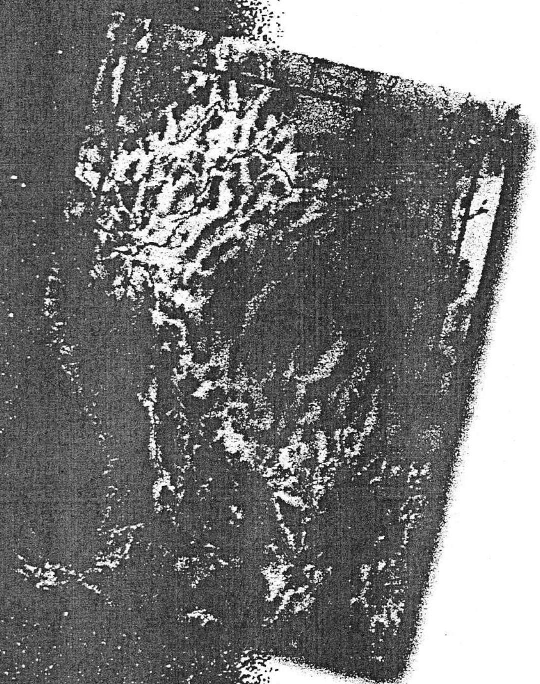
f= papers.

99-ES

ISSSR

SYSTEMS AND SENSORS FOR THE NEW MILLENNIUM



LAS VEGAS, NEVADA

THE TROPICANA HOTEL

OCTOBER 31 – NOVEMBER 4, 1999

TUTORIALS / SYMPOSIUM PAPERS
SYMPOSIUM POSTERS / WORKSHOPS / EXHIBITS

ISPRS

International Society for
Photogrammetry & Remote Sensing

An Efficient Storage, Analysis, and Retrieval System for SeaWiFS Data using SeaDAS, Sem-ODB technology, and the World Wide Web*

Naphtali Rishe, Martha E. Gutierrez, Alejandro Gonzalez, Debra Lee Davis
High Performance Database Research Center
NASA Regional Application Center
Florida International University, Miami, Florida
hpdrc@cs.fiu.edu

ABSTRACT

Ocean Color is important in the study of the ocean's primary production of phytoplankton and its global biogeochemistry. Studying the growth of marine phytoplankton is useful in order to understand the ocean's role in the global carbon cycle.

Remote sensing data is becoming very important for marine and coastal environments. Over the years the remotely sensed data collected have grown exponentially, necessitating information-processing technologies to store and retrieve this data efficiently. The High-Performance Database Research Center (HPDRC) at the School of Computer Science at Florida International University has developed the Semantic Database System Sem-ODB that is efficient in storing and retrieving non-conventional data such as spatial data.

We use Sem-ODB to store SeaWiFS data. In addition, we have developed a suite of tools to analyze Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data and facilitate its storage and retrieval to and from Sem-ODB. These can then be queried using the World Wide Web or NASA's analysis software, SeaDAS. Through either of these methods, algorithms can be applied to the data to give scientists a platform on which to study ocean color on a global basis.

1. INTRODUCTION

The influx of satellite imagery has caused the scientific community to explore more efficient ways of storing and retrieving the data. One such satellite imagery is referred to as Ocean Color Imagery and the instrument currently used to acquire it is called the Sea-viewing Wide Field-of-view Sensor (SeaWiFS). The SeaWiFS instrument is aboard the OrbView-2 satellite.

The semantic database technology (Sem-ODB), developed at the High Performance Database Research Center (HPDRC), provides efficient storage and access to the data.
As a first step in database design, a semantic schema must be developed which describes the SeaWiFS Data through objects and relations, in a way that is easily comprehended. The SeaWiFS semantic database can be queried through the World Wide Web (WWW). The SeaWiFS Data Analysis System (SeaDAS), a NASA image analysis package for processing SeaWiFS data, was integrated with Sem-ODB technology to make the storage and retrieval of various levels of processed data more accessible and to take advantage of Sem-ODB's unique features. Sem-ODB in conjunction with the SeaWiFS project is available to the scientific community for study of ocean color data.

