WaveformECG: A Platform for Visualizing, Annotating, and Analyzing ECG Data

Raimond L. Winslow, **Stephen Granite**, and **Christian Jurado**

Johns Hopkins University

Abstract

The electrocardiogram (ECG) is the most commonly collected data in cardiovascular research because of the ease with which it can be measured and because changes in ECG waveforms reflect underlying aspects of heart disease. Accessed through a browser, WaveformECG is an open source platform supporting interactive analysis, visualization, and annotation of ECGs.

The electrocardiogram (ECG) is a measurement of time-varying changes in body surface potentials produced by the heart's underlying electrical activity. It's the most commonly collected data in heart disease research. This is because changes in ECG waveforms reflect underlying aspects of heart disease such as intraventricular conduction, depolarization, and repolarization disturbances, [1,2] coronary artery disease,[3] and structural remodeling.[4] Many studies have investigated the use of different ECG features to predict the risk of coronary events such as arrhythmia and sudden cardiac death, however, it remains an open challenge to identify markers that are both sensitive and specific.

Many different commercial vendors have developed information systems that accept, store, and analyze ECGs acquired via local monitors. The challenge in applying these systems in clinical research is that they're closed and don't provide APIs by which other software systems can query and access their stored digital ECG waveforms for further analyses, nor the means for adding and testing novel data-processing algorithms. They're designed for use in patient care, rather than for clinical research. Despite the ubiquity of ECGs in cardiac clinical research, there are no open, noncommercial platforms for interactive management, sharing, and analysis of these data. We developed WaveformECG to address this unmet need.

WaveformECG is a Web-based tool for managing and analyzing ECG data, developed as part of the CardioVascular Research Grid (CVRG) project funded by the US National Institutes of Health's National Heart, Lung, and Blood Institute.[5] Users can browse their files and upload ECG data in a variety of vendor formats for storage. WaveformECG extracts and stores ECGs as a time series; once data are uploaded, a browser can select, view, and scroll through individual digital ECG lead signals. Points and time intervals in ECG waveforms can be annotated using ontology from the Bioportal ontology server operated by the National Center for Biomedical Ontology (NCBO),[6] and annotations are stored with the waveforms for later retrieval, enabling features of interest to be marked and saved for others. Users can select groups of ECGs for computational analysis via multiple algorithms, and analyses can be distributed across multiple CPUs to decrease processing

time. WaveformECG has also been integrated with the Informatics for Integrating Biology and the Bedside (I2B2) clinical data warehouse system.[7] This bidirectional coupling lets users define study cohorts within I2B2, analyze ECGs within WaveformECG, and then store analysis results within I2B2.

WaveformECG has been used by hundreds of investigators in several large longitudinal studies of heart disease including the Multi-ethnic Study of Atherosclerosis (MESA),[8] the Coronary Artery Disease Risk in Young Adults (CARDIA),[9] and the Prospective Observational Study of Implantable Cardioverter Defibrillators (PROSE-ICD)[10] studies. A public demo version is available for use through the CVRG Portal. All software developed in this effort is available on GitHub, under the CVRG project open source repository, with an Apache 2.0 license. Instructions for deployment of the WaveformECG tool is available on the CVRG wiki.

WaveformEC G System Architecture

WaveformECG is accessed via a portal developed using Liferay Portal Community Edition, an open source product that lets developers build portals as an assembly of cascading style sheets (CSS), webpages, and portlets (see Figure 1).[11] Liferay was extended to use Globus Nexus, a federated identity provider that's part of the Globus Online family of services,[12] for authentication and authorization. Users login to WaveformECG with their Globus credentials or credentials from other identity providers (such as InCommon or Google) linked with their Globus identity. Following authentication, users can access four separate portlet interfaces: upload, visualize, analyze, and download. We developed several backend libraries and tools supporting these interfaces that enable storage, retrieval, and analysis of ECG time-series data, metadata, and their annotations.

