever more important for locating and accessing information of all kinds. The schema will be used by geographic information users to add metadata in a consistent and verifiable form to data being created and to evaluate quickly and accurately the data being selected from other sources. Software developers will use the schema to provide applications that provide consistent methods of handling metadata."

Geospatial metadata defines the following characteristics of dataset[25]:

- Structure
- Terms and definition of data description
- Allowed data range or class
- Which metadata is mandatory, mandatory if applicable, or optional.

Besides the defining standard in above criteria, a metadata standard is also designed to be human readable and also machine readable. Current standard metadata format is XML based which is used in FGDC Remote Sensing Extension for FGDC CSDGM Version 2, and ISO 19139 (XML Implementation of ISO 19115). There are several approach of spatial data organization based on their metadata (Li et al., 2005; Bose and Frew, 2004; Ruan et al., 2006; Sukmayadi et al., 2004)

---

[24] As listed in http://www.fgdc.gov/dataandservices/fgdcmeta, last accessed Oct 2009
[25] See also: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf. and http://csi.cgiar.org/metadata/Metadata_Why.asp and http://en.wikipedia.org/wiki/Metadata

The following sections will discuss the FGDC and ISO standard metadata more in detail. These two standards are widely used and compatible one to the other. With a specific tool, a metadata from one format can be converted to fit the other.

3.3 FGDC Metadata

The Federal Geographic Data Committee (FGDC) is an interagency committee in the United States of America, which is directed to coordinate the National Spatial Data Infrastructure (NSDI) of USA. It encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data. The FGDC develops geospatial data standards for implementing the NSDI, in consultation and cooperation with state, local, and tribal governments, the private sector and academic community. Remote sensing data is one of major types of geospatial data that the federal agencies collect, archive, distribute, and use[26].

The Content Standard for Digital Geospatial Metadata (CSDGM), Version 2 (FGDC-STD-001-1998) is the US Federal Metadata standard. The Federal Geographic Data Committee originally adopted the CSDGM in 1994 and revised it in 1998. According to Executive Order 12096 all Federal agencies are ordered to use this standard to document geospatial data created as of January, 1995. The standard is often referred to as the FGDC Metadata Standard and has been implemented beyond the federal level with State and local governments adopting the metadata standard as well.

The Organization of FGDC Metadata Standard can be shown in the following Figure 3.2. The FGDC metadata standard contains approximately 300 data elements. Among those 300 elements, 199 elements are attributes tag; the remainder are grouping elements (compound elements) which give an overall structure to the standard. Not all of these 199 elements are considered "mandatory". Around 100 elements are considered "mandatory if applicable", and the rest are optional to provide structured places for information that would otherwise become lost in generalized comment fields[27]. Figure 3.2 shows a main structure of the standard, and it also shows an example of mandatory parts. The FGDC standard is divided into sections and each data element has a number to ease in human navigation.

FGDC metadata is accepted and used by many governmental and private institutions and can be applied using popular GIS Software. Bakosurtanal in Indonesia has also decided

[26] http://www.fgdc.gov/dataandservices/fgdcmeta, last accessed Oct 2009.

[27] http://www.fgdc.gov/dataandservices/fgdcmeta, last accessed Oct 2009.

to use FGDC metadata standard as Indonesian standard for digital spatial metadata. A task force has been established to implement this standard into Indonesian National Spatial Data Infrastructure.

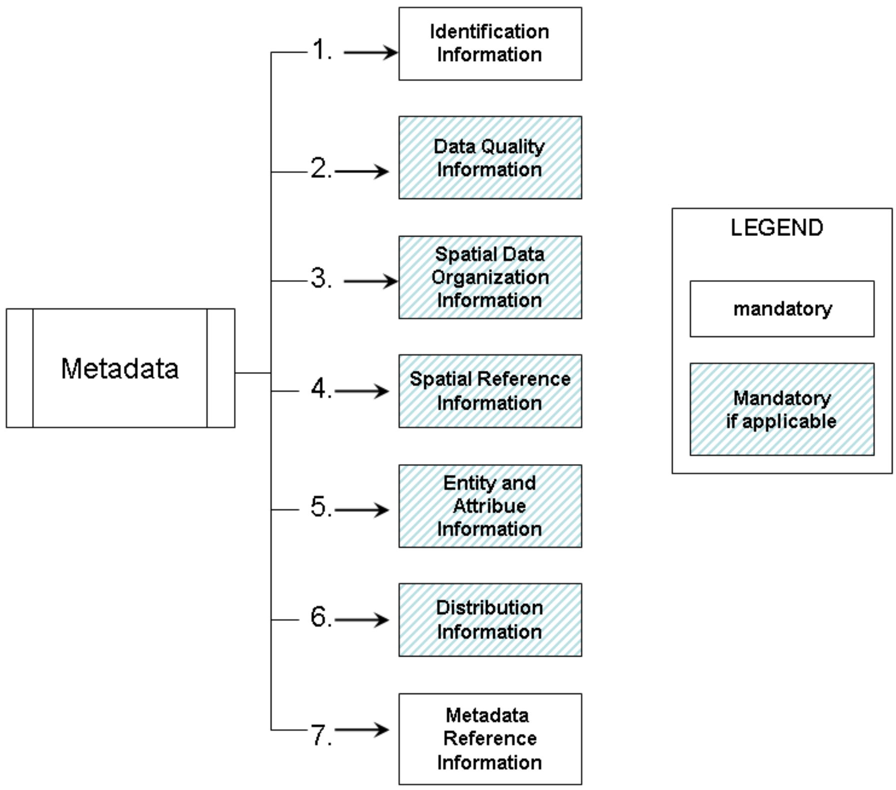


Fig 3.2: FGDC schematic diagram (FGDC, 2000)

3.4 ISO 19115 Metadata

The ISO Technical Committee 211 (TC211) builds the 19100 series of geo-information schemata that allows geographic information to be integrated with the information technology. ISO 19115 identifies a set of core metadata from the many metadata elements it defines. There is a set of required elements, most of which are essentially catalogue elements that identify the dataset, tell the user what it is about, and tell the user how to get it. A few elements are mandatory under certain conditions. All dataset metadata must include these elements. A set of optional elements, whose use will increase interoperability is defined. The following clauses describe the core metadata in detail.

Table 3.2: ISO 19115 metadata content (ISO TC211 Scope, 2007)

(M: mandatory, C: mandatory under certain conditions, O: optional)

Core metadata for geographic datasets	obligation	UML hierarchy
Dataset title	(M)	(MD_Metadata > MD_DataIdentification.citation > CI_Citation.title)
Dataset topic category	(M)	(MD_Metadata > MD_DataIdentification.topicCategory)
Abstract describing the dataset	(M)	(MD_Metadata > MD_DataIdentification.abstract)
Dataset reference date	(M)	(MD_Metadata > MD_DataIdentification.citation > CI_Citation.date)
Dataset language	(M)	(MD_Metadata > MD_DataIdentification.language)
Metadata point of contact	(M)	(MD_Metadata.contact > CI_ResponsibleParty)
Metadata date stamp	(M)	(MD_Metadata.dateStamp)
Dataset character set	(C)	(MD_Metadata > MD_DataIdentification.characterSet)
Geographic location of the dataset (by four coordinates or by geographic identifier)	(C)	(MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or EX_GeographicDescription)
Metadata language	(C)	(MD_Metadata.language)
Metadata character set	(C)	(MD_Metadata.characterSet)
Dataset responsible party	(O)	(MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty)
Additional extent information for the dataset (vertical and temporal)	(O)	(MD_Metadata > MD_DataIdentification.extent > EX_Extent> EX_TemporalExtent or EX_VerticalExtent)
Spatial resolution of the dataset	(O)	(MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)
Spatial representation type	(O)	(MD_Metadata > MD_DataIdentification.spatialRepresentatio

		nType)
Reference system	(O)	(MD_Metadata > MD_ReferenceSystem)
Lineage	(O)	(MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)
Distribution format	(O)	(MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)
On-line resource	(O)	(MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Metadata file identifier	(O)	(MD_Metadata.fileIdentifier)
Metadata standard name	(O)	(MD_Metadata.metadataStandardName)
Metadata standard version	(O)	(MD_Metadata.metadataStandardVersion)

Considering the similarity and exchangeability between ISO 19115 and currently used FGDC Metadata Standard, Bakosurtanal has evaluated the current implementation of spatial metadata in Indonesia to conform the usage of this ISO Standard.

3.5 Remote Sensing Metadata Extension

ISO 19115 and its implementation do not provide all the metadata elements which are needed to describe geospatial imagery. Currently, three standards are available to provide additional information for imagery like remote sensing metadata:

- FGDC Remote Sensing Extensions, FGDC-STD-012-2002 (FGDC, 2002).
  The purpose of this *Extensions for Remote Sensing Metadata* (hereafter *Remote Sensing Extensions*) is to provide a common terminology and set of definitions for documenting geospatial data obtained by remote sensing, within the framework of the *FGDC (1998) Content Standard for Digital Geospatial Metadata* (hereafter *FGDC Metadata Content Standard* or simply base standard). Creating these *Remote Sensing Extensions* will provide a means to use standard FGDC content to describe geospatial data derived from remote sensing measurements. This standard is intended to support the collection and processing of geospatial metadata for data derived from remote sensing. It is intended to be usable by all levels of government and the private sector. The standard is not intended to reflect an implementation design. An implementation design requires adapting the structure and form of the standard to meet application requirements.

- ISO 19115-2 Geographic Information Metadata Part 2: Extensions for imagery and gridded data[28]

  This standard extends the ISO metadata standard to incorporate elements that are needed for the description of imagery and gridded data. The most recent working draft (WD) includes elements in many areas from the Remote Sensing Extension:

  - Information on the algorithm and/or processing used to generate the data
  - Conditions during acquisition
  - Dimensions and positions of individual cells
  - Information that is not in ISO 19115 about the wavelength bands
  - Identification information for missions, platforms, and sensors

  Additional individual elements from the Remote Sensing Extensions that would naturally have belonged to existing classes in ISO 19115 have been added.

- ISO 19130 Geographic Information —Sensor and data model for imagery and gridded data

  The scope of this standard includes (Di, 2003):

  1. Specifying a sensor model describing the physical and geometrical properties of each kind of photogrammetric, remote sensing and other sensors that produces imagery data; and
  2. Defining a conceptual data model that specifies, for each kind of sensor, the minimum content requirement and the relationship among the components of the content for the raw data that was measured by the sensor and provided in an instrument based coordinate system, to make it possible to geo-locate and analyze the data.

3.6 ISO 19139 Geographic Information Metadata XML Schema Implementation

Early drafts of ISO 19115 contained implementation information in the form of an annex with a Document Type Description (DTD). However, as ISO 19115 was developing, practice in the usage of XML was shifting from the use of DTDs to XML schema. Rather than delay progress of the ISO standard any further, it was decided to produce a

[28] Can be obtained in: http://www.iso.org/iso/catalogue_detail.htm?csnumber=39229

separate technical specification with an XML schema that could be used to generate XML for ISO 19115 compliant metadata. This project was approved as ISO 19139.

FGDC, together with Intergraph, develop a crosswalk which bridge the FGDC Metadata Standard with the ISO 19115/19139. With this crosswalk, metadata created using FGDC Standard can be used further in FGDC or ISO System Standard, without recreating metadata.

3.7 Development of Metadata Standards

Liping Di (Di, 2003) from the Laboratory for Advanced Information Technology and Standards (LAITS), Georg Mason University, describes the development of remote-sensing related standards at U.S. Federal Geographic Data Committee (FGDC), the Open GIS Consortium (OGC), and the International Organization for Standardization (ISO) Technical Committee 211 (TC 211). The paper introduces the individual remote sensing standards developed or being developed by those organizations. It also discusses the relationship among those standards and how to use the standards in the development of interoperable, standard-compliant geospatial data and information systems for remote sensing. Di also developed a prototype of geospatial web service system for remote sensing data which implements several possibilities for interoperability.

4 Database Systems for Spatial Data Organization

4.1 Overview

The discussion database system here is to find a suited system for organizing metadata of remote sensing data. Since remote sensing metadata contains spatial and temporal information, a spatially-enabled database system is highly considered. This chapter will explore currently existed spatially-enabled database system to support the proposed spatial data handling. Building a prototype of the spatial data handling will be useful to verify and evaluate the proposed method. The proposed method will be explained in the chapter 5, where the evaluation including the prototype of the implementation will be discussed in chapter 6.

To build a spatially-enabled database system for organizing raster data, several aspects of computer environment should be supported, which are:

Spatial-enabled Database
This database is designed to be part of or support the spatial data infrastructure where spatial data can be in form of vector data and raster data.

XML support
The database system will be used to organize the remote sensing metadata which is stored in XML-format, conforming FGDC Standard Metadata or ISO-19115 / ISO-19139 Metadata

Web-enabled application
The interoperability of spatial data in Indonesian National Spatial Data Infrastructure relies on internet connections among distributed data stakeholders and users. The use of web-enabled applications should be considered.

Visual interface for user query
Although information of spatial coverage of each dataset in metadata is limited to coordinate values of a bounding-box, an interactive map for selection allows better interactions and user-friendly queries.

The following section discusses the current support or availability in spatial databases or spatial data management, and the current progress in implementation of national spatial infrastructure in Indonesia.

4.2 Spatial Database Features

A spatial database is an extension of object-relational DBMS, or an additional engine that is embedded to a DBMS to support spatial objects. With this extension, a database system supports additional features to store and query data which are related to space or geometry such as: point, line and polygon.

A spatial database which is an extension of a relational database management system (RDBMS), inherits the currently standard query language, SQL, and other RDBMS features like indexing. Both indexing and SQL are optimized to support spatial indexing and spatial query.

The Open Geospatial Consortium (OGC)[29], as an international voluntary consensus standard organization, has introduced many standards in spatially-enabled database. A spatial feature in spatial database is defined in the Simple Feature SQL including point, line-string, polygon, multipoint, and multi-polygon. Several examples of typical spatial query, which have been introduced and standardized by OGC are (OGC, 2009): spatial measurements, spatial functions, spatial predicates, constructor functions and observer functions. OGC also introduces several consensuses which are supported by many commercial and open source database systems, for example:

XML and GML

Extensible Markup Language (XML) is a flexible text format derived from SGML (ISO8879). It is human-readable and machine-readable, and commonly used for data exchange[30]. There are many XML-based language have been developed so far, mainly in the web application. XML-based format is also widely used in document formats for office application. Through this popularity, most of the current office application support the XML-based document formats. Geographic Markup Language (GML) is XML grammar to express geographical features. The work in the field of GML and XML in spatial database is also related to current standard ISO TC-211, ISO 19115, ISO 19139 and FGDC

[29] http://www.opengeospatial.org/
[30] http://www.w3.org/XML/

Metadata Standard. Many proposals of the use of XML in spatial data handling has been introduced (Bose et al., 2004; Jea et al., 2008; Nance and Hay, 2005)

Web Map Service (WMS)[31]

Web Map Service (WMS) is a standard protocol for serving maps over the internet protocol (TCP/IP). This service uses a server, called map server to provide maps from a spatial database. WMS is widely supported by open source software like Mapserver and commercial database software like ArcGIS, MapInfo, GeoMedia and Oracle.

4.3 Comparison of Spatial Database Software

As listed in OGC, there are currently hundreds of spatially-enabled database systems or similar products (OGC, 2009). The following Table 4.1 shows several examples of spatially-enabled database systems:

Table 4.1: Popular spatial database systems

Company	**Products**	**Remarks**
ESRI	ArcGIS, ArcSDE, ArcGIS Server	Server, Client
IBM Corporation	IBM DB2 Spatial Extender	Server, Client
ORACLE	Oracle Spatial	Server, Client
FAO	GeoNetwork Opensource	Server, Client
AUTODESK	AutoCAD Map, MapGuide	Server, Client
Bentley Systems	Bentley Map, Geo Web Publisher	Server, Client
ERDAS	ERDAS Apollo Server	Server, Client
Intergraph	GEOMedia	Server, Client
UMN Mapserver	Mapserver	Server, Client
Refractions	PostGIS + PostGRESQL	Server
SUN	MySQL	Server

The following sections discuss the current state of the most popular spatial database software, i.e: ESRI Spatial Data Engine, Oracle Spatial, and Open source PostGIS.

[31] http://portal.opengeospatial.org/files/?artifact_id=4756

4.3.1 ESRI ArcSDE[32]

ESRI's announced Spatial Database Engine (ArcSDE) claims to meet the needs of the next generation of geographic data users. This geographic database product provides a spatial extension to an underlying commercial Relational Database Management System (RDBMS), thereby enabling all data (spatial and non-spatial) to be stored within a single RDBMS.

Until the release of ArcGIS 9.1, ArcSDE can be used within popular commercial RDBMS such as Oracle, Microsoft SQL Server, Informix and IBM DB2 (ESRI SDE, 2009; Murray and Lutz, 2009). While the RDBMS software is used to keep track of the tables and records contained in the database, ArcSDE enables the organizing of geographic data which comprises several tables. The use of spatial-enabled database is performed with ESRI SDE Application Programmer's Interface (API). ArcSDE serves data to ESRI Desktop GIS application where the spatial analysis can be performed. Since this architecture is designed to be used with ESRI Desktop GIS, the directly use of Java Application through Java Database Connection is not suitable.

Starting from ArcGIS 9.2, ArcSDE is introduced as "ArcSDE technology" which is an integrated part of ArcGIS Server as a core element of ESRI GIS solution. ArcSDE is no longer being sold as a separated product.

4.3.2 Oracle + Oracle Spatial 11g

Oracle Spatial is an integrated set of functions and procedures that enables spatial data to be stored, accessed, and analyzed quickly and efficiently in an Oracle database (Oracle Spatial 11g, 2009). Spatial data represents the location characteristics of objects in which they are exist. Oracle Spatial provides SQL schema and functions that facilitate the storage, retrieval, update, and query of collections of spatial features in a database. It includes of the following components (Oracle Spatial 11g, 2009):

- A schema (MDSYS) that prescribes the storage, syntax, and semantics of supported geometry data types
- A spatial indexing mechanism
- A set of operators and functions for performing area-of-interest queries, spatial join queries, and other spatial analyses operations
- Administrative utilities
- A topology data model

[32] http://www.esri.com/software/arcgis/arcsde/index.html

- A network data model
- A GeoRaster feature to store raster image and gridded data and its associated metadata.

An example of an implementation of Oracle Spatial is the Topographic Mars Information System (TMIS) by Dorninger (2003), where a database is used to organize the remote sensing dataset and its metadata. TMIS uses an own metadata structure in its Meta Data Catalogue (MDC). XML was tested to be implemented within this system, but it has been dropped later due to its inefficiency. The simple ASCII list is used inside an Oracle RDBMS. Therefore, there is no discussion about remote sensing metadata standard in TMIS. The result is good, but the complexity of Oracle Spatial makes it not really flexible to be customized (Dorninger, 2003).

Oracle Spatial 11g extension provides a robust foundation for complex geospatial applications that require more spatial analysis and processing in Oracle Database. Oracle Spatial 11g is declared by Oracle as a complete geospatial data management platform for the requirements of any spatial information system. It supports all major spatial data types and models, addressing the challenging business-critical requirements from the public sector, defence, logistics, energy exploration, and business geographic domains.

Oracle Spatial 11g features include (ORACLE Spatial 11g, 2009):

- Powerful linear referencing system
- Over 400 Spatial functions such as centroids and aggregate functions (e.g. unions and user defined aggregates)
- GeoRaster data type that natively manages georeferenced raster imagery (e.g., satellite imagery, gridded data) in Oracle Database 11g
- Support for more file formats for loading and exporting, more metadata and data types (new in Oracle Spatial 11g)
- Enhanced ease of use, reliability, manageability (new in Oracle Spatial 11g)
- Java API (new in Oracle Spatial 11g Release 2)
- A data model to store and analyze network (graph) structure
- Load-on-demand for very large spatial networks (new in Oracle Spatial 11g)
- Advanced analysis and modeling features: database handling of user or application attributes, path arithmetic support (new in Oracle Spatial 11g)
- A data model and schema to persistently store and update topology
- Spatial analytic functions
- 3-dimensional data type support for terrain and city models and virtual worlds, support for LIDAR-based map production (new in Oracle Spatial 11g)

- Spatial web services support (WFS 1.0, WFS-T 1.0, CSW 2.0, OpenLS 1.1, web services security) (new in Oracle Spatial 11g)
- Support for SQL/MM Spatial types and operators*** (new in Oracle Spatial 11g)
- Spatial Java API

Oracle Spatial has lots of user worldwide, building a community named Independent Oracle Users Group (IOUG)[33]. This community helps each member to enhance their skill and collaboration in using Oracle products and sharing their practices with Oracle Database.

4.3.3 PostgreSQL[34] + PostGIS[35]

PostGIS is developed by Refractions Research Inc as a spatial database technology research project. Refractions is a GIS and database consulting company in Victoria, British Columbia, Canada, specializing in data integration and custom software development. PostGIS supports a range of important GIS functionality, including full OpenGIS support, advanced topological constructs (coverages, surfaces, and networks), desktop user interface tools for viewing and editing GIS data, and web-based access tools.

PostGIS is an extension to the PostgreSQL object-relational database system which allows Geographic Information System's objects to be stored in the database. Early versions of PostGIS used the PostgreSQL R-Tree indexes. However, PostgreSQL R-Trees have been completely discarded since version 0.6, and spatial indexing is provided with an R-Tree-over-GiST scheme[36].

The GIS objects supported by PostGIS are a superset of the "Simple Features" defined by the OpenGIS Consortium (OGC). As of version 0.9, PostGIS supports all the objects and functions specified in the OGC "Simple Features for SQL" specification. PostGIS extends the standard with support for 3DZ, 3DM and 4D coordinates.

---

[33] http://www.ioug.org/
[34] http://www.postgresql.org/
[35] http://postgis.refractions.net/
[36] PostGIS manual: http://postgis.refractions.net/docs/index.html, last accessed 30 October 2009

Currently there are some examples of using of the combination of PostGIS + PostgreSQL and MapServer to develop a web-based GIS, as listed in PostGIS webpage[37], for example:

- ImageFinder[38]. Previously named GlobeXplorer, this system currently serves over 1 million image requests a day on average (with occasional peaks of over 5 million requests). The data preparation and content management systems use PostGIS.

- GeoStore by Infoterra[39]. The GeoStore uses PostgreSQL/PostGIS as the database backend, UMN Mapserver for map rendering duties, and a variety of bespoke applications for data loading, order fulfillment, and access. The data managed in GeoStore now include 30TB of aerial, satellite, and vector data directly online; 1000TB of data on near-line robotic tape storage; and 600 million rows of Ordnance Survey topographic data for the UK.

- Vessel Detection System (VDS) by EU Joint Research Centre[40]. VDS combines information from the Radarsat-1 and ENVISAT synthetic aperture radar (SAR) satellites, merchant marine vessel transponders and fisheries vessel transponders to create an operational picture of known and unknown vessels in European waters. SAR satellites are very good at sensing vessels, because metal ships are very effective radar backscattering targets while water is not.

4.3.4 OGC and WFS Conform Comparison

The Open Geospatial Consortium (OGC) specifies many specifications for development and implementation of Geographic Information System including Web Feature Service (WFS). The WFS specification defines interfaces for describing data manipulation operations of geographic features. It uses HTTP (web) as the distributed platform, and XML-based Geography Markup Language.

The following Table 4.2 provides a summary of the capabilities of Postgis, Oracle Spatial and ArcSDE in relation to Web Feature Server Requirements (Garnett and Owens, 2003).

