

The CenSSIS Image Database

Huanmei Wu, Becky Norum, Judith Newmark, Betty Salzberg,
Carol M. Warner, Charles DiMarzio, David Kaeli
Center for Subsurface Sensing and Imaging Systems
Northeastern University
Boston, MA USA

Abstract

The CenSSIS Image Database is a scientific database that enables effective data management and collaboration to accelerate fundamental research. This paper describes the design and use of a state-of-the-art relational image database management system, accessible through a standard web-browser interface. The application utilizes a robust security architecture and is designed for efficient data submission. Our database query engine provides complex query capabilities to facilitate fast and efficient data retrieval. The system offers a highly extensible metadata schema, with the option of storing data within a hierarchical format.

1. Introduction

The Center for Subsurface Sensing and Imaging Systems (CenSSIS), a National Science Foundation Engineering Research Center, seeks to revolutionize our ability to detect and image biomedical, environmental, and civil objects or conditions that are underground, underwater, or embedded within cells or inside the human body. Our unified, multi-disciplinary approach combines expertise in wave physics, sensor engineering, image processing and inverse scattering with rigorous performance testing to create new sensing prototypes that are transitioned to industry for further development.

A major barrier facing 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. A centralized database system to store, organize and retrieve subsurface imaging data is key to addressing these challenges.

The creation of the CenSSIS web-enabled image database (CenSSIS-DB) has had a dramatic impact upon the Center's mission. Its major contribution has been to provide a common model for scientific data sharing. The CenSSIS-DB also facilitates data collection by providing a framework for experimental annotations and the assignment of variables. These data management tools remove the burden of managing images and image metadata from the scientist.

In addition to serving as a data management tool, the CenSSIS-DB promotes the sharing of data between research groups to enable team-based solutions. It forces consensus on data and imaging standards within the CenSSIS community, thus facilitating the development of toolboxes and other data management tools that bridge traditional disciplinary boundaries. The CenSSIS-DB also provides a valuable resource for CenSSIS educational initiatives by providing real data for students to use in the classroom.

This paper is organized as follows. Section 2 discusses related work. Section 3 is a brief introduction to the embryo viability research project, which is supported by the CenSSIS-DB. Section 4 presents details of the data model design, and section 5 describes our implementation of the CenSSIS-DB. Section 6 demonstrates the query and submission capabilities of the CenSSIS-DB. We summarize and discuss future directions in section 7.

2. Related work

There is a considerable amount of existing work on scientific databases that provide support for research efforts. Here we describe some bioinformatics databases (most of which contain images) and some non-bioinformatics scientific image databases.

Well-known and publicly available bioinformatics databases include the NCBI [15] and the NCI [16] databases. NCBI is a national resource for molecular biology information, and includes a genetic sequence

database (GenBank) [3], human genes and genetic disorder database (OMIM) [14], and molecular modeling database (MMDB) containing 3-D macromolecular structures [27]. The CBCTR database [13] in NCI provides clinical data for specimens distributed to clinical researchers. Examples of other bioinformatics databases include PDB [26], BioSig [17, 18], CCDB [6], and ECHBD [8, 9]. PDB (Protein Data Bank) is the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data. BioSig is an imaging bioinformatics system for studying phenomics and intracellular signaling. CCDB populates immunocytochemical data on the 3-D distribution of cellular constituents within realistic compartmental neuronal models. ECHBD is optimized for viewing and analyzing brain specimens. Most of these bioinformatics databases contain images.

Non-bioinformatics scientific image databases include SkyServer [24], SIMBAD [20], ADEPT [12, 21] and DIMES [31]. SkyServer provides online access to the public Sloan Digital Sky Survey [25] data. The SIMBAD astronomical database provides basic data, cross-identifications and a bibliography for astronomical objects outside the solar system. ADEPT is a distributed digital library of spatial map sets covering most of the world and includes images from satellite, space shuttle, aerial, and other sources. DIMES is an earth science data system that accepts metadata submissions in any valid XML format, thus placing no restrictions on metadata entries.

Since the CenSSIS-DB contains both bioinformatics and non-bioinformatics image data, while designing our system, we studied databases from both domains. We looked at each of the systems in order to ascertain their functionality and limitations, focusing upon their data models and user interfaces, to determine best practices as we developed our own system.

3. Embryo Viability Research

The classes of imaging problems addressed in CenSSIS include medical, environmental, biological, and civil applications. Many of these problems come from some of the most pressing societal issues in these fields: breast cancer detection, bridge deck assessment, cardio-vascular plaque imaging, landmine detection, embryo viability, and coral reef assessment. In this section, we describe one of these problems: embryo viability research. We will use this example frequently in the paper.

