University of California
Larry L. Sautter Award Submission

Epilepsy Phenome Genome Project - Innovation in Clinical Study Informatics at the University of California, San Francisco

Submitted By:

Mr. Michael Williams, MA
University of California
Executive Director, Information Technology
UC San Francisco Diabetes Center, Immune Tolerance Network
and
Chief Information Officer
UC San Francisco Neurology, Epilepsy Phenome/Genome Project

Telephone: (415) 860-3581
Email: mwilliams@immunetolerance.org

Date Submitted: Friday, May 23, 2008

Table of Contents

1. Project Title

Epilepsy Phenome Genome Project –

Innovation in Clinical Study Informatics at the University of California, San Francisco

2. Submitter's Details

Mr. Michael Williams, MA
University of California
Executive Director, Information Technology
UC San Francisco Diabetes Center, Immune Tolerance Network
and
Chief Information Officer
UC San Francisco Neurology, Epilepsy Phenome/Genome Project
Telephone: (415) 860-3581
Email: mwilliams@immunetolerance.org

3. Names of project leader(s) and team members

Informatics Project Team

Mr. Gerry Nesbitt, MBA PMP
Director of Bioinformatics, UCSF

Mr. Kevin Miller
Data Manager and Senior Developer, UCSF

Mr. Alan Carpenter
Senior Developer, UCSF

Mr. Harry LeBlanc
Senior Data Architect (Contract), UCSF

Project Sponsors

Mr. Michael Williams
Chief Information Officer, EPGP

Dr. Daniel Lowenstein, M.D.
Department of Neurology at UCSF, Director of the UCSF Epilepsy Center

Ruben Kuzniecky, M.D.
Professor of Neurology, Comprehensive Epilepsy Center, NYU Medical Center

Margaret Jacobs
Program Director, NIH/NINDS Extramural Research Program, Neuroscience Center

4. Project Significance

In May 2007, a team of US scientists at 13 epilepsy centers led by UCSF and NYU received a grant of $15M to study the complex genetic factors that underlie some of the most common forms of epilepsy. The study, known as the "Epilepsy Phenome/Genome Project" (EPGP), is funded by the National Institute of Neurological Disorders and Stroke, and brings together over 50 researchers and clinicians from 15 medical centers around the country. The success of this study relied heavily on the advanced web-based informatics solutions developed by the EPGP informatics team at UCSF. Nearly one year after study commencement, these informatics solutions have demonstrated to the NIH/NINDS the objective indicators that EPGP will be successful and has assured continued funding to find better treatments and cures for epilepsy.

5. Project Description

5.1. What is EPGP?

In May 2007, a team of US scientists in 13 epilepsy centers led by UCSF and NYU received a grant of $15M to study the complex genetic factors that underlie some of the most common forms of epilepsy. The study, known as the "Epilepsy Phenome/Genome Project" (EPGP), is funded by the National Institute of Neurological Disorders and Stroke, and brings together over 50 researchers and clinicians from 15 medical centers around the country.

The EPGP study is being led by UCSF's Dr. Daniel H. Lowenstein. Dr. Lowenstein is Professor and Vice Chairman in the Department of Neurology at the University of California, San Francisco (UCSF), Director of the UCSF Epilepsy Center, and Director of Physician-Scientist and Education Training Programs for the UCSF School of Medicine. He was also a recent president of the American Epilepsy Society. Dr. Lowenstein was instrumental in establishing and nurturing the vision for the EPGP study to help further his life-long commitment to discovering better therapies and new cures for epilepsy.

The Epilepsy Phenome/Genome Project (EPGP) is a large-scale, national, multi-institutional, collaborative research project aimed at advancing our understanding of the genetic basis of the most common forms of idiopathic and cryptogenic epilepsies and a subset of symptomatic epilepsy; i.e. epilepsies that are probably related to genetic predispositions or developmental anomalies rather than endogenous, acquired factors such as CNS infection, head trauma or stroke. The overall strategy of EPGP is to collect detailed, high quality phenotypic information on 3,750 epilepsy patients and 3,000 controls, and to use state-of-the-art genomic and computational methods to identify the contribution of genetic variation to: 1) the epilepsy phenotype, 2) developmental anomalies of the brain, and 3) the varied therapeutic response of patients treated with AEDs.

5.2. Finding better therapies and cures to Epilepsy

Epilepsy is among the most common neurological disorders in the world, affecting one in every 100-200 people. Although heredity has been known since antiquity to cause epilepsy, the progress to date in identifying the genetic basis of epilepsy has been limited to the discovery of single gene mutations that cause epilepsy in relatively rare families. For the more common types of epilepsy, heredity plays a more subtle role, and it is thought that a combination of mutations in multiple genes likely determine an

individual's susceptibility to seizures, as well as the responsiveness to antiepileptic medications.

The approach to teasing apart the more complicated genetic factors in epilepsy requires a very large number of patients whose epilepsy has been extremely well-characterized. To this end, the EPGP investigators will be enrolling 3,750 patients and 3,000 controls over the course of the study. Details about seizure types, EEGs, imaging studies, and effects of treatment will be collected and archived in a central data repository, and all participants will be asked to submit a sample of blood or saliva as a source of their DNA. (All the clinical information and the DNA samples will be de-identified so that it cannot be traced back to a specific individual.) Once this first phase of the study is completed, Dr. Neil Risch and colleagues at the UCSF Institute for Human Genetics, along with researchers at Emory University, will carry out "whole genome scans" and look for potential connections between patterns of DNA sequences and specific characteristics of epilepsy in the study population.

The long-term goal of EPGP is to identify potential molecular targets that could be the basis of much more specific and effective treatments for patients who have epilepsy, and the prevention of epilepsy in those at risk. The EPGP represents a unique attempt to capitalize upon the recent advances in high-throughput, genomic techniques and the wealth of information accumulated through bioinformatics, and apply this knowledge and these methodologies to epilepsy. The most significant and immediate benefits to be gained from the EPGP include:
1) creation of a detailed phenotype database;
2) maintenance of a permanent repository of DNA and cell lines linked to the phenotype data;
3) support for an infrastructure that will facilitate basic and clinical studies of seizure disorders for many years to come.

Furthermore, the EPGP will conduct groundbreaking research that will characterize the clinical, electrophysiological, and neuroimaging phenotypes of 3,750 patients with discrete subtypes of idiopathic-generalized, focal, or severe early-onset pharmacoresistant epilepsy.