The upload and visualize portlets utilize an open source distributed storage system running on Apache Hadoop and Hbase known as the open time-series database (OpenTSDB),[13] a database system optimized for storage of streaming time-series data. OpenTSDB sits on top of Apache HBase,[14] an open source nonrelational distributed (noSQL) database developed as part of the Apache Software Foundation Hadoop project (https://hadoop.apache.org). Apache Zookeeper (https://zookeeper.apache.org) serves as the synchronization and naming registry for deployment. This configuration allows HBase to be deployed across multiple servers and to be scaled to accommodate high-speed, real-time read/write access of massive datasets—an important consideration because WaveformECG is being extended to accept real-time ECG data streams from patient monitors. Reported ingest rates for OpenTSDB in many different applications range from 1 to 100 million points per second. OpenTSDB defines a time series as a set of time-value pairs and assigns a unique time-series identifier (TSUID) to each time series. OpenTSDB also supports execution of aggregation functions on query results (a query result is a read of time-series data). Examples of aggregators include calculations of sums, averages, max-min values, statistics, and custom functions.

OpenTSDB provides access to its storage and retrieval mechanisms via RESTful APIs.[15] With this capability, other software systems can query OpenTSDB to retrieve ECG datasets. The open source relational database system PostgreSQL[16] maintains file-related information

and other metadata. PostgreSQL is also used for portal content management (user identities, portal configuration, the Liferay document and media library [LDML], and so on), storage of all uploaded ECG data files in their native format, and other ECG metadata (sampling rate, length of data capture, subject identifier, and so on).

Data Upload and Management

WaveformECG can import ECG data in several different vendor formats, including Philips XML 1.03/1.04, HL7aECG, Schiller XML, GE Muse/Muse XML 7+, Norav Raw Data (RDT), and the WaveForm DataBase (WFDB) format used in the Physionet Project.[17] In addition to the ECG time series, Philips, Schiller, and GE Muse XML files also contain results from execution of vendor-specific proprietary ECG analysis algorithms, metrics on signal quality, and other data. These data are also extracted and stored.

Within the upload interface (Figure 2a), users can browse their file system to locate folders containing ECG data. Files are selected for upload by clicking the "choose" button or by dragging and dropping files into the central display area (Figure 2a).

Clicking the "upload all" button initiates transfer of data from the user's file system to WaveformECG. WaveformECG automatically determines each file's format and follows a multistep procedure for storing and retrieving data (Figure 2b). Completion of these steps is used as an indicator of progress. Progress information is displayed in the right-most portion of the upload interface, under the "background queue" tab. In the first step, the system checks to make sure that all required files have transferred from the local source to the host. While most formats only have one file per dataset, some formats split information across multiple files. Figure 2a shows this for s0027lre, a WFDB format ECG dataset. s0027lre's data is packaged in three different files, with dataset metadata in the header (.hea) file, and time-series data in the other two (.dat and .xyz). In this example, WaveformECG has fully received the .hea file, but the .dat and .xyz file transfers are still in progress. The progress bar for dataset s0027lre is empty, and the phase column of the background queue displays "wait" because these data files are still being transferred. Once each ECG file transfer to the service is complete, they're stored in their native format in the LDML. Files at this stage of the workflow have a progress bar at 40 percent completion, with "transferred" displayed in the "background queue" area. The folder structure within the LDML corresponds with that of the folders created by the user in the upload interface. WaveformECG displays this folder structure on all screens where the user interacts directly with their uploaded files.

Once transfer is complete, WaveformECG spawns a separate process to extract each ECG time series for storage in OpenTSDB. A single ECG file contains signals from multiple leads, and a time series for each lead signal is extracted and labeled with a unique TSUID. Once this step is complete for all leads of an ECG waveform, the progress bar moves to 60 percent completion, with "written" displayed in the "background queue" column.

Following completion of writing, another background process is spawned to extract ECG waveform analysis results in the files. Each result is labeled with an appropriate ontology term, selected from the NCBO Bioportal Electrocardiography Ontology (http://

purl.bioontology.org/ontology/ECG) by storing the ontology ID along with the result. WaveformECG bundles this information, along with the subject identifier, the format of the uploaded ECG dataset, and the start time of the time series itself, and writes labeled analysis results into OpenTSDB. Once this is completed, the progress bar moves to 80 percent, with "annotated" displayed in the "background queue" phase column.