1.1 HPDRC

The High Performance Database Research Center (HPDRC) is a division of Florida International University (FIU), School of Computer Science. It conducts research on database management systems and various applications, leading to the development of new types of Database Management Systems (DBMS), new database techniques, and the refinement of existing ones.

HPDRC's largest project is the development of Sem-ODB, a prototype massively parallel Semantic/Object Oriented DBMS. Our system is useful for most typical database applications, as well as for specialized domains such as Earth Sciences [RISH95].

1.2 RAC

The NASA Regional Applications Center (RAC) at FIU is a subdivision of HPDRC. NASA has

* This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095, NAS5-97222, and NAG5-6830) and NSF (CDA-9711582, IRI-9409661, HRD-9707076, and ANI-9876409).

p. 214-218

established approximately 15 RAC's across the country at this time. The RAC Program was initiated by NASA Goddard Space Flight Center's (GSFC) Applied Information Sciences Branch, to extend the benefits of its information technology research and cost-effective system development to a broader user community. The RAC objectives are based on the goal of fostering the use of environmental and Earth resource data by regional institutions. The ultimate goal of the RAC is to establish a fundamental set of remote sensing technologies that can be assembled by a specific user community, to meet the information needs of that community.

2.0 SEM-ODB

Sem-ODB, HPDRC's Semantic DBMS, is based on the Semantic Binary Model. In the Semantic Binary Model, the information is represented by logical associations (relations) between pairs of objects and by the classification of objects into categories. The Semantic Binary Model is a natural and convenient way of specifying the logical structure of information and for defining the concepts of an application's world. It is represented in the form of a semantic binary schema [RISH92].

2.1 DESCRIPTION

The semantic database models are potentially more efficient than the conventional models for two main reasons. The first is that all the physical aspects of the representation of information by data are invisible to the user and the second is that the system knows more about the meaning of the user's data and about the meaningful connections between such data. The first reason creates a potential for optimization by allowing more changes without affecting the user programs. The second allows this knowledge to be utilized to organize the data so that meaningful operations can be performed faster at the expense of less meaningful operations [RISH92].

The mathematical abstraction of the relational model has allowed the introduction of powerful and easy-to-use languages for retrieval and updates of databases. The semantic model offers a higher degree of abstraction, which results in more concise user programs, speedier processing (due to optimization), and a wealth of other features. Relational databases are good for general conventional database applications. However, in situations where the structure of information is complex, or where greater flexibility is required (objects with unknown identifiers, or objects moving from one category to another, etc.), or where non-conventional data is involved (spatial data, long text, images, etc.), semantic databases need to be considered.

The efficient retrieval and updates are a requirement of the semantic database. Requests are maximized by decomposing queries into atomic retrieval operations and each atomic retrieval request normally requires only one disk access. A transaction is composed of a set of facts. These facts can state that the objects belong to a category, they can state that there is a relationship between objects or they can be fact relating objects to data, such as numbers, texts, dates, images, etc. HPDRC's Semantic DBMS contains semantic facts and inverted semantic facts. This fact inversion scheme assures efficiency of queries including range queries and content access and also exhibits low entropy of data blocks, which facilitates compression [RISH92].

2. INTRO TO OCEAN COLOR DATA

SeaWiFS is on board OrbImage's OrbView-2 satellite. The SeaWiFS sensor is used to measure global ocean color. Little changes in ocean color allow scientists to determine the number of phytoplankton in the oceans. The reason why these microscopic marine plants can be detected is because the SeaWiFS sensor measures the world's oceans in the visible spectrum. The color varies with concentration of chlorophyll, and other plant pigments in the water. The more of these pigments in the water, the greener the water. This information has scientific and practical applications. Ocean color is important to the oceanographic community for the study of ocean primary production and global biogeochemistry.