5.3. Informatics was Key to winning EPGP Funding

The EPGP consortium, led by Dr. Lowenstein, commenced work on preparing the study proposal back in 2003 through funding by various foundations. Funding based on the initial proposal submission in 2006 was declined due to weaknesses in the Informatics infrastructure and plans for sample management, at which point the consortium needed to revamp its informatics strategy. At that time, Dr. Lowenstein approached the Executive Director of IT at UCSF's Immune Tolerance Network which had a very mature bioinformatics and IT infrastructure supporting 30+ large-scale clinical trials at multiple sites, and who had effectively solved the informatics issue already. With the

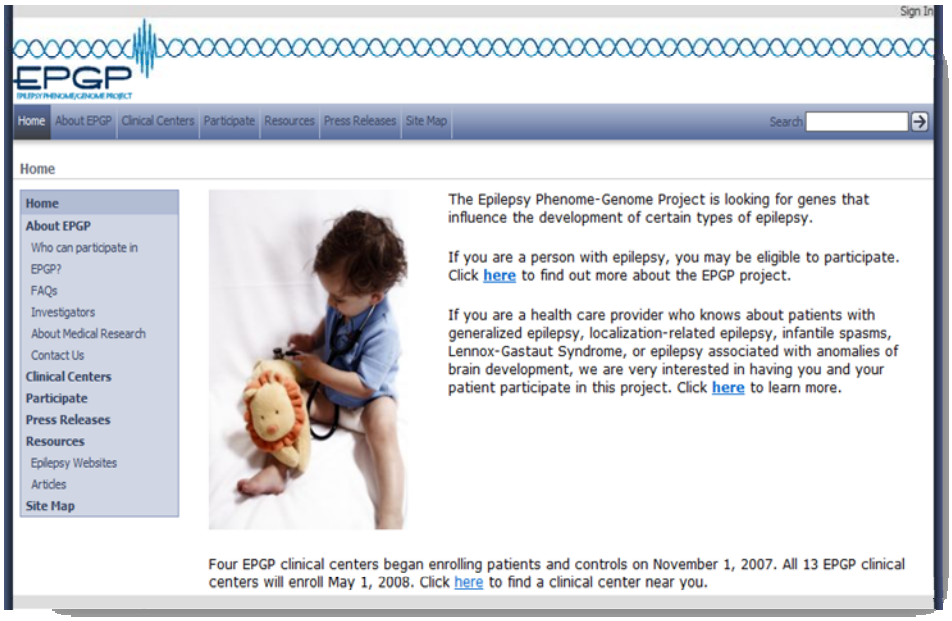
new informatics strategy formulated, the revised EPGP proposal was resubmitted and funding was received allowing the EPGP study to commence in May 2007.

5.4. The ARCAMIS Platform

The clinical management suite developed by the EPGP Informatics team is called the ***Academic Research Computing & Analysis Managed Informatics Services*** (ARCAMIS) Platform. It provides the following functional modules:

EPGP Public Website

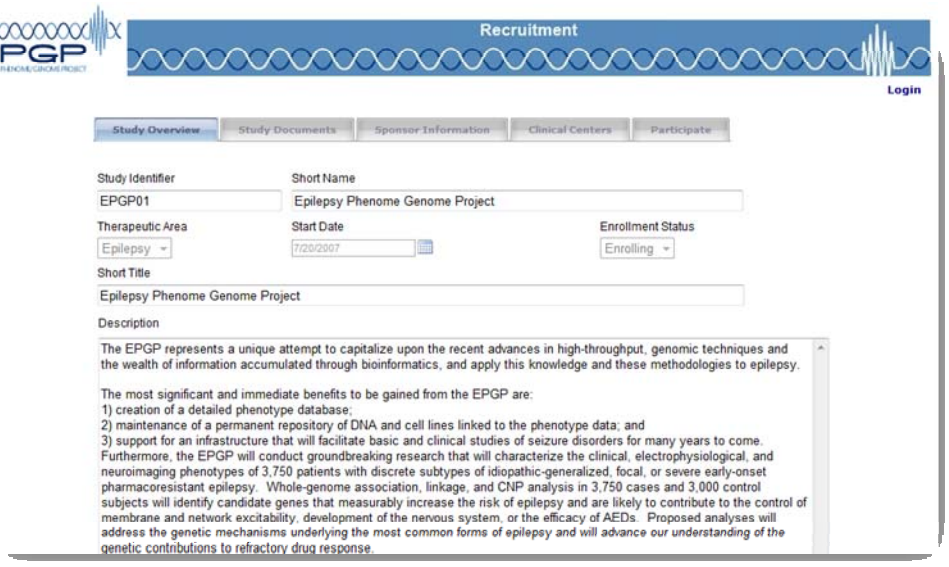
The EPGP Public Website, http://www.epgp.org/ has been up and running since early 2007. It has enabled potential participants, principal investigators, family physicians and the general public to view information of the EPGP study, such as the type of participants EPGP is seeking and the clinical sites involved in the study.

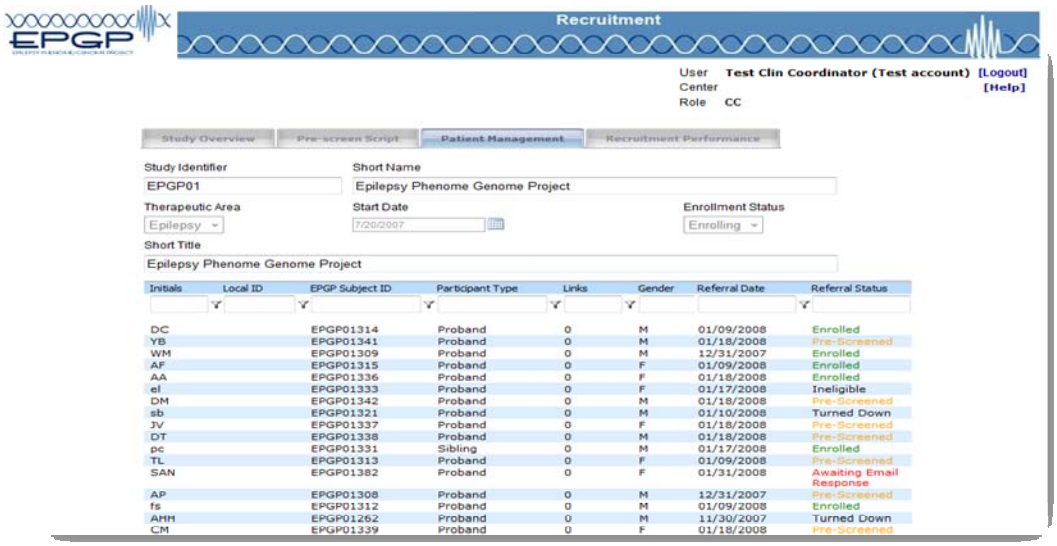


Patient Recruitment

The EPGP project aims to recruit 3,750 epilepsy patients and 3,000 controls at thirteen academic epilepsy Clinical Centers (CCs) over a five-year period. It was vital that potential participants have the ability to view information on the EPGP study, and register their interest in participating in the study. It also needed to provide clinical centers to track and manage potential and actual participants.

The Patient Recruitment application, available at https://recruitment.epgp.org/, has been in production since September 2007 and has attracted more than 200 recruits.