---

[37] PostGIS Case Study: http://postgis.refractions.net/documentation/casestudies/, last accessed 30 October 2009
[38] http://browse.digitalglobe.com/imagefinder/main.jsp
[39] http://www.geostore.com/geostore4/WebStore?xml=geostore4/xml/application.xml
[40] http://ipsc.jrc.ec.europa.eu/showdoc.php?doc=promotional_material/VDS.pdf&mime=application/pdf

Table 4.2: Comparison of PostGIS, Oracle and ArcSDE
related to Web Feature Server requirements

WFS Requirements	PostGIS	Oracle	ArcSDE
Feature storage	✓	✓	✓
Feature Schema	✓	✓	✓
Feature ID Generation			
Transaction WFS Requirements	**PostGIS**	**Oracle**	**ArcSDE**
Transaction Support	✓	✓	✓
Locking Support	**PostGIS**	**Oracle**	**ArcSDE**
Per feature locks			✓
Lock persistence			✓
Lock timeout			
Simple Feature Support	**PostGIS**	**Oracle**	**ArcSDE**
Simple Properties	✓	✓	✓
Multiple Geometries per Feature	✓	✓	✓
Geometry Support	**PostGIS**	**Oracle**	**ArcSDE**
Simple Feature Geometry Model	✓	✓*	✓
Well Known Text representation	✓	✓	✓
Well Known Binary representation	✓	✓	✓
Bounding Box	✓	✓	✓
Spatial Relationships	✓	✓	✓**
Spatial Analysis	✓	✓	✓**
Geotools2	**PostGIS**	**Oracle**	**ArcSDE**
DataSource	✓	✓	
Database Connection	**PostGIS**	**Oracle**	**ArcSDE**
JDBC Support	✓	✓	
JDBC Geometry Data Type Extension	✓	✓	
Custom Connection API			✓

* Oracle does not provide a representation of the Shape Geometry

** ArcSDE Geometric functions may be performed on client side

4.4 Current Progress in Indonesia

Bakosurtanal and participating institutions meet regularly in developing the Indonesian National Spatial Information System. The development which is started in 2004 using MS Access for maintaining poverty database, currently includes much more applications and employs a SQL server as an engine, a Web Mapping Service, and metadata standard. Indonesian government endorsed this idea by Presidential Rules no 85/2007 about Indonesian National Spatial Data Network (Jaringan Data Spatial Nasional)

The following Figure 4.1 shows the proposed network configuration and the access to the Indonesian National Spatial Information System.

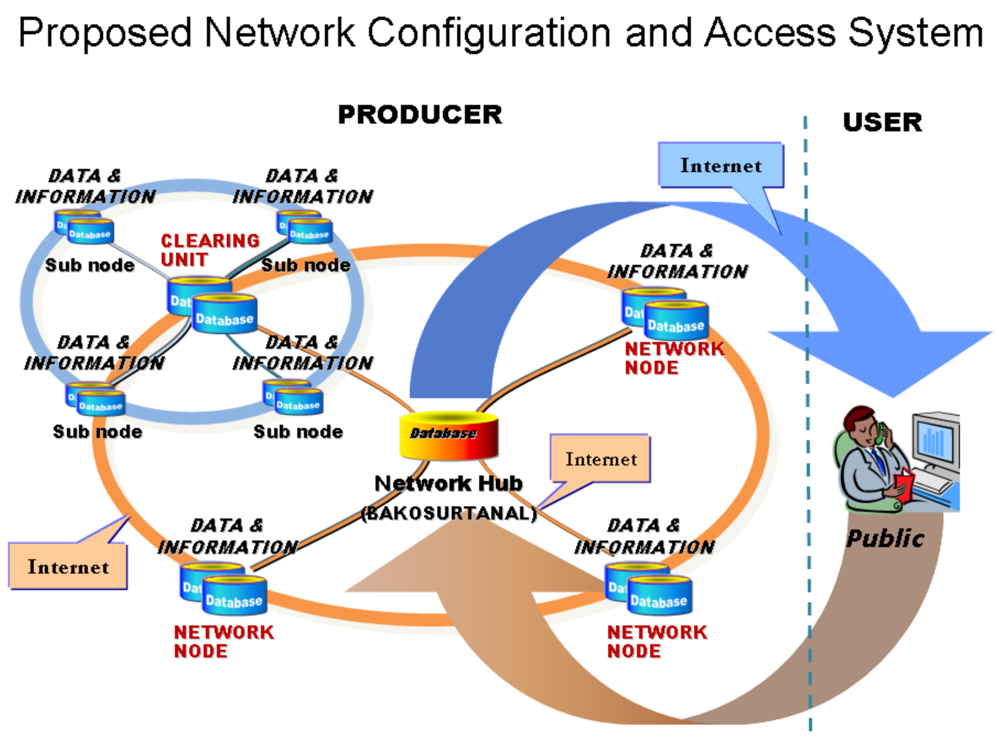


Fig 4.1: Indonesian National Spatial Information System (Bakosurtanal, 2008)

The guidelines for the development of metadata for spatial data have been brought out by the Clearinghouse Technical Working Group Bakosurtanal in 2005. The guidelines emphasize the use of FGDC Metadata Standard and Z39.50 as a standard protocol for computer-to-computer information retrieval.

The Indonesian Geospatial Data Clearinghouse (Puntodewo and Nataprawira, 2004) has been established and involves all geospatial data stakeholders in Indonesia. This clearinghouse is also connected to the Asia Pacific Spatial Data Infrastructure (APSDI) and Global Spatial Data Infrastructure.

The Indonesian Spatial National Data Infrastructure employs Oracle Database. This solution has been chosen, because of the Oracle Database as a commercial product has a support with tutorial, assistance, and it is mature compared to open source database like MySQL or PostgreSQL. To increase the interoperability of the National Spatial Catalogue System, a standard metadata structure has been considered, and FGDC

Metadata Standard was introduced. The possibility of data exchange between FGDC Metadata Standard and ISO 19115 opens the possibility to adopt the ISO-19115 as a new metadata standard for spatial data. For remote sensing data, the next possible standard is defined as a remote sensing extension and introduced both in FGDC and ISO 19115/ 19139.

5 Image Data Handling for Indonesian National Spatial Data Infrastructure

This chapter discusses the actual status of Indonesian national spatial data infrastructure as of 2009 and our proposed spatial data handling which shares the same standard and data as used in current system.

In compliance to the recommendations resulted in the Indonesian Spatial Data Infrastructure (ISDI) Coordination Meeting in 2006, Indonesia starts the National Geospatial Data Clearinghouse. This clearinghouse helps all geospatial data stakeholder in Indonesia to easily find, evaluate, and obtain many kinds of geospatial data. The following figure 5.1 shows the user interface of this clearinghouse.

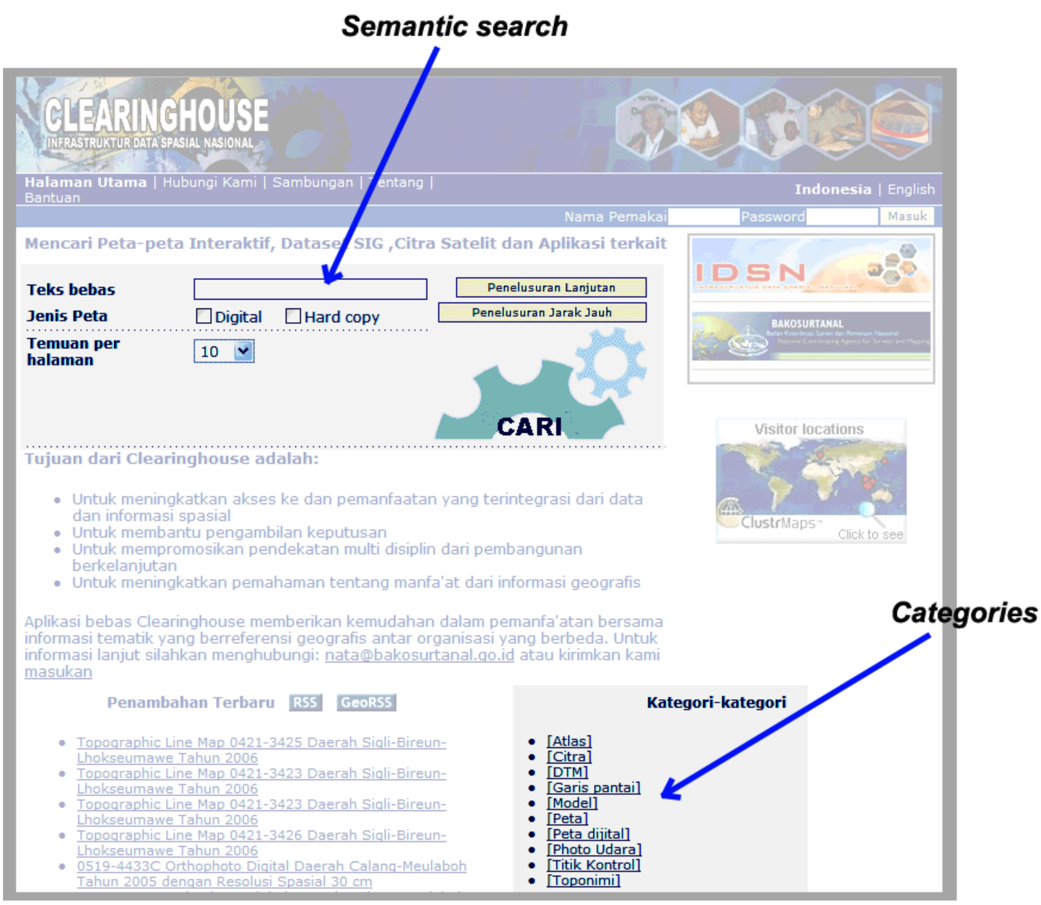


Fig 5.1: Indonesian Clearinghouse

The main screen of Indonesian clearinghouse consists of two main query supports. The first part is a text search entry to find a matching text inside the metadata. The second part is a list of categories which can be used to filter the query into a certain type of spatial data. The next section discusses the existing system, and proposes a methodology for remote sensing data handling.

5.1 Protocols and Standards in Indonesian National Spatial Data Infrastructure

Like many other mapping institution, Bakosurtanal started to produce spatial data in a certain data format which is limited by software availability. Some example of data format which is commonly used includes:

- Vector data: ESRI coverage, ESRI Shapefile, AutoCAD DXF, AutoCAD DWG
- Raster data: Raw Bitmap, ERDAS image, ERMapper dataset
- Database: Oracle, dBASE / FoxBase, MS Access

Although the data can be converted from one format to another, it is not easy to coordinate the variety of data formats from data producers since each data format has its own specification and requirement. Standardization is a key issue for developing a new platform of a geospatial data clearinghouse.

Since the existence of many spatial data formats cannot be avoided, a standard format for metadata should be used. To develop metadata for the National Geospatial Data Clearinghouse, Indonesia adopts the FGDC metadata standard called 'Content Standard for Digital Geospatial Metadata', FGDC-STD-001-1998, as the national content standard of metadata (BAKOSURTANAL, 2005).

FGDC has extended the above standard to FGDC-STD-012-2002 in order to improve remote sensing data handling. This extension will provide a means to use standard FGDC content to describe geospatial data derived from remote sensing measurements. Several new elements are placed within the structure of the base standard, allowing the combination of the original standard and the new extensions to be treated as a single entity.

FGDC-STD-001-1998 is consistent with the ISO 19115 Metadata Standard. ISO has also a recommendation to establish a remote sensing extension, which is later known as ISO 19139. This ISO standard is similar to FGDC-STD-012-2002.

The Indonesian National Geospatial Data Clearinghouse adopts the protocol Z39.50 as its data network protocol. The use of this protocol enables each data producer to share its own spatial data through the clearinghouse data server using a Z39.50 supported client.

The main purpose of this clearinghouse is providing a catalogue system of Indonesian spatial data. It uses a geographical coordinate system (i.e. latitude and longitude coordinates in decimal degrees) for referencing data to a geographical area.

5.2 Input and Output Specifications

As mentioned earlier, there are many formats and types of remote sensing data. Therefore, these data cannot be included directly in a spatial database management system. Data can be georeferenced or not georeferenced. In the first case they are usually referenced to geographical coordinates or UTM coordinates, in the later case they are un-rectified raw data. Remote sensing data in Indonesian National Spatial Data Infrastructure (IDSN) are stored mostly in ERDAS, ER-Mapper or ENVI file format, where for various spatial analyses the UTM projection is commonly used.

As a coordinating agency, Bakosurtanal needs a certain standard to maintain the collection of various format spatial data. These data can be unrectified or they have a certain projection. To make data collection understandable to the system, a strict standard rules should be applied. The standard rule applies to the structure of metadata, and also to the standard of attribute entries. The bounding box coordinates in longitude-latitude coordinate system, for example, is used to define the spatial properties of each dataset, even if the dataset is stored in a different coordinate system.

Bakosurtanal introduces a clearinghouse with FGDC metadata standard as a format for data interchange among data providers. The current clearinghouse system supports keyword query to search the spatial data collection and produces the list of datasets. It has a simple user interface, where user can input keyword strings for query, or select a list of metadata based on its category, such as air photo, satellite imagery or thematic maps. The output consists of a list of datasets with a short description (abstraction) and search keywords. By opening the metadata from this output, user can get more information about the data provider to get the actual spatial data. This query system uses text as its input parameter and produces a list of metadata. No graphical output has been provided yet.

By November 2009, there were only less than 20 metadata for raster datasets available[41]. This very limited data shows the early stage in the development of raster data management in Indonesian National Geospatial Data Clearinghouse.

This research uses current Indonesian geospatial catalogue system, and the proposed system will use the same format of metadata as used by this clearinghouse. All fields in currently available metadata files are kept unchanged to ensure the compatibility of the proposed system with the existing one. An access to the original dataset is not necessary to be established within this research.

It is very useful to define a rule or strategy for dataset name and keep the filename unique, for example by including the processing level, area code, and a counter to produce a unique name for each dataset. We will use the template name DB0-xxxxxxxx, DB1-xxxxxxxx, DB2-xxxxxxxx and DB3-xxxxxxxx for dataset name in our conceptual design.

5.3 Conceptual Design

In many surveying and mapping institutions, the use of remote sensing data grows rapidly. A catalogue system has been introduced to organize remote sensing data and other raster data. To enable the organization using a catalogue system, the following characteristics have been identified:

- Each of remote sensing data or raster data including its metadata is identified as a dataset. In a catalogue system, dataset is usually represented by a metadata file.
- A collection is formed from all datasets which are organized within a catalogue system
- Each dataset has a unique identity. This unique identity is usually implemented by assigning a unique file name for each dataset in the collection; therefore a naming rule is required.
- Metadata files are stored locally at the same place with the catalogue system while their corresponding raster data can be distributed in other places or other

storage media. For this purpose, a link is established in a metadata to locate its original location.

- Catalogue system provides user query to through the computer network or internet and provide a list of dataset which suit certain characteristics as requested.
- Catalogue system may contain datasets with various type of sensor and various level of processing. We use the term heterogeneous collection for a collection like this.

Each dataset in a collection is managed as an independent entity by current catalogue systems. It is a challenge to find and organize the relationship among datasets if any. There are several perspectives on how one dataset can be related to the other. From the image processing perspective, the similarity of image contents can be used as a parameter to find a relationship. From the data provider perspective, the information about airstrip for airphotos or satellite tracks for satellite imagery can be used as an indicator that one dataset is related to the other. In a heterogeneous collection, the processing steps of creating one dataset from the others can be used to find a relationship between two datasets.

Reconstructing the creation history of all datasets in a collection is possible only if the processing step of the creation of a dataset is recorded. Recording processing steps in a collection can be done in a separate file or it is recorded in each corresponding dataset. The proposed method will use the information of the creation process which is recorded in each. The creation process information contains the list of used datasets, also called source datasets, among others. Although this list is limited to the list of direct sources of the corresponding datasets, the reconstruction as proposed by this research will produce a complete list of direct and indirect sources of each dataset. Finding the sources of each dataset opens the possibility to reconstruct the processing steps among existing datasets

This proposed system requires a data structure to represents the processing step among datasets in the collection. This structure should support the following outputs:

a. For each dataset, a list of all dataset which are involved directly or indirectly in **creating** that dataset will be created,
b. For each dataset, a list of all dataset which are created directly or indirectly **using** that dataset will be created,

[41] Result from Clearinghouse query with the category “Citra“ (Image)

c. To support a future enhancement, several attributes are recommended to be supported. These attributes includes: bounding box coordinates, cloud covers, band / wavelength information, spatial pixel resolution, platform and instrument information, processing algorithm, and thematic layer information. These attributes are supported by FGDC Standard Metadata and ISO 19115 Standard.

As mentioned before, many spatial catalogue systems exist currently, which provide variety of feature for user queries. The following discussion will explore the existing catalogue systems to find whether they support any kind of relationship among datasets.

5.3.1 Features of Existing Solutions

Remote sensing data providers use catalogue system to provide information about their data collection to public. The detail of information in the catalogue system and how they are presented are varies from one catalogue system to the other. Most of the existing catalogue systems facilitate user queries through internet and create a list of data which meet the given criteria. User can refine the query to find a specific data, or choose from a list based on the detailed information of the data. The detailed information of a selected dataset is usually used to determine whether this data fulfils the user need. As an alternative to the selected data, a catalogue system can also provide a list of other data which are similar to the selected one based on a certain criteria, for example a similarity in spatial coverage.

When a data has been selected from a collection, it is a challenge to evaluate the previous usage of that data if it was used to create other data, or evaluate the creation process of it if it is created from other data. To enable this evaluation, a catalogue system should be able to maintain the relationship among data within the collection based on the creation process of each data.

This section focuses on our evaluation of the currently available geospatial catalogue system. Several popular online spatial databases or remote sensing catalogue systems have been chosen based on their state-of-the-art method of organizing remote sensing data. We try to find whether the existing solution consider the importance of relationship among datasets if any.

5.3.1.1. Microsoft TerraServer

The TerraServer-USA Web site[42], formerly Microsoft TerraServer, is claimed as one of the world's largest online databases. It provides free public access to a vast data store of maps and aerial photographs of the United States. TerraServer is designed to work with commonly available computer systems and Web browsers over slow speed communications links. Maps and images in Microsoft TerraServer are supported by USGS. The word “Terra” in its name refers to the 'earth' or 'land' and also to the terabytes of images stored on the site.

TerraServer system core consists of MS Server 2003, MS SQL Server 2000, MS .NET Framework, and Open GIS implementation of Web Map Server.

Users can only perform queries based on coordinate values or address keywords, which both represent locations, and they can choose between two options of image: aerial photos or topo-maps. A commonly used graphical user interface allows setting of the zoom level or panning the selected area. The displayed image on screen is the actual output which can be downloaded or ordered.

This solution is very limited since it can only address queries based on location and the output is just an image displayed on screen, similar to Google Maps. There is not any list of remote sensing data or metadata displayed as result.

GeoFUSE[43] from GeoEye also provides the similar way of Image Catalogue as TerraServer as explained above. GeoFUSE uses Google Maps and Google Earth as its platform for online map display, and ESRI ArcGIS Server as its geo-database engine. This catalogue supports image queries based on specified location as in TerraServer, but it provides a list of remote sensing as an output, from which the user can get more details and image preview.

5.3.1.2 NOAA Satellite Imagery catalogues

The National Oceanic and Atmospheric Administration (NOAA) Comprehensive Large Array-data Stewardship System (CLASS) is NOAA's premier on-line facility for the distribution of NOAA and US Department of Defense (DoD) Polar-orbiting Operational Environmental Satellite (POES) data and derived data products. CLASS is operated by

---

[42] http://terraserver-usa.com/
[43] http://geofuse.geoeye.com/landing/, last accessed 1 Oct 2009

the Information Processing Division (IPD) of the Office of Satellite Data Processing and Distribution (OSDPD), a branch of the National Environmental Satellite, Data and Information Service (NESDIS).

CLASS[44] formerly named as NOAA Satellite Active Archives, provides spatial catalogue system for many remote sensing data including NOAA AVHRR, SAR, and GOES. Query is performed using a common graphical user interface, with additional search keys and options. The output of this catalogue is a list of remote sensing data. User can download the data, or order it.

This catalogue system has an advanced feature through which the user can interact or access directly satellite imagery. Search facility in this catalogue system can answer the most of the user's needs. Typically, three common questions are supported: *where* (spatial boundary), *when* (time frame), and *what* (data type, satellite code, direction of acquisition).

Since most of the provided data are raw satellite imagery, this catalogue system has no option to organize relationship among their dataset collections. In raw images, only overlapping between two adjacent datasets can be used as a hint of a relationship. This catalogue system has a limitation that it does not provide metadata in its output.

5.3.1.3. USGS EarthExplorer

The USGS EarthExplorer[45] is an online spatial catalogue system, similar with NOAA CLASS as described above. Most of the user query facilities are similar with what provided by NOAA CLASS. The main difference between these two providers is their collections. USGS provides remote sensing data from several sources including Landsat imagery, Radar imagery, NOAA AVHRR, and aerial photography. User can not interact with metadata in this catalogue system.

5.3.1.4. FAO GeoNetwork

The catalogue system[46] developed by FAO gives 3 basic keys: *what, where* and *when*. User can browse selected topics including: administrative and political boundaries, agriculture, applied ecology, biological and ecological resources, climate, fisheries and aquaculture, forestry, hydrology and water resources, land cover and land use, population and socio-economy indicators, soils and soil resources, and topography. With

---

[44] http://www.class.noaa.gov
[45] http://edcsns17.cr.usgs.gov/EarthExplorer
[46] http://www.fao.org/geonetwork/srv/en/main.home

advanced options, user can activate a dialog screen which has options for spatial boundaries (bounding box), selection by county name, temporal selection, and also selection from a list of data providers. GeoNetwork categorizes this advanced dialog as *what*, *where* and *when* options. The "*what*" section provides user free-text query of image content. This entry will be used as a search key for any metadata entries. The "*where*" section provides user a spatial query. User can define the location-based query using coordinate input, regional name, or graphical browser. The "when" section provides a temporal query. There are several additional filters given on the bottom right of the query screen for more options.

As described above, this catalogue system support inter-operability with other data providers and form a network of spatial data catalogue. We find this method as an interesting platform to be explored. Moreover, FAO provides the use of this Catalogue System as an Open Source application, named GeoNetwork Opensource. Due to the open source characteristics, there is a chance to extend the functionality of this catalogue system with our proposed data organization. The following Figure 5.2 shows the user interface of FAO GeoNetwork.

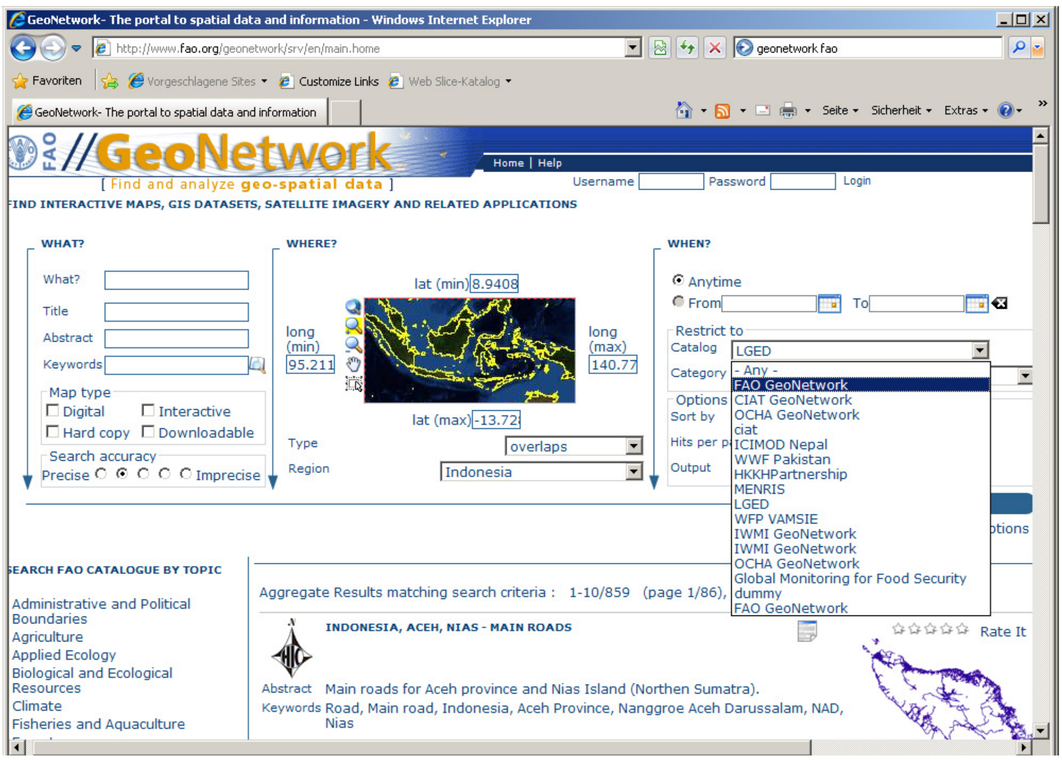


Fig 5.2: User interface of FAO GeoNetwork Catalogue (FAO)

This catalogue system is implemented based on image metadata which has its advantages and also limitation. The main advantage of developing a remote sensing catalogue based on metadata is the portability and broad semantic search keys, but it lost the capability to perform spatial analysis or image processing.