Phytoplankton contains chlorophyll, a green pigment. Near coastlines, chlorophyll-dissolved organic material, and suspended sediments from rivers and lagoons affect the color of the ocean. By observing the color of different parts of the oceans, scientists can measure the amount of these materials in the water.

Measuring these materials is important in understanding the role of the oceans in the carbon cycle. Phytoplankton removes carbon dioxide from the atmosphere for internal use, and scientists want to measure this exchange of carbon dioxide, and its effect on the Earth.

Since the industrial revolution, an increase of carbon dioxide led scientists to believe that the Earth's Temperature is rising. Using SeaWiFS, scientists can have a more accurate predication to the response of the global climate to changes such as carbon dioxide. The

SeaWiFS project is part of NASA's Mission to Planet Earth (MTPE), which is designed to look at our planet from space to better understand it as a system in both behavior and evolution.

The SeaWiFS Sensor works as a follow-on sensor to the Coastal Zone Color Scanner (CZCS), which ceased operations in 1986. SeaWiFS brought the ocean color scientific community a welcomed and improved renewal of the ocean color remote sensing capability lost in CZCS.

The SeaWiFS orbiting sensor can view every square kilometer of cloud-free ocean every 48 Hours, therefore this ocean data retrieved by SeaWiFS is a valuable tool on obtaining the abundance of phytoplankton on the global scale.

3. THE SEAWIFS INSTRUMENT

The SeaWiFS Sensor contains 8 Bands:

Band	Center Wavelength (NM)
1	412
2	443
3	490
4	510
5	555
6	670
7	765
8	865

The SeaWiFS Sensor has 1 km and 4 km resolutions of phytoplankton and marine primary productivity using visible and near infrared imagery.

Field of View:	Nominally 1.6 mRads
Side-To-Side Scan Angle:	+/-58.3 Degrees
Tilt:	20 Degrees forward or aft

SeaWiFS will include periodic sweeps of both an onboard solar diffuser plate and the moon. The diffuser will provide short-term checks, and the lunar will monitor long-term trends in the instrument's radiometric stability.

4. SEAWIFS ALGORITHMS

Many Algorithms can be applied to the SeaWiFS Data. SeaWiFS is able to detect cloud and ice at high latitudes. SeaWiFS is also able to classify Coccolithophore Blooms in ocean color imagery. SeaWiFS can produce images compatible and related to CZCS imagery, and can classify Chlorophyll-*a*, which is the most appropriate measure of viable phytoplankton biomass (i.e. phytoplankton which are actively growing).

The SeaWiFS data is currently being distributed by the NASA Goddard Distributed Active Archive Center (DAAC), to researchers for educational purposes. It is distributed in multiple levels of processing: level 0, level 1A, level 2 and level 3. The differences between these is described below:

Level 0 - has raw radiance counts from all bands as well as spacecraft and instrument telemetry.

Level 1A - is Level 0 data, with appended and reformatted calibration, navigation, selected spacecraft telemetry .

Level 2 - is generated from Level 1A data and it applies sensor calibration, atmospheric correction, and bio-optical algorithms (such as Chlorophyll-A and CZCS)

Level 3 - is a composite of many images corresponding to a period of one-day, 8 days, a month, or a year.

Every data set above Level 0 is stored in NCSA's developed HDF algorithm (Hierarchical Data Format). HDF is a self-describing extensible file format using tagged objects that have a special meaning. The idea is to store both, a known format description and the data in the same file. This format is widely used by NASA's DAAC as the general format for satellite digital data [NCSA93]. This format is so popular because the same format can be used to describe many different data sets that have different meanings.

A semantic database representation of this model was developed for the Level 1A dataset. Level 1A was chosen as the source for the SeaWiFS database. All of the other formats above Level 1A are processed using parameters, given by the user. Since we have a Level 1A Database, we can perform the calibration and bio-optical algorithms after retrieval, thereby reducing redundancy in the database.