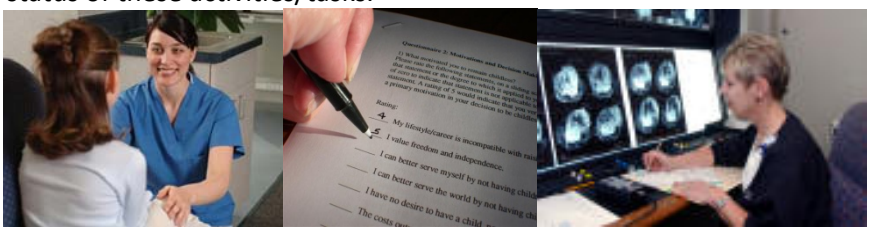




Participant Activity Tracking

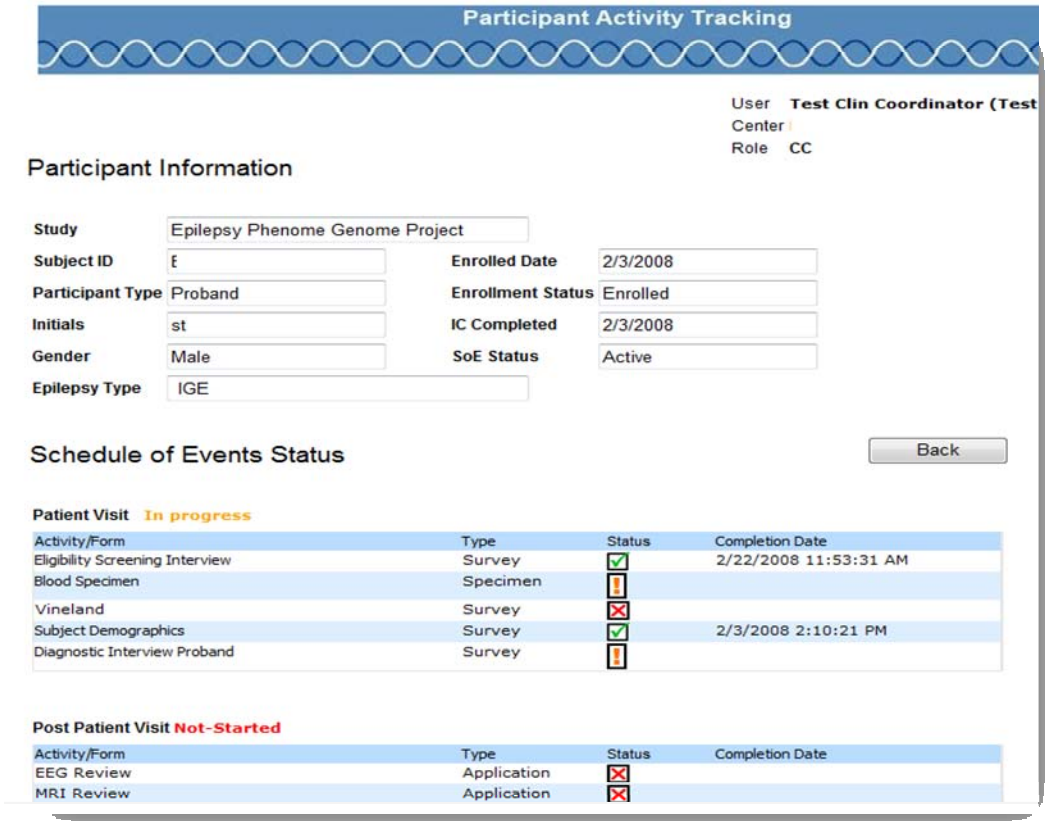
Which such throughput of participants at each clinical center, there was a need to develop a simple, easy to use web application to enable study coordinators at clinical centers manage participant visits and record participant questionnaires. Participant activities include is a group of one or more tasks that are planned (or have been performed) for a participant on a clinical study, such as completing an informed consent, taking a blood specimen or completing a phenotypic questionnaire. The study nurse required the ability to view a list of all the

participants at a specific clinical center and the tasks associated with each participant, and the completion status of these activities/tasks.



In addition to tracking participants and the study activities, the participant activity tracker allows clinical sites to record and track adverse events and issues electronically. An Adverse Event is any change in health that occurs in a person after he or she enrolls in a clinical study. The adverse event module allows the study coordinator record any impacts on the participant's health so that the relevant authorities are notified and the appropriate action is taken. The issues management module handles all issues concerning the EPGP study. Issues can be related to the protocol (such as protocol deviations), questions on the data collected, manual of operations, specimen tracking irregularities, subject related, procedures issues, etc. This module allows the study coordinator review the issues they submitted or are assigned to them.

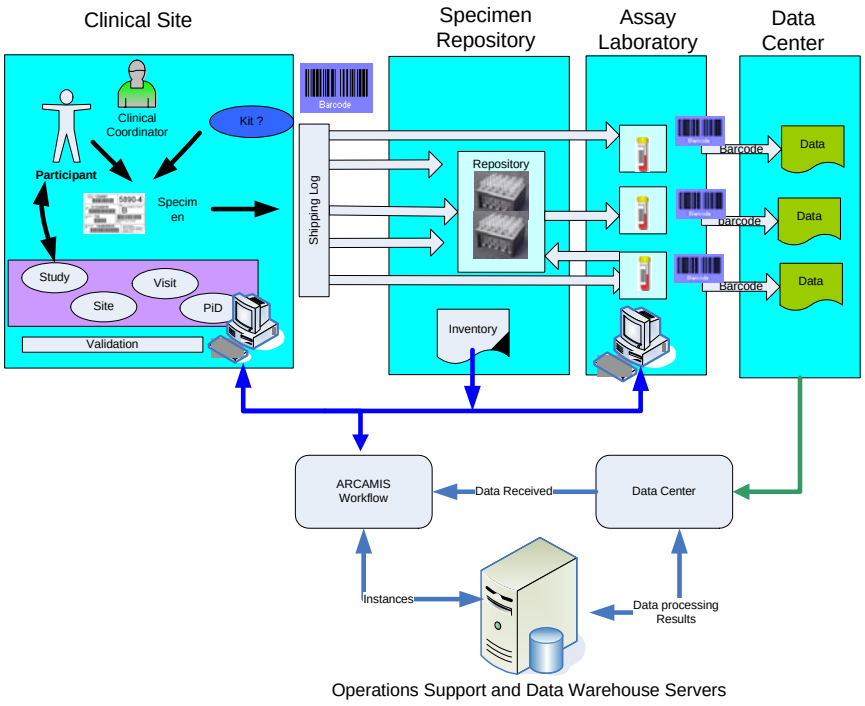
The Participant Activity Tracking application, available at https://participants.epgp.org, has been in production since November 2007.



Specimen Management

The specimen tracking system is an easy to use web application to manage the lifecycle of the specimens collected from subjects from the moment the blood is drawn or saliva is collected, to the moment is arrives at the core. It provides a complete specimen tracking application will help EPGP clinical sites and laboratories automate the rules governing specimen identification and routing, track specimens from the time they are drawn until ultimate disposition, maintain a history of the processing of all specimens and their derivatives, and automatically alert personnel to specimen processing problems or the occurrence of significant specimen processing events.