3.1. Introduction to Embryo Viability Research

Since the first success of in vitro fertilization (IVF) [22], more than one million IVF babies have been born. However, the success rate of producing live babies after IVF is

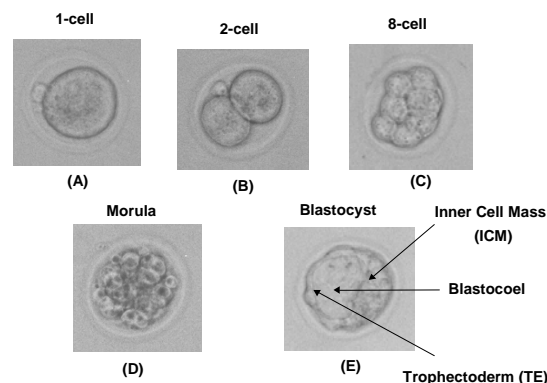


Figure 1. Mouse Preimplantation Embryos.

only around 25%. To increase the rate of success, more than one embryo is transferred to the mother, leading to a higher chance of twins and triplets. Multiple births cause an increased chance of perinatal morbidity and mortality, in addition to physical and emotional stress on the parents. The low rate of IVF success is largely due to the inability to properly assess embryo quality using present biochemical, genetic, and imaging methods. Accurate evaluation of oocytes (unfertilized eggs) and embryos is a key to IVF success. Embryo viability research is aimed at developing genetic, immunological, and imaging methods to assess preimplantation embryo health so that only one embryo needs to be transferred back to the mother.

Women undergoing IVF are treated with hormones to stimulate production of up to 20 oocytes per hormone cycle. After fertilization and several days of culture, the "best" embryos must be chosen for transfer back to the mother. The preimplantation period of development starts at the moment of fertilization and continues until implantation of the embryo into the uterine wall, which is 5-6 days in humans and 4-5 days in mice. During the preimplantation period the embryos are free-floating in the reproductive tract. This makes it possible to remove the oocytes or embryos from their mothers and subject them to well-controlled experimental protocols. A mouse model system can be employed to carry out these controlled studies. The stages of mouse preimplantation embryo development are shown in Figure 1.

3.2. Parameters for the Evaluation of Preimplantation Embryos

IVF clinics evaluate oocytes and embryos based on morphology. The embryos are given a grade, which is based on overall morphology and the number of cells in each embryo. This grade is then used to calculate the total number of embryos to transfer back to the recipient mother. If

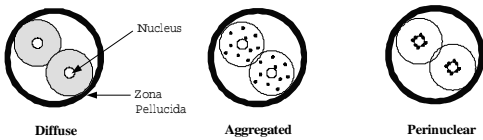


Figure 2. Patterns of Mitochondrial Distribution in Two-Cell Embryos.

one embryo could be identified with a very high grade, then single embryo transfer would be possible. The following sections discuss two parameters that we are presently using in a mouse model system to define embryo grade: 1) mitochondrial distribution and 2) the number of cells in an embryo. Current research is focused on measuring these two parameters using non-invasive imaging methods that produce a large number of images that need to be stored and annotated in the CenSSIS-DB.

3.2.1 Mitochondrial Distribution

There is a growing body of literature that suggests that the number and distribution of mitochondria in oocytes and embryos are directly related to their potential to give rise to live offspring [1, 2, 5, 19, 28, 30]. Mitochondrial distribution changes with development of the embryo. Figure 2 illustrates three patterns of mitochondrial distribution that can be seen in oocytes or two-cell embryos. It remains to be determined which of these patterns is associated with healthy embryos. Mitochondria can be visualized using fluorescent stains and a laser scanning confocal microscope (LSCM), which is unsuitable for use in an IVF clinic because of toxicity issues. Non-invasive techniques for imaging mitochondrial distribution could have a significant impact on the ability to identify viable, healthy embryos.

3.2.2 Cell Counting in Embryos

The number of cells in an embryo is another parameter that is related to embryo viability [28, 29]. A larger number of cells at a given point in time during post-fertilization indicates a faster cleavage rate. During the early stages of preimplantation development, the cells in the embryo are large and easy to count. By the blastocyst stage, as shown in Figure 1 (E), cells are differentiated to form two tissue types. The outer cells form the trophectoderm (TE), and the inner cells form the inner cell mass (ICM). The number of cells in the ICM is related to embryo viability, with higher cell numbers correlated with increased IVF success. However, it is very difficult to count the cells due to the complexity of the embryo at this stage. Fluorescent stains, such as Hoechst stain, can be used to label and count cell nu-

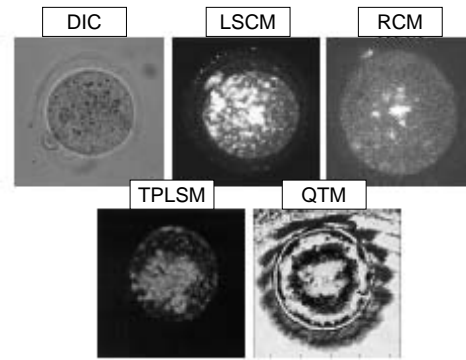


Figure 3. Oocyte Images with Five Different Microscopic Modalities.

clei using a fluorescent microscope, but as stated above for stained mitochondria, there are major toxicity issues. No method currently exists for counting the number of cells in a late preimplantation stage embryo without the use of dyes or invasive imaging methods.