5.3.1.5. Comparison of Surveyed Catalogues

A possibility to reconstruct the relationship among datasets does not appear in the current catalogue system even if there are many parameters in metadata which can be used to reveal the relationship among dataset. Parameters which can be considered as hints of common senses among datasets are:

a. *Spatial coverage (outer polygon and bounding box).* This parameter is a most used parameter in dataset queries. It does not mention the relationship among the selected datasets, but it shows that the selected datasets have a common spatial coverage, and might help the user by narrowing down the number of datasets to be selected or analyzed.
b. *Time stamp or temporal information.* If there are too many datasets which resulted from query based on spatial coverage, filtering based on temporal information (time-stamp) may help user to focus on a certain time frame collections. This temporal information can also reveal the time-series datasets.
c. *Special Interest or Analyses Category*. High level logical information can be stored in metadata to specify the range of analyses which can be supported by the dataset, for example forestry, agriculture, geology, meteorology, urban, cadastral.
d. *Spectral band.* Although this filtering is not common, it might be used to limit the datasets to certain purposes of analyses. For example, for vegetation analyses red band and NIR band is required.

Current online remote sensing catalogue system or spatial data inventory system based mainly on above parameters. The popular remote sensing data source organizes the dataset collection based on type of analyses, spatial coverage and temporal interval. Other catalogue system introduces a simpler logical query; user should define “what” and “where” parameter to execute the inventory query. This “what” and “where” is another representation of type of analyses and spatial coverage.

Each output dataset of above catalogue system is an independent dataset which meets the search criteria. This is good enough for a company which produces remote sensing data, but it is not suitable for an organization body which coordinates the collection of survey and mapping data like Bakosurtanal. The currently implemented data clearinghouse by Bakosurtanal supports query based on keyword only, for instance:

region name. There is no possibility to reconstruct the creation process or relationship within the currently available collection.

From above four catalogues, only NOAA catalogue gives a clue of relationship by giving a footprint map of selected area containing all datasets which cover the study area, but those footprints (boundaries) are selected based on bounding box input, they are not related directly with the selected dataset.

5.3.2 Dataset Relationship and Level of Processing

As discussed in the previous section, there are several attributes or parameters which can be used to determine whether one dataset is related to another. Many institutions have a collection of remote sensing data which are ranged from raw data to derived data. For this kind of collection, a relationship among dataset can be defined based on the derivation or creation history. A relationship between two datasets based on the creation history exists if a dataset is used directly or indirectly in a creation process of another.

Creating a final map product from a series of raw data usually requires several steps of processing. For each of these steps an output data is created using certain analysis or algorithm. Considering its processing step, the un-rectified or raw data is not derived from other data thus assigned as the lowest level of data, and the final map product as the highest level of data. Further on, based on the processing level as introduced by Amhar (Amhar, 2001), raw dataset is classified as level-0, its rectified dataset is classified as level-1, an analytic or thematic map which are derived from rectified image is classified as level-2, and the final product or presentation map is classified as level-3.

The following Figure 5.3 is an example of how a presentation map (atlas) is usually produced. The naming structure for remote sensing data below is based on Indonesian National Spatial Data Infrastructure, as discussed in section 2.5.

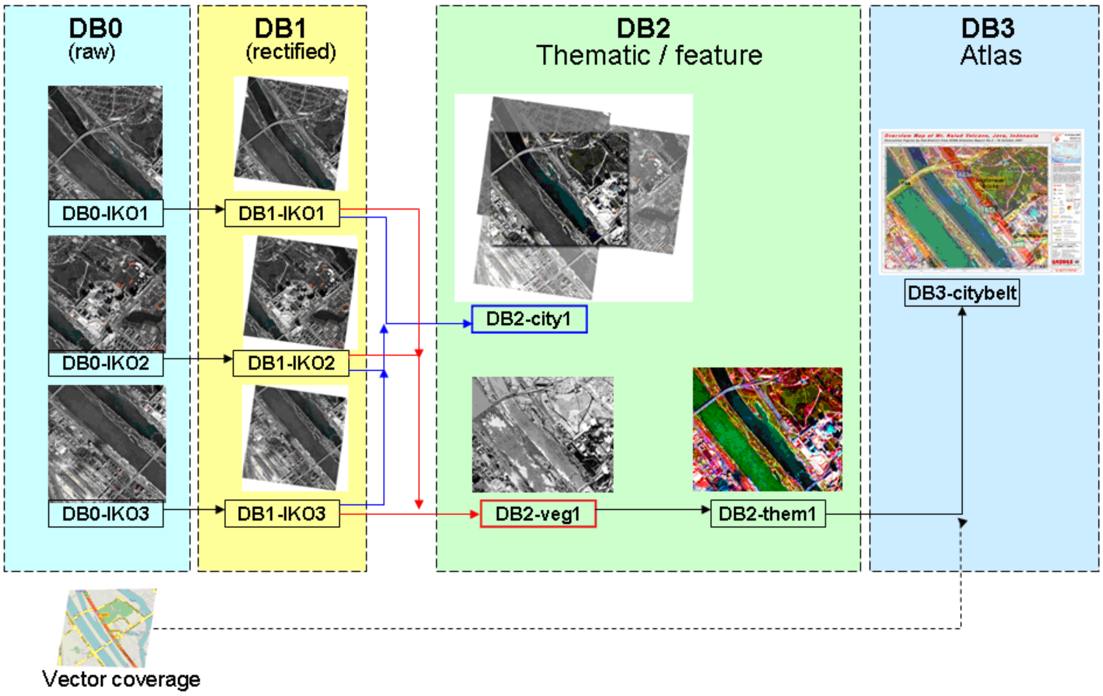


Fig 5.3: Datasets dependencies and processing diagram

This example shows the following datasets: There are a series of raw datasets for a certain study area, named DB0-IK01, DB0-IK02, DB0-IK03; and a vector coverage for geo-referencing. A rectified image is then created for each dataset. The corresponding rectified datasets are named: DB1-IK01, DB1-IK02, and DB1-IK03.

A mosaicking process is applied to above three rectified datasets to create a new rectified dataset with a rectangular coverage boundary, and its visible-color layer is selected. This new dataset is created for an urban map and named DB2-city1. Another output from this mosaicking process is also created using a combination of certain layers to produce a NDVI dataset named DB2-veg1 a shown in the above diagram.

From the DB2-veg1 dataset, a new thematic-map of vegetation classification DB2-them1 is generated. It is then combined with a vector data to create a new vegetation atlas named DB3-citybelt.

Arrow lines in the above diagram indicate the processing direction. The processing history is obtained by following the arrow lines in the opposite way. In practice, only processing history can be recorded in each dataset. Each dataset can have more than one source datasets. The following Table 5.1 shows the corresponding source lists of above datasets.

Table 5.1: Link table

Dataset	Source dataset	Remarks
DB0-IKO1	(none)	Raw dataset
DB0-IKO2	(none)	Raw dataset
DB0-IKO3	(none)	Raw dataset
DB1-IKO1	DB0-IKO1	1:1 geo-rectification
DB1-IKO2	DB0-IKO2	1:1 geo-rectification
DB1-IKO3	DB0-IKO3	1:1 geo-rectification
DB2-city1	DB1-IKO1, DB1-IKO2, DB1-IKO3	Example of mosaicking
DB2-veg1	DB1-IKO1, DB1-IKO2, DB1-IKO3	Example of NDVI analysis
DB2-them1	DB2-veg1	Example of GIS output
DB3-citybelt	DB2-them1	Example of Presentation Map

Based on above table, a backward tracing can be performed to trace back the history of each dataset. For example:

DB3-citybelt is created
from ***DB2-them1***,
from ***DB2-veg1***,
from (***DB1-IKO1***, ***DB1-IKO2***, and ***DB1-IKO3)***,
from (***DB0-IKO1***, ***DB0-IKO2***, and ***DB0-IKO3)***.

With this concept, a complete history or creation process of each dataset can be reconstructed without a complex structure. Each dataset refers, through its metadata, to just a list of source datasets used for its creation. The link to source datasets is recorded in metadata just one time when the dataset is created.

FGDC Metadata standard opens the possibility to include a dataset source list into a metadata. Moreover, in essence, metadata may answer who, what, when, where, why, and how about every facet of the data that are being documented. This means, we can insert more specific field into metadata to enhance the query and analysis. Information in metadata is extracted from many parameters by the data provider. These parameters can be considered as static information, consisting of spatial information, spectral information, and institution or data provider. These static parameters are very useful as input parameters in image query or catalogue system.

5.4 Data Model for Dataset Relationship

The conceptual design of dataset organization as discussed in the previous section requires a data model to represent the processing steps which are extracted from source links information in the metadata. The main properties of a processing step are the existence of a pair datasets and a flow of processing between them. The following characteristics are considered to model processing steps within a collection:

- One dataset can have one or more source datasets
- One dataset can be used as a source for creation of one or more datasets
- Only a maximum of one processing step may exist between a pair of datasets.
- A processing step represents the flow of creation process from a source dataset to its corresponding derived dataset. This flow has a direction but there is no value assigned to it.
- The flow of creation process is usually exist from a lower level of dataset to a higher level of dataset, and from an older dataset to a newer one, therefore a cyclic or recursive flow does not exist.
- Only the list of used datasets (source datasets) exists in the metadata of each dataset.

By representing a node for each of dataset in a collection as given in the Figure 5.3 and drawing all the source links as listed in Table 5.1, the complete creation history within a collection can be reconstructed. Figure 5.4 shows the flow of creation history within this collection.

Each node in the following Figure 5.4 represents a dataset, and each arrow represents a pointer to a source dataset. There exist datasets which are not pointed by other dataset, which are DB2-city1 and DB3-Citybelt in above figure. Each of these nodes is visualized like a root in a tree data structure. There exist datasets which do not point to other datasets, which are DB0-IKO1, DB0-IKO2. Those nodes are visualized like leaves in a tree or graph data structure.

Using graph data representation, each dataset can be assigned to a *node*, and a relationship to other raster data can be assigned to a *vertex*. This relationship represents the pointer to source dataset(s), therefore only one direction exists.

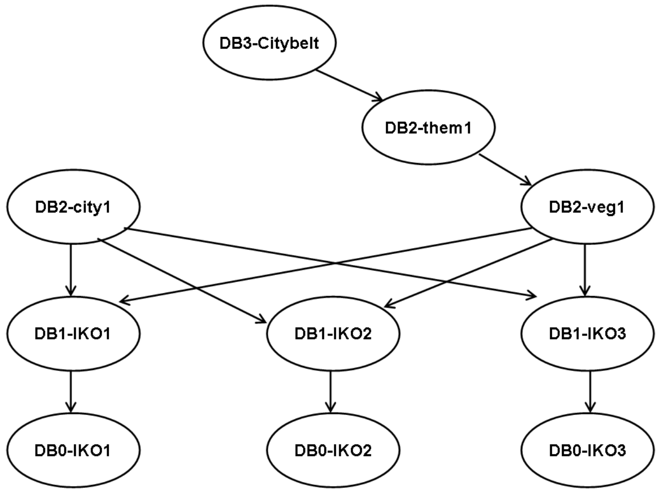


Fig 5.4: Representation of the creation history within a collection

Tree data structure is a graph structure where no recursive path is allowed, which is also one of the characteristic of the relationship to be modelled. The next section will discuss the tree data structure and evaluate the possibility to use it as a data model to represent remote sensing data collection with its relationship. Another data structure which should be discussed more in detail is a non-recursive directed graph or directed acyclic graph (DAG).

5.4.1 Tree Data Structure

The following discussion focuses on the basic rules of a tree data structure (Cormen et al., 2000, pp.200-218; Knuth, 1997), and how it can be applied to the organization of image dataset.

Theorem 1:

A node is a structure which contains a certain value. Each node in a tree has zero or more child nodes below it. A node that has a child is called the child's parent node (or ancestor node, or superior). A node has only one parent at most.

Each dataset in our image organization can be considered as a node, and each source of this dataset is considered as its child node. Each dataset has zero or more child nodes like a tree structure, but in our case, each node can have more than one parent which is **not suitable** with this definition of tree.

Theorem 2:

Nodes always have references to their children, but not always to their parent.

Since each dataset has only references to their children (sources), not to their parent (output dataset), the theorem is fulfilled

Theorem 3:

Nodes that do not have any children are called leaf nodes or they are also referred to as terminal nodes.

Raw datasets have no source data. In our case, raw datasets are terminal nodes. Theorem 3 is fulfilled.

Theorem 4:

The height of a node is the length of the longest downward path to a leaf from that node. The height of the root is the height of the tree. The depth of a node is the length of the path to its root

Reconstruction of the processing step as shown in Figure 5.4 connects the topmost dataset to its direct source(s), and also to its indirect sources which finally reach the raw datasets. The number of steps within a path from topmost dataset to its raw dataset can be considered as the height of the node. Calculating the height of a node is important to know the maximum length of the path of creation process. Therefore, theorem 4 is fulfilled.

Theorem 5:

The topmost node in a tree is called the root node. Being the topmost node, the root node will not have parents. It is the node at which operations on the tree commonly begin (although some algorithms begin with the leaf nodes and work up ending at the root). All other nodes can be reached from it by following edges or links. (In the formal definition, each such path is also unique). In diagrams, it is typically drawn at the top. In some trees, such as heaps, the root node has special properties. Every node in a tree can be seen as the root node of the sub-tree rooted at that node.

There may be many datasets in the data collection which are considered as topmost nodes. DB2-city1 and DB3-Citybelt nodes in Figure 5.4 are examples of topmost nodes within the collections. Considering the theorem 5, our data collection **does not** suit tree structure.

As a conclusion, our organization of raster datasets does not fulfil three theorems (1,4,5) for a tree data structure. Thus, this data structure can not be used in our case.

5.4.2 Directed Acyclic Graph

A ***directed graph*** or also known as ***digraph*** is a pair (V, E), where V is a finite set representing vertices (also called nodes), and E is a binary relation of V which represents adjacency of that relation, usually named as edges (Cormen et al., 2000). An edge E (x,y) in digraph has a direction from node x to node y. For node x, this edge is an outbound edge, and for node y, this edge is an inbound edge. A ***directed acyclic graph*** (**DAG**) is a directed graph with no cycle. Directed acyclic graph always has a node which has only outbound edge (known as source), and a node which has only inbound edge (known as sink).

All of the properties of DAG as explained above are also exist in the relationship among dataset of remote sensing data collection, based on their creation history. DAG always has a source node which is exactly represent the raw dataset, and DAG always has a sink node which is represent the final output product, for example a presentation map. The creation process in a remote sensing collection has a direction from source node to its derived node which is also suit the edge of DAG. Therefore, this DAG is used to model the dataset relationship.

Several algorithms which are commonly used in directed acyclic graph data structure can be applied to organize dataset history, for example topological ordering, traversal, and shortest path (Bang-Jensen and Gutin, 2008). In the following, they are discussed in more detail.

Topological ordering

Topological ordering of a directed acyclic graph (DAG) is a linear ordering of its nodes in which each node comes before all nodes to which it has outbound edges. Every DAG has one or more topological ordering.

An early method of topological ordering, introduced by Kahn (1962), works by finding a list of "start nodes" which have no incoming edges and inserts them into a set. This method is also used to verify whether a graph is acyclic or not. An alternative algorithm, known as depth-first search, introduced by Tarjan (1976), is used to create a linear list from a directed acyclic graph by looping through each node of the graph, in an arbitrary order, initiating a depth-first search that terminates when it hits any node that has already been visited since the beginning of the topological sort. The depth-first-search algorithm can be listed as follows (Corrnen et al., 2000):

```
L ← Empty list that will contain the sorted nodes
S ← Set of all nodes

function visit(node n)
    if n has not been visited yet then
        mark n as visited
        for each node m with an edge from n to m do
            visit(m)
        add n to L

for each node n in S do
    visit(n)
```

This algorithm can be used to find the highest level of dataset node in the dataset history.

Traversal

Traversal is the facility to move through a structure visiting each of the vertices once in order to obtain a list of all nodes. This algorithm can be used to calculate all paths that direct to a selected node, and also all paths that come from there.

In our data structure, we use traversal to list all the related nodes of a selected node (i.e. directly and indirectly used). By creating a topological order as discussed above, we can make a list of nodes, which efficiently subtract related nodes from the collection into a short list, and perform a sub-list (traversal) within this list.

Shortest path

If digraph is weighted (or known as weighted digraph) then shortest path from a given starting vertex to certain selected node can be calculated, for example using the Dijkstra’s algorithm.

Calculating shortest path is required if more than one paths exist between a pair of nodes. This calculation is designed to be used in a weighted graph, where each edge has a certain value and this value is interpreted as a distance between two nodes. Although finding a shortest path is relevant the dataset relationship, no weighting value is assigned to every edge. An edge in is used to represent a pointer to the source dataset, and this pointer does not has any weighting value. All edges can be considered to have a length of one forming an un-weighted directed acyclic graph.

Network

A directed graph may be used to represent a network of process; in this formulation, data enters a process through its incoming edge(s) and leaves the process through its outgoing edge(s), which could be the incoming edge(s) for subsequent process(es).

Processing remote sensing data to create a thematic map has a similar routine as above concept. Main input for remote sensing processing is a raw data from remote sensing sensor, e.g. Satellite imagery or airphoto. One or more raw image(s) used as input, and processed into one or more intermediate output, and then processed to create a final output. However, in organizing remote sensing data, we do not consider the process relation as mentioned above as a dynamic process. In a common network using digraph, an algorithm is used to re-calculate the processing inside network, if one or more input parameter is changed. In our design, we do not consider the “weight” of each relationship, therefore nothing will be changed in our system if certain information in node is changed, except if the structure is changed (i.e. link to other node).

5.5 Data Structure Implementation and Its Complexity

From the illustration of relationship among datasets in figure 5.4, the appropriate data structure is needed to represent the relationship. Two common data structure models can be implemented to represent that relationship, which are adjacency matrix and adjacency list. Adjacency matrix, for example (Cormen et al, 2000, pp. 527) , is a two-dimensional matrix, where each entry for every cell in the matrix requires only one bit representation, for example “1” for related, and “0” for not-related. The adjacency list is a list of existing edges only. This data structure consists of a list of vertex pairs if they are related in the graph representation. Therefore, each relationship requires a pair of containers / fields to record the vertex names.

5.5.1 Adjacency Matrix versus Adjacency List

Adjacency matrix and adjacency list data structures are used to represent a graph. The following discussion will compare these two data structure and evaluate which one suit the dataset relationship graph.

Adjacency matrix

From a relationship graph as shown in Figure 5.4, an adjacency matrix can be built as shown in the following Figure 5.5. One should keep in mind that the relationship graph is a directed graph.

As shown in Figure 5.5, raw datasets (DB0x) have no source dataset, which make all cells for DB0x rows have values of zero. Rectified dataset (DB1x) usually has only one source dataset among all dataset collections which produce only one cell with a value of one among others. All datasets which are not used by any other dataset have their column filled with all zeroes (e.g. DB2-city and DB3-Citybelt). The directed graph structure makes the upper triangle of the matrix filled with all zeroes. All of these conditions show that the matrix is sparsely filled, thus the use of adjacency matrix is very inefficient.

USED DATASET

DATASET

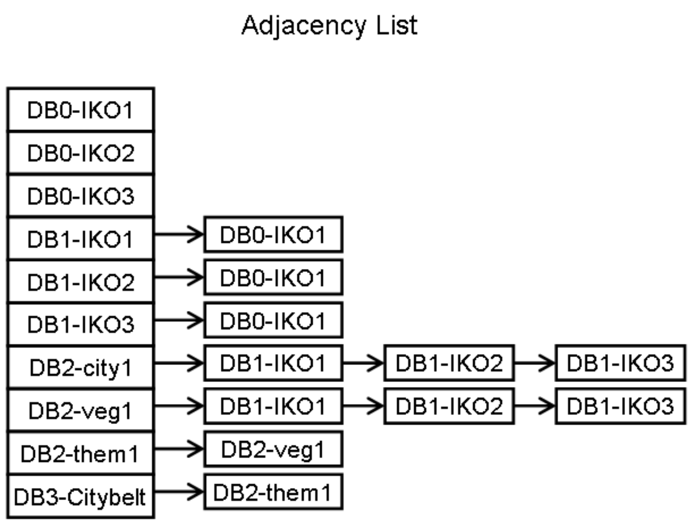
Used / By	DB0-IKO1	DB0-IKO2	DB0-IKO3	DB1-IKO1	DB1-IKO2	DB1-IKO3	DB2-city1	DB2-veg1	DB2-them1	DB3-Citybelt
DB0-IKO1	0	0	0	0	0	0	0	0	0	0
DB0-IKO2	0	0	0	0	0	0	0	0	0	0
DB0-IKO3	0	0	0	0	0	0	0	0	0	0
DB1-IKO1	1	0	0	0	0	0	0	0	0	0
DB1-IKO2	0	1	0	0	0	0	0	0	0	0
DB1-IKO3	0	0	1	0	0	0	0	0	0	0
DB2-city1	0	0	0	1	1	1	0	0	0	0
DB2-veg1	0	0	0	1	1	1	0	0	0	0
DB2-them1	0	0	0	0	0	0	0	1	0	0
DB3-Citybelt	0	0	0	0	0	0	0	0	1	0

Fig 5.5: Implementation of an adjacency matrix

As discussed in (Cormen et al., 2000, pp 217), adjacency matrix is designed to store a graph structure in a compact way, where each edge can be represented with a binary digit. However, it has also several disadvantages or even limitations which make it unsuitable for our purpose. If an edge has a certain weighting or embedded value, a binary representation is not suitable. The implementation of binary value for each cell requires a complex processing in current computer architecture where the smallest data container is usually byte or word. Furthermore, the extraction of the adjacency of one vertex requires an examination of all values in one row, again causing some inefficiency.

Adjacency list

An adjacency list represents a graph by keeping a list of every existing edge only. The following Figure 5.6 shows the implementation of an adjacency list for the graph in Figure 5.4. There are eleven edges in this example graph which requires also eleven records in an adjacency list. Each of this record contains a pair of nodes which are connected by the corresponding edge.



Adjacency Table

#	Source	Target
1	DB1-IKO1	DB0-IKO1
2	DB1-IKO2	DB0-IKO2
3	DB1-IKO3	DB0-IKO3
4	DB2-city1	DB1-IKO1
5	DB2-city1	DB1-IKO2
6	DB2-city1	DB1-IKO3
7	DB2-veg1	DB1-IKO1
8	DB2-veg1	DB1-IKO2
9	DB2-veg1	DB1-IKO3
10	DB2-them1	DB2-veg1
11	DB3-Citybelt1	DB2-them1

Fig 5.6: Implementation of adjacency list

While adjacency matrix has always n x n matrix, adjacency list has a maximum number of n(n-1)/2 elements in case all nodes are connected to each other in a directed graph. If the number of source for each dataset is relatively small compared to the total number of datasets, then adjacency list will be more efficient than adjacency matrix.