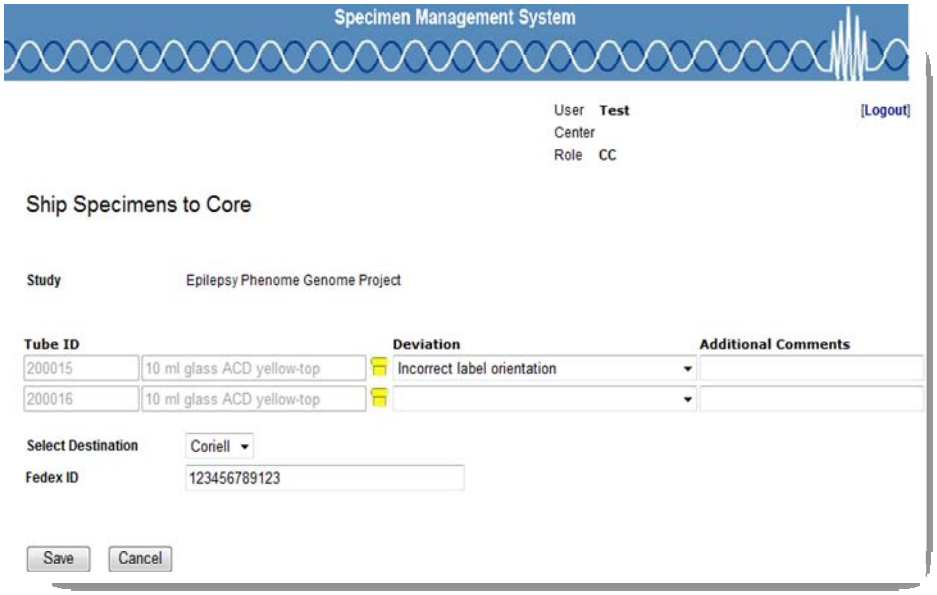
The Specimen Management module, available at https://specimens.epgp.org, has been in production since November 2007.



The ARCAMIS application is based on a workflow engine and extended through MS .Net platform, providing a fully functional base with configuration capabilities that allow rapid integration of EPGP processes and requirements. Technical design and development processes are consistent with data collection systems used throughout the EPGP, such that ARCAMIS is seamlessly integrated into EPGP bioinformatics and information management systems. Data are linked to complete specimen histories via a bar-coding system in use throughout the specimen management process.

a. ordering of specimen collection kits
b. site kit inventory
c. prepare for visit module which allows tubes to be uniquely associated with one participant

d. specimen shipment module which tracks the specimen from the site to the repository and from the repository to the core
e. view specimen inventory which allows the site and clinical operations manager view specimens residing at the core



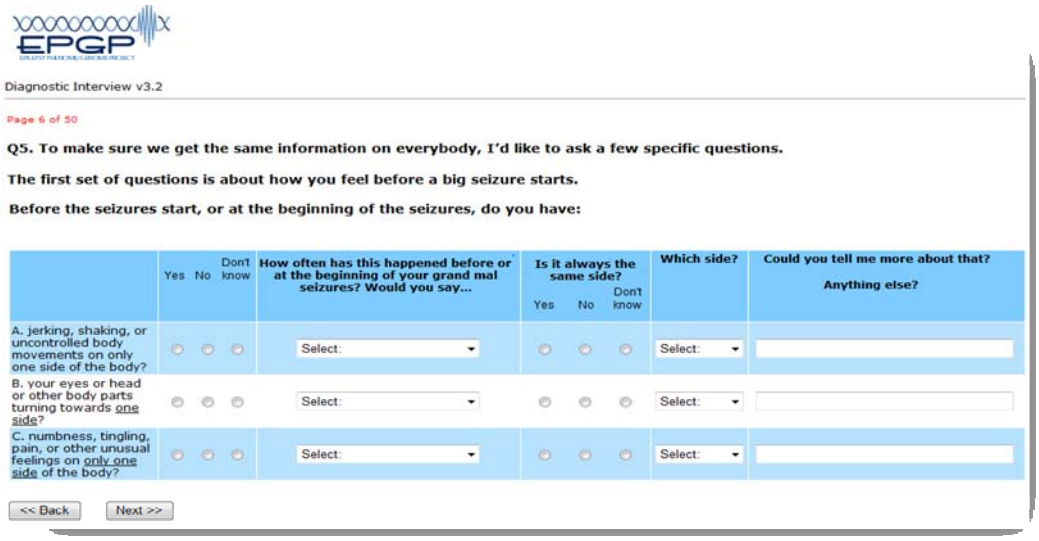
Phenotypic Survey Engine

The EPGP study required an advanced web-based solution for gathering phenotypic data. The survey engine selected all. Another key achievement was integrating the survey engine with the EPGP BioMed database so that the participant's Schedule of Events is updated on completion of specific surveys and the responses collected can be integrated of the phenotype tables in the EPGP BioMed database.

The phenotypic data collection instruments designed using the survey tool include:

1. Brief Eligibility Screen
2. Brief Medical Record Review
3. Patient Activities
4. Eligibility Screening Interview
5. Adult Depression Screen
6. Subject Demographics
7. Diagnostic Interview
8. Patient Not Present
9. Surgical form
10. Morphology Record Abstraction Checklist
11. Medical Record Abstraction
12. Final diagnosis form

13. Pharmacogenomics Form
14. Pharmacogenomics 1yr Form
15. EEG inclusion criteria
16. EEG component evaluation
17. EEG ictal evaluation
18. MRI form--Site Version
19. MRI form--Core Version
20. Patient Present, 1 Year Follow Up
21. 1-year follow up form



EPGP

Diagnostic Interview v3.2

Page 6 of 50

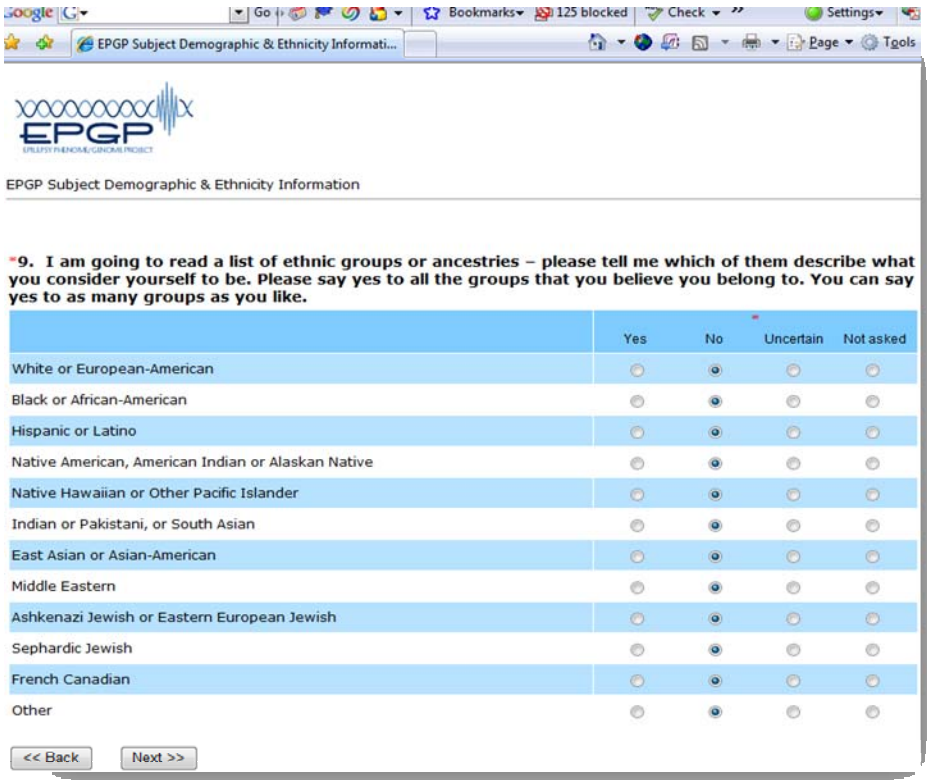
Q5. To make sure we get the same information on everybody, I'd like to ask a few specific questions.