3.3. Imaging Modalities for Embryo Viability Research

Several technologies have been used to image oocytes and embryos. Light microscopy and fluorescent imaging have been widely applied. For example, a scoring system based on morphological characteristics has been defined with differential interference contrast (DIC) microscopy [29]. Mitochondrial distribution has been visualized using fluorescent stains such as Mitotracker and a laser scanning confocal microscope (LSCM). The number of cells in the ICM of an embryo has been visualized using fluorescent stains for cell nuclei such as Hoechst. However, no existing microscope can distinguish embryo health without compromising viability. For example, DIC is unable to count the number of cells in embryos beyond the 8-cell stage. Imaging with high powered microscopes or with staining is toxic, invasive and damaging to the embryo. New state-of-the-art instruments are being developed in order to non-invasively assess embryo viability, including the quadrature tomographic microscope (QTM) and the Keck 3-D fusion microscope (3DFM), described below.

The QTM is a new imaging system that has been developed in CenSSIS [7, 11, 23]. It combines conventional light microscopy with interferometric quadrature detection techniques to yield a simultaneous measurement of both the amplitude and phase of the electric field. The low power laser of the QTM enables noninvasive imaging of biological samples without causing physiological damage. The QTM

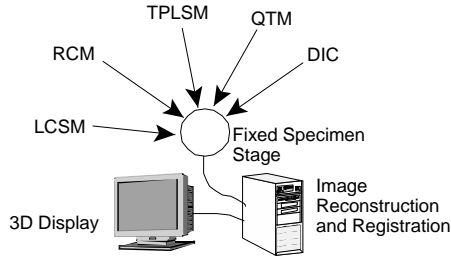


Figure 4. The 3D Fusion Microscope.

can distinguish normal from fragmented oocytes and can image internal structure in oocytes that is not visible with other techniques. Shown in Figure 3 are oocyte images obtained with five microscopic techniques: Differential Interference Contrast (DIC), Laser Scanning Confocal Microscopy (LSCM), Reflectance Contrast Microscopy (RCM), Two Photon Laser Scanning Microscopy, and Quadrature Tomographic Microscopy. These five modalities will be combined in a single microscope, the 3DFM. The QTM can distinguish normal from fragmented oocytes and can image internal structure in oocytes that is not visible with other techniques. The use of this information can help select the best oocytes and embryos to use in IVF clinics.

The 3DFM, shown in Figure 4, combines five imaging modalities into a single platform so that specimens can be evaluated at the same time and at the same place with five complementary imaging methods. The images produced by these modalities are registered and fused for three-dimensional display and analysis. This instrument allows accurate spatial registration among different microscopic modalities, so that features absent in one modality can be identified with structures that are only apparent using another modality.

3.4 Image Management

A goal of embryo viability research is to distinguish healthy from unhealthy oocytes and embryos in order to increase pregnancy rate and reduce birth defects in IVF clinics. The embryo viability research is now in its data gathering and data analysis stage. Each year, more than 20 GB of image data will be produced. In addition, some of these images will be disseminated to other research groups for processing and statistical analysis. Currently, embryo images are stored on CD-ROMs and metadata is recorded by hand in laboratory notebooks. However, due to the rapid in-

crease in data acquisition rates and the resulting abundance of data, alternative image management methods are needed. A centralized image database system such as the CenSSIS-DB will provide an improved data storage and retrieval system. The CenSSIS-DB provides efficient storage, searching, and exchange of large quantities of both raw and processed image and sensor data. The CenSSIS-DB hierarchical data scheme fits nicely into an experimental framework and facilitates image searching/browsing and storage.

4. Data Model Design

The broad research base of the CenSSIS community created a significant challenge in satisfying user requirements. We began by incorporating several key image modalities in our model. However, it is likely that additional image types will be identified. Therefore, while developing the data model and choosing a database system, flexibility, extensibility and reliability were our key considerations.

4.1. Characteristics of the image data

A CenSSIS image data set is defined as consisting of two parts. The first is an image data file, provided in either binary or ascii text format. The second is information about the image data set (i.e., metadata). The CenSSIS-DB supports a range of standard image formats, including JPEG, FIG, TIFF, MATLAB, and DICOM. Data sets are classified into categories according to imaging methods and modalities. Every data set belongs to a unique category.

Metadata is data about data. Every image data set is accompanied by a generic set of metadata, including the date the image was acquired, who acquired it, and a textual description about the image. In addition, each category includes an additional set of metadata fields which are specific to the imaging method. This additional metadata describes the conditions under which the data set is acquired, the instrument(s) used for data acquisition, and whether the data set consists of raw or processed data. The following section discusses our data model in more detail.

4.2. Data model

The multi-disciplinary nature of CenSSIS has led to a range of metadata schemas and data formats, as discussed above. In addition, new categories are continually being defined as new imaging methods are developed and additional data set types are submitted to the CenSSIS-DB. Our data model design accommodates these diverse requirements while illustrating the commonalities between disparate data sets.