Adjacency list has also an advantage in finding relationship among vertex since it records only list of existing edges. This functionality is a key purpose in our data handling strategy. From the evaluation of values in the adjacency matrix of our sample graph (see Figure 5.5) and the number of records of adjacency list (see Figure 5.6), It is clear that the adjacency list has more advantages to be implemented.

Figure 5.6 above shows also the table structure of the implementation of adjacency list. In the basic concept of an adjacency list, a single linked list is used for keeping all edges of a vertex. Alternatively this structure can be implemented by storing each edge as a record in a table, thus make it a linear list. Using a table for storing an adjacency list has another advantage. If the edge has additional attributes or properties, the table can be extended by adding more fields.

Multiple Disjoint Graphs

In literatures (Cormen et al., 2000; Knuth, 1997, pp. 309), the forest data structure is introduced for multiple trees which have their own root nodes. In our case, the directed acyclic graph is used, but multiple graphs may exist in the collection. The example of directed acyclic graph as shown in Figure 5.4 may be just one part of many disjoint graphs of that kind, forming a set of graphs similar to the forest data structure.

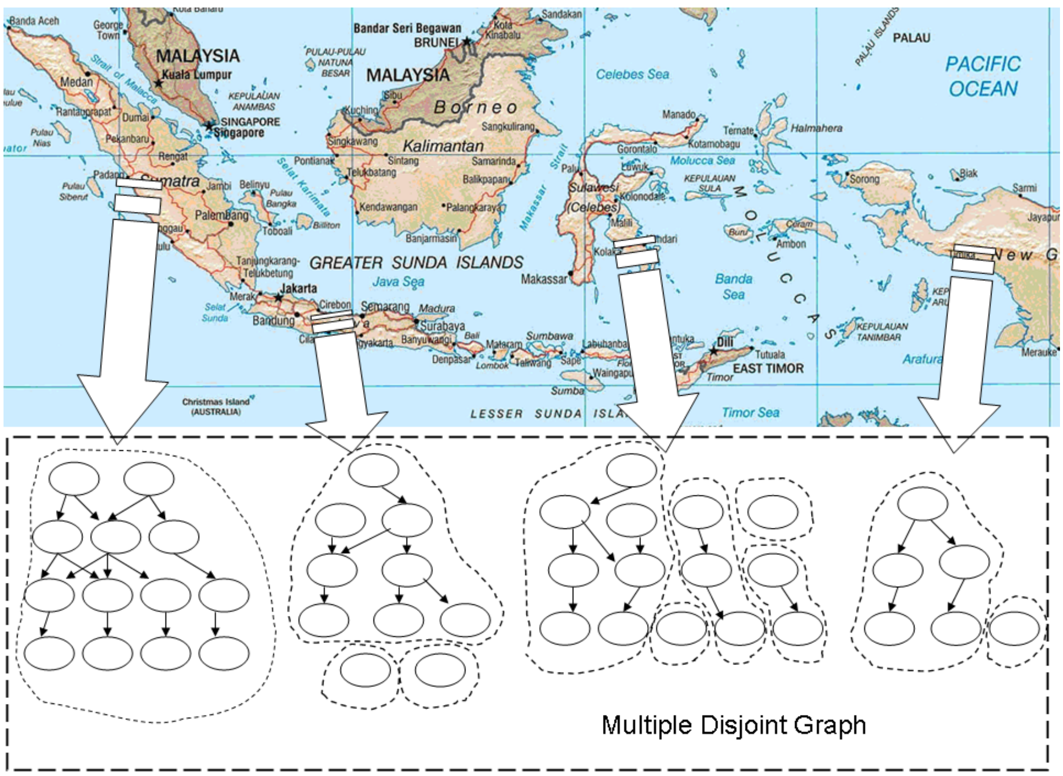


Fig 5.7: Collection is divided into multiple disjoint graphs

The collection may be partitioned into many dataset groups if there is a separate usage of datasets. Separation or partitioning of datasets into groups is usually based on a separation of spatial or temporal coverage. Each dataset which is not used by other dataset forms a disjoint graph also.

In an archipelago country like Indonesia, a continuous remote sensing coverage for all islands throughout the country is not possible. Continuous or neighbouring spatial data coverage is usually limited to each island. The above Figure 5.7 shows how multiple disjoint graphs are required to cover an archipelago country like Indonesia.

5.5.2 Analysis of Complexity

The adjacency list as described in the previous section has been chosen to handle the relationship among datasets. This adjacency list, as shown in Figure 5.6, represents a direct relationships (two datasets which are connected directly), and a linear table is used to record each of those relationships. The number of records which is required to store or manage this adjacency list is equal to the number of links in adjacency list. Using the topological ordering algorithm, a list of indirect relationships (two datasets which are connected through one or more other datasets) for each dataset can be derived. These indirect relationships can be inserted into the table to cover the complete list of history of every dataset in a collection. This section will discuss the complexity of data handling by this method. This investigation needs the maximum number of records which are required to store or manage the relationship in a linear table. For this theoretical "worst case" scenario, the following considerations and assumptions are used:

- The data collection contains multiple disjoint graphs which represents a collection of datasets with their history links.
- Datasets are divided into 4 levels based on their processing level (see section 5.3.2).
- Level 0 datasets (raw data) have no source dataset.
- Each level 1 dataset has one source dataset.
- Level 2 and level 3 datasets have the same number of source datasets.
- For each dataset, a list of all direct relationships (direct sources) and indirect relationships (indirect sources) will be recorded

The following notations are used for the calculation:

A = total number of level 0 datasets
B = total number of level 1 datasets
C = total number of level 2 datasets
D = total number of level 3 datasets
n = number of source datasets for each dataset of level 2 and level 3
R = maximum number of records

The maximum number of records R, which is the total number of direct plus indirect relationship, is calculated using the following rules:

- In level 0, datasets are raw datasets which have no source dataset. Therefore, the number of records in level 0 is 0.
- In level 1, each dataset has a one source dataset (which is its raw dataset), and there are B datasets. Therefore, the number of records in level 1 is B.
- In level 2, the total number of datasets is C, where each dataset has a maximum of n direct links to its source dataset in level 1. From each of its level 1 source dataset, there is one link which points to its source. Therefore, each level 2 dataset has n direct sources and n indirect sources, and the number of records in level 2 is C (n+n) or 2Cn.
- In level 3, the total number of dataset is D. Each dataset has n direct links to its source datasets in level 2, and each of the level 2 dataset has (n+n) source links which are recorded as indirect links. Since each of the level 3 dataset has n level 2 datasets, the number of its indirect records is n (n+n) or $2n^2$. The number of records in level 3 is D (n + $2n^2$) or Dn + $2Dn^2$

The maximum number of records R is the total number of records in all levels:

$$R = Dn + 2Dn^2 + 2Cn + B + 0$$

The following scenario will be used to estimate the maximum number of records if this method is implemented to handle the remote sensing data collection of Indonesia.

Total land area of Indonesia is 1.9 million km^2 with urban areas cover about 15% of it. Most of the volume of remote sensing data covers land area. Each remote sensing data is assumed to have an overlapping area of about 25% with its neighbours. The following Table 5.2 shows a rough estimation of the total number of remote sensing images which are required to cover Indonesia. To cover the land area with 20-30 meter resolution images such as Landsat or SPOT, a composition of more than 600 images are required. To cover details of urban areas with 1-2 meter resolution images, a composition of more than 3,600 images are required. This estimation is calculated for land cover mapping purpose only. The complexity will be higher for supporting other type of mapping, such as oceanic mapping, digital elevation model, weather analysis, dynamic remote sensing acquisition (e.g., weather analysis, crop simulation), and temporal analysis.

Table 5.2: Number of images to cover Indonesia

Coverage + 25%	Area (km2)	Number of Images
Land Area	2,375,000	600 *)
Urban Area	356,250	3,600 **)
TOTAL		4,200

*) Using SPOT, 60 km x 60 km coverage area, 20 m resolution

**) Using Ikonos, 10 km x 10 km coverage area, 1 m resolution

The total of 4,200 images from above table shows the estimation number of remote sensing images to cover the Indonesian archipelago for land cover mapping purpose using two types of sensor. The remote sensing data collection may contain also the multi-temporal data and data from other sensors like aerial photography. Using this consideration, a total number of 10,000 is realistic. This number of dataset is used as estimation for raw data (level 0). The number of dataset in level 1 (rectified data) is assumed to be the equal to 10,000, and this number is used as a maximum number of datasets in level 2 and level 3. Assuming the average number of four source datasets for each dataset in level 2 and level 3, the maximum number of records is:

$$
\begin{aligned}
R &= Dn + 2Dn2 + 2Cn + B \\
&= 10{,}000*4 + 2*10{,}000*4^2 + 2*10{,}000*4 + 10{,}000 \\
&= 40{,}000 + 320{,}000 + 80{,}000 + 10{,}000 \\
&= 450{,}000
\end{aligned}
$$

These 450,000 records of adjacency list is the worst case scenario for a collection with 40,000 datasets. Both 40,000 datasets (and their corresponding metadata) and the adjacency list table require a certain amount of storage. The following simulation will compare the size of original datasets collection, metadata collection, and the size of adjacency table.

Assuming the size of each of the remote sensing data as shown in Table 5.2 is 150 Megabytes; the total required storage for keeping these 40,000 data is about six Terabytes. The proposed remote sensing data handling uses metadata as input. With an average size of 15 kilobytes for each metadata file, the total size of the metadata collection is 600 Megabytes. Assuming a record size of 32 byte for storing the relationship in an adjacency table, the size of a history table for the above mentioned 450,000 records is 14,400,000 bytes (13.73 MB). This number is far below the limit of table size in database management system software. The following Table 5.3 shows estimation of the file size of each type of data type as discussed above.

Table 5.3: File types and their size

Type of File / Table	Number of Record / File	Unit Size	Total Size (MB)
Remote sensing original data	40,000	150 MB	6,000
Remote sensing metadata file	40,000	15 kB	600
Metadata master table	40,000	1 kB	40
History table	450,000	32 bytes	13.73

As shown in above table, the number of maximum records to handle the history of remote sensing data in Indonesia is much bigger than the number of datasets (remote sensing data), but its file size is much smaller which is feasible and efficient for implementation.

5.6 Optional Attributes from FGDC Metadata

The above described concept shows how history links and relationship among remote sensing datasets can be constructed using source list tags which are extracted from metadata. By managing several additional remote sensing key features, a better catalogue system or knowledge-based remote sensing data query can be developed. Considering the characteristic of remote sensing data and common image processing algorithms which are usually applied, several fields from metadata are chosen and extracted for further use:

Table 5.4: Additional attributes from remote sensing metadata

Field name	**Remarks**
Boundary (bounding box)	4 corner-point coordinate
Visible band	Exist or Not exist
NIR band	Exist or Not exist
FIR band	Exist or Not exist
SAR band	Exist or Not exist
Spatial Resolution	Value in meter
Cloud cover	In %
Platform and Instrumentation Information	Sensor information
Band Number	Value / code
Band wavelength	Value of wavelength
Thematic Layer	Description if exist
Pixel resolution	Pixel size in meter
Process date	Date & time of processing
Process description.	Phrase text (algorithm)

Although the implementation of this possible extension is not covered directly within this dissertation, we recommend the extraction of above additional fields and store them inside the database. It is also possible to search those fields directly from metadata files, but storing them in a database makes the searching more efficient.

5.7 Implementation

It is a challenge to prove the proposed method by developing a prototype of its implementation. The core part of this prototype is a database system which uses metadata as its data sources and provides the processing history of a selected data input. There are several choices of software environments which can be chosen to develop this prototype since the conceptual design is independent of any software providers. Web based applications are commonly used in geo-spatial catalogue systems or other system with internet based data sharing. The web-based application is an example of Service Oriented Architecture (SOA), where several type of services work together based on protocol and standard rather than using specific sharing codes (e.g. Windows API). To develop a prototype implementation of the proposed data handling, we incorporate several applications, as shown in a diagram below:

Source data for this proposed system is a collection of remote sensing metadata, which uses FGDC standard stored in a XML text file. We assume that the required attributes exist in the metadata file. Since the standard format for metadata and catalogue protocol is used, there exist solutions for creating and editing the metadata if required. An online solution named MERMAid (Metadata Enterprise Resource Management Aid), provided as a free service by NOAA.GOV, and USGS TKME Metadata Editor can be used for this purpose.

The core part of this prototype, as shown in Figure 5.8, is a geo-database and a web-based application. The geo-database is used to store and process data from remote sensing data collection, while the web-based application is used to provide user interface and user queries. A module for XML parsing is required to read attributes from metadata files and it stores the extracted attributes in the database record.

The following sections discuss the development of a prototype using the currently available geo-database. Firstly, we design and develop a prototype using an integrated software package ArcGIS. As one of a very popular and state-of-the-art GIS Solution, ArcGIS from ESRI supports all requirements that we need, i.e.: XML catalogue system, spatially enabled database system, graphical user interface, and programming language. Secondly, we design and develop a prototype using a set of open source software. This second prototype uses PostgreSQL as its database management system, a MapServer for Windows as its web-based map server, Apache[47] as its webserver and PHP scripts for its programming.

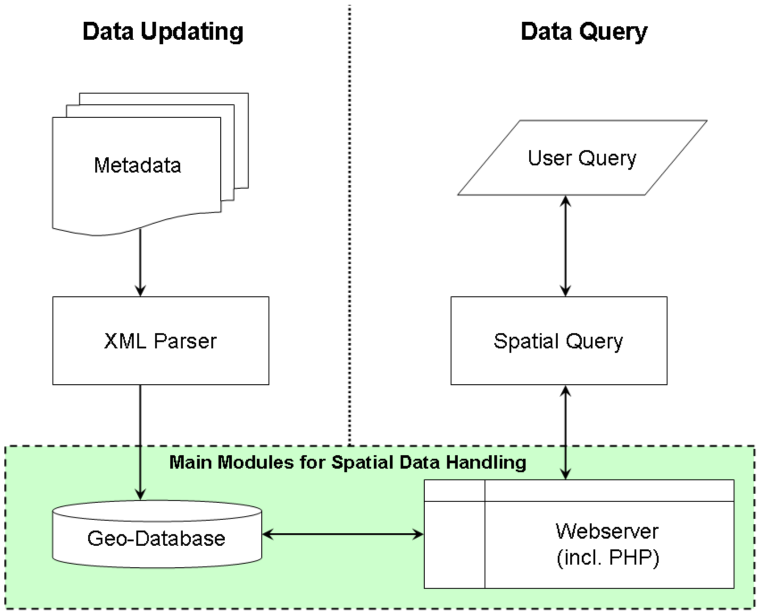


Fig 5.8: Main modules for the prototype

5.7.1 Prototype Design with ArcGIS

The geo-database in ArcGIS by ESRI is a collection of geographic datasets in various formats. Fundamental datasets for theArcGIS geo-database are: feature classes, raster datasets, and tables. ESRI provides several integrated tools and services within ArcGIS to develop a web-based geo-database including the catalogue system. The current version of ArcGIS, version 9.3 (as of October 2009), has a great conformance with OGC and ISO, which enables sharing data with other GIS systems. The following discussion is about the design and implementation of a geo-database and catalogue system with XML metadata in ArcGIS Version 8.1.

We incorporated several modules in designing and implementing our prototype with ArcGIS. Since our focus is organizing raster datasets using their metadata which are stored in XML format, we did not use the whole design concept of geo-database as

[47] The Apache Foundation, http://www.apache.org, last accessed March 2008

provided by ESRI. ArcCatalog is implemented as an application to deal with metadata. This application supports metadata ISO and FGDC standard, and also supports the search protocol Z39.50, which is also a requirement for our implementation.

To share the content of this implementation through the web, ArcIMS is used. Most all of the preparation steps and executions steps can be done using interactive user interface of ArcCatalog and ArcIMS server. This user interface can also be used to create additional relationship between datasets by updating dataset’s properties.

To customize the application, a module is created using Visual Basic programming scripts which are supported by ArcGIS. ESRI has a meta-language named ArcXML (or simply AML) as a protocol to communicate with ArcIMS Server. This XML-based script enhances the possibility to access and modify the configuration files of datasets, and enables more interactions with map visualizations.

Our prototype of remote sensing data handling can be set up using several modules which are integrated in or part of ArcGIS desktop package as listed above. From the evaluation of our prototype, ArcGIS has full support of spatial metadata handling and provides wide possibility of customization. On the other hand, there are several drawbacks which are also considered, as follows:

- ArcGIS is designed as a complete desktop GIS solution which supports sophisticated visualization. With an additional functionality of ArcIMS and ArcSDE to enable web-catalogue system, the system requirement is rather heavy compared to the dimension of the developed module. We do not use the functionality of ArcSDE in this prototype since spatial analysis and feature processing are not required.
- After a test module has been implemented, an independent module is required to migrate or integrate this solution into current development in Indonesia. This additional effort makes it unsuitable to develop this remote sensing data handling using ArcGIS geo-database.

5.7.2 Prototype Design with PostgreSQL

The second prototype uses open source software from several providers to implement all the required modules. Compared to our first prototype which is build using ArcGIS, this second prototype is rather complicated in its preparation. In ArcGIS, only a one-step installation is required to setup the environment before developing the prototype. In an open source environment, we should choose and install an application for each database system, spatial-module support for database, web-server, and web-map service.

For a general purpose web-enabled database system, a Linux-Apache-MySQL-PHP (LAMP) solution is commonly used. Another open source database system, PostgreSQL, has a spatial support using a spatially-enabled solution from PostGIS. With this reason, we decided to use PostgreSQL instead of MySQL. PostgreSQL Version 8.2.9 is used with PostGIS Version 1.3 under Windows XP Operating System as our geo-database solution. This open source geo-database works in Linux environment also for a complete open source system.

As introduced in section 5.4, a database is designed to handle remote sensing metadata files and to produce a table containing history links among datasets. Since all fields to be stored are fixed in size and type, a relational database is suitable. After database normalization, we have three main tables to handle the data:

- The first table is a master data of remote sensing dataset files, this table records: Dataset ID, Dataset Name, Dataset Physical Location, Dataset Boundary Coordinate, and other attributes.
- The second table is a relationship table which record the 1:N relationship among datasets.
- The third table is the complete history list, which is derived from the second table above.

The following Figure 5.9 shows tables structure of this database.

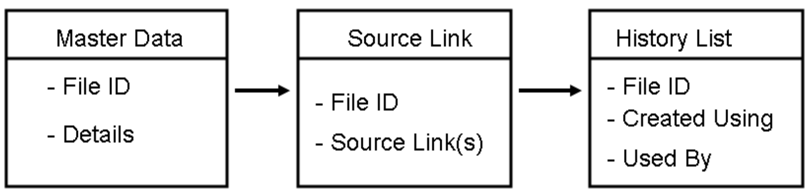


Fig 5.9: Geo-database table structure of the prototype

Several functions are defined for managing the data within the database, including:

Data initialization

This module initializes graph class and database, starts map server and prepares user interface to receive input from user.

XML parsing

This prototype uses existing remote sensing metadata from the currently available catalogue system provided by Bakosurtanal. The existing metadata can be used directly with this proposed system without any modification, and metadata created for this system can also be used by existing system without prior modification.

The following metadata structures are parts of FGDC Remote Sensing Extension metadata. Tags written in highlight are selected for our database implementation. Each of the tags may have sub-sections. The history link of each dataset is extracted from a tag named <SrcUsed>, which is located inside a sub-section named <ProcStep> (processing step). This <ProcStep> tag is located inside a sub-section named lineage which is located in *Data Quality Information* section. The following structure shows the hierarchy in the first level only.

Selected fields (written in boldface) in the Identification Information section:

Identification_Information =

Dataset_Identifier +

Citation +

Description +

Time_Period_of_Content +

Status +

Spatial_Domain +

0{Processing_Level}n +

Keywords +

0{***Platform_and_Instrument_Identification***}n +

[Band_Identification|

Thematic_Layer_Identification] +

Access_Constraints +

Use_Constraints +

(Point_of_Contact) +

(1{Browse_Graphic}n) +

(Data_Set_Credit) +

(Security_Information) +

(Native_Data_Set_Environment) +

(1{Cross_Reference}n) +

0{***Aggregation Information***}n

Selected fields (written in boldface) in the Data Quality section:

Data_Quality_Information =

0{Attribute_Accuracy}1 +

Logical_Consistency_Report +

Completeness_Report +

0{Positional_Accuracy}1 +

Lineage +

(Image_Quality) +

(Acquisition_Information) +

(***Cloud_Cover***)

Besides the history link which is extracted from the XML tag <Srcused> in metadata file, several fields are also extracted for the future use of this application. The following Table 5.5 shows selected objects within remote sensing metadata which are recommended for further implementation of remote sensing data handling.

Table 5.5: Target fields

Object name	**Field Name**	**Description**
Platform_and_Instrument_Identification	Plainsid	Satellite / Aerial Instrument detail
Band_Identification	Bandidnt	Contain Band_ID and Band_Measurement_Mode_ID
Band_ID	Bandid	Band code or number
Band_measurement_Mode_ID	Bmmodid	Band wavelength
Thematic_Layer_Identification	Thelayid	Number of kinds of geospatial information represented by the dataset.
Aggregation_Information	Agginfo	Describe the possible link to other datasets
Process_Date	Procdate	Temporal Information
Algorithm_Description	Algodesc	Describe the methode to develop this dataset
Source_Used_Citation_Abbreviation	Srcused	Link to source / previous dataset
Pixel_Resolution	Pixlreso	Ground resolution
West_Bounding_Coordinate	Westbc	Bounding box boundary
East_Bounding_Coordinate	Eastbc	Bounding box boundary
North_Bounding_Coordinate	Northbc	Bounding box boundary
South_Bounding_Coordinate	Southbc	Bounding box boundary

Graph creation

This part is considered as the core part of this prototype. In this step, we organize the relationship information which is extracted from *SrcUsed* tag into a directed graph and implement it using several tables in a database. Other tags are also read from the metadata. In this prototype, the recoded attributes are limited to dataset id, dataset name, dataset linkage, bounding box coordinate, and cloud cover value. Verification is process applied when importing a new metadata. Each metadata files should be imported only once to avoid inconsistency. The following program code shows the main module of this graph creation.

```
$dbTable = "dtset_link";

                    $dbField[0] = "link_id";
                    $dbField[1] = "link_nama";
                    $dbField[2] = "link_linkage";
                    $dbField[3] = "link_west_bounding";
                    $dbField[4] = "link_east_bounding";
                    $dbField[5] = "link_north_bounding";
                    $dbField[6] = "link_south_bounding";
                    $dbField[7] = "link_cloud_cover";

                    $sql = "select link_id from gis.dtset_link where
                            link_nama = ".QuoteValue(DPE_CHAR,$linkNama);
                    $rs = $dtaccess->Execute($sql,DB_SCHEMA);
                    $dataLink = $dtaccess->Fetch($rs);
                    $linkId = $dataLink["link_id"];

                    $isSave = ($dataLink["link_id"]) ? false : true;
                    if($isSave) $linkId = $dtaccess
                       ->GetNewID("dtset_link","link_id",DB_SCHEMA);
                    $dbValue[0] = QuoteValue(DPE_NUMERIC,$linkId);
                    $dbValue[1] = QuoteValue(DPE_CHAR,$linkNama);
                    $dbValue[2] = QuoteValue(DPE_CHAR,$hasilOnlink[0]);
                    $dbValue[3] = QuoteValue(DPE_NUMERIC,$hasilWestBound[0]);
                    $dbValue[4] = QuoteValue(DPE_NUMERIC,$hasilEastBound[0]);
                    $dbValue[5] =
QuoteValue(DPE_NUMERIC,$hasilNorthBound[0]);
                    $dbValue[6] =
QuoteValue(DPE_NUMERIC,$hasilSouthBound[0]);
                    $dbValue[7] =
QuoteValue(DPE_NUMERIC,substr($hasilCloud[0],0,-1));

                    $dbKey[0] = 0; // -- set key for  where clause
                    $dtmodel = new
DataModel($dbTable,$dbField,$dbValue,$dbKey,DB_SCHEMA);

                    if ($isSave) {
                         $dtmodel->Insert() or die("insert error");
                    } else $dtmodel->Update() or die("insert error");

                    unset($dtmodel);
                    unset($dbField);
                    unset($dbValue);
                    unset($dbKey);

                    // ---- insert history ----
                    $dbTable = "dtset_link_history";

                    $dbField[0] = "link_hist_id";   // PK
                    $dbField[1] = "link_hist_parent";
                    $dbField[2] = "link_hist_child";


for($i=0,$totParent=count($hasilSrcused);$i<$totParent;$i++){
                         unset($dataLink); unset($dataHist);
```

Input-output query

As introduced, our proposed system will create two types of lists: A list of datasets which are created using selected dataset, and a list of datasets which are used to create the selected dataset. For this purpose, the user interface of our prototype should give the user a possibility to choose the selected input. Assuming the collection may contain a lot of datasets and distributed through a wide geographic area; user interface in this prototype contains the option to localize the study area. This localization can be performed by using a zooming feature in the graphical user interface, or by defining a bounding box with coordinate values as its parameters. Based on the bounding box which is defined either from text-entry or map selection, a query will be performed based on coordinate values. The following Figure 5.10 shows the input query screen of this prototype. Several standard mapping features are also provided like zoom, pan and selection.

