

January 01, 2007

The CenSSIS web-accessible image database system

Furong Yang
Northeastern University

David Kaeli
Northeastern University

Recommended Citation

Yang, Furong and Kaeli, David, "The CenSSIS web-accessible image database system" (2007). *Research Thrust R3 Presentations*. Paper 16. <http://hdl.handle.net/2047/d10009188>

This work is available open access, hosted by Northeastern University.



The CenSSIS Web-accessible Image Database System

Staff : Furong Yang (fuyang@ece.neu.edu, ECE, NU)
Faculty: Prof. David Kaeli (kaeli@ece.neu.edu, ECE, NU)



This work was supported in part by Gordon-CenSSIS, the Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821)

Abstract

The Gordon-CenSSIS Web accessible Image Database System (CenSSIS-DB) is a scientific database that enables effective collaborative scientific data sharing and accelerates fundamental research. We describe a state-of-the-art system using the Oracle RDBMS and J2EE technologies to provide remote, Internet based data management. The system incorporates efficient submission and retrieval of images and metadata, indexing of metadata for efficient searching, and complex relational query capabilities.

1. Challenges and Significance

A major barrier facing Gordon-CenSSIS researchers is the storing, indexing, and sharing of subsurface image and sensor data. The geographical separation between and the diverse disciplines of CenSSIS members make collaboration a particular challenge. In addition, scientific disciplines such as biology and the earth sciences have recently been generating data at enormous rates, making it difficult for scientists to track and organize these vast repositories. The development of a centralized database system to store, organize and retrieve subsurface imaging data is key to addressing these challenges.

A centralized image database system has several benefits. First, it facilitates data collection for individual members by providing a framework for experimental annotations and variables. Also, it provides a valuable resource for the educational initiatives of Gordon-CenSSIS by providing real data for students to use in the classroom. Thirdly, it minimizes the required effort of individual Gordon-CenSSIS members to manage data sets, freeing their time for analysis and research. Fourth, it forces a consensus on data and imaging standards within the Gordon-CenSSIS community. These standards will then facilitate the development of CenSSIS toolboxes and other data management tools. CenSSIS-DB has advantages over other scientific databases available, it is web accessible, requires no client, provides powerful database service and capabilities, and it is extremely flexible to manage various types of research data.

2. Technical Approach

2.1 Data Model

Our key considerations in developing a data model and choosing a relational database system were flexibility, extensibility, and reliability. The broad research base of the CenSSIS community requires that a number of different types of image data are generated, each with unique metadata characteristics.

We have identified a set of common characteristics to be included with all data sets – these are the metadata for all categories. Category refers to the image type. Then we add additional metadata required for a particular category.



Figure 1. Partial CenSSIS-DB Data Model

Two relationships are critical in our model, and are a key to its understanding. The first is the relationship between the DATA entity and its subtypes, represented by an "IS-A". This design allows us to extend the DATA entity attributes by creating subtypes with minimal redundancy. This design also makes our model flexible and extensible, since we can create new subtypes quickly without negatively impacting the model.

The second interesting relationship is between DATA and DATA_RELATIONS entities. This is a bill-of-materials (BOM) data structure, used to represent a data hierarchy. We created a DATA_RELATIONS entity containing the attributes of parent and child to associate data sets with one another. This allows us to generate an unlimited number of relationships between data entities and thus allow clients to organize data sets into collections .

2.2 System and Software Architecture

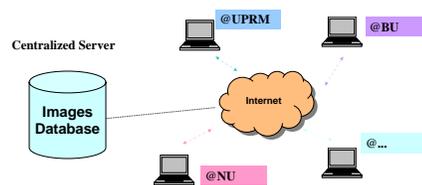


Figure 2. Centralized Architecture, Web accessible

CenSSIS-DB architecture (Figure 3) is divided into components, include:

- A user interface written in HTML/JSP
- Java™ source code, Java™ Servlets, Enterprise Java Beans™ (EJB), JDBC, Java™ Server Pages (JSP)
- Metadata stored in a relational database system (Oracle)
- Image and data files stored on a separate file server and referenced by pointers in the relational database system.

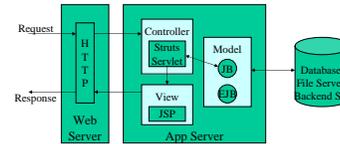


Figure 3. CenSSIS-DB System Architecture. The client can interact with the system via a HTTP connection. Web pages are generated using HTML and Java Server Pages™ which interact with the JavaBeans™ in order to retrieve and submit data. The Controller Servlet interacts with the Oracle database using JDBC™ and can also access the file server where images are stored.

2.3 File Server

The binary data files could be stored in the database itself or in a separate file system. There were several compelling reasons to store them in a separate file system, with links to the data stored with the descriptive metadata. Such as:

- Storage of binary data is not standardized across relational database systems.
- File server is more reliable to store relatively large amounts of binary data.
- Easily accessible to other tools that need to manipulate the data.
- The size of the file containing the metadata is smaller and searches will be more efficient

2.4 Security

The security of CenSSIS-DB is of special concern because it is world wide web accessible. The search and retrieval areas of the system are publicly accessible. Some Gordon-CenSSIS clients, however, need to restrict access to their data sets. Not only do we need to restrict access of particular data sets; but also we want to be able to provide restricted access for the submission of data in order to minimize the need to curate data. A client must select an access permission level when submitting a data set.