Our approach was to first separate the metadata from data files. Metadata is stored in a relational database man-

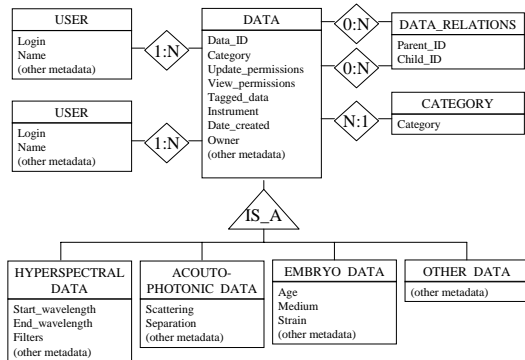


Figure 5. The CenSSIS Data Model.

agement system (Oracle), while data files are stored on a separate file server (see section 5.2). Metadata and data files are linked by a unique image identification number generated for each data set.

Figure 5 presents a simplified entity-relationship (ER) data model of our metadata schema. Each box in the diagram corresponds to an entity (i.e., table) in the database. Entities are related by relationships, represented by solid lines. Different types of relationships are represented by the diamonds between entities.

A set of common characteristics have been identified as the core metadata set. These characteristics are stored within the DATA entity and included with all data sets. Core metadata fields include a unique data set identification number, viewing and updating permissions (see section 4.3.3), and instrumentation, owner, and tagging information (see section 4.3.4).

Category-specific metadata is stored in entities having an IS_A relationship with the DATA entity. For example, an EMBRYO_DATA entity is a subtype of the DATA entity. It stores metadata specific to the embryo viability group (e.g., embryo age and mouse strain). A row in the EMBRYO_DATA table correlates to a row in the DATA table using the unique data set identification number. This design allows us to extend the DATA entity for new image types with minimal redundancy and without impacting the existing database schema.

In order to create hierarchies of data sets, we designed a bill-of-materials data structure using the DATA and DATA_RELATIONS entities. This data structure is a many-to-many recursive relationship that allows us to generate an unlimited number of relationships between data entities and thus allow clients to organize data sets into collections (see Figure 6).

<i>Security Levels</i>	<i>Users</i>
Public	Anyone with a web browser
CenSSIS	Registered CenSSIS-DB users
Project	A predefined group of users
Client	A registered client

Table 1. CenSSIS-DB Security Levels.

4.3 New and Enhanced Features of the Data Model

The data model discussed above has enhanced the CenSSIS-DB’s functionality in the following ways.

4.3.1 Hierarchical Collections

The ability to organize data sets into collections and projects is critical to our clients. The data model described above enables clients to create hierarchies of data sets or collections, organizing them into experiments or projects. As illustrated in Figure 6 (c), a client can create a root node, such as a project (nodes 100, 400 and 500). Within this node, the client can add additional nodes, such as folders (node 210) or data sets (nodes 200, 220). This hierarchy is potentially limitless, offering each client the flexibility of naming and organizing his or her own data sets, while adhering to the general metadata schema described above.

For example, the QTM constructs a 3D model of the internal structure of a cell from multiple raw images. Each raw image corresponds to one data set; however, data sets can be linked using the collection representation to capture data relationships (in this case, the set of raw images used to construct the single 3D model).

4.3.2 A Project Centered Environment

The CenSSIS-DB also provides researchers a collaborative research environment, based upon the BioCoRE system [4]. In addition to using our public query engine, clients can create and manage projects, represented by an online web page. A project can be thought of as a shared repository. Project owners can add or remove additional CenSSIS-DB members to a project in order to share data sets and information.

4.3.3 Security Model

The metadata schema provides a strong security model. In order to gain submission (i.e., write) access to the CenSSIS-DB, a client must obtain an account the host system. Registered clients can submit new data, update data, and manage instrument profiles. They can also create and join projects.

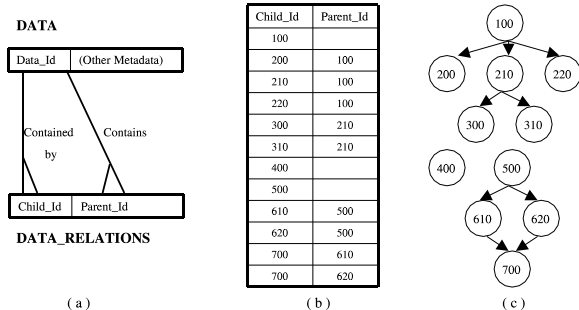


Figure 6. Representation of the Bill of Materials Data Structure in the CenSSIS-DB Data Model. (a) Relationships Between DATA and DATA_RELATIONS. (b) Sample Entries in the DATA_RELATIONS Table. (c) A Hierarchical Representation of the Data in (b).

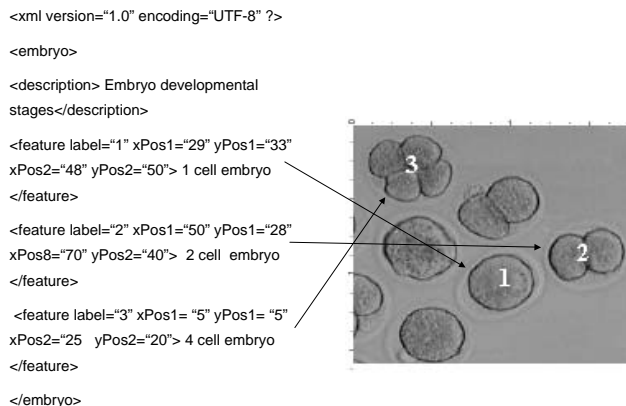


Figure 7. XML Image Tagging.