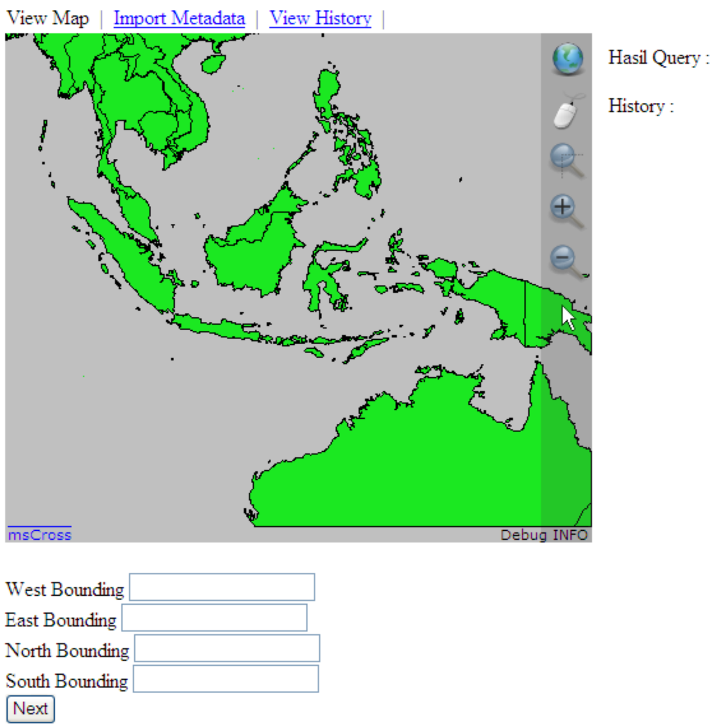


Fig 5.10: Client user interface of the prototype system

Dependent to the size of the given bounding box and the number of remote sensing data, which lie within the bounding box, a list of datasets will be created. The user chooses one dataset from this list to define the selected dataset input.

The following module shows the first level query, which display one or more dataset within selected geographic extend.

```
<?php
     require_once("root.inc.php");
     require_once($APLICATION_ROOT."library/xmlReader.cls.php");
     require_once($APLICATION_ROOT."library/dataaccess.cls.php");
     require_once($APLICATION_ROOT."library/datamodel.cls.php");

     $dtaccess = new DataAccess();

     if($_POST["btnProses"]){
          $sql = "select a.link_nama
                                   from dtset_link a
                                   left join dtset_link b on a.link_id =
b.link_id
                                   and (b.link_east_bounding <
".QuoteValue(DPE_NUMERIC,$_POST["west_bound"])."
                                   or b.link_west_bounding >
".QuoteValue(DPE_NUMERIC,$_POST["east_bound"])."
                                   or b.link_north_bounding <
".QuoteValue(DPE_NUMERIC,$_POST["north_bound"])."
                                   or b.link_south_bounding >
".QuoteValue(DPE_NUMERIC,$_POST["south_bound"])."
                                   ) where b.link_id is null";
          $rs = $dtaccess->Execute($sql,DB_SCHEMA);
          $dataLink = $dtaccess->FetchAll($rs);
     }
?>

<!DOCTYPE HTML "//-W3C//DTD HTML 4.01 Transitional//EN">
<html>
<head>
<TITLE>.:: Welcome ::.</TITLE>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>

<body>
<table width="100%" border="0" cellpadding="0" cellspacing="0">
    <tr >
        <td width="20%">
                      <a href="index.php">View Map</a>  |
                      Import Metadata  |
                      <a href="task.php">View History</a>  |
              </td>
    </tr>
</table>

<table width="100%" border="0" cellpadding="1" cellspacing="1">
     <tr>
          <td align="left" colspan=2 >Find Dataset within boundary</td>
     </tr>
</table>
<BR />

<form name="frmEdit" method="POST" action="<?php echo
$_SERVER["PHP_SELF"]?>">
       <table width="100%" border="0" cellpadding="0" cellspacing="0">
       <tr>
              <td>
                     West Bounding  <input type="text" size=5
name="west_bound" id="west_bound" value="<?php echo $_POST["west_bound"];?>"
/>    
                     East Bounding  <input type="text" size=5
name="east_bound" id="east_bound" value="<?php echo $_POST["east_bound"];?>"
/>    
                     North Bounding  <input type="text" size=5
name="south_bound" id="south_bound" value="<?php echo
$_POST["south_bound"];?>" />    
                     South Bounding  <input type="text" size=5
name="north_bound" id="north_bound" value="<?php echo
$_POST["north_bound"];?>" />    
```

```
                     <input type="submit" name="btnProses" id="btnProses"
value="Show DataSet" />
              </td>
       </tr>
       </table>
</form>

<br />
<?php if($_POST["btnProses"]){
       echo "Hasil Query : <BR>";
       for($i=0,$n=count($dataLink);$i<$n;$i++) {
              echo $dataLink[$i]["link_nama"]."<BR>";
       }

       if($n==0) echo "Not Found";
 } ?>
 <script>document.frmEdit.west_bound.select(); </script>
</body>
</html>
```

After one dataset is selected, the next module will calculate the history of the corresponding datasets, displaying the list of datasets, which are created using the selected dataset, and a list of datasets, which are used to create the selected dataset. The following module shows process displaying history lists for the selected dataset:

```
<?php
     require_once("root.inc.php");
     require_once($APLICATION_ROOT."library/xmlReader.cls.php");
     require_once($APLICATION_ROOT."library/dataaccess.cls.php");
     require_once($APLICATION_ROOT."library/datamodel.cls.php");

     $dtaccess = new DataAccess();

     if($_POST["btnDelete"]){
              $sql = "delete from gis.dtset_link where link_id =
".$_POST["link_id"];
              $dtaccess->Execute($sql,DB_SCHEMA);
              unset($_POST["link_id"]);
       }

     $sql = "select link_id, link_nama from gis.dtset_link order by link_id";
     $rs = $dtaccess->Execute($sql,DB_SCHEMA);
     $dataLink = $dtaccess->FetchAll($rs);

     if($_POST["btnProses"] && $_POST["link_id"]){
          $sql = "select gis.expand_link(".$_POST["link_id"].")";
          $dtaccess->Execute($sql,DB_SCHEMA);

          $sql = "select * from gis.dtset_link where link_id =
".$_POST["link_id"];
          $rs = $dtaccess->Execute($sql,DB_SCHEMA);
          $dataHist = $dtaccess->Fetch($rs);

          $sql = "select distinct child_name from gis.vdtset_predecessor
where id_link_root = ".$_POST["link_id"];
          $rs = $dtaccess->Execute($sql,DB_SCHEMA);
          $dataPre = $dtaccess->FetchAll($rs);

          $sql = "select distinct child_name from gis.vdtset_successor  where
id_link_root = ".$_POST["link_id"];
          $rs = $dtaccess->Execute($sql,DB_SCHEMA);
          $dataSuc = $dtaccess->FetchAll($rs);
     }
?>

<!DOCTYPE HTML "//-W3C//DTD HTML 4.01 Transitional//EN">
<html>
```

```
<head>
<TITLE>.:: Welcome ::.</TITLE>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>

<body>
<table width="100%" border="0" cellpadding="0" cellspacing="0">
    <tr >
        <td width="20%">
                     <a href="index.php">View Map</a>  |
                     <a href="import.php">Import Metadata</a>  |
                     View History  |
              </td>
    </tr>
</table>

<BR />

<form name="frmEdit" method="POST" action="<?php echo
$_SERVER["PHP_SELF"]?>">
       <table width="50%" border="0" cellpadding="0" cellspacing="0">
       <tr>
              <td align="left" width="20%"> Select Dataset</td>
              <td>
                     <select name="link_id" id="link_id">
                            <?php for($i=0,$n=count($dataLink);$i<$n;$i++){ ?>
                                   <option value="<?php echo
$dataLink[$i]["link_id"];?>" <?php
if($_POST["link_id"]==$dataLink[$i]["link_id"]) echo "selected"; ?>><?php
echo $dataLink[$i]["link_nama"];?></option>
                            <?php } ?>
                     </select>
                     <input type="submit" name="btnProses" id="btnProses"
value="View History Link" />
              </td>
       </tr>
       </table>
</form>

<br />
<?php if($_POST["btnProses"]){ ?>

<form name="frmDelete" method="POST" action="<?php echo
$_SERVER["PHP_SELF"]?>">
<?php echo $dataHist["link_nama"]; ?> was created using :
<?php
     for($i=0,$n=count($dataPre);$i<$n-1;$i++){
          echo $dataPre[$i]["child_name"].", ";
     }
     echo $dataPre[$i]["child_name"];
?>

<br />

<?php echo $dataHist["link_nama"]; ?> was used by :
<?php
     for($i=0,$n=count($dataSuc);$i<$n-1;$i++){
          echo $dataSuc[$i]["child_name"].", ";
     }
     echo $dataSuc[$i]["child_name"];
?>

<br />
Number of Used : <?php echo $n; ?>
<br />
Online Linkage : <?php echo htmlentities($dataHist["link_linkage"]); ?>
<br />
West Bounding : <?php echo $dataHist["link_west_bounding"]; ?>
<br />
```

```
East Bounding : <?php echo $dataHist["link_east_bounding"]; ?>
<br />
North Bounding : <?php echo $dataHist["link_north_bounding"]; ?>
<br />
South Bounding : <?php echo $dataHist["link_south_bounding"]; ?>
<br />
Cloud Cover : <?php echo $dataHist["link_cloud_cover"]; ?> %

<BR/>
<input type="submit" name="btnDelete" id="btnDel" value="Delete Link" />
<input type="hidden" name="link_id" value="<?php echo
$dataHist["link_id"];?>" />
</form>
<?php } ?>
</body>
</html>
```

5.8. Keys of Proposed System

This new methodology has been tested with the following environment and principles:

Environment

Type of source data: remote sensing metadata in XML file structure of FGDC Standard.

Approach used: directed acyclic graph, represented with adjacency-list, graph-traversal, topological ordering to find relationship or history of each dataset.

Computer environment: Windows XP, PostgreSQL, Apache, PHP, Mapserver for Windows

Pre-requirements

Each dataset, which has its processing step recorded in metadata, should have one or more tags <SrcUsed> which record a list of used datasets. This field will be used as a pointer to build the relationship among datasets in a catalogue system. Other fields or attributes as listed in Table 5.4 are also used to enhance the functionality of developed application.

Principle 1: DAG representation with PostgreSQL

A directed acyclic graph (DAG) and an adjacency-list is implemented within a database. There are two main tables created in this step: vertices table which contains information from each dataset, and adjacency table which stores all pointers created from their <SrcUsed> field.

An edge is implemented for each relationship between two dataset.

Principle 2: Constructing remote sensing dataset relationship

Several sub-modules required for this step:

2.1. Inserting new vertex from XML metadata

Algorithm:

```
Read dataset
     Extract related fields
     Open vertex table
Check dataset key within current list of vertex
     If dataset already exist then update the existing fields
            Else create new vertex
```

2.2. Creating edge / adjacency from *source* pointer

Algorithm:

```
Read dataset
Extract SrcUsed field(s)
Open edge table
Select edge record from edge table where origin of SrcUsed exist,
If record exist then update the existing fields
else
   insert a new record in edge table
   open vertex table
   insert new record in vertex table
```

2.3. Preparing a module to remove a vertex. This module is required for user-defined deletion and an automatic deletion in a validation process.

Algorithm

```
Get vertex key
Open vertex table
Find key in vertex
If found then delete record
Open edge table
Find key in edge table as origin field
If found then delete record
Find key in edge table as destination field
If found then delete record
```

2.4. Validating graph structure. This part will verify the consistency of DAG structure by removing self loop and removing duplicate edge keys.

Algorithm

```
Open edge table
Select edge where origin = destination
If found then delete record

Open edge table
Find duplicate records
If found then remove duplicates
```

Principle 3: DAG depth calculation, traversing backward and forward

This part will generate a list of forward trace which is used to list all of datasets which are used directly or indirectly as source datasets, and a list of backward trace which are used to list all of datasets which are directly or indirectly produced from selected dataset.

Calculation of DAG depth

This procedure is required to find the limit of dataset levels, which is also used as an upper limit for calculating loop.

Algorithm:

```
  Open adjacency-list (edge) table
  Set depth = 0
  Select first record
  Set counter = 1
  Traverse adjacency-list, increment counter for every forward step
If counter > depth then update depth = counter
Reset counter=1 and repeat traverse starting from the next record.
```

Notes:

The depth of the history path for a level three dataset in our four-level structure is normally three, since usually the processing direction proceeds directly from one level to the other. In practice, the path can be longer if there are datasets which are created using other datasets from the same level. In this case, the depth of the history path may reach four or five. It is very unlikely that the depth is greater than five except a cyclic structure exists. To avoid an infinite loop due to an invalid recursive structure, a threshold number of five is used in the test application as a maximum possible depth.

Forward traverse

This procedure will find direct and indirect source of selected dataset / vertex.

Algorithm

```
Input dataset key (from user interface)
```

```
Find record in vertex table
If found then
    Open edge / adjacency table
    Find record in edge table where source = key
    Traverse adjacency-list, exit traverse if max depth reached
```

Backward traverse

This procedure will find direct and indirect output of selected dataset / vertex.

Algorithm

```
Input dataset key (from user interface)
Find record in vertex table
If found then
    Open edge / adjacency table
    Find record in edge table where destination = key
    Traverse adjacency-list, exit traverse if max depth reached
```

5.9 Analysis of Proposed System

As discussed in previous sections, our proposed system reconstructs the processing history of the data collection with the help of metadata into a directed acyclic graph, which is implemented using adjacency list. From this reconstruction, we can evaluate the processing steps of each remote sensing dataset, including which datasets were used by the selected dataset and which datasets were created using the selected dataset. In the following organization of remote sensing data will be compared with current spatial data catalogue systems and current proposals from other researchers.

Comparison with ArcCatalog

ESRI ArcCatalog provides a catalogue system for spatial data including raster data. There are many functions within this catalogue system including conversion tools and map projection tools, and those, which help users to integrate and share their spatial data collections. This catalogue system supports the current metadata standard from ISO and FGDC. Although a module named “Set Data Source” exists, it does not deal with relationship among datasets; this option is used to establish a link between the metadata field and its original dataset. Until version 9.2, ArcCatalog does not have the capability to construct the processing, even if processing steps are recorded in the metadata. In general, ArcCatalog from ESRI concentrates more on tools and functionalities to deal with dataset contents.

Comparison with FAO GeoNetwork Opensource

GeoNetwork implements ISO19115/19139 Geographic Metadata, Z39.50 and OGC-WMS standard among others. Compared to ArcGIS, this Open Source Catalogue system has the capability to organize spatial data collections which are stored decentralized. This catalogue system has its strength in interoperability among many spatial data providers and spatial data users through web-interface. In general, GeoNetwork brings spatial data in a regional or global cooperation, mainly for sharing information, but the access to the original dataset is very limited. This catalogue system has also a capability to edit metadata file.

Currently, this catalogue system can not organize relationships among remote sensing datasets, but its functionalities can be extended since it is an open source system. The hardware and software specification of this catalogue system is compatible with our prototype, therefore our proposed method would be implemented within FAO GeoNetwork Opensource catalogue system.

Comparison with satellite data providers' web catalogues

There are several examples of web catalogue systems for remote sensing data, for instance from NOAA, EOPortal, ESA, NASA and USGS. Most of these catalogue services limit their user interface to browse the collections based on several scopes, i.e.: spectral information, temporal information, and spatial information. Compared to our proposed system, these services concentrate more in information, which is important for remote sensing analysis, for example information about wavelength, type of sensor, cloud covers, spatial resolutions, geometric and radiometric corrections, and other remote sensing specific.

Comparison with other methods proposed by scientists

Several scientists also discussed issues relating to retrieval of remote sensing data. Bojinski et. al. (Bojinski, et al, 2002) introduced SPECCHIO, a web-accessible spectrum database for administration and storage of heterogeneous spectral data. Dorninger (2003) introduced TMIS, which organizes topographic data of Mars, based on live-metadata (updated from the data provider) and local dataset collections. Guo (Guo et al., 2003) developed ICEAGE, system for interactive clustering and exploration of large and high-dimensional geodata.

Huang (Huang et al., 2006) develop a sample application of Spatial Metadata Service System (SMSS) based on Z39.50 protocol, named "ZSpatial Meta Service" for an implementation of interoperability of several nodes in China. Shi (Shi et al., 2009) builts a database system for archiving and managing remote sensing data using Oracle

Database and ArcSDE based on web service technology. Ruan (Ruan et al., 2006) used ontology approach to develop semantic based image retrieval in remote sensing archive.

There are many proposals related to the development of query by image content (Seidel et al., 1998), feature extraction (Schroeder et al., 1997) and image mining (Hsu et al., 2002; Li, 2004) which can be integrated to our proposed system. Most of those researches will produce certain information which is extracted from remote sensing data. By including that information into metadata, a metadata-based image catalogue system can support broader possibility of user query.

6 Validations and Testing

This chapter evaluates the designed system and its prototype of the implementation. The specification of the prototype environment will be described in section 6.1. Currently, the number of actual metadata in the real catalogue system provided by Bakosurtanal is very limited; therefore hundreds of metadata are generated from existing metadata to simulate the real system. Many types of possible errors in these metadata have also been generated to evaluate the behaviour of the prototype in case of invalid metadata. The developed prototype is also tested with many possible logical errors to simulate the real implementation. The main result of this designed system is a list of all datasets, which were used as sources by a selected dataset, and a list of all datasets, which were created from the selected dataset, thus providing a kind of genealogical tree of the data collection.

Tracing the processing history of each dataset or, in other words, analysing the genealogical tree opens a possibility to evaluate a sort of relationship between two datasets based on their common sources. Section 6.5 is dedicated to the evaluation of this relationship, which should be expressed here by the term *relatedness*, derived from the notion of their more or less intensive involvement of common sources within the proposed DAG.

6.1 System Preparation

A computer application under Windows operating system has been developed as a prototype as described in chapter 5. The system environment for running this application includes:

- PostgreSQL Database Version 8.2.9 with PostGIS module
- Apache Webserver version 2.2.9
- PHP version 5.2.6 with Zend Optimizer
- Mapserver for Windows version 4
- Web browser as a client

Since there is no special requirement for the above mentioned geo-database and web-server installation, a standard installation as described in the installation manual can be pursued. Maptools.org (http://www.maptools.org/ms4w/) provides a Mapserver package for Windows, which includes all of above software. This package is a simple and straight-forward way to setup that environment.

To edit remote sensing metadata, a program from USGS named TK Metadata Editor (TKME) is used. This program can be downloaded as freeware from USGS website (http://geology.usgs.gov/tools/metadata/). The Microsoft XML Editor can also be used for this purpose. We use this XML editor to insert the <SrcUsed> tag and its value(s).

The developed application is installed as a root of webserver. There are several modules which are used, including main program, XML parser, dataset query and its user interface, and query processor.

The program is tested with a hardware configuration as follow: Intel Centrino Core2 Duo T5500, 2048MB DDR2 memory, Intel GMA950 graphic, and 500 GB harddisk.

6.2 Preparation of Metadata Files

The actual remote sensing metadata from Bakosurtanal clearinghouse is not suitable for this validity process since the number of existing data is very limited (less than 20), and the processing steps of creating those datasets are not available. A number of remote sensing metadata files are derived from existing Bakosurtanal metadata, and modified to simulate the proposed system. The following steps show the data preparations:

- Existing remote sensing metadata are copied from Indonesian Geospatial Data Clearinghouse. Currently only about 20 remote sensing metadata can be downloaded from Indonesian Geospatial Data Clearinghouse. Although these metadata do not contain any specific information about processing history, they are needed to verify the compatibility of our implementation with the existing catalogue system.

- Existing remote sensing metadata are edited using MERMAid or MS XML Editor to insert remote sensing extensions and additional fields. The main purpose of this step is specifying the history link field as a new tag named <SrcUsed>. We insert several other elements as listed in Table 5.3.

- Edited metadata is parsed to check for possible errors (using MERMAid or TKME). The example of metadata file, which is proposed to be used with this system, can be seen in Appendix A.

- Several hundreds XML metadata files are generated from above existing 20 metadata files. Bounding box coordinate values, processing step, and other element values are filled to simulate the use of this application in the future.

In an ideal situation, every metadata file contains valid information about the generation process of the respective dataset, including the name of the used sources. In practice, we should consider inconsistencies or mistyped values. In Table 6.1 few conditions are listed which could appear.

Since this prototype does not interact directly with the raster dataset, we do not need to import or include the raster dataset within our database system.

After above preparation steps have been completed, the remote sensing metadata is imported into the database. The existing XML formatted metadata files are imported one-by-one until all remote sensing datasets are listed in the database. A simple query can be used to display all imported datasets.

Table 6.1: Typical errors which are included in the sample of XML metadata

Elements	**Value(s)**	**Expected system response**
<SrcUsed>	blank	Considered as a raw image (created by none)
<SrcUsed>	Points to other datasets in cyclic way, for example: dataset A points to B as its source, dataset B points to C as its source, and dataset C points to A as its source.	Endless loop or stack overflow
<SrcUsed>	Invalid value, points to non-existing dataset	Failed to process the query
<SrcUsed>	Points to itself, similar to cyclic	Endless loop or stack overflow
Bounding coordinate	Invalid values or blank	Dataset can not be located or can not be displayed in an expected area of interest

As shown in above Table 6.1, several invalid values have been introduced in metadata to verify the validation process during import. If the <SrcUsed> tag has no value, then the system can not determine whether the dataset is actually a level 0 dataset (raw data), or a dataset, which has no recorded history available. A value of <SrcUsed>, which point to itself, can be removed during the import process, while an endless loop caused by a cyclic structure can be avoided by limiting the depth of the path. An invalid value in <SrcUsed>, which points to a non-existing dataset, can not be detected during the import process. A validation process is required after all metadata have been imported to list all datasets, which point to non-existing datasets. The result of those invalid inputs within this prototype will be explained later in Table 6.2.

6.3 Metadata Query

The proposed catalogue system is now ready to be tested. A graphical user interface (GUI) can be used to perform a query based on an area of interest. Several common GUI tools are provided, such as zoom-in, zoom-out, pan, reset zoom, and select. The user can also enter specific values for a bounding box by inputting coordinate values in decimal degrees system.

The following Figure 6.1 shows the sample of a GUI of this application, and a result of a simple query.