The database is designed using the Level 1A HDF format specifications. A scientific data set is a set containing a multidimensional array and information describing the data in the array. A Vgroup is a general grouping structure containing any kind of HDF Object that a user wishes to include.

The SeaWiFS Level 1A global attributes sets are mission and documentation, data time, data quality, file metrics, scene coordinates.

The SeaWIFS Level 1A Vgroups that are a function of the number of scanlines in the file are Scan-Line attributes, raw SeaWiFS data, converted telemetry, navigation, sensor tilt, and calibration. In the semantic database schema we have similar categories, and relations between the categories. [SEAWI95]

The Mission Documentation Set is separated into two categories MISSION_PRODUCT, and STATION. This is to allow for two data sets from different stations to be linked (by the relation processed_at). The Category DATA_TIME (which corresponds to the HDF equivalent), is related to MISSION_PRODUCT by ingest_time. Data quality and file metrics are all attributes in the category SCENE.

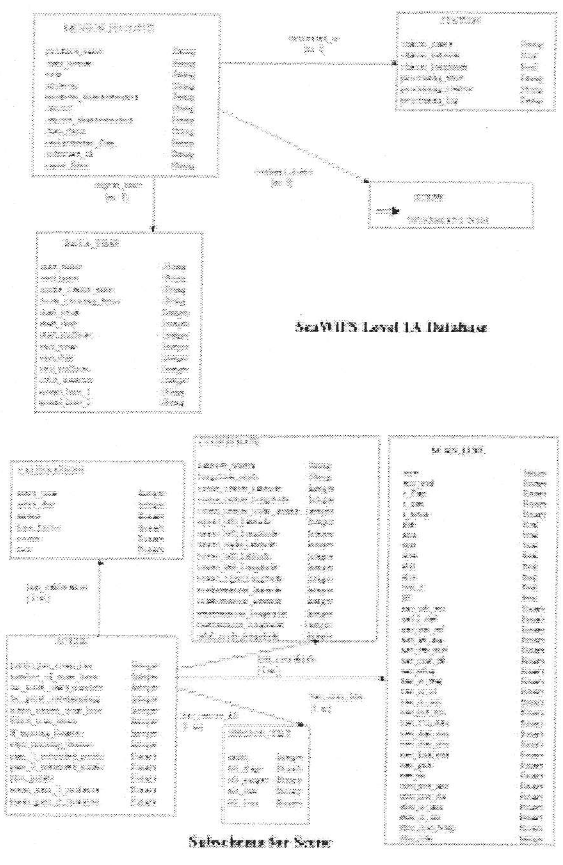


Figure 1 – SeaWiFS Level 1A Semantic Schema

5. INTRODUCTION TO SEADAS

SeaDAS (SeaWiFS Data Analysis System) is an image analysis package for the processing, display, analysis, and quality control of SeaWiFS data products. [SEAD99] SeaDAS is programmed in the C language and Research Systems' Interactive Data Language (IDL).

IDL is a complete environment for the interactive analysis and visualization of data. This is a timesaving alternative to programming in FORTRAN or C. Using IDL, tasks which require days or weeks of programming, can be accomplished in hours. Creating a user interface, and exploring data sets is very easy and accomplished with a few commands.

The Seadas-Sem-ODB Integration enhances the functionality of SeaDAS by allowing a user to randomly access different levels of data, to query the data arbitrarily, it allows parallelization, and a plethora of database services such as backup and recovery. Sem-ODB provides a clean, modular, and simple programming interface.