WaveformECG must be able to maintain a connection with the original uploaded ECG files, the stored time-series data, file metadata, analysis results, and manual annotations made to ECG waveforms. To do this, the OpenTSDB TSUID is stored in PostgreSQL. Once this is done, the progress bar moves to 100 percent, and "done" is displayed in the "background queue" phase column.

Data Analysis

ECG analysis algorithms are made available for use in WaveformECG as Web services. The analyze interface (Figure 3a) uses libraries for Web service implementations of ECG analysis algorithms accessed through Apache Axis2.[18]

Analysis Web services are developed by using Axis2 for communicating with the compiled version of the analysis algorithm. Axis2 is an open source XML-based framework that provides APIs for generating and deploying Web services. It runs on an Apache Tomcat server, rendering it operating-platform- independent. Algorithms developed using the interpreted language Matlab can be compiled using Matlab Compiler (www.mathworks.com/products/compiler/mcr/?requestedDomain=www.mathworks.com) and executed in Matlab Runtime, a stand-alone set of shared libraries that enables the execution of compiled Matlab applications or components on computers that don't have it installed. An XML file is developed that defines the service, commands it will accept, and acceptable values to pass to it. In a separate administrative portion of the analyze interface, a tool allows administrators to easily add algorithms implemented as Web services to the system. Upon entry of proper algorithm details and parameter information, WaveformECG can invoke an algorithm that the administrator has deployed. This approach simplifies the process of adding new algorithms to WaveformECG.

Figure 3a shows the analyze interface and Figure 3b shows the associated processing steps. Users select files or folders from the file chooser on the left; multiple files can be dragged and dropped into the central pane. Placing a file in that pane makes that file available for analysis by one or more algorithms, listed in the bottom center pane. Clicking the checkbox on an algorithm entry instructs the system to analyze the selected files with that algorithm. The checkbox at the top of the algorithm list allows a user to toggle the selection of all the available algorithms. All available algorithms have default settings—some have parameters that can be set via the "options" button, but all parameters set for an algorithm will be applied to all files to be analyzed. Upon selection of files to be processed and the algorithms with which to process them, the user clicks the "start analysis" button, which creates a thread to handle the processing. The thread dispatches a RESTful call to OpenTSDB to retrieve all the data requested. Depending on the algorithms chosen, the thread writes the data into the necessary formats required by the algorithms (for example, algorithms from the

PhysioToolkit[19] require that ECG data be in the WFDB file format). The thread then invokes the requested algorithms on the requested data. As long as the analyze screen remains open, the background queue will be updated, incrementing the number of algorithms that have finished processing. Upon processing completion of all selected algorithms for one file, the phase will update to "done" in the background queue.

Analyses of ECGs can provide information on the heart's both normal and pathological functions. The lead V3 ECG waveform in Figure 4 shows body surface potential (uV; ordinate) as a function of time (Sec; abscissa) measured over a single heartbeat. The ECG P, Q, R, S, and T waves are labeled. The P wave reflects depolarization of the cardiac atria in response to electrical excitation produced in the pacemaker region of the heart, the sinoatrial node. Onset of the Q wave corresponds to onset of depolarization of the cardiac interventricular septum. The R and S waves correspond to depolarization of the remainder of the cardiac ventricles and Purkinje fibers, respectively. Ventricular activation time is defined as the time interval between onset of the Q wave and the peak of the R wave. The T wave corresponds to repolarization of the ventricles to the resting state. The time interval between onset of the Q wave and completion of the T wave is known as the QT interval and represents the amount of time over which the heart is partially or fully electrically excited over the cardiac cycle. The time interval between successive R peaks is the instantaneous heart rate. Abnormalities of the shape, amplitude, and other features of these waves and intervals can reflect underlying heart disease, and there has been considerable effort in developing algorithms that can be used to automatically analyze ECGs to detect these peaks and intervals. Table 1 lists the algorithms available in the current release of WaveformECG.[20,21]

Data Visualization

The visualize interface lets users examine and interact with stored ECG data. This feature also provides a mechanism for manually annotating waveforms. When the user selects a file to view in the visualization screen, it initially displays the data as a series of graphs, one for each lead in the dataset (15 leads for the GE MUSE dataset shown in Figure 4).