The first set of questions is about how you feel before a big seizure starts.

Before the seizures start, or at the beginning of the seizures, do you have:

	Yes	No	Don't know	How often has this happened before or at the beginning of your grand mal seizures? Would you say...	Is it always the same side? Yes	No	Don't know	Which side?	Could you tell me more about that? Anything else?
A. jerking, shaking, or uncontrolled body movements on only one side of the body?	○	○	○	Select:	○	○	○	Select:	
B. your eyes or head or other body parts turning towards one side?	○	○	○	Select:	○	○	○	Select:	
C. numbness, tingling, pain, or other unusual feelings on only one side of the body?	○	○	○	Select:	○	○	○	Select:	

<< Back Next >>



Barcode Printing and Scanning

Barcode printing and scanning solution deployed to all sites. Biological specimens are collected in barcode-labeled primary collection tubes by clinic personnel. These collection tubes are provided by the EPGP in visit-specific kits. Each kit is barcode labeled and the information concerning the tube types and numbers included in that kit is pre-loaded into the ARCAMIS system. When clinical personnel scan the kit barcode and enter the individual patient ID (PID), that patient is immediately linked to all kit components.

Shipment Report

Your shipment has been confirmed. An email has been sent to the destination core informing them that a specimen has been sent via Fedex (ID: 543245678)

Study:	**EPGP01 Epilepsy Phenome Genome Project**
Investigator:	**Dr Dan Lowenstein, MD**
Collaborator:	
Clinical Center:	**NYU Medical Center**
Study Coordinator:	**Nr. Joan Powell**
	Email: nyu@epgp.org, Tel: (123) 432 2345
Collection Date:	**9/12/2007**
Destination Core:	**Coriell Repository**
Shipment Date:	**9/13/2007**
Fedex Tracking ID:	**543245678**
EPGP Subject ID:	**EPGP0112345**
Visit:	**1**

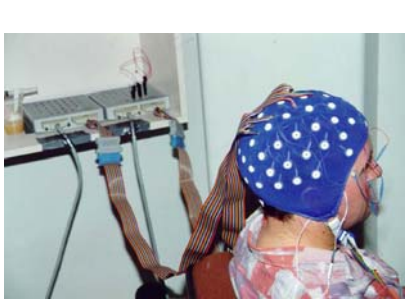
Subject ID: EPGP0112345

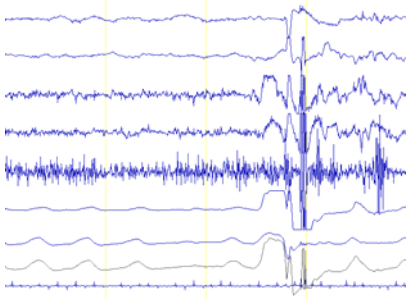
Shipment Contents:

Tube ID	Tube Description	
200003	10ml glass ACD yellow-top <5ml of blood was collected	Tube ID: 200003
200004	10ml glass ACD yellow-top	Tube ID: 200004

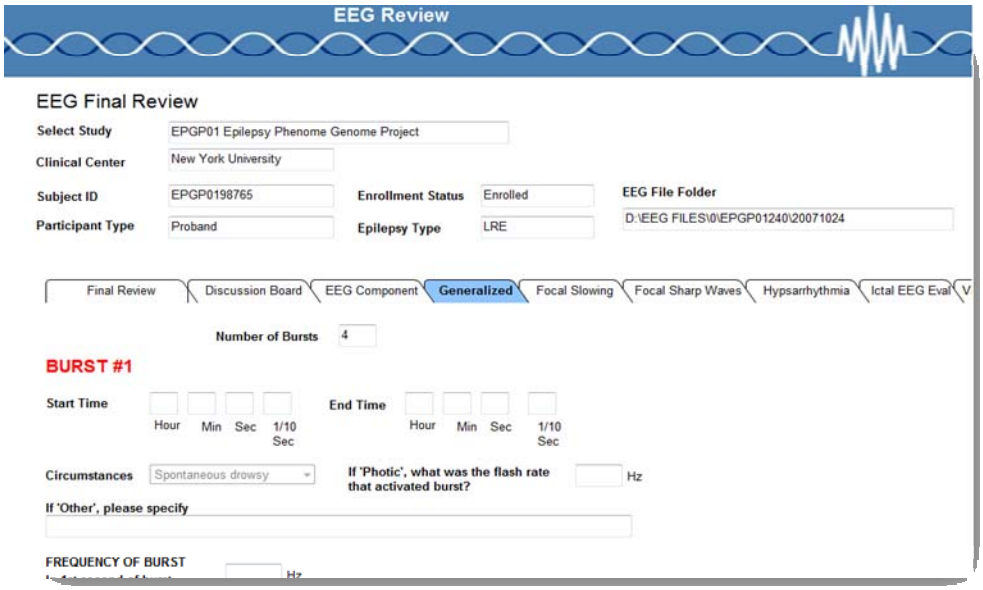
EEG Review

EEG, abbreviation for electroencephalograph, is a diagnostic test which records and measures electrical activity in the brain, or "brainwaves". This is important for the detection and diagnosis of epilepsy. EPGP's EEG Review application enables the neurophysiology core members assess, review and score subject's EEGs.