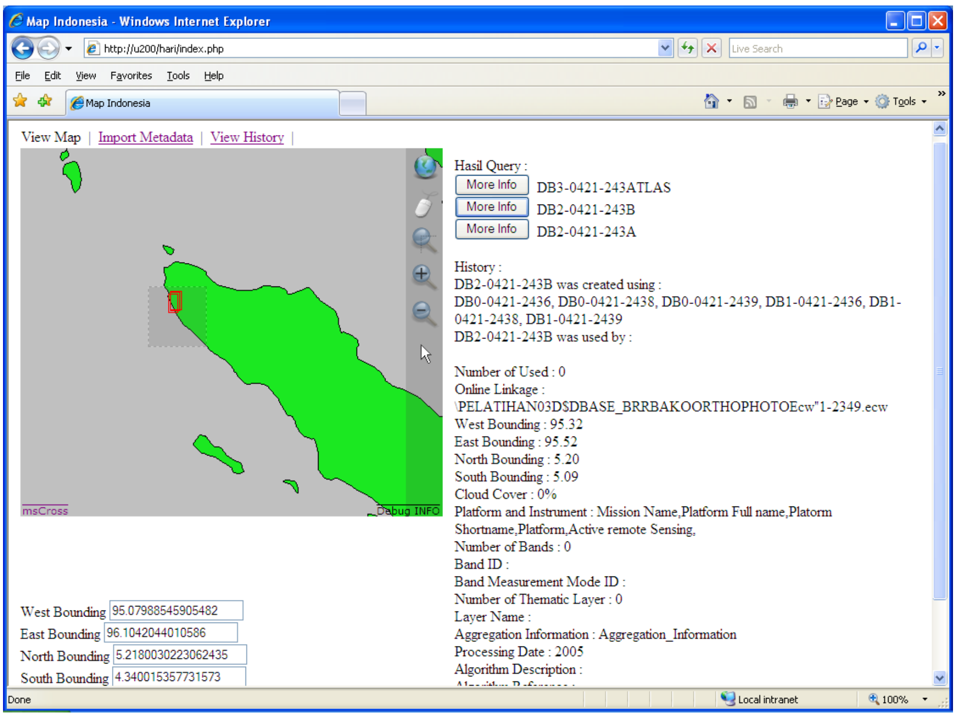


Fig 6.1: Example of the user interface of the proposed system

The output of this application shows two aspects of remote sensing datasets. Firstly, the static data are listed directly extracted from the dataset, for example the bounding box value, original dataset name and location, spectral information, and data quality information. Secondly, it shows the analytical information calculated by the core of this application, such as the list of datasets used to create the selected dataset, the list of datasets already created using this selected dataset, and how frequently this dataset has been used.

For verification, a study area has been selected. After this step, a list of datasets has been created with a button for more information next to each list element. This “More Info” button can be used to display selected attributes from the dataset. There are many fields which are prepared for future use. In current metadata, these fields have no data. The following list is an example of the output from a selected dataset:

DB2-0420-392D was created using:
DB0-0420-3926B, DB0-0420-3927B, DB0-0420-3928B, DB1-0420-3926B, DB1-0420-3927B, DB1-0420-3928B

DB2-0420-392D was used by :

Number of Used : 0
Online Linkage : \PELATIHAN03D$DBASE_BRRBAKOORTHOPHOTOEcw"1-2349.ecw
West Bounding : 95.212264
East Bounding : 95.262125
North Bounding : 5.468245
South Bounding : 5.425254
Cloud Cover : 0%
Platform and Instrument :
Number of Bands :
Band ID :
Band Measurement Mode ID :
Number of Thematic Layer :
Layer Name :
Aggregation Information :
Processing Date :
Algorithm Description :
Algorithm Reference :
Pixel Resolution :
Grid orientation :

The aggregated information from this query output can also be used for advanced data query, which is not included in this prototype, for example:

- Finding the dataset which is frequently used
- Finding the dataset which is not yet used
- Finding the dataset which has a specific attribute
- Finding the dataset which is suitable for specific analysis

6.4 Evaluation

6.4.1 Critical Conditions and Their Solutions

Several inconsistencies or critical conditions have been found during our validation process as listed in the following Table 6.2. Most of these inconsistencies or conditions can be avoided or fixed.

Table 6.2: Inconsistencies or critical conditions

Case	**Result / Solution**
1. Attribute errors	
Field <SrcUsed> is empty	Considered as a raw image
Field <SrcUsed> is invalid	Ignored, replaced with blank in output
Field <SrcUsed> points in a cyclic way	Depth of path should be limited
Field <SrcUsed> points to itself	Similar to the cyclic case
Bounding coordinate values invalid or blank	Ignored
2. Logical errors	
A dataset has been removed from collection, producing missing link	Logical solution is rebuilding the DAG
The same dataset file is imported more than once	The second attempt is ignored.

For each of the above listed critical conditions the following paragraphs explain the solution in more detail.

Attribute Error: Required source field is blank

The system uses a list of source datasets for each individual dataset. However, in raw images no processing step has been applied; therefore the list of source datasets is empty or does not exist. In another case, even datasets in level 1, 2 or 3, which are expected to have been derived from other datasets, may have missing metadata about their history. In order to maintain this condition, datasets with no information about their source are considered as raw images, or new datasets, which were not derived from others. Therefore, the system does not indicate an error and accepts them to be included in the collection.

Attribute Error: Required source field is invalid

If the system can read a value from <SrcUsed> field, it is difficult to directly decide (if the insertion of datasets has not been finished yet) whether the value is valid or not. All values which point to unlisted datasets could be considered as invalid, but possibly a currently unlisted dataset has not been inserted yet into the collection. The validation process can not start before the importing process has been completed. After a query the decision can be made whether a requested dataset exists in the collection or not. In the latter case, the result of the query will be replaced by blank or ignored, while the value of the field remains unchanged.

Attribute Error: Dataset points to itself

In our remote sensing data handling, we do not support a cyclic structure; therefore we use the directed acyclic graph (DAG). A cyclic structure still occurs if a dataset erroneously points to itself, or it may occur indirectly after a series of links through other datasets, where the final link points back to the first dataset, as for example A → B → C → A.

A cyclic structure can be avoided, as already mentioned before by using a counter to limit the depth of history path thus avoiding an infinite loop. An appropriate threshold could be 6. Another possibility would be the comparison of the time stamp (i.e. creation time) of a selected dataset and of the source dataset . The link can be considered as invalid if the time stamp of the selected dataset is older than that of its source the selected dataset. This approach is not tested implemented in this prototype since datasets are generated, therefore the and thus the given time stamp does not represent the real creation time.

Attribute Error: Wrong values of boundary coordinates

Coordinate values which are given in the metadata are used to display footprints of the datasets in the user interface map. These coordinate values allow user navigation through the collection and locating a study area. After a dataset has been selected for processing, boundary coordinates are not used anymore, therefore the validity of coordinate values does not influence the creation of dataset history lists.

Logical Error: Missing link due to deletions

When a dataset is inserted into a collection and its source pointer addresses a dataset which does not exist, it is called a "missing link". Several possibilities may lead to this problem, for example:

a. Removed: Addressed dataset was deleted from collection
b. Uncompleted: Addressed dataset has not been included in the collection yet
c. Invalid value: Source pointer is invalid.

In this prototype, a pointer to a non-existing dataset does not produce an error to the system, and it can not be detected during inserting or editing process. A validation process is required to rebuild the relationship structure, and verify whether addressed source datasets exist or not. The pointer to the missing dataset can be removed from output by comparing the pointer to a list of existing datasets.

Logical Error: Dataset is entered multiple times

Each dataset-ID (or dataset name in our proposed system) is considered as a unique entry, and it is used in the database as a primary key. Therefore, entering duplicate dataset names is considered as a human error, and it can be trapped at the input process.

To update the dataset, which is already imported into the database, the recorded dataset should be deleted prior to the importing the newer version of it.

The same error condition may occur in the less likely case where for one dataset more than one metafile exists. This can, for instance, happen if metadata are updated and saved with a different name. Both files refer to same dataset. As long as the contents of metadata are not compared this error cannot be detected.

6.4.2 Evaluation of Adjacency Matrix versus Adjacency List

With the very limited number of actual remote sensing metadata from Bakosurtanal, a real complexity can not be tested or evaluated. To simulate the actual condition, several hundreds of sample metadata were generated. The number of records in the database has been compared with the calculated maximum number of records as explained in section 5.5. The result is shown in the following Table 6.3

Table 6.3: Comparison of number of records

Number of sample data	Maximum number of records as calculated in section 5.5*)	Actual number of records
20	318	17
100	6570	180
400	55560	710

*) maximum number of source dataset used. The threshold for the maximum depth is set to 6

The actual number of records as shown in table 6.3 is far less than the maximum number of records as calculated in section 5.5. This comparison shows that the actual adjacency matrix is sparsely populated and therefore, as an optimal solution the implementation of an adjacency list is justified.

6.5 The Relatedness of a Pair of Datasets

With the creation of history links in the data collection using the method as shown previously, all datasets will be organized according to a directed acyclic graph data model. Sometimes it might be of interest to a user to derive the relationship between two nodes (i.e. data sets) within the graph, thus expressing how close they are to each other from point of view of data generation. Relevant aspects could be the total number of nodes, which were involved in the generation of the two nodes under investigation, and the number of nodes, which were common in the generation history, and possibly the length of the shortest path from one node the other by following the directions given by the directed graph. Another notion could be based on the investigation of the genealogy of the nodes by assigning genes to the original data sets and by observing the inheritance. The latter ("biological") approach will not be pursued any more here. This sort of relationship will be called *relatedness* in the following. It is important to mention explicitly, that the *relatedness* between two data sets does not indicate any relationship regarding data contents, it just shows whether the partner data set has been used somewhere in the more recent or further past of the processing history and whether more or less other data sets were also involved.

Using this model, further evaluation concerning properties of *relatedness* can be performed. Since this graph data model is created based on the pointers to the sources of each dataset, eventually the edges of the graph represent the path of the processing history.

In general, there are three main conditions for *relatedness* between two datasets. Using the dataset names as shown in Figure 6.2 as an example, those three conditions are:

- There is a processing link from the first to the second dataset. The *relatedness* can be trivially evaluated using the algorithm explained in chapter 5.8, subsection "Forward Traverse". An example of this condition is the relationship between the datasets R02 and R22.
- There is no processing link from the first dataset to the second dataset, which means the first dataset is not listed as a source dataset for the second dataset, or vice versa, but these two datasets have a common source dataset. In this case, they are linked indirectly. The relatedness of these datasets can be evaluated by involving other datasets. An example of this condition is the relationship between the datasets R21 and R22.

- There is no processing link from the first dataset to the second, and they do not have any common source datasets. A relationship does not exist. Examples of this condition are represented by the dataset pair (R22, R23) and (R11, R12)

It is likely that a collection of datasets does not form a single graph data model, as it is the case, for instance, in Indonesia. Figure 5.9 in section 5.5 shows the groups of remote sensing datasets which forms many graph structures with at least one member in each group. With the directed acyclic graph data model this grouping can be proved by evaluating the *degree* of every node in the collection. Degree of a node is defined by the total number of its *neighbours*, i.e. vertices directly connected to it (Diestel, 2005). In a directed acyclic graph, the *degree* of a node can be calculated as *in-degree* where only the number of links to the selected node is calculated and *out-degree* where only the number of outbound links from that node is calculated.

From the evaluation of the degree values of every node, it can be concluded that the data collection in our sample has many datasets with a degree value of zero which means this collection does not form a network data structure, therefore, further analysis which usually applied for network analysis like centrality, can not be performed. In practice, remote sensing datasets which are not used for any processing will have a value of out-degree equal to zero, and raw dataset or dataset with no source information will have a value of in-degree equal to zero. From the evaluation of the degree of each node as described above, all neighbours of each node can also be listed, and each of these neighbours can be considered as a closest dataset relative to the selected node.

The next evaluation step is to find how close one dataset to another dataset based on their processing steps. In a directed acyclic graph which is implemented here, a *path* from one dataset to another dataset can be tracked from the history list. The number of steps in a path from one dataset to another one represents the *distance* between them. Each dataset which is listed as a direct source for its pair dataset will have the distance value of one, and this is the closest neighbourhood relationship between two datasets. The higher value of *distance* represents the longer path between them.

Considering the possibility that more than one path exists between two nodes, a shortest path (minimum distance) will be used as a distance between two nodes. The shortest path between two nodes is calculated using existing algorithms (Bang-Jensen, 2008) where each link/edge in this graph has a constant value of one as its weighting value. The distance from one node to another node shows how many steps of processing have occurred in the direction of processing at its shortest path. Using the above shortest-path algorithm, the distance value can not be calculated if two nodes do not have a one-way-path from one node to the other, even if both nodes shares many common input datasets. The following figure 6.2 shows an example of datasets collection with their creation links. From this collection, a relatedness of a pair of datasets R22 and R31 will be evaluated.

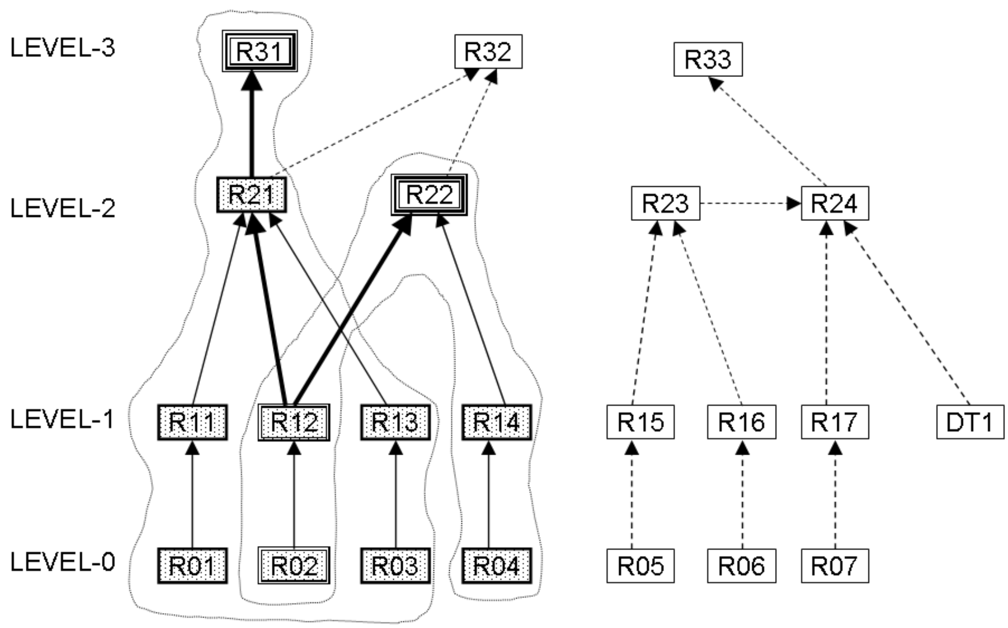


Fig 6.2: Neighbourhood evaluation of dataset collection

From a visual approach of Figure 6.2, we can evaluate the neighbourhoods and clustering of this collection, and find that there are two main clusters of datasets, and if we choose one node as our reference node, for example R22, we can then evaluate how close another node to this node is. With a visual evaluation, we can see that node R14 is closer to R22 than R15, since R14 is a direct input for R22, and R15 does not have any path linked to R22. Let us now compare how close the relatdness between R31 and R22 is, and compare it with the relatedness between R23 and R22. Even though there is no path from R31 to R22 (and vice versa), and also there is no path from R23 to R22 (and vice versa), it is visually clear that R31 is closer to R22 and R23 is located in a different group of processing paths. From this point of view, we want to find parameters which can be calculated to prove that R31 is closer to R22 in a comparison with the relatedness of R23 and R22.

The main rule of the calculation is using only the *inbound path* of each node, which represents a list of source datasets. This rule has its strength that the selected dataset is then chosen as a tree-root of its own creation process, and the future processing which include this node can be eliminated from the calculation. By applying this rule, any additional processing steps in the future which include this dataset as its data source will not affect this neighbourhood calculation. Figure 6.2 shows the sub-tree which is created from R31, where R31 is now a root of it; and a sub-tree which is created from R22, where R22 is now a root of it. From this step, a node R32 can be eliminated from further calculation since it is not part of sub-tree R31 and sub-tree R22.

From sub-tree R31 and sub-tree R22, the sub-node list which is also the source dataset list for each R31 and R22 can be created, and by comparing these two lists, a list of common datasets can be extracted if there is an intersection between them. This intersection can be interpreted that there is a certain connection between both root datasets (ie. R31 and R22) even if there is no path from R22 to R31 or vice versa. The strength of the connection between R22 and R31 can not be determined with the number of common nodes only, since this calculation is based solely on the processing path, without knowing the algorithm of the creation process. Since each of the common nodes has a path to both evaluated datasets (R22 and R31), the shortest distance from this common dataset to R22 and to R31 can be calculated. The distance from R22 to R31 is then calculated by combining the shortest distance from the common node to them.

From the calculation of the *distance* between two datasets as described above, an approach to evaluate the neighbourhood or clustering has been made. Applying this calculation to the three conditions as described in page 82 resulting in three neighbourhood conditions between two datasets:

- If the first dataset is listed as a source dataset of the second dataset (or vice versa), distance value will be equal or greater than one, with the smaller value represents the closer neighbourhood.
- If the first dataset is not listed as a source dataset of the second dataset and vice versa, but they have common datasets, distance value will be equal or greater than two, with the smaller value represents the closer neighbourhood
- If the first dataset is not listed as a source dataset of the second dataset and vice versa, and if they do not have common dataset, then the distance value can not be calculated, and these two datasets are considered independent of each other or belong to a different cluster / group.

Using the above explanation, distance between two datasets can be calculated even if there is no processing path from one dataset to its pair, and this distance value can be used as a main parameter of the neighbourhood between two datasets. From the above explanation, it is also important to find the common dataset of a pair of datasets if there is no direct link between them. The higher number of common dataset shows more possibilities of processing path between a pair of dataset. The following discussion shows how to calculate the distance between a pair of datasets R31 and R22, and evaluate the neighbourhoods of this pair of datasets.

Step 1: Find a list of source datasets that are used by each dataset.

The result for R31 is: R21, R11, **R12**, R13, R01, **R02**, and R03; a total of 7 datasets

The result for R22 is: **R12**, R14, **R02**, and R04; a total of 4 datasets

Step 2: Find common datasets that are members of both lists resulting from step 1.

The result is: R02 and R12 (2 datasets).

If the evaluated dataset is listed in the list of source dataset, then it is considered as one additional common dataset.

Step 3: Find the total used datasets by combining the two lists resulting from step 1 and taking off the duplicate datasets as resulted in step 2.

The result is (7+4 – 2) : 9 dataset.

Step 4: Find the distance from each of common dataset to R31 and R22, and choose the shortest distance.

From R02, the distance to R31 is 3 steps, and distance to R22 is 2 steps, total = 5 steps

From R12, the distance to R31 is 2 steps, and distance to R22 is 1 step, total = 3 steps

The shortest distance is from R31 to R22 is 3 steps.

From 4 steps above, several parameters of neighbourhood between R31 and R22 can be calculated which are:

Number of source of R31: 7 datasets
Number of source of R22: 4 datasets
Total common dataset: 2 datasets
Total used datasets: 9 datasets
Shortest distance: 3 steps

From the comparison of source dataset lists, R31 is not used by R22 as its source and vice versa (see step 2).

Using the above algorithm / steps, several pairs of datasets are evaluated including the datasets pair R22 and R23, and the result is shown in Table 6.4. This table shows the three parameters as a result of the evaluation of neighbourhood of a pair of datasets and an additional column shows whether the dataset is used as a source dataset by its pair (the first dataset is used by the second dataset, or the second dataset is used by the first dataset).

Table 6.4: Calculation of neighbourhood parameters of several pairs of datasets

First Dataset	Second dataset	Shortest distance	Total used dataset	Total common dataset	Used by its pair
R21	R22	2	8	2	No
R21	R23	n/a	10	0	No
R21	R31	1	7	7	Yes
R21	R01	2	6	1	Yes
R01	R02	n/a	0	0	No
R11	R12	n/a	2	0	No
R22	R31	3	9	2	No
R22	R23	n/a	8	0	No

Shortest distance between two datasets can only be calculated if at least one common source dataset is found. It is considered as not applicable (displayed as n/a in Table 6.4) if there is no common source dataset. The value of distance between two datasets is calculated based on the number of step which is linked between those two datasets, and only the creation procedure is considered. Table 6.4 above shows that a pair of R31 and R21 has a shortest distance of 1, which means both datasets are very closely linked to each other. This table also shows that dataset R21 and R22 are not used by each other, but they have two common source datasets and the distance between them has a value of 2.

To use the value from the calculation of neighbourhood parameter of a pair of dataset with this method, several remarks should be considered:

1. The method uses only links to the past of a respective dataset. It does not take into consideration any process where this dataset may be or may have been involved later on. Thus the calculated values are independent of any future extension of the collection.

2. The value of the shortest distance represents the minimum number of edges between two nodes. This rule does not put the proportion of each source or any algorithm into its consideration.

3. The combination of parameters of this neighbourhood evaluation of a pair of dataset shows how common their used dataset is and it can be used to group or classify datasets in a collection based on their processing steps. This grouping functionality can be used to divide datasets in a collection based on several reference datasets.

4. The content of dataset does not take part in this calculation; therefore this neighbourhood evaluation does not reflect the dataset similarity based on image content.

Table 6.4 shows the list of values or parameters which are considered as properties of neighbourhood between a pair of datasets. This combination of numbers and logical value should give a hint of relatedness and how tight this relatedness. If the value of relatedness can be assigned as a logical “Yes” or “No”, where “Yes” means there is relatedness and “No” means there is no relatedness, then the value of common dataset (column 5) can fulfil this calculation by assigning the following rules for each of datasets pair:

```
If (common datasets = 0) then Relatedness = “No”
      Else Relatedness = “Yes”
```

This logical value of relatedness can be used to find a group of cluster within the datasets collection, for example by finding all datasets in the collection which has relatedness value of “Yes” with a certain reference dataset.

The above logical expression will produce the same value for a dataset which is a direct neighbour (strictly connected) and a dataset which is not directly related (less connected). The new rule should be formulated to measure the relatedness using all of possible parameters and producing a certain value. This measure should express whether there is a more or less strict connection between two datasets in their processing history. For the next discussion, the following notation will be used:

Я = Relatedness between a pair of datasets

Δ = Shortest distance (topological distance)

ΣS = Total number of sources (Used datasets)

ΣC = Total number of common sources (common datasets)

P = Used by partner (logical)

The strong argument for the relatedness is the number of common datasets (ΣC). The higher the number of common datasets, the relatedness (Я) is considered higher also. This shows the direct relation between relatedness and the number of common datasets. However, the ratio between common datasets (ΣC) and total used datasets (ΣS) shows a better approach for relatedness since it gives the sense of the proportion of participation in the creation process. This calculation will produce a magnitude of value between 0 and 1. To shift this value into a better range similar to percentage, from 0 to 100, the above ratio is multiplied by 100.

The third parameter, topological distance (Δ), is inversely related with the relatedness. The smaller the distance the higher is the relatedness. The eventual value for relatedness is therefore directly related to the mentioned ratio and inversely related to the topological distance. The additional value P (used by partner) can be neglected since the value actually has been used in the calculation of common datasets. The following formulation is then used to measure the relatedness (Я):

$$Я = 100\ (\Sigma C/\Sigma S)/\Delta$$

The following Table 6.5 shows the calculation of relatedness (Я) to the list of selected pairs as used in Table 6.4

Table 6.5: Calculation of relatedness of several pairs of datasets

First Dataset	Second dataset	Shortest distance (Δ)	Total used dataset (ΣS)	Total common dataset (ΣC)	Relatedness (Я)
R21	R22	2	8	2	12.5
R21	R23	n/a	10	0	0
R21	R31	1	7	7	100
R21	R01	2	6	1	8.33
R01	R02	n/a	0	0	0
R11	R12	n/a	2	0	0
R22	R31	3	9	2	7.41
R22	R23	n/a	8	0	0

By comparing the calculated relatedness from above table 6.5 with the visualization of datasets creation process as shown in figure 6.2, there are several remarks can be listed:

3. The higher relatedness value shows the higher similarity of their creation process
4. The value of relatedness of zero shows there is no similarity in the creation of those datasets. This condition is also applied when the relatedness of two raw datasets (level-0 datasets) is calculated.
5. There are several pairs of datasets which belong to the same group but the relatedness is zero (i.e. R01 and R02, R11 and R12)
6. Datasets belonging to distinct groups have a relatedness of zero (i.e. R23 and R21, R23 and R22)

This evaluation showed that this value meets the intended requirements. The more detailed calculation of relatedness from Figure 6.2 can be shown in Appendix B.