Application-level security measures are administered by assigning permissions to all data sets and projects. Upon submission of a data set, a client must assign read/write permissions to that data set. These permissions are listed in Table 1. Project permissions are administered by the project owner(s). The project owner(s) can add new members to a project to allow privileged access to a project. Additional security measures are applied within our system architecture and are discussed in section 5.3.

4.3.4 Image Tagging

We are currently developing a capability to allow a client to tag areas of an image. Upon selecting an area of the image, a popup window will show an enlarged and detailed view of that area, with a text field for the client to fill out (i.e., tag). Tags are stored in an XML format.

Tagging images allows researchers to identify regions and annotate them for future reference. In addition, tags will be stored, indexed and queried as part of the metadata schema, enhancing the query capabilities of the system. Figure 7 gives a simple example of how we can use image tagging to label embryo developmental stages.

This tagging feature will be invaluable to both researchers and clinicians. It will provide researchers with enhanced means of gathering, assessing and collating relevant data, and assist clinicians in diagnostic assessment and feature identification.

5. Implementation of the CenSSIS-DB

The implementation of the CenSSIS-DB can be divided into the following three sections – overall hardware and

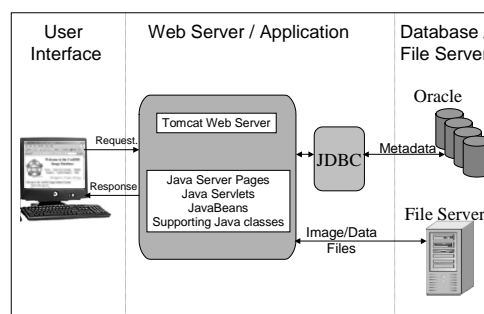


Figure 8. The CenSSIS-DB System Architecture.

software architecture, file server organization, and system-level security measures.

5.1. Architecture

The CenSSIS-DB system has been implemented with a three-tier design as the underlying architecture [10]. The three layers include the web-application front-end, business and transaction logic middle-tier, and the data access layer, as shown in Figure 8.

Application Front-End: The application front-end provides access using a standard web browser. The application presents a series of dynamically generated HTML screens that are driven by the application logic.

Application Logic: The middle tier contains the application logic and the web server. The application is written

using JavaBeans, Java Servlets, Java Server Pages (JSP), and Java utility classes. The web server delivers dynamically generated web pages to the client and parses client requests. In addition, this layer contains the appropriate drivers to connect to the data access layer.

Data Access Layer: The data access layer is comprised of a relational database management system (Oracle) and a file server. The database is accessed from the application-level Java classes through a Java Database Connectivity (JDBC) driver. The file server is mounted upon the application server to allow direct system access to all directories and files.

This multi-tiered architecture provides high performance, flexibility, and maintainability, while hiding application complexities from the client.

5.2. File Storage

Data files could be stored in the database itself or in a separate file system. There were some compelling reasons to store them in a separate file system, with links to the data stored with the descriptive metadata. First, many of the data files are in binary format and storage of binary data is not standardized across relational database systems. Second, by keeping the image data files on a file server, they remain easily accessible to other tools that may be needed to manipulate the data. Third, the image data will consume gigabytes of space, so disentangling the metadata from the data files results in much more efficient I/O access, versus interleaving the metadata in the image file.

5.3. System Security

The standard web protocol, HTTP (hypertext transfer protocol), transmits requests and responses in plaintext over the Internet. Transmissions can be captured and easily read by third parties. In order to prevent this, we are using a secure web server with HTTPS (secure hypertext transfer protocol).

HTTPS creates a secure connection between a web client and web server, ensuring that all transmissions are encrypted. Thus, if transmissions are captured, they cannot be read by a third party. HTTPS is a standard extension for secure data transmission over the World Wide Web.

6. Using the CenSSIS-DB

The CenSSIS database can be accessed through a web browser independent of the geographical location (<http://censsis-db1.ece.neu.edu>). In this section, we describe the CenSSIS-DB query engine, data submission and provide query examples (see Figure 9).

6.1. The CenSSIS-DB Query Tool

Clients can retrieve image/data sets using the CenSSIS-DB GUI query tool on the CenSSIS-DB website. The interface provides both SQL and forms-based query modes. The SQL query mode is designed for clients familiar with SQL syntax. It allows them to create and run SQL queries on the database. The forms-based query mode provides comprehensive search capabilities through a menu of options, without the need to know any SQL syntax. Several key query types were identified for our forms-based query mode:

ID Search: The simplest way to search for an image is by the image identification number, a unique number assigned to each image data set upon submission.

Keyword Search: A text search option is available which scans the descriptive metadata fields for the entered keyword(s).

Complex Queries: For each metadata field, the query engine provides either an empty text box or a dynamically generated dropdown menu in a web form. As illustrated in Figure 9 (a), clients can select multiple search options and the query engine dynamically builds an SQL query. Searches may be specified using the logic operators AND or OR, to perform a more complex query.