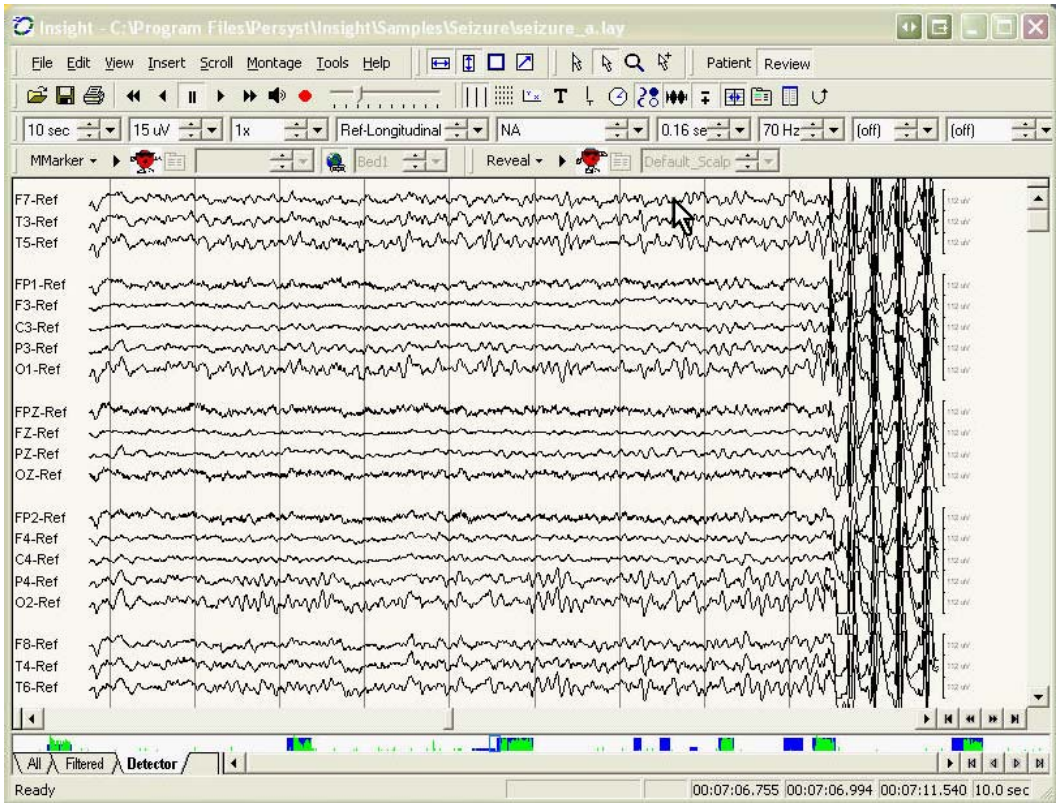




The EEG Review application was launched in March 2008 and is available at https://eeg.epgp.org.

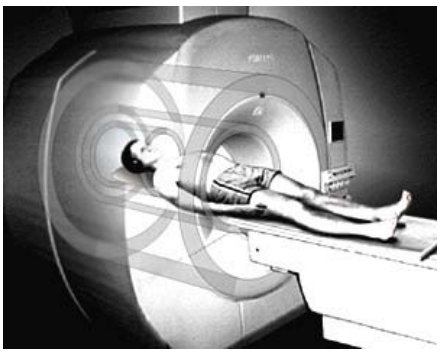


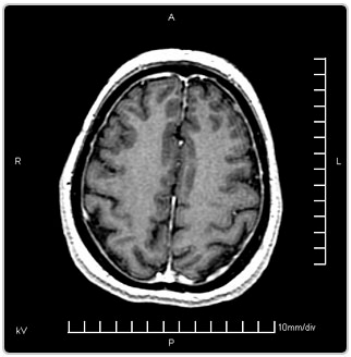
In order to allow EEG Reviewers view EEG files, we have implement Insight EEG Viewer from Persyst. The unique feature of this viewer is the ability to view EEGs from multiple vendor formats, a key requirement for EPGP as clinical centers will have heterogeneous EEG machines.



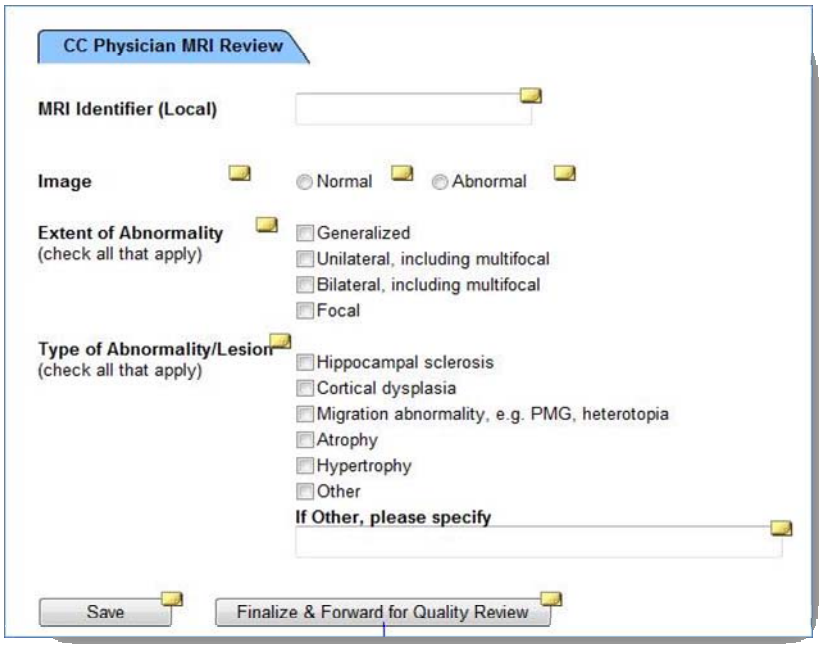
MRI Review Module

MRI , an abbreviation for Magnetic Resonance Imaging, is a brain scanning technique that generates cross-sectional images of a human brain by detecting small molecular changes. MRI scans reveal a contrast between normal and abnormal tissues. The image produced is similar to those generated by CT scans. The EPGP MRI Review application enables imaging core members assess, review and score subject's MRIs.

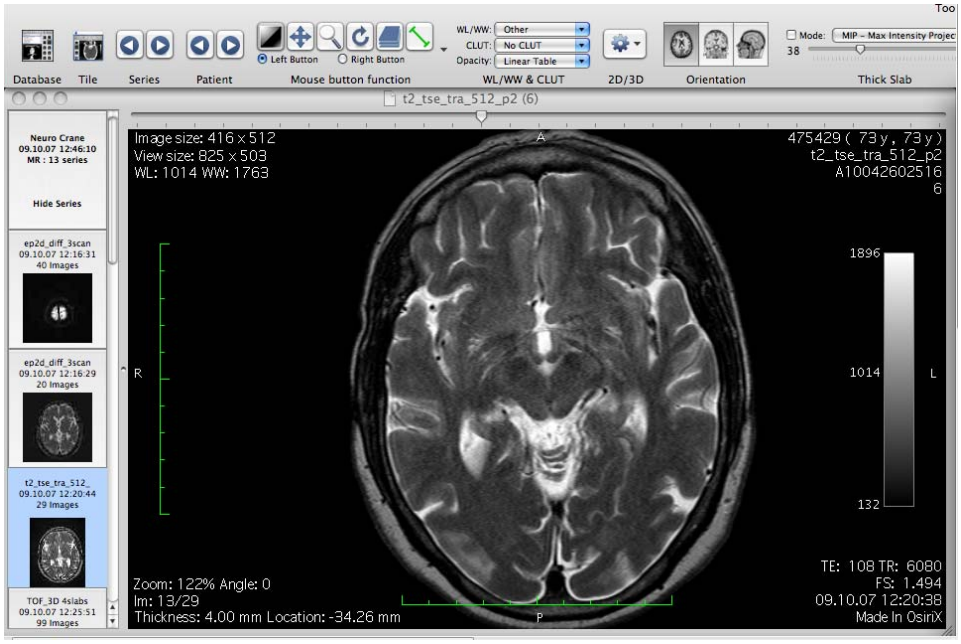




The MRI Review application was launched in March 2008 and is available at https://mri.epgp.org.