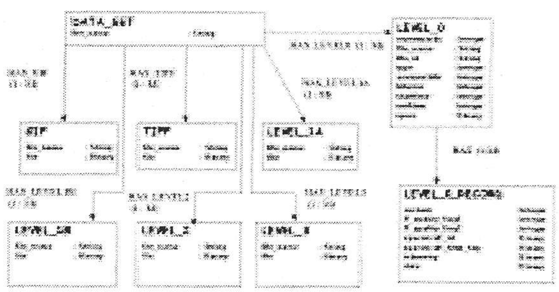
SeaDAS is constantly changing, and many of the algorithms it incorporates are complex and required years of manpower to develop. For this reason the Sem-ODB – SeaDAS integration was developed with as little modification, of the core SeaDAS files, as possible and by adding new IDL functions, in order to facilitate the upgrade from one version to the next. This also allows the functionality of SeaDAS to remain the same which enhances the power of Sem-ODB.

The modules where the Sem-ODB options were incorporated were those that SeaDAS uses to input and output files from. This module is then wrapped around in new IDL Code that allows for a choice of the original functions (to/from the File System), or to input/output from Sem-ODB. If the user chooses Sem-ODB, control is then passed to the Sem-ODB IDL code which loads files to and from the database. In order to be processed, the exact file is reproduced from the database and passed along to SeaDAS for processing. To the SeaDAS software, the database interaction is completely transparent. SeaDAS only sees the file (in its original format). Once the file is processed, it is then put into Sem-ODB. The advantages, are that this allows the database to be queried from other sources including SeaDAS. If the user chooses the file system, control is than passed into the original SeaDAS functions.

This simple change allows any SeaDAS upgrade to also be incorporated to the SeaDAS – SemODB software without any changes. Any change of processing functions, need only be applied to the Sem-ODB input/output wrapper.

6.1 SEM-ODB - SEADAS DATABASE SCHEMA

In its original form, the Seadas – Sem-ODB database is composed of a subschema for every dataset processed by Seadas. In order for Seadas to process a data file correctly, an exact duplicate of the original data should be reproduced from this database.



SEADAS DATABASE

Figure 2 – SeaDAS – Sem-ODB database schema

To do this every data set is defined as a binary file. This allows for the integration into SeaDAS. When SeaDAS needs to process a file, it regenerates the raw file from one attribute and one category from the database. This, however, works well in the case of SeaDAS, but for other applications this process can be time-consuming. Therefore, each category is than organized into a subschema corresponding to the processing level defined by the HDF format. For example the Level 1A Category, which contained the file alone, would now contain the database discussed in Section 4.

There are 7 different data formats that can be queried (GIF, TIFF, LEVEL_1A, LEVEL 1B, LEVEL 2, LEVEL 3, and LEVEL 0). Each of them contain the filename, however LEVEL_0 has been expanded to show the as it should be.

These methods are used so that as SeaDAS changes, and interfaces are added, the semantic database integration will require minimal change and the functionality can be used regardless of which SeaDAS version the user is running.

7.0 CONCLUSION

Semantic database technology is an effective tool for storing and manipulating ocean color data. With Sem-ODB, a SeaWiFS HDF data set can be easily converted to a semantic database category. Compression techniques can than be used to store data more compactly. This database can than be queried through the WWW, and through NASA's own image processing software, SeaDAS. Sem-ODB provides the features needed to facilitate ocean color imagery storage, retreival and analysis as well as for other remote sensed imagery data.

8.0 REFERENCES

[NCSA93] National Center for Supercomputing Applications, NCSA HDF Specfication and Developer's Guide Version 3.2, University of Illinois at Urbana-Champaign, Sep. 1993.

[RISH92] N. Rishe, Database Design: The Semantic Modeling Approach, McGraw-Hill, 1992.

[RISH95] N. Rishe, W. Sun, D. Barton, Y. Deng, C. Orji, M. Alexopoulos, L. Loureiro, C. Ordonez, M. Sanchez, A. Shaposhnikov, "Florida International University High Performance Database Research Center," SIGMOD Record, 1995, vol. 24, no. 3, pp. 71-76.

[SEAWI95] SeaWiFS Project, SeaWiFS Products Specification, May 1995.

[SEAD99] SeaDAS Group, SeaDAS 3.2 Users Guide, 1999