Subtree Search (Hierarchical View): Clients select a data set as a root element and the search returns a tree representation, the root's children. This representation can be expanded and reduced upon request. This presents data sets in a hierarchical format that is easy to navigate.

While building a query, the client can indicate whether results should be presented individually or aggregated. Also, clients may opt to view thumbnails of the data sets within a result set. In a future release, clients will also have an option of adding data set(s) to a *shopping cart* for later review and download.

6.2. Data Submission

The CenSSIS-DB provides straight-forward uploading and downloading capabilities. Clients choose to create a new project or add data to an already existing project. For added convenience, existing files can be used to specify all default settings for future submissions.

Submission is done using a menu-based web page. The client fills out a web form with the appropriate metadata, attaches the relevant file(s) and submits the request.

The metadata passes through a syntax check based upon expected values. If there are no errors, the metadata and an index to the corresponding files are stored in the database and the attached file(s) is stored on the file server.

Upon submission, a data set is immediately available for retrieval. However, the new submission is noted with a conditional tag to alert clients that the data set has not

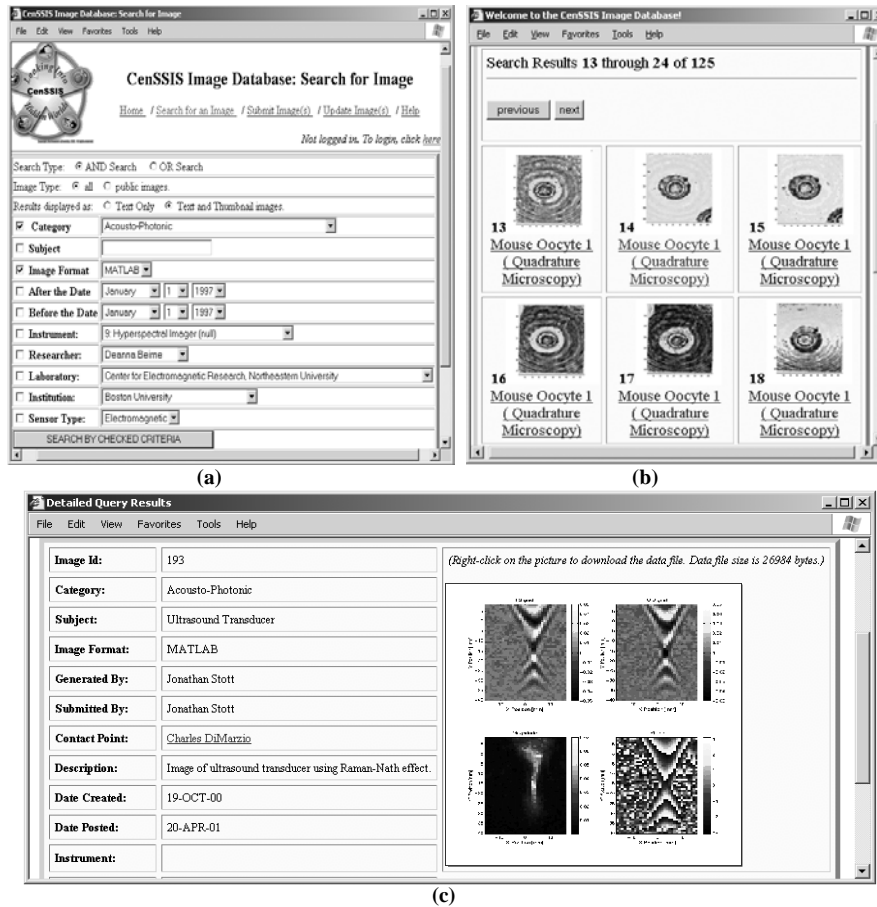


Figure 9. The Execution of a Complex Query and Its Results. (a) Query, (b) Result, (c) Detail of One Result

yet been reviewed by the administrator. Once approved, the conditional tag is removed and the data set is available unconditionally. The approval process was implemented in order to meet two conflicting system requirements. First, clients need to be able to share very large data sets quickly. These files are often too large for standard email servers; the CenSSIS-DB provides an online repository to facilitate sharing. However, in order to ensure system integrity, we need to have control over and be able to remove problematic submissions. Therefore, we adopted the above approval process to provide the maximum amount of flexibility and security.

6.3. Database Query Examples

Our query engine provides rapid searching of data by dynamically building database queries to match client requests, as discussed in Section 6.1. As an example how the database is used, this section presents sample queries frequently requested by embryo viability researchers and briefly discusses how the queries are handled by the query engine.

Some sample queries include:

- Q1.** Find images based on user name, experiment name, or date.
- Q2.** Find images based on mouse strain.
- Q3.** Find images based on the stage of embryonic development.

- Q4.** Find images based on the instrument used.
- Q5.** Find images based on a combination of the above parameters (Q1 - Q4).
- Q6.** Find the annotations for an image, given the image identification number.
- Q7.** Find the experimental protocols used to produce an image.
- Q8.** Find the raw data for a processed image and the series of images leading to the final image.
- Q9.** Find “matched” images of a single sample taken using different modalities.
- Q10.** Find images demonstrating similar patterns or characteristics within a modality.
- Q11.** Find images demonstrating similar patterns or characteristics across multiple modalities.