7 Conclusions and Future Works

7.1 Summary and achievements

Organizing remote sensing data in databases has been the main issue for many years. Basically all providers of remote sensing satellite imagery have their own catalogue system, which helps the prospective users to quickly find their requested datasets among the huge pool of images acquired over decades. Search criteria are predominantly the geographical region of interest, the type of sensor, time of acquisition, cloud cover and possibly certain pre-processing levels, which may range from raw images, to system corrected images, to geo-referenced images and sometimes also to derived products, such as digital terrain models, vegetation maps, soil moisture products, etc.

The data organisation in this work is intended to add other criteria. High level data have always been derived from low level data, and therefore, there is a logical connection from a resulting product to its source data. Creating some structure among the data by keeping the links to their genealogical history, provides the users with an added value in terms of quality management and quality assurance. They are given the possibility to trace back an existing product to its roots, thus enabling the access of original data if required for some reason. By setting up the history links each dataset can be easily assigned the information about its involvement in a creation process of other existing datasets.

This thesis has been initiated by the situation in Bakosurtanal, the national mapping agency of Indonesia, and therefore the research activities have definitely been driven by Indonesian interests. A certain basic structure, introducing product levels dependent on the order of processing steps, which were applied to generate a certain product, has already been built up for data in Indonesia and the decision has been made, that metadata have to be consistent with existing standards, although further developments and implementations of databases have not been pursued consequently. This work adopts the existing national decisions and tries to develop and thoroughly investigate an appropriate structure of data organisation which allows efficient data storage and queries.

It has to be emphasized, that the main objective of this work was not the development of methods for efficient storage of mass data like remote sensing imagery. The methods developed and investigated are applied to metadata. It was of interest to show how metadata for a collection of spatial datasets need to be extended in order to leave unchanged existing information, to be compliant with standards, to remain open for future extensions, and to add further information which is required to build up a graph structure and which optimally represents the data history with respect to their applied processing steps. It turned out that an acyclic directed graph (DAG) is best suited for the required sort of organisation, which represents the relationship between the datasets in the data collection.

Chapter 5 investigates the realisation of the proposed DAG, which could be an adjacency matrix or an adjacency list. The comparison came to the conclusion, that an adjacency list is more suited because of the sparse population of the matrix. In a next step, the existing datasets, which are available as an unsorted collection, need to be organised according to the chosen graph model by analysing the metadata. They contain information about the immediate source data, which were involved in the creation process of their respective dataset. Eventually the graph represents the entire processing history of the data collection, thus for each element a list of datasets used for its creation, and a list of datasets created from it can be derived. The graph, therefore, also provides an appropriate means for evaluating a measure for the relations among the datasets, which is called here relatedness.

Chapter 6 is dedicated to the evaluation of the design and to finding suitable parameters for calculating the relatedness. Although there may be certain crucial situations, where metadata contain incorrect entries, the concept proved feasible to be realised. As far as relatedness is concerned, this measure should express whether there is a more or less strict connection between two datasets in their processing history. It is obvious that datasets, which are common in their history, are important and therefore, the number of common sources serves as one of the parameters. This very strong argument for relatedness is reduced by the other datasets which have been used in the generation history. The ratio between the number of common datasets and the number of total datasets expresses in a better way the relatedness. As a further parameter the topological distance between the datasets has been introduced. The smaller the distance the higher is the relatedness. The eventual value for relatedness is therefore directly related to the mentioned ratio and inversely related to the topological distance. The evaluation showed that this value meets the intended requirements.

By using metadata as the source of information this method becomes independent of the file format in which the data are stored. In addition, employing FGDC standard, which is also compatible with ISO standard, allows the realisation of this method for many spatial data catalogue systems or spatial data collections. Chapter 6 also shows that the designed system can be realized without high demand for specific computer hardware and software, which is also a key consideration in our hypothesis (section 1.3).

A logical categorization of spatial datasets based on processing levels introduced by Amhar (2001), is used in this data model. However, this categorization is not a prerequisite. Therefore, our proposed method can also be implemented in other spatial dataset collections even if the logical categorization is different as long as the model of a directed acyclic graph is fulfilled and metadata conform to FGDC standard.
There are several steps required to organize spatial data from their original dataset. Firstly, metadata should be created for each dataset, these metadata are then imported into a database, where the processing history is extracted and an adjacency list is created. Based on this adjacency list, a further analysis can be performed, for example finding the list of source dataset of a selected dataset, or calculating the relatedness of a pair of datasets. An update in a metadata files does not automatically update the respective information in the database, which may lead to inconsistency between the contents of the database and the data in the collection. A procedure is required which automatically triggers updating the database when changes to the collection are made, which needs to be considered if the method will be implemented in practice.

From the above summary one can clearly see, what the achievements are:

- From a collection of datasets a structured relationship based on the creation process could be established, which provides an excellent overview of how the data depend on each other and – in particular for geo-services – it also provides many advantages like better possibilities for quality assessment and quality assurance. Additionally, statistics become available, which show how often a dataset participated in the derivation process of later datasets
- By having available a measure for the interrelationship of dataset pairs, it is now possible to estimate how independent they are as far as the involvement of common datasets used for their creation is concerned. This aspect may also be important for quality assessment. It may also serve as a sort of segmentation process as it indicates. In case where the relatedness is zero, those datasets belong to distinct groups.
- Since the methods are based on standards for metadata a later implementation is highly independent of hardware platforms and software environment. Therefore, all theoretical investigations can be easily transported into practice.
- Although the early idea in this thesis is an organization of remote sensing data, the proposed method relies on the metadata, therefore it can also be used for organizing vector data or any other type of spatial data or a mixture of many type of spatial data.

7.2 Future Work

As discussed in the previous section, this method is based on a directed acyclic graph. In order to avoid in advance infinite loops caused by possible cyclic structures, currently a threshold value terminates the tracing process. Setting a threshold appeared to be helpful so far, although it is not the optimal solution. Since an assumption was needed, which defined the possible maximum number of processing steps in a group of datasets, finding the correct result may even be prohibited. By terminating the tracing process, the information whether a cyclic structure actually exists, gets lost. If the threshold is too low, not all sources can be found, and if the threshold is too high, loop processing and duplication cannot be avoided.

Therefore, future research should focus on developing methods with an extended capability of the tracing process, which should be able to clearly show whether a cyclic structure exists. If a cyclic structure is detected, it may either be caused by wrong data or it may indicate that the assumption of a DAG is wrong. In the first case data must be corrected, in the latter case further investigations of the entire collection are necessary.

The discussion of the relationship among datasets as in section 6.5 focuses on the evaluation of the strength of the correlation between two datasets. The DAG structure which is applied in this method opens a possibility to evaluate the relative position of a node within the processing steps where it is involved, for example how many steps exist "before" the dataset and how many steps exist "after" the dataset. Analyzing this relative position may be used as one indicator to predict the level of processing of each dataset, which is also useful for organizing the dataset based on a certain processing level grouping.

An indicator for relatedness has been introduced. The calculated value will not change if in future a new dataset is created using the datasets of the collection. Even though this method has its strength, it neglects the fact that datasets which produce a common output have a certain similarity or relatedness. Currently raw datasets, for instance, always have an indicator of relatedness of zero to other raw datasets, even if they are used together to produce a higher level dataset. The data model used by this dissertation can be used for the calculation of the relatedness based on the common output. Further investigations are required to evaluate and define this relatedness and its parameters.

Relationship among remote sensing datasets can be extended based on other logical relation. In our research, we organize remote sensing data collection based on processing steps and use these processing steps as a key for relationship. In practice, there are also several keys which can be used to link one dataset to another, for example:

Dataset correlation based on time series acquisition

Time-series of satellite data or aerial photos is not linked by default; they are covered by query as independent datasets in user defined study area. To deal with time-series data like this, a new container might be defined if required.

Dataset correlation based on image content

Many researchers have proposed methods for feature extraction from remote sensing data. By defining a standard list of extracted features, and put this information in image metadata, a module can be implemented to extract information from a remote sensing data, and update its metadata with the standard-coded feature, and this new field can be considered for a future expansion of user query based on image content.

This research is initiated from the case of organizing remote sensing data, which lead into an investigation for supporting many specific aspects of remote sensing data and how usually they are stored. As pointed out, the proposed method can be used to organize any kind of spatial data through their metadata, including vector data. In principle, this method can be used for organizing vector data based on their creation history. However, specific aspects of vector data have not been evaluated in this research and they require further investigation. A further evaluation is also required if a mixture of many type of spatial data will be organized using this method.

Appendix A: Sample Metadata

The following meta-code is an example of spatial metadata from Indonesian clearinghouse with several additional fields (tags) which are inserted as proposed in this research. This XML file conform FGDC Standard Content for Digital Geospatial Metadata.


```xml
<?xml version="1.0" encoding="UTF-8" ?>
 <metadata>
 <idinfo>
 <citation>
 <citeinfo>
 <origin>BAKOSURTANAL</origin>
 <pubdate>2005</pubdate>
 <pubtime>Unknown</pubtime>
 <title>Orthophoto_0421-2349.ecw</title>
 <edition>1</edition>
 <geoform>raster digital data</geoform>
 <serinfo>
 <sername>raster orthophoto data with 30 cm spatial resolution</sername>
 <issue>STG BRR 2005</issue>
 </serinfo>
 <pubinfo>
 <pubplace>BRR NAD-Nias, Jl. Ir. M. Thaher No. 20 Lueng Bata, Banda Aceh,
Nanggroe Aceh Darussalam</pubplace>
 <publish>Satuan Tugas Geospasial (Geospatial Task Force) BRR-
BAKOSURTANAL</publish>
 </pubinfo>
 <onlink>\\PELATIHAN03\D$\DBASE_BRRBAKO\ORTHOPHOTO\Ecw\0421-
2349.ecw</onlink>
 <lworkcit>
 <citeinfo>
 <origin>BAKOSURTANAL</origin>
 <pubdate>2005</pubdate>
 <pubtime>Unknown</pubtime>
 <title>Digital_Photo_0421-2349</title>
 <edition>1</edition>
 <geoform>remote-sensing image</geoform>
 <serinfo>
 <sername>Digital Aerial Photo for Topographic Mapping with 1:10.000, 1:5.000
and 1:2.000 Scale</sername>
 <issue>STG BRR 2005</issue>
 </serinfo>
 <pubinfo>
 <pubplace>BRR NAD-Nias, Jl. Ir. M. Thaher No. 20 Lueng Bata, Banda Aceh,
Nanggroe Aceh Darussalam</pubplace>
 <publish>Satuan Tugas Geospasial (Geospatial Task Force) BRR-
BAKOSURTANAL</publish>
 </pubinfo>
 </citeinfo>
 </lworkcit>
 </citeinfo>
 </citation>
 <descript>
 <abstract>Data is raster orthophoto with 30cm spatial resolution taken in June
2005 by NORAD under the supervision of BAKOSURTANAL (National Coordinating
Agency for Surveys and Mapping).</abstract>
```

<purpose>To provide data for topographic mapping around NAD Province. The topographic maps derived from this orthophoto are line maps with 1: 2.000 scale, especially dedicated for Rehabilitation and Reconstruction of Aceh-Nias after tsunami disaster.</purpose>
</descript>
<timeperd>
<timeinfo>
<sngdate>
<caldate>2005</caldate>
<time>unknown</time>
</sngdate>
</timeinfo>
<current>publication date</current>
</timeperd>
<status>
<progress>In work</progress>
<update>As needed</update>
</status>
<spdom>
<bounding>
<westbc>95.20801</westbc>
<eastbc>95.250164</eastbc>
<northbc>5.500161</northbc>
<southbc>5.458169</southbc>
</bounding>
</spdom>
<keywords>
<theme>
<themekt>Tri - Service Spatial Data Standard</themekt>
<themekey>Topography</themekey>
</theme>
<place>
<placekt>Geographic Names Information System</placekt>
<placekey>Banda Aceh, Nanggroe Aceh Darussalam</placekey>
</place>
<temporal>
<tempkey>2005</tempkey>
</temporal>
</keywords>
<accconst>The aerial photo can be accessed on request to BRR or BAKOSURTANAL.</accconst>
<useconst>Data may not be used, resold or passed on to any third party for commercial gain. Users agree to include data citation as follow: Aerial photo provided by BAKOSURTANAL.Perjanjian Penggunaan Data 1. Data yang tersedia dalam perjanjian ini diberikan tanpa biaya apapun, dan hanya boleh dipakai untuk kegiatan-kegiatan yang berkaitan dengan upaya rehabilitasi dan rekonstruksi NAD-Nias yang disetujui oleh BRR. Semua kegiatan tersebut ditujukan dalam rangka capacity building dan untuk mencapai tujuan pengembangan infrastruktur data spasial nasional dan provinsi.2. Yang dimaksud dengan data dalam perjanjian ini adalah data elektronik photo udara ortho-rectified dan segala bentuk data turunannya hasil kerjasama Pemerintah Norwegia dan Pemerintah Indonesia yang diwakili BAKOSURTANAL, diperkuat dengan kesepakatan kerjasama (MOU) BRR dan BAKOSURTANAL.3. Data sebagaimana tersebut dalam poin 2 tidak boleh digunakan dan atau diperjualbelikan untuk mendapatkan keuntungan komersial dan tidak boleh dipindahtangankan kepada pihak ketiga atau kegiatan lain untuk mendapatkan keuntungan komersial.4. Meskipun pihak penyedia data telah melakukan segala upaya untuk menjamin kualitas data, penyedia data tidak bertanggung jawab terhadap segala kesalahan yang diakibatkan dalam penggunaan data. Pengguna data diharapkan dengan sangat untuk melaporkan segala permasalahan atau ketidakwajaran data kepada penyedia data. Penyedia dan distributor data akan memberikan tanggapan kepada pengguna data dalam waktu

yang tidak terlalu lama.5. Pengguna data menyetujui untuk mencantumkan sumber data, lihat poin 6 di bawah, dalam semua produk cetak, laporan atau publikasi elektronik (termasuk internet) berkaitan dengan penggunaan data, termasuk pencantuman logo BAKOSURTANAL dan BRR.6. Cara penulisan pencantuman sumber data : "Foto udara ini disediakan oleh BAKOSURTANAL, hibah dari Pemerintah Norwegia, didistribusikan oleh Satuan Tugas Geospasial BRR-BAKOSURTANAL."7. Deng an menandatangani surat perjanjian ini maka pengguna data memahami dan menyadari akan hak BAKOSURTANAL sebagai pemilik data dan pengguna wajib mentaati segala peraturan dan perundang-undangan yang berlaku berkaitan dengan data digital dan cetak.</useconst>
<ptcontac>
<cntinfo>
<cntperp>
<cntper>Dr. Ir. M. Darmawan, M.Sc.</cntper>
<cntorg>BAKOSURTANAL-BRR</cntorg>
</cntperp>
<cntpos>Head of Geospatial Task Force</cntpos>
<cntaddr>
<addrtype>mailing and physical address</addrtype>
<address>BRR, Jl. Ir. M. Thaher No. 20 Lueng Bata</address>
<city>Banda Aceh</city>
<state>Nanggroe Aceh Darussalam</state>
<postal>23247</postal>
<country>Indonesia</country>
</cntaddr>
<cntvoice>+6281316640018</cntvoice>
<cntvoice>0651-636666</cntvoice>
<cntfax>0651-637777</cntfax>
<cntemail>drmoel@yahoo.com</cntemail>
<hours>08.00 - 17.00 Local Time</hours>
<cntinst>Please make an appointment before coming.</cntinst>
</cntinfo>
</ptcontac>
<browse>
<browsen>0421-2349.ecw</browsen>
<browsed>The data format is an ER Mapper Compressed File</browsed>
<browset>Compressed File of ERMapper (ecw)</browset>
</browse>
<datacred>Pusat Pemetaan Dasar Rupabumi dan Tata Ruang BAKOSURTANAL</datacred>
<secinfo>
<secclass>Restricted</secclass>
</secinfo>
<native>Microsoft Windows XP Version 5.1 (Build 2600) Service Pack 2; ESRI ArcCatalog 9.0.0.535; ERMapper 6.4</native>
<crossref>
<citeinfo>
<origin>BAKOSURTANAL</origin>
<pubdate>2006</pubdate>
<pubtime>Unknown</pubtime>
<title>Topographic Line Map 0421-2349</title>
<edition>1</edition>
<geoform>map</geoform>
<serinfo>
<sername>Topographic Line Map with 1:10.000 Scale</sername>
<issue>STG BRR 2006</issue>
</serinfo>
<pubinfo>
<pubplace>BRR NAD-Nias, Jl. Ir. M. Thaher No. 20 Lueng Bata, Banda Aceh, Nanggroe Aceh Darussalam</pubplace>

<publish>Satuan Tugas Geospasial (Geospatial Task Force) BRR-BAKOSURTANAL</publish>
</pubinfo>
</citeinfo>
</crossref>
</idinfo>
<dataqual>
<lineage>
<srcinfo>
<srccite>
<citeinfo>
<origin>BAKOSURTANAL</origin>
<pubdate>2005</pubdate>
<pubtime>Unknown</pubtime>
<title>Digital_Photo_0421-2349</title>
<edition>1</edition>
<geoform>remote-sensing image</geoform>
<serinfo>
<sername>Digital Aerial Photo for Topographic Mapping with 1:10.000, 1:5.000 and 1:2.000 Scale</sername>
<issue>STG BRR 2005</issue>
</serinfo>
<pubinfo>
<pubplace>BRR NAD-Nias, Jl. Ir. M. Thaher No. 20 Lueng Bata, Banda Aceh, Nanggroe Aceh Darussalam</pubplace>
<publish>Satuan Tugas Geospasial (Geospatial Task Force) BRR-BAKOSURTANAL</publish>
</pubinfo>
</citeinfo>
</srccite>
<typesrc>DVD</typesrc>
<srctime>
<timeinfo>
<sngdate>
<caldate>June 2005</caldate>
</sngdate>
</timeinfo>
<srccurr>publication date</srccurr>
</srctime>
</srcinfo>
<procstep>
<procdesc>Aerial photo has been rectified using photogrammetry technique and proceed with mosaicking technique to produce orthophoto.</procdesc>
<procdate>2005</procdate>
- <proccont>
<cntinfo>
<cntperp>
<cntper>Dr. Ir. M. Darmawan, M.Sc.</cntper>
<cntorg>BAKOSURTANAL-BRR</cntorg>
</cntperp>
<cntpos>Head of Geospatial Task Force</cntpos>
<cntaddr>
<addrtype>mailing and physical address</addrtype>
<address>BRR, Jl. Ir. M. Thaher No. 20 Lueng Bata</address>
<city>Banda Aceh</city>
<state>Nanggroe Aceh Darussalam (NAD)</state>
<postal>23247</postal>
<country>Indonesia</country>
</cntaddr>
<cntvoice>0651-636666</cntvoice>
<cntvoice>+6281316640018</cntvoice>

<cntfax>0651-637777</cntfax>
<cntemail>drmoel@yahoo.com</cntemail>
<hours>08.00 - 17.00</hours>
<cntinst>Please make an appointment before coming.</cntinst>
</cntinfo>
</proccont>
</procstep>
<procstep>
<srcused>None</srcused>
</procstep>
</lineage>
<cloud>0 %</cloud>
</dataqual>
<spdoinfo>
<direct>Raster</direct>
<rastinfo>
<rasttype>Pixel</rasttype>
<rowcount>15427</rowcount>
<colcount>15457</colcount>
<vrtcount>1</vrtcount>
</rastinfo>
</spdoinfo>
<spref>
<horizsys>
<planar>
<gridsys>
<gridsysn>Universal Transverse Mercator</gridsysn>
<utm>
<utmzone>46</utmzone>
<transmer>
<sfctrmer>0.999600</sfctrmer>
<longcm>93.000000</longcm>
<latprjo>0.000000</latprjo>
<feast>500000.000000</feast>
<fnorth>0.000000</fnorth>
</transmer>
</utm>
</gridsys>
<planci>
<plance>row and column</plance>
<coordrep>
<absres>0.300000</absres>
<ordres>0.300000</ordres>
</coordrep>
<plandu>meters</plandu>
</planci>
</planar>
<geodetic>
<horizdn>D_WGS_1984</horizdn>
<ellips>WGS_1984</ellips>
<semiaxis>6378137.000000</semiaxis>
<denflat>298.257224</denflat>
</geodetic>
</horizsys>
</spref>
<distinfo>
<distrib>
<cntinfo>
<cntperp>
<cntper>Dr. Ir. M. Darmawan, M.Sc.</cntper>
<cntorg>BAKOSURTANAL-BRR</cntorg>

</cntperp>
<cntpos>Head of Geospatial Task Force</cntpos>
<cntaddr>
<addrtype>mailing and physical address</addrtype>
<address>BRR, Jl. Ir. M. Thaher No. 20 Lueng Bata</address>
<city>Banda Aceh</city>
<state>Nanggroe Aceh Darussalam</state>
<postal>23247</postal>
<country>Indonesia</country>
</cntaddr>
<cntvoice>+6281316640018</cntvoice>
<cntvoice>0651-636666</cntvoice>
<cntfax>0651-637777</cntfax>
<cntemail>drmoel@yahoo.com</cntemail>
<hours>08.00 - 17.00 Local Time</hours>
<cntinst>Please make an appointment before coming.</cntinst>
</cntinfo>
</distrib>
<resdesc>Orthophoto_0421-2349</resdesc>
<stdorder>
<digform>
<digtinfo>
<transize>22.117</transize>
</digtinfo>
</digform>
<ordering>PROSEDUR PELAYANAN PRODUK SURVEI DAN PEMETAAN FORMAT CETAKAN BRR 1. Secara Langsunga. Pengguna (User) datang langsung ke Satuan Tugas Geospasial di lingkungan BRR NAD-Niasb. Mengisi lembar kunjungan (formulir A-1) dan menyampaikan kebutuhan data c. Mengidentifikasi dan melihat contoh produk yang diperlukand. Mengisi formulir pesanan yang telah disediakan (B1) dengan mencantumkan nama dan alamat serta keperluan penggunaan datae. Menyerahkan formulir yang telah diisi dan disetujuif. Menerima produk yang telah dipesan dan menandatangani tanda terima2. Secara Tertulisa. Mengirim permohonan pelayanan produk melalui surat, faximile atau email dengan mencantumkan jenis produk, nomor lembar/nama tema, skala dan jumlah masing-masing produk yang dipesanb. Mengambil pesanan sesuai dengan jadwal yang disepakatiPROSEDUR PELAYANAN PRODUK SURVEI DAN PEMETAAN FORMAT DIGITALa. Menyampaikan permohonan secara tertulis melalui surat, faximile atau email yang ditujukan kepada kepala BAKOSURTANAL, cq. Kepala Pusat Pelayanan Jasa dan Informasi dengan tembusan kepada Koordinator Satuan Tugas Geospasial di lingkungan BRR NAD-Nias dengan mencantumkan jenis produk. Nomor, skala, jumlah, lokasi data dan keterangan lain yang diperlukan.b. Menandatangani Surat Perjanjian Penggunaan Data/Lisensi khususnya untuk data digitalc. Produk pesanan dapat diambil langsung oleh pemesan di BRR NAD-NiasPROSEDUR PELAYANAN FOTO UDARA DAN PETA CITRAa. Mengirimkan Surat Permohonan untuk mendapatkan Security Clearence penggunaan foto udara/citra satelit ke : Kepala Badan Pelaksana Rehabilitasi dan Rekonstruksi NAD-Nias, dengan tembusan kepada Kepala BAKOSURTANAL u.p. Koordinator Satuan Tugas Geospasial di BRR NAD-Niasb. Mencantumkan dengan jelas nomor foto dan jumlah foto yang diperlukanc. Pesanan dapat diambil di Satuan Tugas Geospasial di BRR NAD-Nias</ordering>
</stdorder>
</distinfo>
<metainfo>
<metd>20070125</metd>
<metc>
<cntinfo>
<cntorgp>
<cntorg>BAKOSURTANAL BRR</cntorg>
<cntper>Teguh Mulyadi, S.T.</cntper>
</cntorgp>
<cntpos>Assistance Supervisor of Geospatial Task Force</cntpos>

