A Neuroinformatics Database System for Disease-Oriented Neuroimaging Research[1]

Stephen T.C. Wong, PhD, PE, Kent Soo Hoo, Jr, PhD, Xinhua Cao, PhD, Donny Tjandra, MSc, J.C. Fu, PhD, William P. Dillon, MD

Clinical databases are continually growing and accruing more patient information. One of the challenges for managing this wealth of data is efficient retrieval and analysis of a broad range of image and non-image patient data from diverse data sources. This article describes the design and implementation of a new class of research data warehouse, neuroinformatics database system (NIDS), which will alleviate these problems for clinicians and researchers studying and treating patients with intractable temporal lobe epilepsy. The NIDS is a secured, multi-tier system that enables the user to gather, proofread, analyze, and store data from multiple underlying sources. In addition to data management, the NIDS provides several key functions including image analysis and processing, free text search of patient reports, construction of general queries, and on-line statistical analysis. The establishment of this integrated research database will serve as a foundation for future hypothesis-driven experiments, which could uncover previously unsuspected correlations and perhaps help to identify new and accurate predictors for image diagnosis.

Key Words. Databases; clinical informatics; epilepsy.

© AUR, 2004

How do I retrieve and process all my patients' data for analysis so I can do this research study efficiently? This is a question that every clinical researcher asks when embarking on a new clinical study. As every clinician knows, retrieving patient data such as diagnostic reports, clinical test results, radiologic images, and laboratory test results from the various hospital departments is a time-consuming process. Administrative and technical obstacles stand in the way, and when the obstacles are finally overcome, the clinician is faced with the dilemma of managing, storing, and maintaining the data on their own. These difficulties are faced by everyone conducting retrospective or prospective research studies that integrate diagnostic, therapeutic, and outcomes data from multiple hospital departments. We describe in this article the design and implementation of a research data warehouse, neuroinformatics database system (NIDS), to alleviate these problems for neuroradiologists, neurologists, neuropsychologists, and neurosurgeons studying and treating patients with intractable temporal lobe epilepsy.

Academic medical centers and teaching hospitals generally have three main goals: to provide clinical care, to train new clinicians, and to conduct biomedical research. Investigating brain structure and function requires the integration of information from the level of the gene to the

Acad Radiol 2004; 11:345–358

[1] From the Departments of Neurology, Radiology, and Neurosurgery, University of California, San Francisco, CA (K.S.H., D.T., W.P.D.); the Center of Bioinformatics, Harvard Center of Neurodegeneration and Repair, Harvard Medical School, the Department of Radiology, Brigham & Women's Hospital and the Center for Bioinformatics, Harvard Center for Neurodegeneration and Repair, Harvard Medical School, Goldenson Bldg, Room 524, 220 Longwood Avenue, Boston, MA (S.T.C.W., X.C.); and the Department of Industrial Engineering, Da-Yeh University, Taiwan, Republic of China (J.C.F.). Received August 15, 2003; revision requested September 16; received in revised form October 10; accepted October 12. Supported in part by grants from the National Institutes of Health, National Institute of Neurological Disorders and Stroke, and National Institute of Mental Health, a National Library of Medicine/National Institute of Health FIRST award (R29), and Harvard Center of Neurodegeneration and Repair, Harvard Medical School. **Address correspondence to** S.T.C.W. e-mail: stephen_wong@hms.harvard.edu

© AUR, 2004
doi:10.1016/S1076-6332(03)00676-7

level of behavior. At each of these many and diverse levels there has been an explosion of information, with a concomitant specialization of scientists. The price of this progress and specialization is that it is becoming virtually impossible for any individual researcher to maintain an integrated view of the brain and to relate his or her narrow findings to this whole cloth. Although the amount of information to be integrated far exceeds human limitations, solutions to this problem are available from the advanced technologies of computer and information sciences.

Efforts are under way in the clinical and basic science communities to develop information systems that can integrate biomedical data from multiple laboratories and hospitals. There are many clinical neuroscience research groups studying humans and animals using individualized imaging and experimental techniques, so the Biomedical Informatics Research Network is being developed to provide a "protocol" for collaborative research among neuroscientists and medical scientists (1). The American College of Radiology Imaging Network is a complete infrastructure for managing multi-institutional clinical trials of imaging technologies as they relate to cancer (2,3). The Medical Imaging Resource Center of the Radiological Society of North America serves as a central repository for teaching files, clinical trials support, research data, and educational materials (4). All of these efforts are aimed at inter-institutional collaborative research and education.

Picture archive and communication systems (PACS) is one example of how information technology is increasingly being used to support daily operations in today's hospitals. Clinical PACS has greatly enhanced the ability of clinicians to use radiologic images in the delivery of care (5,6). IN addition to supporting basic patient care, the PACS archive enables clinicians to easily retrieve past image datasets for consultation purposes, resulting in the growth of digital image teaching files within the radiology department. At the enterprise level, hospital information systems pre-date PACSs and manage demographic, laboratory test, and diagnostic report data. Other specialty departments such as pathology, neurology, and surgery are also developing their own departmental information systems, and clinical and research groups within departments usually manage their data on some kind of computer-based system. All of these stand-alone systems contain data that is potentially important in conducting clinical research, but they are not interconnected for efficient data retrieval and interrogation.

An entirely new information system built on top of the existing infrastructure needs to be built with the goal of facilitating clinical and biomedical research. Such a system will have new features such as storage and management utilities for preserving meta-data, ie processed data derived from radiologic images such as tumor volumes, vascular permeability factors, etc. The system will need to provide on-line analytical processing of multiple sets of data, including image display, image processing, and statistical analysis.

Our aim is to develop an integrated, structured, and extensible NIDS to support clinical imaging research for neurological diseases and cancers. The system incorporates data from multiple sources to enable the analysis of radiologic, neurologic, psychological, surgical, and pathologic data. These data are retrieved and stored within a clinical research database along with meta-data subsequently generated during the course of the data post-processing and analysis. This type of system differs from clinical diagnostic information systems which are transactional in nature; instead, the NIDS is built using data warehousing techniques that emphasize aggregate analysis of the data instead of emphasizing the query and retrieval of individual pieces of data. We have finished prototyping the system to support image-based epilepsy research and surgical planning and are intended to complete data and analysis tools for additional modules for other disease types: intra-cranial hemorrhage, brain tumors, and Creutzfeldt-Jakob Disease (Fig 1).

The research protocol of the NIDS project is approved by institutional research board (IRB) of our institutions for specific prospective and retrospective studies of patients of certain neurological diseases. The protocol also limits the access of the database only to those participating clinicians and scientists of individual disease centers, eg, epilepsy, multiple sclerosis, and so on. For example, the epilepsy database module has been completed and installed in the Northern California Comprehensive Epilepsy Center of the University of California, San Francisco (UCSF) to support research of clinicians and neuroscientists of that Center. Each disease module of NIDS is assigned with a clinical "owner" who tracks the usage of the system and monitors the adherence to approved research protocols. For instance, the epilepsy module owner is an epileptologist in the Northern California Comprehensive Epilepsy Center. For new research studies outside the existing approved scope by IRB, individual research-