The first six queries (Q1 - Q6) are simple metadata-based queries. By indexing the relevant metadata fields, we provide efficient execution of these queries. The web application generates the appropriate SQL query and forwards it to the database. The database returns a result set which is formatted into a user-friendly web page(s) (see Figure 9 (b)).

Queries Q7, Q8 and Q9 are more complex recursive queries and use the hierarchical data table `DATA_RELATIONS` (Figure 6). The same raw data can produce different processed images depending upon the protocol used; accessing the data set protocols (Q7) is a common client request.

Our clients are also concerned with the history of a data set. An experiment may contain both raw and processed data. Relationships between data sets are represented in the `DATA_RELATIONS` table. A raw data set can be thought of as a parent, and the parent can have many children (i.e., processed data sets). Also note that a single child can be produced by multiple parents (see Figure 6, node 700). By searching on the `Parent_Id(s)` of a processed data set, we can retrieve the raw data sets that generated the processed data.

The 3DFM generates multiple images from a single sample. Query 9 (Q9) represents a typical request to see all data generated for a particular sample. By defining a collection (e.g., embryo sample 5), and adding all data sets generated for this sample, this collection can later be retrieved using, again, the relationships represented in the `DATA_RELATIONS` table.

Queries Q10 and Q11 are content-based queries, which pose questions about image content that may not be in the original image metadata. For example, automatic identification of healthy mitochondrial distribution patterns at different stages of embryo development would be a noninvasive method for identifying viable, healthy embryos. Also, the retrieval of data sets based upon the stage of embryo development as identified by morphological characteristics or cell counts would facilitate the identification of viable em-

bryos. Content-based retrieval is presently under development in the CenSSIS-DB.

7. Summary

We have successfully developed a web-based database system for storing, organizing, and retrieving subsurface sensing and imaging data. Our CenSSIS-DB adopts a flexible and extensible framework to handle the complex and diverse metadata of different imaging applications. Our design includes security functions in order to protect privileged information.

Our major contribution is providing a common model for scientific image data sharing and management. We worked closely with biological researchers and CenSSIS engineers to develop a flexible and extensible system to meet current and future research storage requirements. The CenSSIS-DB provides researchers with the ability to share and search on image data efficiently. Our system also enables CenSSIS to develop solutions to problems using real data, as well as to develop new solutions that bridge traditional disciplinary boundaries. The CenSSIS-DB is presently online and being populated with a diverse set of subsurface sensing and imaging data sets. Several research groups from different universities and research hospitals are using the database. The success of the CenSSIS-DB is only possible through the combined efforts of many individual researchers working toward a common vision. We will continue to collaborate with other CenSSIS members to broaden the scope of our data collection.

Future research topics include new indexing technology and new toolkit development. New designs for views, indices and access functions will be incorporated to provide better answers and allow for a wider range of query capabilities. Specifically, we will focus on content-based indexing and retrieval (CBIR), multidimensional database indexing and content-based image tagging and searching.

Different software tools have been developed to manifest one central theme of CenSSIS – the concept of “Diverse Problems/Similar Solutions.” Incorporating these tools within the CenSSIS-DB will accelerate CenSSIS research. In addition, we are developing new tools to ensure transparent image format interchange and developing an advanced graphical interface to allow researchers to annotate and query parts of any image.

8. Acknowledgements

This work was supported in part by NIH grants HD39215 and HD40309 to CMW, NSF grant IIS-0073-063, the W.M. Keck Foundation, and CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the

Engineering Research Centers Program of the NSF (Award Number EEC-9986821).