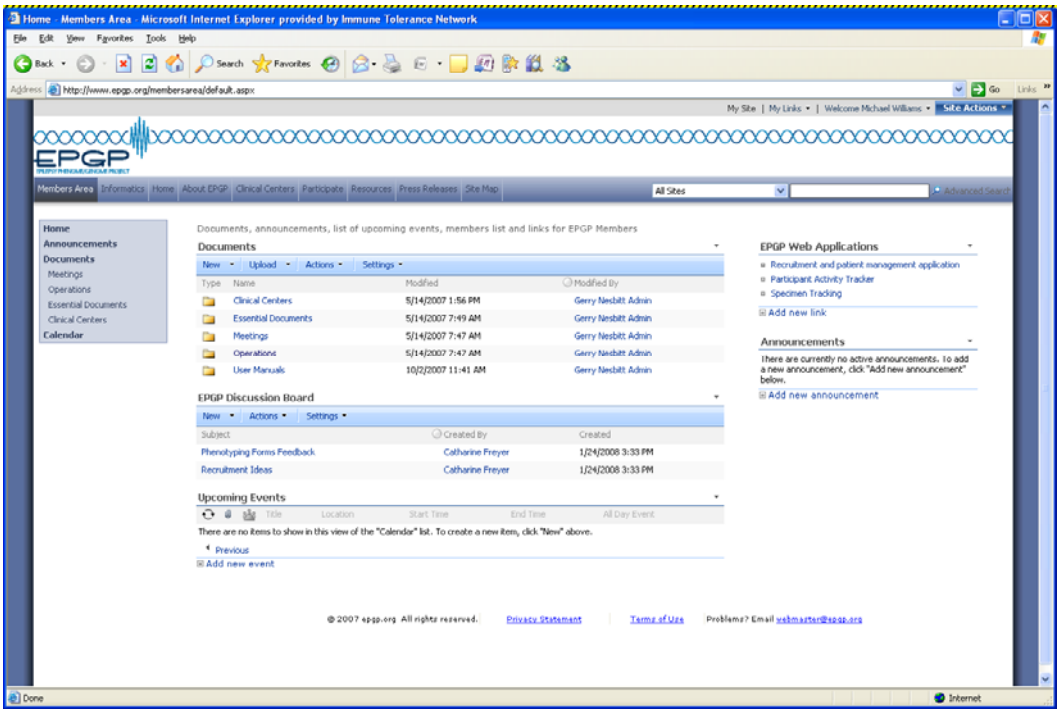


The MRI reviewers will use OsiriX as the DICOM viewer which has been specifically designed for navigation and visualization of multi-modality and multi-dimensional images. Osirix is an image processing software dedicated to DICOM images produced by medical equipment (MRI, CT, PET, PET-CT) and confocal microscopy (LSM and BioRAD-PIC format). It can also read many other file formats: TIFF (8,16, 32 bits), JPEG, PDF, AVI, MPEG and QuickTime. It is fully compliant with the DICOM standard for image communication and image file formats.



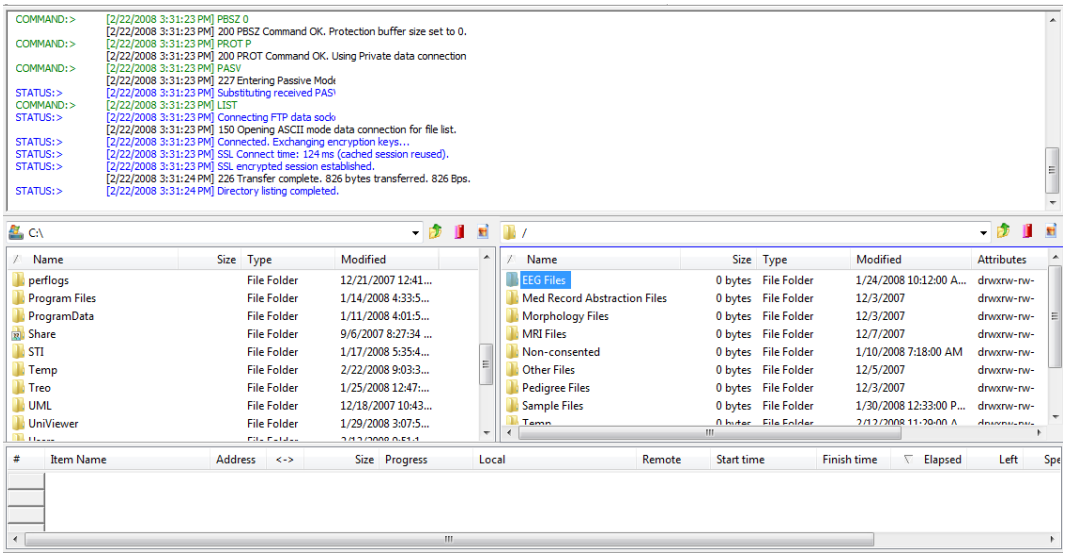
Web Portal for Consortium Members

A website dedicated to team based communications was essential to the EPGP study. The web portal for consortium members provides access to essential documents in a central location, fully versioned and controlled. In addition, the web portal provides access to a discussion board, issue management, contacts lists, announcements and links to launch EPGP web applications.



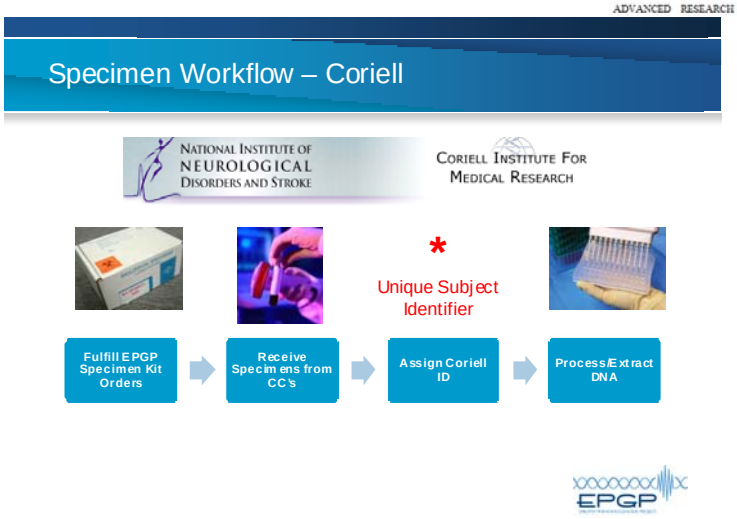
Secure File Transfer for EEG, MRI and Morphology files

EPGP's secure FTP server allows clinical sites to upload and provide access to EEGs, MRI and morphology files. The security implemented includes both Implicit SSL over port 990 and Explicit SSL over port 21.