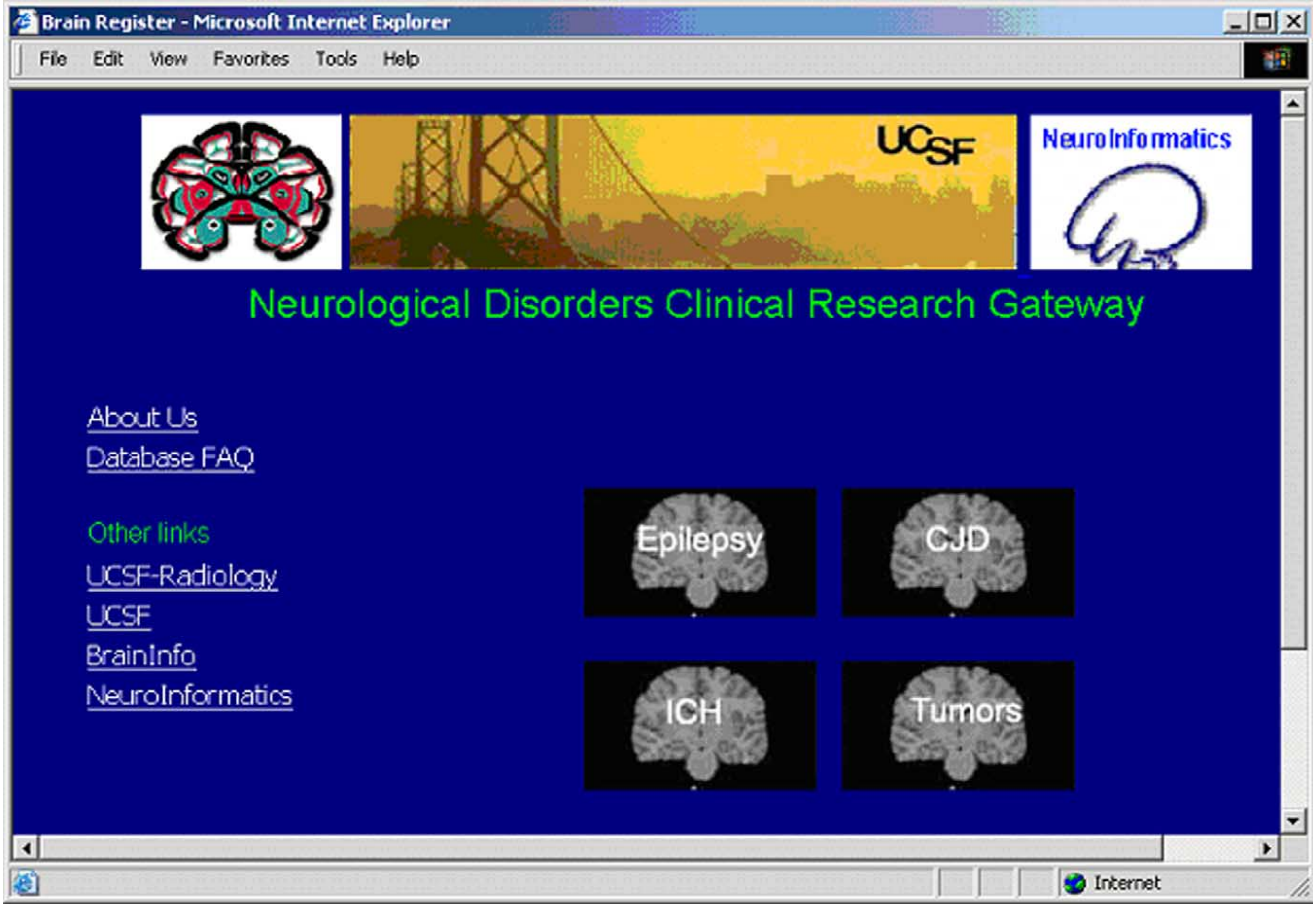

Figure 1. This splash page is the starting point for navigating the UCSF neuroinformatics database system. Clinical researchers select which disease domain they wish to investigate using the NIDS. Username and password authentication protect the links to the respective disease categories. Regular change of user passwords is enforced in this database system.

ers will have to apply for the IRB approvals for those studies.

MATERIALS AND METHODS

Architecture

The NIDS architecture is the key technological innovation that enables a clinician to gather, proofread, analyze, and store clinical data for research purposes. The software architecture consists of four tiers as adopted in most data warehouse systems: web clients, NIDS servers or application servers, data staging area, and backend clinical data sources (Fig 2). The back-end data sources are the actual clinical data repositories located in various hospital departments. The hospital information system (HIS) at UCSF is known as the STOR/CDS (Summary Time Oriented Record and Clinical Display System) and serves as a repository for clinical data generated from several source systems, including laboratory, pathology, transcription, radiology, EKG, microbiology, and pulmonary function lab. The UCSF STOR/CDS is not a comprehensive medical record, so we are also required to interface to other departmental data sources. Magnetic resonance imaging (MRI) and MRSI (magnetic resonance spectroscopic imaging) data sets are retrieved from the radiology department PACS using a DICOM client. However, positron emission tomography (PET) images are not currently recorded in the PACS, necessitating direct interface with the nuclear medicine department's PET database which uses proprietary software for archiving the PET sinograms and reconstructed image data sets. The neurosurgery department records presurgical and surgical information in a Microsoft Access database (Microsoft Corp., Redmond, WA) which is exported directly into the

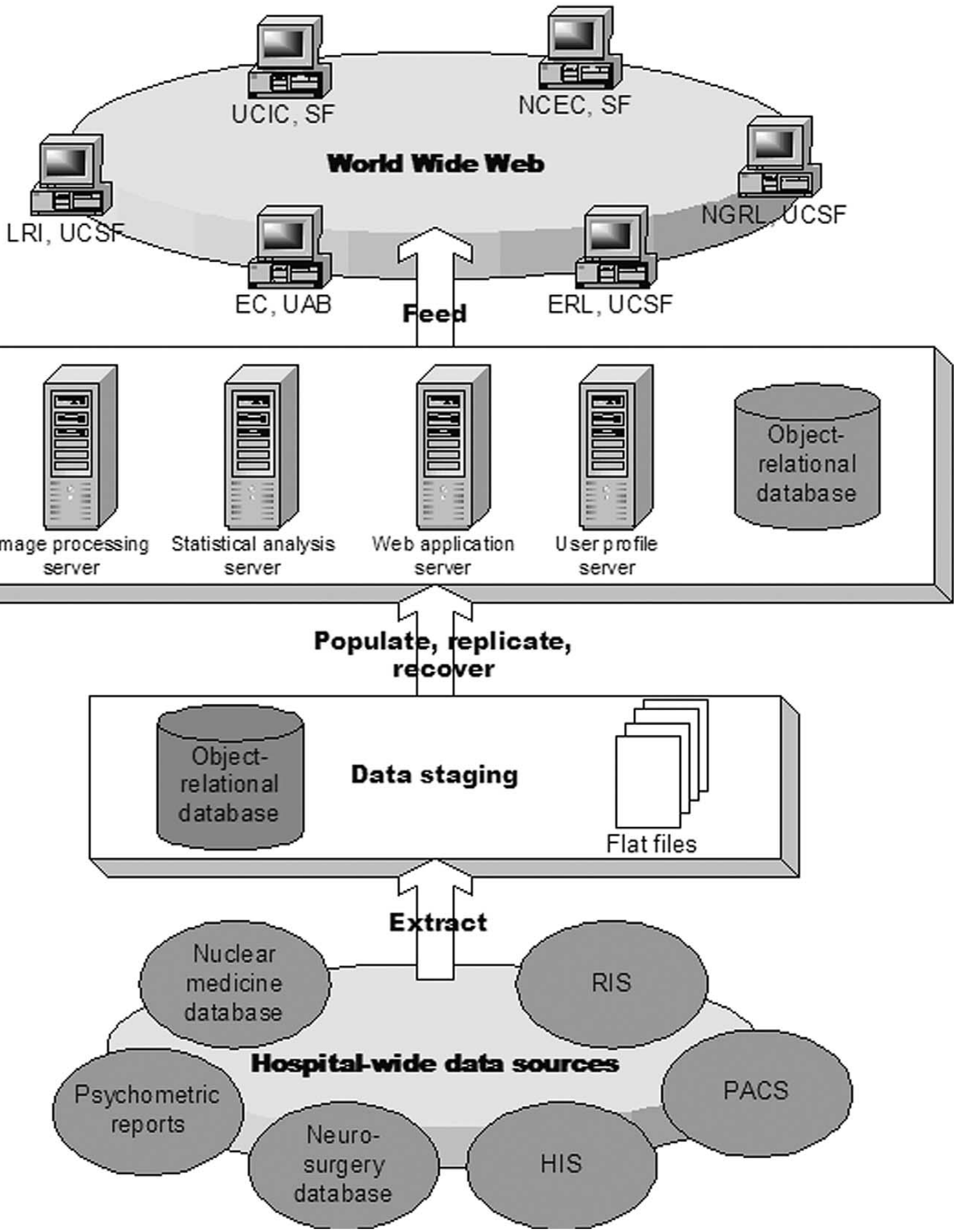

Figure 2. The four-tiered architecture of the neuroinformatics database system.