- Public - anyone in the world with a web browser
- Gordon-CenSSIS - registered CenSSIS users
- Client - a registered client
- Group - a predefined group of users

A data set can have different permission levels for view or update. This functionality allows Gordon-CenSSIS members to create online communities where they can share privileged information. For general public, only public data is searchable and downloadable.

2.5 Client Metadata Format

The application can gather metadata of an image through data submission and saves it to tables in the database, or user can generate metadata xml file with xml reader program provided by the application, and use one set of metadata for multiple sets of images when submit data to the database, thus helps client with quick and easy data submission.



Figure 4. A Sample Data Submission Page

3. Applications of CenSSIS-DB

3.1 Data Submission

Data submission is a critical challenge for CenSSIS-DB. Clients can choose to create a new data collection or add data to an already existing collection. The client can save some existing files as default settings for later submission. Upon submission, a data set is available for retrieval immediately. Figure 4 is an example of the submission page.

Multiple sets of images and metadata can be submitted to The database at once by custom application. It can be a very efficient and easy data submission when users have multiple images with one defined sets of metadata. Figure 5 is a screenshot of the custom application.



Figure 5. An Example of a Complex Query

3.2 Hierarchical View

A client can select a data set as a root element and be given a tree presentation of all of its child nodes. This presentation can be expanded and reduced upon request. This is a way to present data sets in a way that is convenient for the client and easily navigable.

3.3 Searching Abilities

key types of queries currently available:

- **ID Search:** Simplest, based upon the image id (assigned uniquely upon submittal for each image).
- **Complex Queries:** Form based, multiple entries from a list of metadata or entered criteria, the criteria can be executed with AND or OR operations.
- **Textual Search:** Search upon keyword and description fields

4. Accomplishments

The database is presently online and being populated with a diverse set of subsurface sensing and imaging data. This year, we focused on attracting new users from CenSSIS community through a series of seminars. By now, we have regular registered users, thousands of image data along with accompanying metadata. Figure 7 is a sample oocyte image and its associated metadata from Dr. Charles Dimarzio's group of NEU.

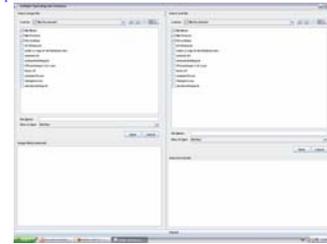


Figure 5. Screenshot of Multiple Data Uploading using Drag and Drop

In 2006, we redesigned the system to allow better access control, so that only public data are searchable and downloadable. Thus users are more willing to store sensitive data. Image tagging is under development which will make image content more useful and searchable. We made efforts in meeting user request and developed customized applications to facilitate the DB use.

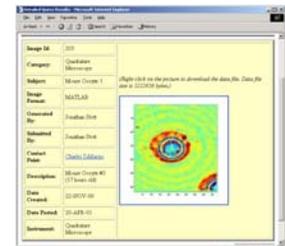


Figure 6. Example of an Oocyte and its Associated Metadata

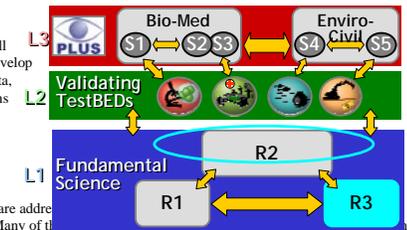
5. Plans

Future research topics include content-based indexing and retrieval (CBIR), data mining on the datasets. We will develop new tools to ensure seamless image format interchange and develop an advanced graphical interface to allow researchers to annotate and query parts of any image.

We will continue to collaborate with other Gordon -CenSSIS members to broaden the scope of our data collection. In addition, we plan to add the functionality to permit automatic submissions of images and data, we also plan to develop advanced searching graphical interface in order to permit data downloading.

6. Relation to Center's Mission

As the Center was being conceptualized, efficient image and sensor data management was identified as one of the Center's seven barriers. Giving researchers the ability to share and search on image data efficiently will then enable Gordon-CenSSIS to develop solutions to problems using real data, and well as to develop new solutions that bridge traditional disciplinary boundaries.



7. Impact/Implications

The class of imaging problems we are addressing are biological, and civil applications. Many of these fields: breast cancer detection, landmine detection, embryo viability and coral reef assessment.

References

1. *A database System to Advance Subsurface Sensing and Imaging.* -H. Wu, B. Norum, D. Kaeli and B. Salzberg, Journal of Subsurface Sensing Technologies and Applications.
2. *The CenSSIS Image Database.* -H. Wu, B. Norum, J. Newmark, D. Kaeli, B. Salzberg, C.M. Warner and C. DiMarzio, Proceedings of the 15th ACM International Conference on Scientific and Statistical Database Management.
3. <http://www.ks.uiuc.edu/research/bioacor>
4. *Sciport Collaboration System.*