<cntaddr>
<addrtype>mailing and physical address</addrtype>
<address>BRR, Jl. Ir. M. Thaher No. 20 Lueng Bata</address>
<city>Banda Aceh</city>
<state>Nanggroe Aceh Darussalam (NAD)</state>
<postal>23247</postal>
<country>Indonesia</country>
</cntaddr>
<cntvoice>081315011167</cntvoice>
<cntvoice>0651-636666</cntvoice>
<cntfax>0651-637777</cntfax>
<cntemail>alfarisi@bakosurtanal.go.id</cntemail>
<hours>08.00 - 17.00 local time</hours>
<cntinst>Please make an appointment before coming.</cntinst>
</cntinfo>
</metc>
<metstdn>FGDC Content Standards for Digital Geospatial Metadata</metstdn>
<metstdv>FGDC-STD-001-1998</metstdv>
<mettc>local time</mettc>
<metextns>
<onlink>http://www.esri.com/metadata/esriprof80.html</onlink>
<metprof>ESRI Metadata Profile</metprof>
</metextns>
</metainfo>
</metadata>

Appendix B: Relatedness Evaluation

This Table shows the evaluation of relatedness of several nodes in Figure 6.2

Relatedness (Я) : [percentage of common datasets (ΣC) with respect to used dataset (ΣS)] divided by topological distance (Δ)

		R01	Я	R02	Я	R11	Я	R12	Я	R21	Я	R22	Я	R23	Я	R31	Я	R32	Я
R01	Δ		∞	∞	0	1	100	∞	0	2	8,33	∞	0,00	∞	0	3	4,76	3	3,33
	ΣS			0		1		1		6		4		4		7		10	
	ΣC			0		1		0		1		0		0		1		1	
	P			N		Y		N		Y		N		N		Y		Y	
R02	Δ	∞	0		∞	∞	0	1	100	2	8,33	2	12,50	∞	0	3	4,76	3	3,33
	ΣS	0				5		1		6		4		4		7		10	
	ΣC	0				0		1		1		1		0		1		1	
	P	N				N		Y		Y		Y		N		Y		Y	
R11	Δ	1	100	∞	0		∞	∞	0	1	33,33	∞	0,00	∞	0	2	14,29	2	10,00
	ΣS	1		5				2		6		5		5		7		10	
	ΣC	1		0				0		2		0		0		2		2	
	P	Y		N				N		Y		N		N		Y		Y	
R12	Δ	∞	0	1	100	∞	0		∞	1	33,33	1	50,00	∞	0	2	14,29	2	10,00
	ΣS	1		1		2				6		4		5		7		10	
	ΣC	0		1		0				2		2		0		2		2	
	P	N		Y		N				Y		Y		N		Y		Y	
R21	Δ	2	8,33	2	8,33	1	33,33	1	33,33		∞	2	12,50	∞	0	1	100,00	1	70,00
	ΣS	6		6		6		6				8		10		7		10	
	ΣC	1		1		2		2				2		0		7		7	
	P	Y		Y		Y		Y				N		N		Y		Y	
R22	Δ	∞	0	2	12,50	∞	0,00	1	50,00	2	12,50		∞	∞	0	3	7,41	1	50,00
	ΣS	4		4		5		4		8				8		9		10	
	ΣC	0		1		0		2		2				0		2		5	
	P	N		Y		N		Y		N				N		N		Y	
R23	Δ	∞	0	∞	0,00	∞	0,00	∞	0,00	∞	∞	∞	0,00		∞	∞	0,00	∞	0,00
	ΣS	4		4		5		5		10		8				11		14	
	ΣC	0		0		0		0		0		0				0		0	
	P	N		N		N		N		N		N				N		N	
R31	Δ	3	4,76	3	4,76	2	14,29	2	14,29	1	100,00	3	7,41	∞	0		∞	2	35,00
	ΣS	7		7		7		7		7		9		11				10	
	ΣC	1		1		2		2		7		2		0				7	
	P	Y		Y		Y		Y		Y		N		N				N	
R32	Δ	3	3,33	3	3,33	2	10,00	2	10,00	1	70,00	1	50,00	∞	0	2	35		∞
	ΣS	10		10		10		10		10		10		14		10			
	ΣC	1		1		2		2		7		5		0		7			
	P	Y		Y		Y		Y		Y		Y		N		N			

Legend

white	Same group, but no relationship at all	yellow	relationship exists
pink	Other group, no relationship at all		ignored; dataset itself

Acronyms and Abbreviations

ANSI	American National Standards Institute
API	Application Programming Interface
APSDI	Asia Pacific Spatial Data Infrastructure
ASCII	American Standard Code for Information Interchange
BAKOSURTANAL	Badan Koordinasi Survey dan Pemetaan Nasional (Indonesian National Coordinating Agency for Surveys and Mapping)
BLOB	Binary Large Object
BRR	Badan Rehabilitasi dan Rekonstruksi (Rehabilitation and Reconstruction Agency for Aceh and Nias area)
DAG	Directed Acyclic Graph or also known as Digraph
DBMS	Database Managemen System
DTM	Digital Terrain Model
DTD	Document Type Description
GIS	Geographic Information System
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
HTML	Hypertext Markup Language
IDSN	Infrastruktur Data Spasial Nasional (National Spatial Data Infrastructure)
IPF	Institute of Photogrammetry and Remote Sensing
ISO	International Organization for Standardization
JPG/JPEG	Joint Photographic Expert Group
MS	Microsoft
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OGC	Open GIS Consortium
PDF	Adobe Portable Document Format
PNG	Portable Network Graphic
RDBMS	Relational Database Management System
SAR	Synthetic Aperture Radar
SISN	Sistem Informasi Spasial Nasional National Spatial Information System
SQL	Sructured Query Language
SPOT	Satellite Pour l'Observation de la Terre
SUTRS	Simple Unstructured Text Record Syntax (data exchange format widely used in Z39.50)

TCP/IP	Transfer Control Protocol / Internet Protocol
TMIS	Topographic Mars Information System
USGS	United States Geological Service
USMARC	Standards for the representation and communication of bibliographic and related information in machine-readable form. Also known as MARC21.
UTM	Universal Transverse Mercator
WFS	Web Feature Service
XML	Extensible Markup Language

Glossary

Catalogue = A list of collections. In this dissertation, collections refer to geospatial data collections, so the term catalogue is used for listing and organizing of geospatial data. A catalogue system uses database management system software as its implementation.

Database Management System (DBMS) = a software package that provides all the functionality required for organization and manipulation of data.

Dataset = A raster data representation, usually it is a combination of a bitmap data file and its metadata file, or a file directory which contains a series of data files which belong to a single spatial data entity. In a directed acyclic graph data model which is used here, each dataset is a node or vertex, and each processing step link is a link or edge. In practical implementation, dataset is a raster data and its attributes are extracted from raster data metadata.

Dataset Level = Categorization of spatial dataset, where the raw dataset is assigned as level-0, the rectified image and DTM is assigned as level-1, GIS and analysed dataset as level-2, and a presentation map as level-3.

Degree = Level a node (dataset), which is calculated from a number of predecessor of a node and number of successor of a node. Degree can be divided into *In-degree* (number of predecessor only) dan *Out-degree* (number of successor only).

Depth of link = Total number of steps of a path from one dataset to another dataset. This value is used in our design to limit the iteration and to avoid recursive / cyclic structure.

Inbound Path = A path (link) which comes from other node to the current node. The node where this link comes from is called predecessor node.

Neighbourhood = In graph theory, neighbourhood (N_G) of a vertex (V) is a list of all adjacency vertices and all edges which connecting the adjacency vertices. In this dissertation, the usage of neigbourhood is extended to the *neighbourhood of a pair of dataset* which is a relationship between two data based on their processing steps in the past (predecessor)

Outbound Path =A path (link) which comes from current node and points to other node. The node where this link points to is called successor node.

Predecessor = Previous. In a directed graph, predecessor node is a node before the current node or a node which has a link points to the current node. In this dissertation, this refers also to source dataset

Processing direction = Similar to creation direction of a dataset. It is a path from a source dataset to its destination dataset. A destination dataset usually has a higher dataset level.

Relatedness = This term is used in this thesis to figure a direct or indirect connection between two datasets based on their creation history. The calculated value of relatedness is ranged from 0 to 100, where 0 means no relatedness, and 100 is the highest possible relatedness (one dataset is created from only one other dataset)

Successor = Next. In a directed graph, successor node is a node after the current node. In this dissertation, this refers to destination dataset.

References:

Amhar, F., 2001. Konsep Terpadu Sistem Basis Data Untuk Infrastruktur Data Spasial Nasional, Forum Nasional Tata Ruang 2001 (*Integrated Concept of Database System for National Spatial Data Infrastructure, National Forum of Spatial Planning*). Bakosurtanal internal report

ANSI, 2008. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Weblink. http://www.ansi.org, last accessed September 2008

BAKOSURTANAL, 2005. Panduan Pembangunan Metadata Data Spasial (Spatial Data Metadata Development Guideline). Technical Working Group Clearinghouse, Marine and Coastal Resource Management, Bakosurtanal, Cibinong – Indonesia. Bakosurtanal Internal Report. **http://www.idsn.or.id/images/stories/artikel/Pembangunan_Metadata.pdf**, last accessed: 30 Jan 2009

BAKOSURTANAL, 2008. Indonesian Agency for Surveys and Mapping,. http://www.bakosurtanal.go.id, last accessed, March 2008

Bang-Jensen, J., Gutin, G.Z., 2008. Digraphs: Theory, Algorithms and Applications. Springer Monographs in Mathematics (2nd ed.), Springer-Verlag, London

Bojinski, S., Schaepman, M., Schläpfer, D., Itten, K., 2002. SPECCHIO: a Web Accessible Database for the Administration and Storage of Heterogeneous Spectral Data. ISPRS Journal of Photogrammetry & Remote Sensing. Elsevier, Vol 57 (2002), pp. 204-211

Bose, R., Frew, J., 2004. Composing Lineage Metadata with XML for Custom Satellite-Derived Products. Proceedings of the IEEE 16th International Conference on Scientific and Statistical Database Management (SSDBM'04), pp. 275-284

Cha C., Moon B., Acharya A., Shock, C., Sussman, A., Saltz, J., 1997. Titan: A High-Performance Remote-Sensing Database. Proceedings of 13th International Conference on Data Engineering, 7-11 April 1997, pp. 375 – 384

Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein C., 2000. Introduction to Algorithms. MIT Press and McGraw-Hill, Section 23.4: Topological sort, pp.485–487.

Coupe, J.M., 2004. Data Handling Challenges for Remote Sensing System. Proceedings of IEEE Aerospace Conference 2004, Vol 4, pp. 2263-2268

Di, L., 2003. The Development of Remote Sensing Related Standards at FGDC, OGC, and ISO TC211. Proceedings of IGARSS 2003, Vol 1, pp. 643-647

Dorninger, P., 2003. A Topographic Mars Information System, Concepts for Management, Analysis and Visualization of Planet-Wide Data. Dissertation - TU-Wien, http://www.ub.tuwien.ac.at/diss/AC04223140.pdf, last accessed 20 Jan 2009

Doucette, P., Agouris, P., Musavi, M., and Stefanidis, A., 1999. Automated Extraction from Aerial Imagery Using Kohonen Learning and GIS Data. Lecture Notes in Computer Science (LNCS), Vol. 1797/1999, pp. 20-33, Springer Verlag – Berlin/Heidelberg.

ESRI, 2009. http://www.esri.com/software/arcgis/arcinfo/success-stories.html, last accessed March 2009

ESRI SDE, 2009. http://www.esri.com/software/arcgis/geodatabase/storage-in-an-rdbms.html, last accessed 20 October 2009

FGDC, 2000. Federal Geographic Data Committee, Content Standard for Digital Geospatial Metadata Workbook (For use with FGDC-STD-001-1998). Washington, DC., May 1, 2000. http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf, last access 20 Jan 2009

FGDC, 2002. Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata (FGDC-STD-012-2002). Washington, DC. http://www.fgdc.gov/standards/projects/FGDC-standards-projects/csdgm_rs_ex /MetadataRemoteSensingExtens.pdf, last accessed 20 Jan 2009

FGDC Metadata Tools, 2008. Federal Geographic Data Committee, http://www.fgdc.gov/metadata/geospatial-metadata-tools weblink, last accessed March 2008.

Garnett, J., Owens B., 2003. Geoserver Database Research. Internal Report, Refractions Research Inc. http://vwfs.refractions.net/docs/Database_Research.pdf, last accessed 30 Jan 2009

GeoNetwork, 2010. FAO GeoNetwork Opensource http://geonetwork-opensource.org/, last accessed March 2009

GeoTIFF, 2009, http://www.gisdevelopment.net/technology/ip/mi03117pf.htm, last accessed March 2005.

Girard, C. M., 2003. Processing of Remote Sensing Data, A.A. Balkema, Lisse

Goodchild, M.F., Zhou, J., 2003. Finding Geographic Information: Collection Level Metadata. Geoinformatica 7:2, Springer Netherlands, The Netherlands, pp. 95-112

Guo, D., Peuquet, D.J., Gahegan, M., 2003. ICEAGE: Interactive Clustering and Exploration of Large and High Dimensional Geodata. GeoInformatica, Springer Netherlands, Vol 7, Number 3, September 2003, pp. 229-253

Hari Ginardi, R.V., 2002. Direct NOAA Imagery Exctraction System for Data Acquisition of Rice Growth Modelling, Journal of GIS, Remote Sensing and Dynamic Modelling" No. 2 Year 2002, pp. 1-23, SEAMEO BIOTROP, ISSN 1412-2049.

Harvey, N. R., Theiler, J. Brumby, S.P., Perkins, S., Szymanski, J.J., Bloch, J.J., Porter, R.B., Galassi, M., Young, A.C. , 2002. Comparison of GENIE and Conventional Supervised Classifiers for Multispectral Image Feature Extraction. IEEE Transaction on Geoscience and Remote Sensing, Vol-40, Issue 2, 2002, pp. 393-404

Holowczak, R.D., Artigas, F.J., Chun, S., Cho, J.S., Stone, H.S., 2002. An Experimental Study on Content-Based Image Classification for Satellite Image Databases. IEEE Transactions on Geoscience and Remote Sensing, Vol 40, No 6, June 2002, pp. 1338-1347

Hsu, W., Lee, M.L, Zhang, J, 2002. Image Mining: Trends and Developments. Journal of Intelligent Information Systems, Springer Netherlands, Vol 19, Number 1, July 2002, pp. 7-23

Huang, Z., Fang, Y., Chen, B., Peng, X., 2006. A Spatial Metadata Service System Based on Z39.50 Protocol. Proceedings of IGARSS 2006, pp. 2829-2831

Ikawati, Y., Setiawati D. R., 2009. Survei dan Pemetaan Nusantara *(National Surveying and Mapping)*, Bakosurtanal, Jakarta, ISBN: 978-602-95542-0-5

ISO TC211 Scope, 2007. http://codazzi4.igac.gov.co/gsdi5/Gsdi5_Cartagena/mich/objetivo2_2001/stn/ISO-TC 211 Scope.htm, last accessed March 2007.

Jea K.F., Chang T.P., 2008. A Lock-based Protocol for XML Metadata Knowledge Bases. ISDA '08. Eighth International Conference on Intelligent Systems Design and Applications, 26-28 Nov. 2008, Vol 3, pp.23-28

Jianting, Z., Gertz, M., Gruenwald, Le., 2009. Efficiently managing large-scale raster species distribution data in PostgreSQL. Proceeedings of the 17th ACM SIG SPATIAL International Conference on Advances in Geographic Information System, pp. 316-325

Kahn, A.B., 1962. Topological Sorting of Large Networks. Communication of the ACM, Vol 5, Issue 11, pp. 558-562

Keating, J.P. 1993. Pitman's Measure of Closeness: A Comparison of Statistical Estimators. Society for Industrial and Applied Mathematics (SIAM), pp. 25-37.

Knuth D.E., 1997. The art of computer programming. 3rd Edition, Addison Wesley

Kwoh, L.K.,1997. A Catalog-Browse System with Quick-Look Images for SPOT, ERS, and RADARSAT Data Archives. Remote Sensing – A Scientific Vision for Sustainable Development. Proceedings of IGARSS 1997, Vol 3, 1997, pp. 1257-1259

LACData, 2009. http://www2.ncdc.noaa.gov/docs/podug/html/c3/sec3-2.htm, last accessed March 2009.

Li, J; Narayanan, R.M., 2004. Integrated Spectral and Spatial Information Mining in Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing, Volume 42, Issue 3, March 2004, pp. 673-685

Li, L., 2001. Distributed Geospatial Data Access on the WWW. MSc Thesis at The University of New Brunswick, http://www.cs.unb.ca/tech-reports/files/TR01_139.pdf, last accessed 30 Jan 2009

Li Q., Wang S., Chen B., 2005. Remote Sensing Image Distribute System Supported by Metadata. Proceedings of IGARSS 2005, Volume 2, 25-29 July 2005, pp. 791-794,

Matindas, R W., Puntodewo, Purnawan, B., 2004. Development of National Spatial Data Infrastructure in Indonesia. FIG Working Week 2004, Athens – Greece. http://www.fig.net/pub/athens/papers/ts02/ts02_1_matindas_et_al.pdf , last accessed 30 Jan 2009

Murray, D., Lutz, D., 2009. ESRI's Spatial Data Engine: A Seamless GIS Solution

http://proceedings.esri.com/library/userconf/proc96/to100/pap094/p94.htm

Nance, K.L., Hay, B., 2005. Automatic Transformations between Geoscience Standards Using XML. Computers & Geoscience 31, Elsevier, Issue 9, Nov 2005, pp. 1165-1174

NOAA Satellite and Information Service, Metadata Tools, 2008. http://www.ncddc.noaa.gov/metadataresource/metadata-tools, last accessed 30 October 2008

OGC, 2009. http://www.opengeospatial.org/resource/products, last accessed 15 March 2009

ORACLE Spatial 11g, 2009. http://www.oracle.com/technology/products/spatial/pdf/11gr2_collateral/spatial11gr2_datasheet.pdf, last accessed 30 October 2009

POSTGRESQL, 2008. http://www.postgresql.org, last access, 3 September 2008

Puntodewo, Nataprawira R.,, 2004. Indonesian Geospatial Data Clearinghouse. 3rd FIG Regional Conference for Asia and the Pacific, Jakarta 2004, http://www.fig.net/pub/jakarta/papers/ts_02/ts_02_2_puntodewo_nataprawira.pdf , last accessed 30 Jan 2009

Richard, J . A., 1986. Remote Sensing Digital Image Analysis: an Introduction, Springer Verlag, New York

Ruan, N.; Huang, N.; Hong, W., 2006. Semantic-Based Image Retrieval in Remote Sensing Archive: An Ontology Approach. Proceedings of IGARSS 2006, pp. 2903 - 2906

Schroeder, M., Seidel, K., Datcu, M., 1997. Gibbs Random Field Model for Image Content Characterization. Proceedings of IGARSS 1997, Vol. 1. pp. 258-260

Schlessinger B., 2005. Imagery Metadata. Open Geospatial Consortium Inc (OGC) Discussion Paper, doc number: OGC 05-015, http://portal.opengeospatial.org/files/?artifact_id=8826 , last accessed 30 Jan 2009

Seidel, K.; Schroder, M.; Rehrauer, H.; Schwarz, G.; Datcu, M., 1998. Query by Image Content From Remote Sensing Archives, Proceedings of IGARSS '98, Vol 1, pp. 393 - 396

Shi J., Chen Y., Liu C., 2009. Database System for Archiving and Managing Remote Sensing Images. First International Workshop on Database Technology and Applications, 25-26 April 2009, pp. 536-538

Sukmayadi D., Amhar F., Adhibroto A., Prijanto A., 2004. Geo-Spatial Database Design and Implementation at the Mapping Center for Topographic & Spatial Arrangement (PDRTR) of BAKOSURTANAL, 3rd FIG Regional Conference, http://www.fig.net/pub/jakarta/papers/ts_02/ts_02_3_sukmayadi_etal.pdf, last accessed 30 Jan 2009

Tarjan, R.E., 1976. Edge-disjoint Spanning Trees and Depth-First Search. Acta Informatica, Springer Verlag, Berlin, Vol 6, #2, pp. 171-185

USGS Metadata Tools, 2008. United States Geological Survey, http://geology.usgs.gov/tools/metadata/ , last accessed September 2008

Vatsavai, R.R, Burk, T.E., Shekhar, S., Hansen, M.H., 2001. An Efficient Query Strategy for Integrated Remote Sensing and Inventory (Spatial) Databases. Proceedings of International Conference on Scientific and Statistical Database Management, SSDBM 2001, pp. 115-123

Vinhas, L, De Souza , R.C.M, Camara, G, 2003. Image Data Handling in Spatial Databases. Brazilian Symposium on GeoInformatics, GeoInfo2003, http://www.geoinfo.info/geoinfo2003/papers/geoinfo2003-16.pdf, last accessed July 2006

Walcher, W., 1997. Design Aspects of Information Systems for Remote Sensing Image Data. TU-Graz Dissertation.

Wang, P., Yang, C., Yu, Z, Ren Y, 2004. A Load Balance Algorithm for WMS [Web Map Service], Proceedings of IGARSS 2004, Vol 5, pp. 2920-2922

Wang, X., Wang, S., Wei, W., 2005. Study on Remote Sensing Image Metadata Management and Issue. Proceedings of IGARSS 2005, Vol 1, pp. 612-615

Yu S., Qiang H., Jing S.,2006. A Framework of XML-Based Geospatial Metadata System. APWeb Workshop 2006, Springer Verlag Berlin Heidelberg

Zhang, J., Hsu, W., Lee, M.L, 2001. An Information-Driven Framework for Image Mining. Lecture Notes in Computer Science: Database and Expert Systems Applications, Springer Berlin, Volume 2113/2001, pp. 232-242

CURRICULUM VITAE

Name	Raden Venantius HARI GINARDI
Date of Birth	18 May 1965
Nationality	Indonesian

Lecturer and Researcher, Faculty of Information Technology, ITS - Surabaya University of Technology (Institut Teknologi Sepuluh Nopember)

EDUCATION

2001 to present	Ph.D Program, Institut of Photogrammetry and Remote Sensing, Vienna University of Technology, Vienna – Austria
1998 to 2000	M.Sc in Information Technology for Natural Resources Management, Bogor Agricultural University, Bogor – Indonesia
1984 to 1991	Engineer in Electrical Engineering, ITS – Surabaya University of Technology (Institut Teknologi Sepuluh Nopember) , Surabaya – Indonesia

RESEARCH

- The Use of IBM-PC as a Real-time Multi-device Controller, 1991, Bachelor Script, Electrical Engineering Dept., Faculty of Industrial technology, IST- Surabaya
- Participation / Speaker in the International Symposium of PERHIMPI (Indonesian Agri-meteorology Association), Bogor – 1999
- Participation / Speaker in the ISPRS Working Group VI International Congress, Bandung 1999.
- "Direct NOAA Imagery Exctraction System for Data Acquisition of Rice Growth Modelling" "Journal of GIS, Remote Sensing and Dynamic Modelling" No. 2 Year 2002, pp. 1-23, SEAMEO BIOTROP, ISSN 1412-2049.

Vienna, 29 July 2010