Table 1.
Examples of Raw Data Fields Extracted into the Neuroinformatics Database System from the PACS, Psychometric, Neurosurgery, and Hospital Information Systems Data Sources

PACS	Psychometric	Neurosurgery	HIS
MRN	MRN	MRN	MRN
study_date	test_date	surgery_date	report_date
exam_code	handedness	side	comparison
study_description	verbal_iq	craniotomy	clinical
x_pixel_size	perf_iq	pathology	impression
y_pixel_size	reyavl_ver	engel_class_6mo	conclusion
no._of_slices	language	engel_class_1y	ref_physician

NIDS's relational database management system (Oracle, Redwood Shores, CA). The Radiology Information System (RIS) used is an IDXRad system (IDX, Burlington, VT), and the psychometric reports containing psychological test results are archived in a traditional paper-based filing system.

Data models are devised to classify, organize, and represent classes of different images and text to support information retrieval by image content for respective disease categories. The NIDS data model is constructed to facilitate both retrospective and prospective studies of patient clinical and research data. All of the clinical operational information systems' data models are based on patient medical record number, and clinical researchers are accustomed to retrieving, organizing, and analyzing clinical data based on medical record numbers and patient names. Therefore, the neuroinformatics data warehouse uses the patients' medical record number as the primary organizational key.

Our philosophy is to use the database to manage all the data, including raw, cleansed, processed, and metadata. Raw data from hospital-wide data sources are manually and semi-automatically extracted into the data staging tier on a periodic basis. The data model contains tables that are defined specifically to receive the raw data in an unaltered state (Table 1). Raw data undergoes a process of cleansing that includes reformatting, recoding, rearranging, and reorganizing; it is at this stage that clinicians ensure the integrity of the data before the process of analysis begins. The cleansed data is loaded into new tables in the third layer, which are designed to directly support query and analysis by the researchers. Researchers then retrieve the data to perform the actual analysis for disease-focused research, resulting in processed data that is saved back to the database into summary tables, thus capturing the results of the analysis for future reference and statistical analysis.

The user profile server controls the high-level interaction between the user and the servers, encodes medical rules for decision aids, and provides interactive workup models of different epilepsy categories. Rule-based modules those encode general and specific medical rules for decision making, case-based modules those aid the navigation of database through similar or reference imaging cases, and model-based modules those provide interactive work-up models of different disease categories closely involving physicians making joint decisions.

The user profile server also manages the customizable user interface for individual clinicians depending on their personal preferences. Users are restricted to accessing only the data to which they have authorization to review. After requests for information are processed by other servers, the processed data is sent to the user profile server for formatting before it is sent to the web client. The user profile server also provides user authentication and audit trail capabilities. Furthermore, database triggers within the database engine keep track of who accesses the database, which rows the insert, which rows they modify, and which rows they delete. Such accountability ensures that a complete history of the database is maintained in case mistakes are made by the users when managing the data. The audit trail is used to restore mistakenly deleted or altered data.

The image processing server provides image processing algorithms for image filtering, co-registration, enhancement, feature extraction, quantitation, and structural/functional mapping. Clinicians can extract salient primitive image features, such as volumes of certain brain regions and sizes of lesions, and key words. In addition, the clinician can also annotate and compose image slices of interest. The extracted image features and pictures of interests are then stored into the NIDS to serve as indexes for content-based image retrieval as well as data points for research analysis.

Data Management

The two main considerations in gathering data for the NIDS have been the type of study and the data source. The data warehouse supports both prospective and retrospective studies, and the latter is much more challenging with respect to gathering and processing data. Retrospective studies include data from patients who were treated using inconsistent diagnostic protocols whereas prospective studies are easier to control and clean since all patients follow the same diagnostic protocol. Nevertheless, clinical researchers are very keen to exploit the power of the data warehouse in doing retrospective studies of large numbers of patients.

The data load, transformation, and cleansing components that interface and acquire relevant patient information from various clinical sources, eg, PACS, RIS, Hospital Information System (HIS), imaging scanner consoles, neurological reports, psychometric reports, neurosurgery database system, and other legacy systems. Because each of the systems was developed in relative isolation, the process of retrieving and processing data into the data warehouse requires a combination of automatic and manual methods. Retrieving images from the PACS is facilitated by the DICOM standard protocol for querying and retrieving. Automatic routines extract DICOM header data such as patient, image, and scanner information, and these data are semi-automatically loaded into the data warehouse using the Oracle Loader utility. Besides the radiology department, other departments are less evolved with respect to their data archival methods. Therefore, much more user effort is required to load, transform, and clean data from the neurology, neurosurgery, and psychology sections. A great deal of effort is expended in transforming the data so that it can be inserted into the data warehouse's data model, and an equal amount of effort is expended in checking the data for accuracy and consistency. In the case of neuropsychology, records are still maintained on paper, so users have to manually enter the psychometric test data using web-interface forms to the data warehouse. In the case of neurosurgery, its data is maintained in Microsoft Access whose export facility along with Oracle Loader make transferring data to the data warehouse more efficient and semi-automatic. All of the data is first archived into an intermediary database for subsequent transforming and cleansing by the specialists using Oracle WebDB utility before final archiving into the data warehouse.

The data warehouse framework is being built upon established standards. The data warehouse is implemented on a Windows departmental server running an Oracle 9 relational database system. PACS images are transferred and stored into the data warehouse using DICOM protocols. Laboratory results, diagnostic reports, and demographic data are transferred from the hospital information system using the HL7 (Health Level 7) standard. The data is queried and presented to the clinicians via web browsers using several Web and database standards including hyper-text markup language (HTML), extensible markup language (XML), Java database connectivity (JDBC), and structured query language (SQL). However, all of this sophisticated technology is hidden from the clinician or researcher behind an easy-to-use Web-based graphical user interface.

Image Processing

Figure 3 illustrates the steps to extract image and textual features from neuroimages. Co-registration of functional PET and T1-weighted MRIs is an essential step toward combining functional information from PET images with anatomic information in MRIs (7). The NIDS interfaces with the public domain Automated Image Registration package (AIR, Version 2.03) as well as several proprietary routines in commercial image analysis packages for registration of neuroimages including PET, MRI, MRSI, and magnetoencephalography (8).

Once image pairs are registered into common space, their biomedical information can be combined, extracted, and quantitated to obtain important information about the location and nature of the epileptogenic zone. For example, the high-resolution MR image provides a basis to anatomically map glucose metabolism. This allows the ability to identify metabolic rates in specific brain regions of interest. The algorithm incorporates a program developed at the Center of Functional Imaging, Lawrence Berkeley Laboratory (Berkeley, CA) (9) in conjunction with the Volumetric Image Display & Analysis project at the University of Iowa Medical College (10) Iowa City, IA to map regions of interest defined on the MRI to the PET sonograms for quantitation of glucose uptake. Figure 4 illustrates such an example taken from a partial epilepsy case. This figure marks down the values obtained from structural/functional image analysis on volumes and flu-