References

- [1] D. Barnett, J. Kimura, and B. D. Bavister. Translocation of Active Mitochondria During Hamster Preimplantation Embryo Development Studied by Confocal Laser Scanning Microscopy. *Dev. Dyn.*, 205:64–72, 1996.
- [2] B. Bavister and J. Squirrell. Mitochondrial Distribution and Function in Oocytes and Early Embryos. *Hum. Reprod.*, 15(Suppl.2):189–198, 2000.
- [3] D. Benson, I. Karsch-Mizrachi, D. Lipman, J. Ostell, B. Rapp, and D. Wheeler. GenBank. *Nucleic Acids Res.*, 30(1):17–20, Jan 2002.
- [4] M. Bhandarkar, G. Budescu, W. Humphrey, J. A. Izaguirre, S. Izrailev, L. V. Kal, D. Kosztin, F. Molnar, J. C. Phillips, and K. Schulten. BioCoRE: A Collaboratory for Structural Biology. *Proceedings of the SCS International Conference on Web-Based Modeling and Simulation*, pages 242–251, 1999.
- [5] P. G. Calarco. Polarization of Mitochondria in the Unfertilized Mouse Oocyte. *Dev. Gen.*, 16:36–43, 1995.
- [6] CCDB. Cell Centered Database at the National Center for Microscopy and Imaging Research (NCMIR). <http://pamina2.sdsc.edu/CCDB/>.
- [7] C. A. DiMarzio, A. Devaney, and S. Lindberg. Optical Quadrature Interferometry Utilizing Polarizing Optics. *U.S. Patent 5,883,717*, March 1999.
- [8] ECHBD. European Computerized Human Brain Database. <http://fornix.neuro.ki.se/ECHBD/Database/>.
- [9] J. Fredriksson, P. Roland, and P. Svensson. Rationale and Design of the European Computerized Human Brain Database System. *SSDBM*, pages 148–157, Jul. 1999.
- [10] J. Gallagher and S. Ramanathan. Choosing a Client/Server Architecture. A Comparison of Two-Tier and Three-Tier Systems. *Information Systems Management Magazine*, 2(13):7–13, 1996.
- [11] Y. Glina, G. Tshirintzis, C. Warner, D. Hogenboom, and C. DiMarzio. On the use of the Optical Quadrature Method in Tomographic Microscopy. *SPIE Proceedings*, (3605):101–106, 1999.
- [12] G. Janee and J. Frew. The ADEPT Digital Library Architecture. *JCDL*, pages 342–350, Jul. 2002.
- [13] H. Marshall, J. Meyer, C. Helms, and H. Donis-Keller. St. Louis Breast Cancer Tissue Registry and the Cooperative Breast Cancer Tissue Registry (CBCTR): a combined resource for marker validation studies. *Poster presented at the 47th Annual Meeting of the American Society of Human Genetics*, Oct. 1997.
- [14] V. McKusick. Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. *Baltimore: Johns Hopkins University Press*, 12th edition, 1998.
- [15] NCBI. National Center for Biotechnology Information. <http://www.ncbi.nlm.nih.gov>.
- [16] NCI. National Cancer Institute. <http://www.nci.nih.gov/>.
- [17] B. Parvin, G. Cong, G. Fontenay, and M. H. Barcellos-Hoff. BioSig, An Imaging Bioinformatic System for Studying Phenomics. *IEEE Computer*, 35(7):65–71, 2002.
- [18] B. Parvin, G. Cong, G. Fontenay, J. R. Taylor, R. Henshal, and M. H. Barcellos-Hoff. BioSig: A Bioinformatic System for Studying the Mechanism of Intra-Cell Signaling. *BIBE*, pages 281–288, 2000.
- [19] P. Reynier, P. May-Panloup, M. Chretien, C. Morgan, M. Jean, F. Savagner, P. Barriere, and Y. Malthiery. Mitochondrial DNA Content Affects the Fertilizability of Human Oocytes. *Mol. Hum. Reprod.*, 7:425–429, 2001.
- [20] SIMBAD. Astronomical Database. <http://simbad.u-strasbg.fr>.
- [21] T. R. Smith, G. Janee, J. Frew, and A. Coleman. The Alexandria digital earth prototype system. *JCDL*, pages 118–119, Jan 2001.
- [22] P. Steptoe and R. Edwards. Birth after the Reimplantation of a Human Embryo. *Lancet*, 2(8085):366, 1978.
- [23] J. Stott, E. Bennett, C. Warner, and C. DiMarzio. Three-dimensional imaging with a Quadrature Tomographic Microscope. *SPIE Proceedings*, 4261:24–32, 2001.
- [24] A. Szalay, J. Gray, A. Thakar, P. Kuntz, T. Malik, J. Rad-dick, S. C., and J. Vandenberg. The SDSS SkyServer Public Access to the Sloan Digital Sky Server Data. *SIGMOD Conference*, pages 570–581, 2002.
- [25] A. S. Szalay, P. Z. Kunszt, A. Thakar, J. Gray, and R. Slutz. Designing and Mining Multi-Terabyte Astronomy Archives: The Sloan Digital Sky Survey. *SIGMOD Conference*, pages 451–462, 2000.
- [26] T.S.Schmidt. Banking on Structures. *BioIT World*, 1(8), Oct. 2002.
- [27] Y. Wang, L. Geer, T. Madej, A. Marchler-Bauer, D. Zimmerman, and S. Bryant. MMDB: 3D structure data in Entrez. *Nucleic Acids Res.*, 30:249–252, 2002.
- [28] C. Warner and C. Brenner. Genetic Regulation of Preimplantation Embryo Survival. *Current Topics in Developmental Biology*, 52:151–192, 2001.
- [29] C. Warner, W. Cao, G. Exley, A. McElhinny, M. Alikani, J. Cohen, R. Scott, and C. Brenner. Genetic Regulation of Egg and Embryo Survival. *Hum. Reprod.*, 13(Suppl.3):178–90, 1998.
- [30] M. Wilding, B. Dale, M. Marino, L. di Matteo, C. Alviggi, M. Pisaturo, L. Lombardi, and G. De Placida. Mitochondrial Aggregation Patterns and Activity in Human Oocytes and Preimplantation Embryos. *Hum. Reprod.*, 16:909–917, 2001.
- [31] R. Yang, X. Deng, M. Kafatos, C. Wang, and X. Wang. An XML-Based Distributed Metadata Server (DIMES) Supporting Earth Science Metadata. *SSDBM*, pages 251–256, June 2001.