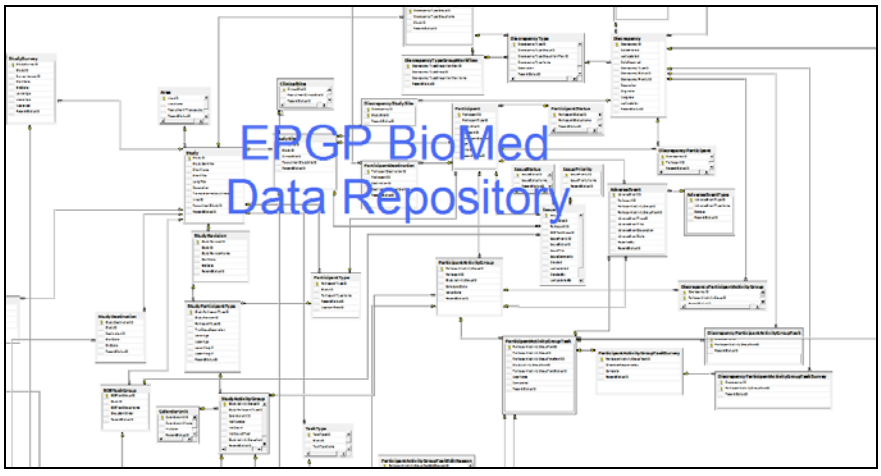
End-user Training

The EPGP informatics team has provided computer based trainbing, user manuals and classroom training materials to ensure the end-users are trained to use the EPGP applications correctly. A 24 hour 1800 Informatics Support Line telephone help desk is available for all site personnel. This help desk is staffed by three permanent EPGP employees to ensure complete coverage for sites that may encounter problems with using the system outside of standard business hours.



The EPGP BioMed Data Repository

Supporting all the EPGP web-based applications is the EPGP BioMed database. This ensures that all applications are integrated in terms of data and security. The database architecture is supported by a physical infrastructure including networking and servers, data management personnel and the operational procedures and processes, designed to support the collection, storage and delivery of data, and to preserve its fidelity, sensitivity and usability. The architecture of the BioMed database, algorithms and procedures for data validation and cleaning and the approaches for data access are consistent with conventional data warehousing methodologies and permit the leverage of industry-proven approaches and tools.



We have established a standard methodology for movement of file sets from external sources into the EPGP Data Center. All data loaded into a staging area and are archived precisely as they were received. This archive is therefore a permanent record of what the core facilities have sent to the EPGP prior to any processing. The archive is maintained on permanent media, readily available to the Data Center processing team. Files are processed out of the archive into a staging queue through a series of validations and verifications to ensure completeness and

compliance with data specifications. These “back office” procedures are used to evaluate, monitor and resolve issues about file consistency, content and the processing status of each upload. The automation of these workflows consists of operational monitoring (*e.g.*, issue tracking and resolution, logging and monitoring of process status) as well as the physical component of data extraction: moving, cleaning, and loading the data into a relational database, which is part of the EPGP Data Warehouse.

We have established our data models based on the assumption that data sources and genotypic analysis technologies will change over time, while presuming that the underlying foundation of a study will remain constant. Our models assume that all clinical trials involve patients who participate in a study, who have patient visits that generate observations and specimens, upon which measurements are made. As we move forward, the ontology underlying our data warehouse will reflect these shared elements and will become the single source from which all data elements will be defined or established. Our current data model will support this planned transition.

The data model of the data repository consists of over 400 relational tables organized to support not only data storage but also operational processing and integration. The data model is highly normalized and optimized for transaction processing and curation.

Security and Privacy Requirements

It was essential that all EPGP informatics applications supported security and privacy standards, including 21CFR11 and HIPAA.

The United States government Food and Drug Administration (FDA) department ruling on Electronic Records and Electronic Signatures (21 CFR 11) states that "the regulations in this part set forth the criteria under which the agency considers electronic records, electronic signatures, and hand-written signatures executed to electronic records to be trustworthy, reliable, and generally equivalent to paper records and hand-written signatures executed on paper". The EPGP applications satisfies all aspects of 21CFR11 including system access, audit trails, trusted 3rd party time, electronic signatures, controls for identification codes/passwords, archiving, procedures and validation.

The EPGP applications conform to all aspects of the Health Insurance Portability and Accountability Act (HIPAA). All Private Health Information (PHI) information is adequately secured and all web based traffic is encrypted using SSL 128bit. Access to certain sections of the website and subject data is highly restricted and will require the user to login to the website. Access to features such as document posting and viewing subject data is restricted based on the user’s role and the clinical center they are associated with.

All applications require the user to login using a single secure login name and password which are controlled by the EPGP Informatics/Infrastructure team. Users are assigned to specific roles and clinical centers, thereby restricting users’ access to specific functionality pertinent to their role and data pertinent to their clinical center.

6. The technology utilized in the project

EPGP uses various software development and web design packages to develop custom applications and websites on the Microsoft Sharepoint 2007 and .NET Platforms. They are used in conjunction with EPGP application development environments including Visual Studio 2005 Professional Edition. This serves as a comprehensive development environment intended for individual developers to build high-performance, multi-tier applications. SharePoint 2007 provides workflow, document management, and various EPGP Platform components.

Other third-party products integrated into EPGP's ARCAMIS suite include CheckBox survey engine, MS Reporting Services, CuteFTP Server, Persyst Insight EEG Viewer and Osirix - DICOM Viewer.

7. The timeframe of implementation

Development of the technologies to support EPGP commenced in May 2007. In terms of the hardware and communications technologies, EPGP was able to leverage the existing sophisticated IT infrastructure at the ITN which provided a computing infrastructure centralized in Tier 1 networked enterprise class datacenters, virtualization, and data consolidation onto a Storage Area Network (SAN) with off-site disaster recovery replication to address these challenges. This meant that EPGP had a fast and reliable communications network, server infrastructure, communication tools up and running in a matter of days!

In terms of software applications, the new informatics team at EPGP had the challenge of designing, developing and implementing web-based applications in a short timeframe for use across the 13 geographically dispersed clinical sites, which included web applications for a public website, subject recruitment, participant activity tracking, specimen tracking, phenotype data collection, EEG Review and MRI review.

Web application	***Completion Date***
Public Website	August 2007
Recruitment	Sept 2007
Specimen Tracking	Jan 2008
Participant Activity Tracker	Jan 2008
EEG Review	Mar 2008
MRI Review	Apr 2008