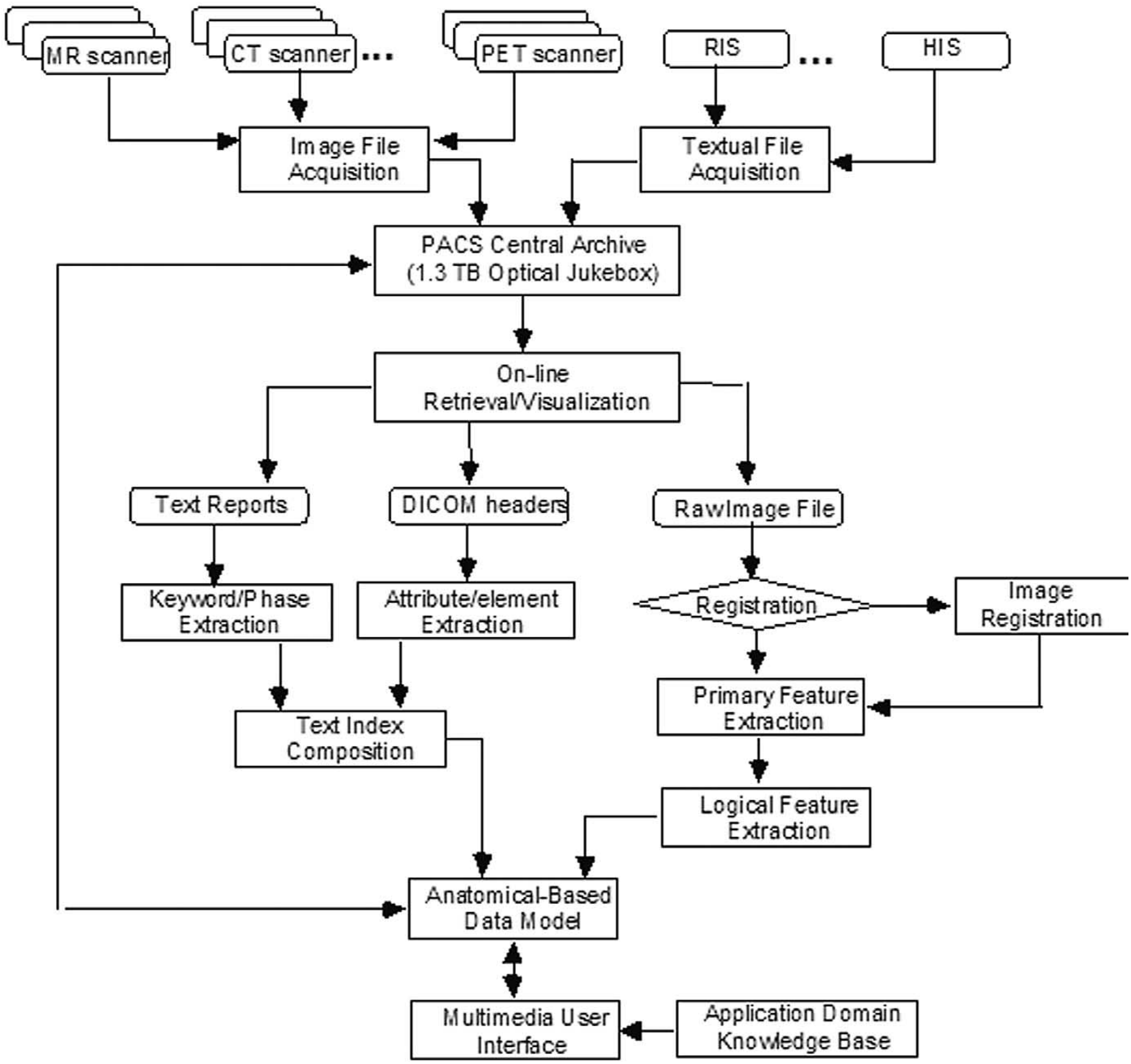

Figure 3. The operational flow of extracting image and text features into an anatomic-based data model for subsequent content-based image indexing in the image data warehouse. Specific knowledge or heuristics is triggered to aid the query and navigation of the medical image database.

orodeoxyglucose (FDG) glucose update per unit volume of hippocampus and amygdala. Such quantitative data, which cannot be obtained in traditional evaluations, provide sensitive and objective localization information.

The correlated image datasets are encoded into the targeted data model to serve as definitive indexing in image query. For example, registering the functional images of PET with the MRI images of the same patient allows the intrinsically better spatial resolution of MRIs to be used in quantitatively analyzing functional information (metabolic count of glucose consumption) of captured PET scans. In epilepsy surgical evaluation, neurologists are able to pinpoint those parts of the brain with suspicious hypometabolic activity and compare their locations with the geometric location of neuronal cell loss in the MRIs. The regions of interest derived from segmenting

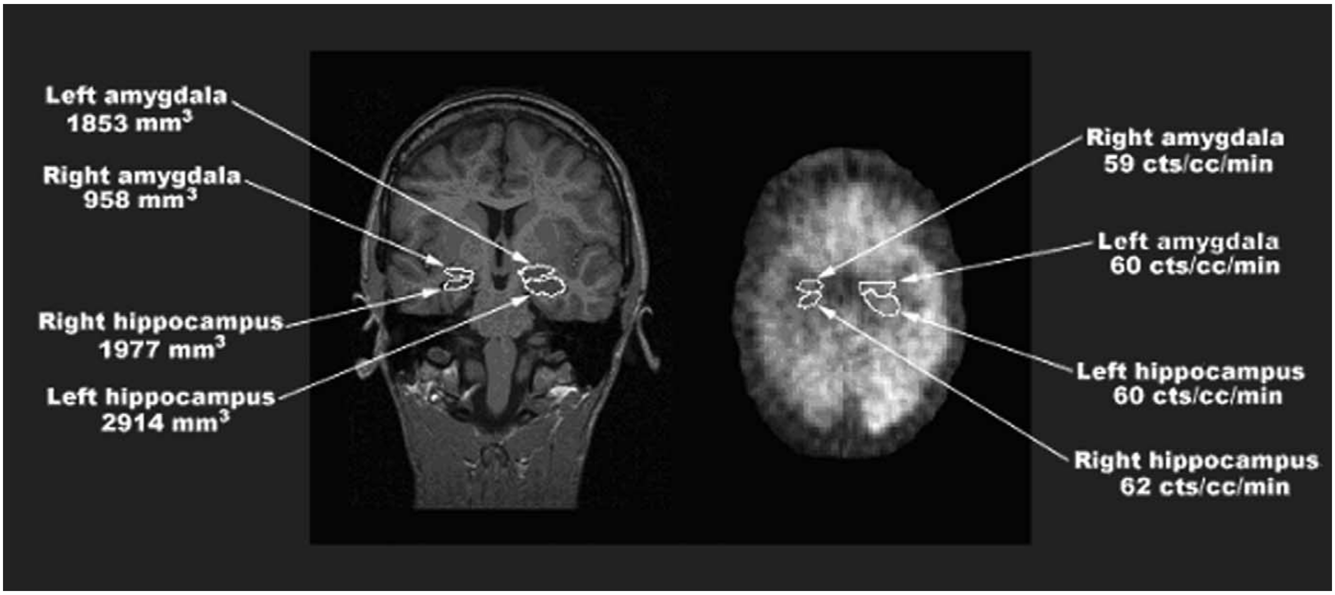

Figure 4. Quantitation of hippocampus and amygdala MRI regions and the use of MRI-PET co-registration to obtain FDG concentration per unit volume in FDG-PET (2-dimensional original slices shown here).

the MRIs are transformed into the PET space for comparison.

Free Text Search

Clinicians can use this interface to browse, sort, retrieve, display, and analyze all the data contained in the data warehouse (Fig 5). With so much data at their disposal, the first step in performing a research experiment is pairing down the data set to isolate a specific cohort of patients based on factors such as age, sex, ethnicity, treatment type, lateralization of epileptogenic focus, or psychometric test scores to mention a few. Clinicians can dispatch a free-text search of the data warehouse by keywords such as "mesial temporal sclerosis" and the data warehouse will return a list of patients with reports containing such a word or phase.

The problem of document storage and retrieval has always been a major issue in information process. One of the most popular techniques of information retrieval has been the use of inverted indices, the method used by most commercial indexing software companies. An inverted index catalogs a collection of objects in their textual representations. Given a set of documents, key words and other attributes (possibly including weights) are assigned to each document. The inverted index is the list of key words and links to the corresponding document. In essence, building a search index is a sorting problem, as the index presents a list of links sorted by word.

The index is normally maintained in a B-tree data structure. This provides great flexibility for manipulation such as insertion, re-position and deletion. More artificial intelligence techniques can be added to strengthen the simple index. Lexical analysis, stop words, stemming, synonym, and weight are just few of the enhancements that can be easily applied the basic inverted index search tools.

Based on the most frequently executed text searches, we have created a FAQ (frequently asked questions) page that presents the user with pull-down menus to construct general queries of the epilepsy database (Fig 6).

Once dispatched, the NIDS sorts through the database and returns a graphical representation of the query results, such as the pie chart below (Fig 7).

The statistical analysis server enables the clinician to sort through thousands of records in the NIDS to make clinical observations such as comparing the concordance between MRI and PET findings of the epileptogenic focus using surgical outcomes as the gold standard. Neuroradiologists scan the diagnostic reports from MRI and PET studies and enter lateralization results into the NIDS. Neurosurgeons enter surgical results into their own departmental information system which in turn updates those results into the data warehouse. This kind of data analysis combines data from different data sources, radiology and surgery, which would be very laborious to accomplish without the NIDS (Fig 8).

Working with psychologists from the Northern California Epilepsy Center (San Francisco, CA), we are pro-

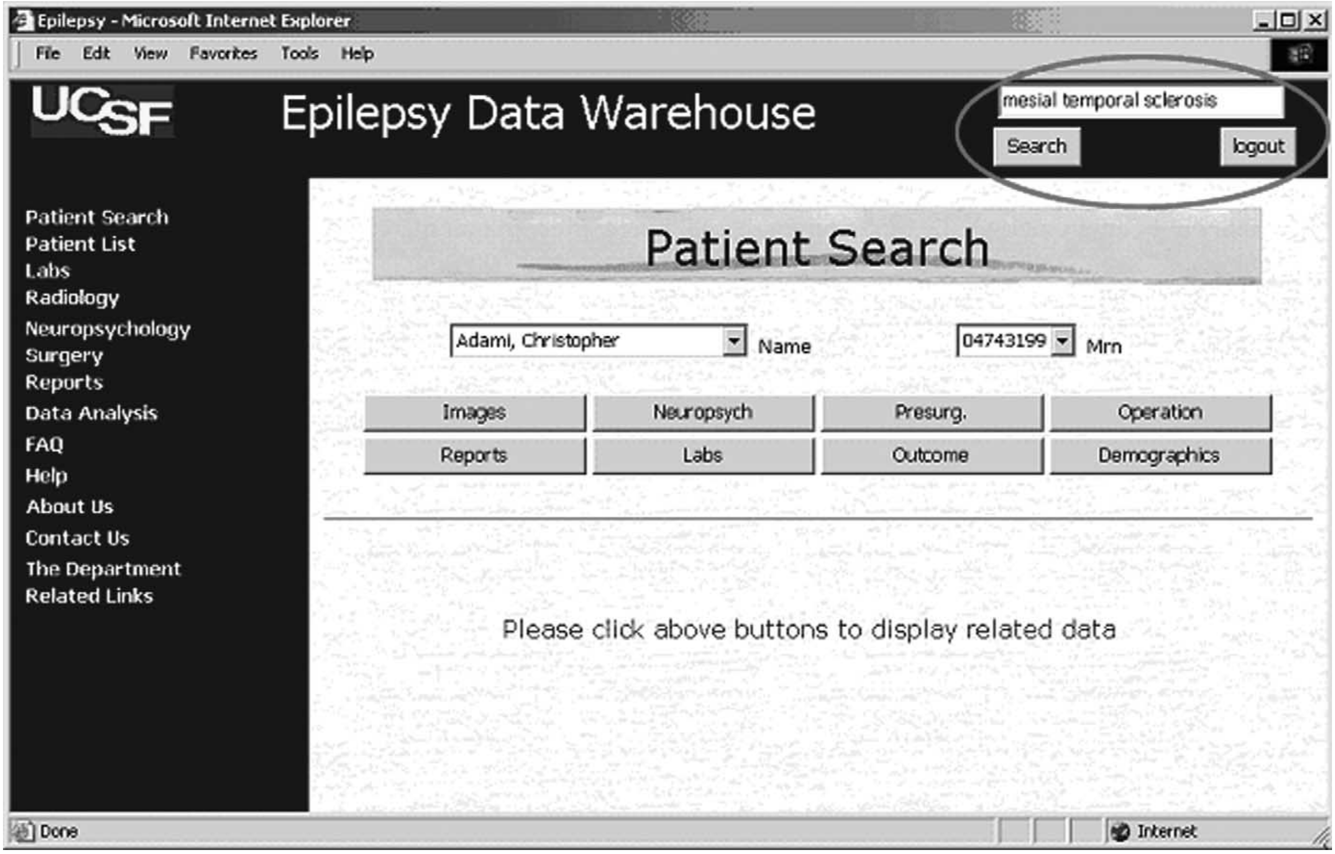

Figure 5. This is the starting point for searching the epilepsy component of the NIDS (12). Note the buttons along the bottom corresponding to the various clinical sources of data. Users can also enter free text into the parametric search field at the upper left corner to search for particular keywords such as mesial temporal sclerosis.

gramming statistical algorithms for analyzing psychometric test results contained in the NIDS, including *t*-tests, analysis of variance, and Pearson product moment correlation. For example, a simple Pearson's correlation coefficient can be calculated for any two measured variables from the NIDS:

$$r = \frac{\Sigma Z_x Z_y}{n}$$

where r is the Pearson's correlation coefficient, Z_x and Z_y are the two measured variables, n is the number of observations. The correlation coefficient indicates the degree of a linear relationship between two parameters stored in the NIDS such as psychometric test "Age" identification and "Volume of Hippocampi" identification (Fig 9). The user simply selects the two variables under consideration from his web client, and the statistical analysis server retrieves the appropriate data from the psychometric test results within the NIDS, calculates the coefficient, and sends the results to the web application server for display at the web client.

Using this interface, a clinician can explore the database for correlations between any two parameters, and as more data is added, the power of the correlations becomes higher. Additional parameters are automatically added to the list of parameters as more data is extracted from underlying data sources, cleansed, and stored into the NIDS.

The NIDS also provides separate user interface to enable direct querying of the underlying neuroinformatics databases and download the queried results into a form of an Excel spreadsheet file for more extensive off-line statistical analysis using commercial statistical packages, such as SAS or SPSS. However, our IRB research protocol does not allow the user to directly query the NIDS in such a manner but submit the query statement to the bioinformatician who is assigned to execute such commands directly to the neuroinformatics database. The query to NIDS database can be via

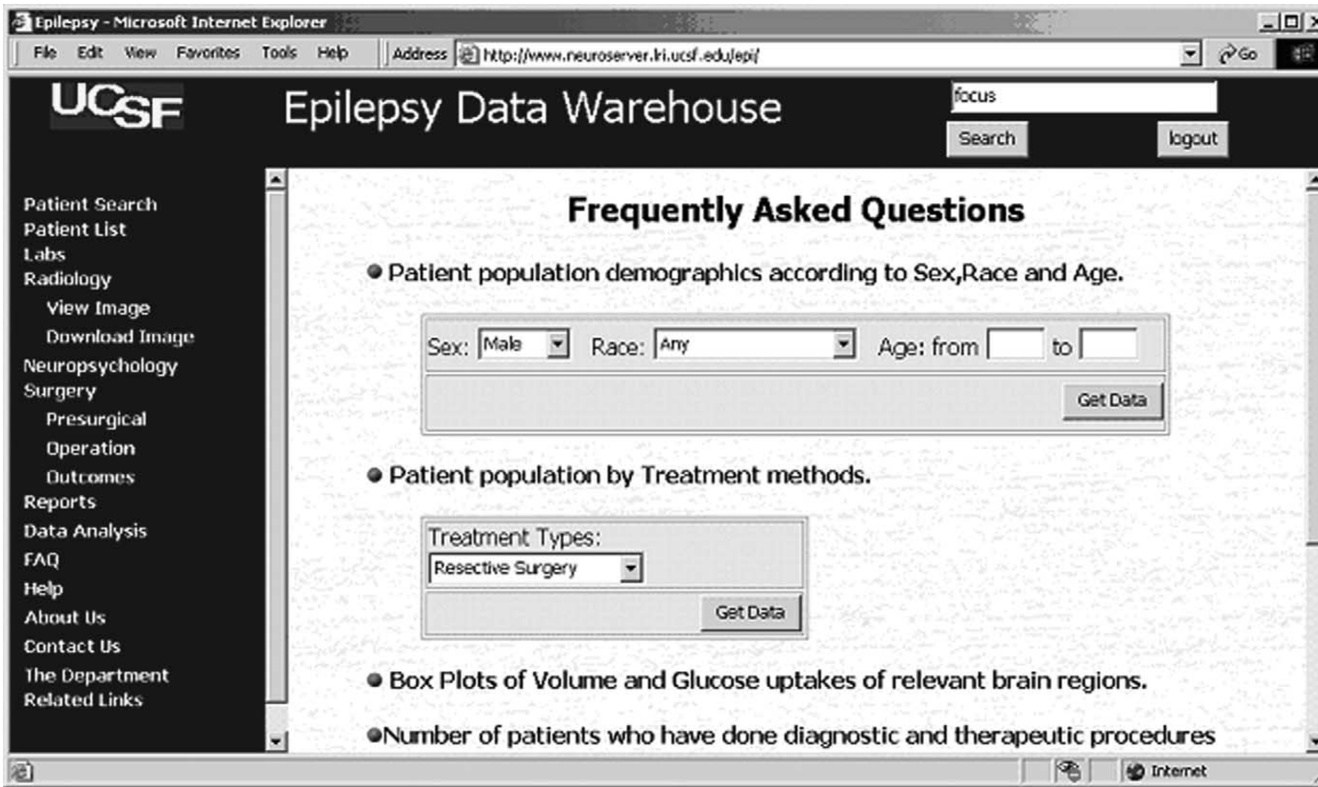

Figure 6. The FAQ interface has pre-defined drop lists and text fields for the user to construct queries to the NIDS. These pre-defined queries are based on the most common ways in which clinicians prefer to classify and organize the patient data.

SQL command line on desktop machines or web based query tools provided by individual database vendors (in our implementation, it is Oracle WebDB utility).

RESULTS

Neuroimaging specialists, psychologists, neurosurgeons, and neurologists from the Northern California Epilepsy Center use the NIDS to conduct clinical research in localizing the epileptogenic focus and as well as biomedical research into understanding brain structure and function. The NIDS is an essential tool because it enables the clinicians to browse all the data in one place, set up experiments, and save the processed data and meta-data in the NIDS for future reference. Currently, there are 696 patient data sets and 24 control data sets in the NIDS, representing the past 17 years of epilepsy surgical candidates from the Northern California Epilepsy Center at the UCSF. This electronic database is one of the largest, comprehensive collections of clinical epilepsy research data in the country. There are 762 MRI data sets and 187 PET datasets with singograms currently stored and analyzed in the epilepsy data warehouse, and more of them can be retrieved from clinical PACS and PET database when needed. Five hundred forty-seven patients have complete HIS data sets resident in the NIDS from the hospital information system, including all demographics, clinical, radiologic, hospitalization, pathology, and microbiology reports for every patient. Seven hundred thirty-two surgical and 394 pre-surgical data sets have been transferred from the neurosurgery database. Finally, 1,154 psychometric test data sets from pre and post-surgical exams have been entered via the on-line Web forms from paper-based files, and each of these data sets consists of about 50 individual test scores for various psychological measures. A small number of video EEG telemetry data sets have also been entered into the NIDS to test new utilities simultaneous for web-based viewing of video, EEG telemetry, radiological images, and laboratory test results (Fig 10).

The epilepsy module of NIDS is completed in May 2002 and is used by clinical researchers of Northern California Comprehensive Epilepsy center. Clinicians are currently car-

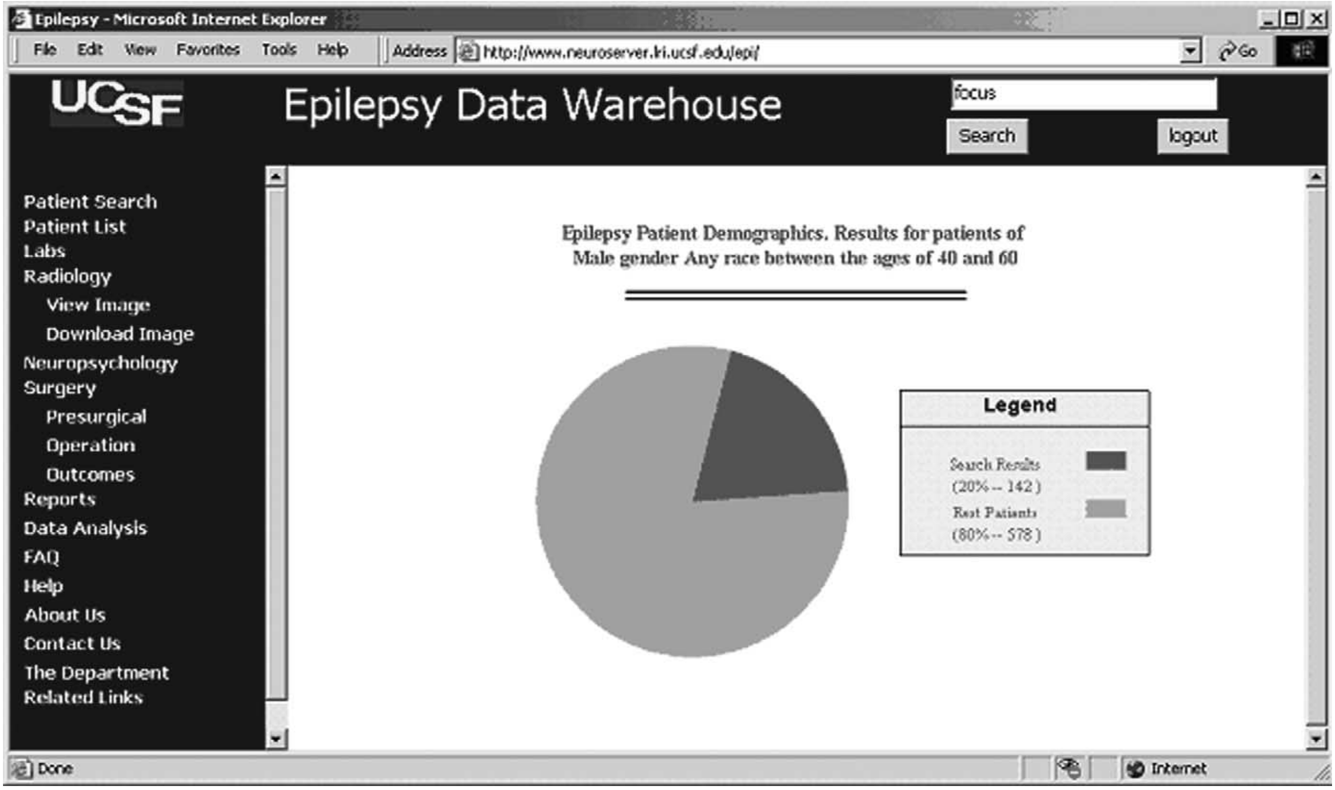

Figure 7. Query results are displayed in graphs such as the pie chart in this figure. The Legend provides specific numerical values.

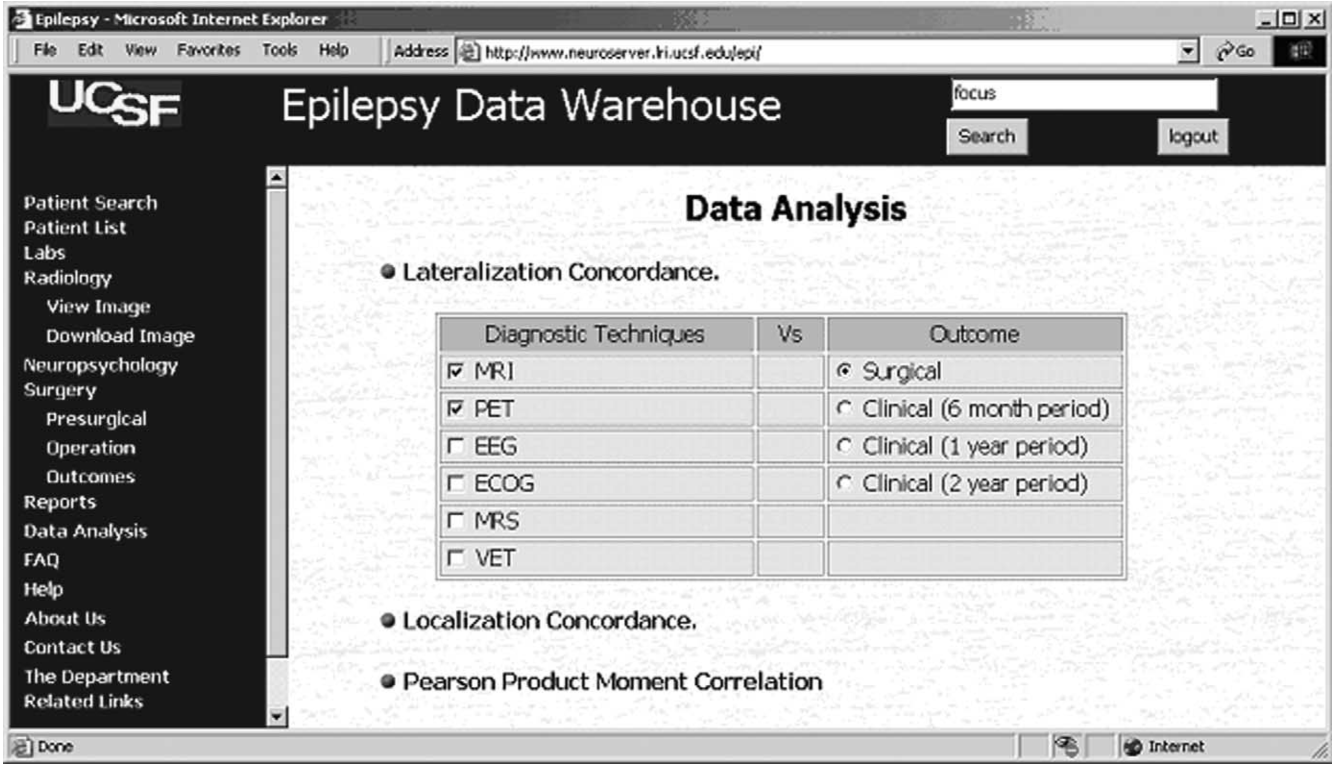

Figure 8. Lateralization concordance comparison can be made between various diagnostic methods and surgical outcomes. The analysis is run against a summary meta-data table especially constructed and coded by the clinical researchers via pre-defined Web forms during the course of data cleansing, ensuring accurate concordance results.

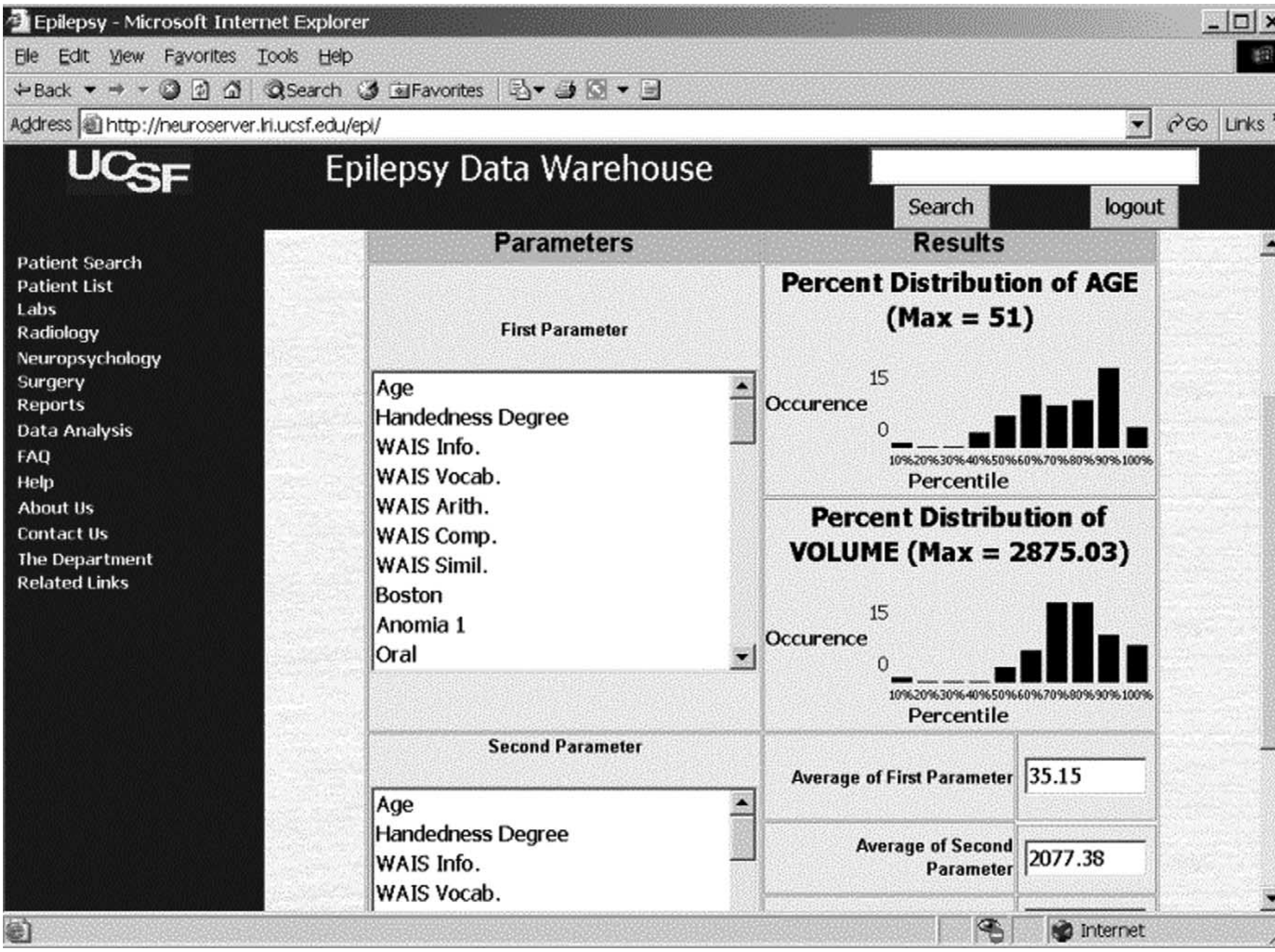

Figure 9. The Pearson Product Moment Correlation is calculated by simply selecting the first and second parameter (age and volume of regions of interest, in this case, Hippocampi, respectively), and clicking the "Compute correlation coefficient" button at the bottom. The results are displayed in the text fields and graphs on the right.

rying out ground-breaking clinical epilepsy research using the contents and data mining tools of the NIDS. One of the most powerful aspects of the NIDS is its ability to support retrospective analysis on hundreds or thousands of epilepsy patients; at least an order of magnitude of subject size more than previously published results. For example, during the pre-surgical evaluation, the neurologist determines the location of the epileptogenic focus and assesses mental function of the patient through a series of clinical and psychometric tests. While administering the WADA test, a psychologist tests the patient's memory to ascertain which side of the brain contributes most to memory. Normally, the good memory side is on the opposite side of where the epileptic seizures originate. However, a certain number of patients have their memory dominant hemisphere ipsilateral to their epileptogenic focus. Most epilepsy centers rule out surgical resection for these patients for fear of severely impairing the patient's memory while operating on their memory-dominant side, but the UCSF epilepsy center does operate on these patients, and therefore UCSF is in a unique position to leverage the power of a large database of epilepsy surgical patients to determine whether operating on these patients has a significant impact on their memory ability. Results on this particular study have been previously published (11). Such a large-scale, retrospective study would have required thousands of hours of tedious data gathering and analysis if not for the NIDS.

Currently, a team of neuroradiologists, neurologists, and neuropsychologists across various locations of our institution are accessing the Web-based NIDS to analyze surgery dates, surgery locations, psychometric test results from before and after the surgery for memory and lan-

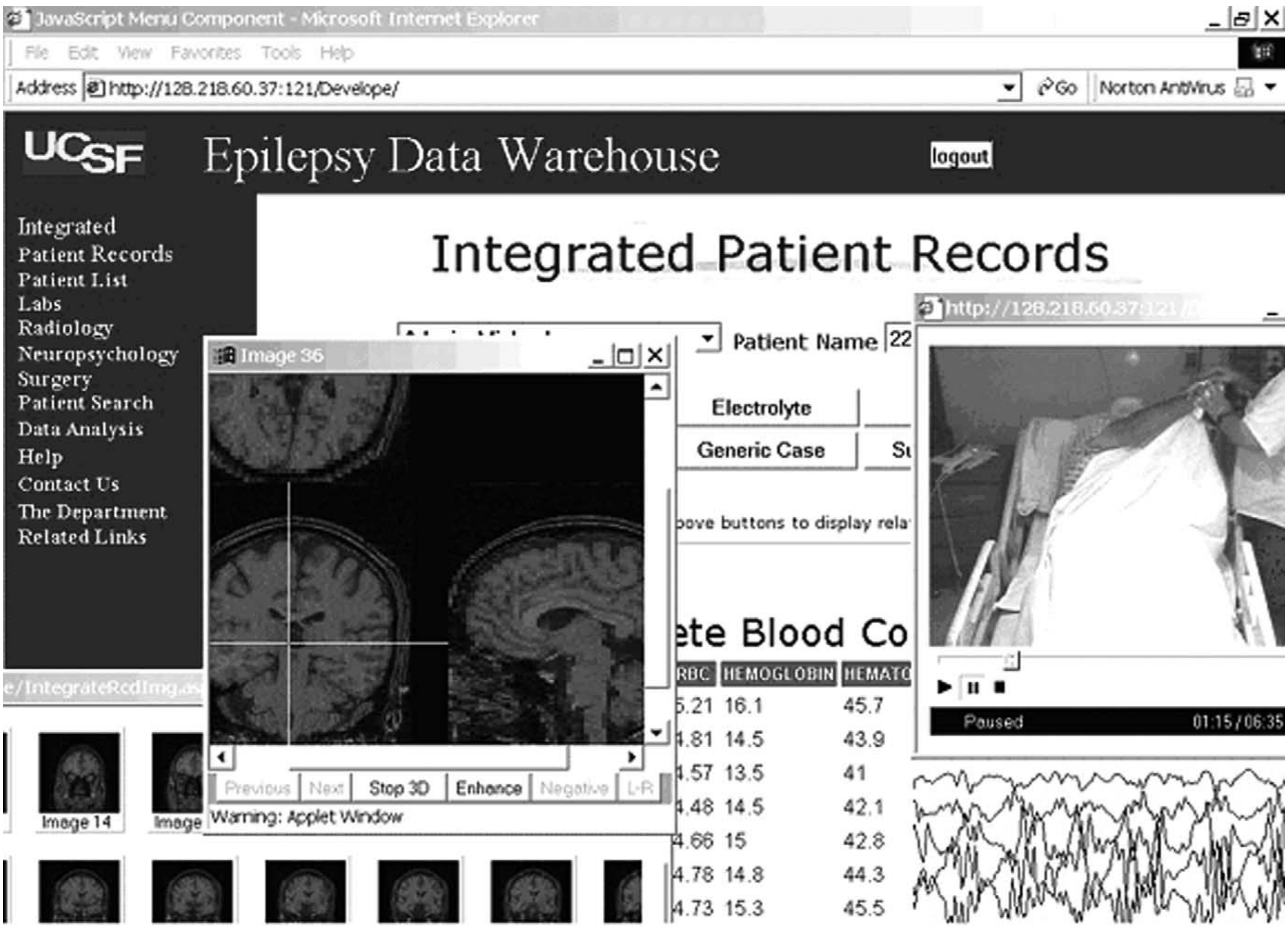

Figure 10. Multiple data sources are accessed and displayed simultaneously, including MRI slices, EEG, laboratory blood counts, and video telemetry.

guage, and MRI images from pre- and post-surgical scans to determine whether UCSF's policy of operating on the memory-dominant side has an effect on surgical outcome as well as post-surgical memory function. All of this data is being retrospectively analyzed, and the results are saved back to the NIDS for future reference.

DISCUSSION

The neuroinformatics database system is an important step toward harnessing the power of the data contained in the distributed information systems of today's hospital environment by creating separate, efficient research archives of integrated data and meta-data. Its primary mission of clinical research enables it to be optimized for online analysis and data mining of large numbers of data sets as opposed to clinical information system whose purpose is to support transactional requests for individual pieces of data. The NIDS provides semi-automatic and automatic facilities for importing raw data from these clinical systems and provides quality assurance utilities for reviewing and cleansing the data. Our experience has shown that reviewing and cleansing the data is an essential but time-consuming process. Despite the best efforts to accurately enter data into clinical information systems, the reality is that there are often missing data points, incomplete descriptions, and even incorrect entries that must be manually corrected by clinicians who have access to accurate data such as original patient medical records. For example, a neurologist has had to manually change the surgery date for 33 patients using the NIDS web interface; these dates were incorrectly recorded in the neurosurgery database, the source of surgery dates for the NIDS. Indeed, the incorrectly entered dates have been reported back to the administrators of the neurosurgery database so that they can update their clinical system with the correct dates, demonstrating an indirect benefit of the NIDS for the underlying data sources.

On the other hand, the design of an integrated research database requires careful consideration and set up to ensure the security and data integrity as well as meeting the requirements of IRB and other patient privacy regulations such as the Health Insurance Portability and Accountability Act (HIPAA). The NIDS resides behind the medical center's firewalls and thus can only be accessed by hospital intranet environment and controlled virtual private networks. The database system also has user authentication and audit trail capability that logs every command and database transaction which can then be queried against later on for performance and security measures. The database system has built-in facilities to automatically mask out and substitute the common elements required by HIPAA (Health Insurance Portability and Accountability Act). However, the participating clinical researchers and imaging specialists consider such a measure greatly reduce the effectiveness of NIDS in conducting clinical research, thus we did not turn on this functionality and opt for IRB-approved studies with particular disease-oriented clinical research groups.

We are presently extending the scope of the data warehouse into new disease types including brain tumor, breast cancer, Creutzfeldt-Jakob disease, and parasitic Ichthyophthirius multifiliis. Each of these new diseases will have their own unique data requirements, but the general framework of the data warehouse will be able to easily accommodate such new data. Furthermore, new utilities for image processing and data analysis are being designed and integrated into the framework. We also intend to include genetics and family information of a selected patient population, when the data become available in a separate research study, for the correlation of genotype information with various image phenotype observations in refractory epilepsy. Finally, we plan to test the scalability of the system in a multi-center image-based and web-enabled clinical trial including thousands of new refractory epilepsy patient datasets. As the data warehouse expands, it will become possible to research diagnostic rules for disease management.

REFERENCES

1. Ellisman MH. Scientific objective of the biomedical informatics research network. Washington, DC: BIRN, National Institutes of Health, 2003.
2. Hillman BJ. The American College of Radiology Imaging Network (ACRIN): research educational opportunities for academic radiology. Acad Radiol 2002; 9:561–562.
3. Hricak H. First open trial of the American College of Radiology Imaging Network: proper imaging approach for invasive cervical cancer. Radiology 2002; 225:634–635.
4. Siegel EL. Medical Imaging Resource Center (MIRC). Oak Brook, IL: Radiological Society of North America. 2003.
5. Huang HK. PACSbasic principles and applications. New York, NY: Wiley, 1999.
6. Arenson RL, Andriole KP, Avrin DE, Gould RG. Computers in imaging and health care: now and in the future. J Digit Imaging 2000; 13:145–156.
7. Kiebel SJ, Ashburner J, Poline JB, Friston KJ. MRI and PET coregistration–a cross validation of statistical parametric mapping and automated image registration. Neuroimage 1997; 5:271–279.
8. Woods RP, Mazziotta JC, Cherry S. MRI-PET registration with automated algorithm. J Comput Assist Tomogr 1993; 17:536–546.
9. Klein GJ, Reutter BW, Huesman RH, Coxson PG, Budinger TF. Volume-of-interest specification on arbitrarily resliced volume datasets. In: Hoffman E, ed. Medical Imaging 1995: Physiology and Function from Multidimensional Images. Newport Beach, CA: SPIE Proceedings; 1995:214–223.
10. Hoffman EA, Gnanaprakasam D, Gupta KB, Hoford JD, Kugelmass SD, Kulawiec RS. VIDA: an environment for multidimensional image display and analysis. SPIE Proceedings 1992;1660:694–711.
11. Wong STC, Hoo KS Jr, Knowlton RC, Laxer KD, Cao X, Hawkins R, et al. Design and applications of a multimodality image data warehouse framework. J Am Med Inform Assoc 2002; 9:239–354.
12. Majors HE, Garcia PA, Hoo KS, Wong STC, Barbaro NM, Laxer KD, et al. Does Wada memory asymmetry predict verbal memory change after temporal lobectomy? Presented at the Annual Meeting of Neurology Association (ANA), San Francisco, CA, 2003.