A 1-mV amplitude with a 200-msec duration calibration pulse is displayed in the left-most panel. Initially, four leads are displayed, but additional leads can be viewed by grabbing and dragging the window scroll bar located on the right side of the browser display. Whenever the cursor is positioned within a display window, its x–y location is marked by a filled red dot, and time-amplitude values at that location are displayed at the bottom of the panel. Cursor display in all graphs is synchronized so that as the user navigates through one graph, others update with it. Lead name and number of annotations for that lead signal are displayed in each graph. File metadata, including subject ID, lead count, sampling rate, and ECG duration, are displayed above the graphs. WaveformECG supports scrolling through waveforms. Clicking on the "next" button at the top of the display steps through time in increments of 1.2 seconds. Users can jump to a particular time point by entering the time value into the panel labeled "jump to time (sec)."

By clicking on a lead graph, users can expand the view to see the data for that lead in detail, including any annotations that have been entered manually. A list of analysis results on the lead are displayed in a table at the left of the view, and the graph is displayed in the center right. In Figure 5a, WaveformECG displays part of the analysis results extracted from a Philips XML upload. While requesting the analysis results displayed, the visualize interface also checks the data originally returned from OpenTSDB to see if any annotations exist for the time frame displayed and, if so, Dygraphs (http://dygraphs.com), an open source JavaScript charting library, renders the annotation display on the screen. Figure 5a shows examples of the two types of supported waveform annotations: *point annotations* are associated with a specific ECG time-value pair—in this case, the time-value pair corresponding to the peak of the R wave, labeled with the number 1—and *interval annotations* are associated with a particular time interval. The user can scroll through the individual lead data using the slider control at the bottom of the display or the navigation buttons. There's also a feature to jump to a specific point in time. Zooming can be performed using the slider bar at the bottom of the screen. To restore the original view of the graph, the user can double-click on it. Manual annotations can be added by clicking in the graph screen.

Data Annotation

Users can manually annotate ECG waveforms to mark features of interest. These annotations are then stored and redisplayed along with the waveform on subsequent visualization. To create an annotation, the user selects a point on the graph or highlights an interval using the mouse. This activates a graphical interface where the user enters the specific details of the annotation (Figure 5b). The system relies on the Bioportal ontology server to select annotation terms. Figure 5b shows the end result of the R peak point annotation listed in Figure 5a, labeled with a 1 on the waveform itself. On the upper- right-hand side of the screen, users can see the time and amplitude values for the point they selected. If it's incorrect, the user can click the "chg" button. The annotation interface will then provide the user with a zoomed-in portion of the graph and view where the current selection is. If they so choose, users can change the point to another one and click "save." If not, they can click "cancel." The annotation interface refreshes and the onset will be updated to the chosen value.

When WaveformECG renders the annotation interface shown in Figure 5b, it dispatches a request to NCBO's Bioportal to retrieve all the root-level terms from the ECG terms view of the ECG ontology in Bioportal. As the ontology itself conforms to the standards of a basic formal ontology (BFO), the ECG terms view provides a less formal display of the terms within. A button at the end of the terms listing in the annotation screen lets users change the display to or from the BFO and terms view of the ECG ontology.

At the top of the term listing is a text box labeled "search for class," which lets a user type a search term. As typing commences, a JavaScript application developed by the NCBO provides a list of terms in the target ontology that match the typed text. The user can then select a term from that list. Upon selection of a term, the lower box in the right-center screen will update with the term and the term definition retrieved from Bioportal. The user can then

enter a comment in the text field below that describes any additional information to be included with the annotation. Upon completion of term selection and comment entry, the user clicks the "save" button. In Figure 5b, this button is grayed out because this figure shows the result of clicking on an existing annotation. This lets users delve into the details of existing annotations and see any comments entered previously in the comment box for annotation.

Integration with the I2B2 Clinical Data Warehouse

WaveformECG has been integrated with the Eureka Clinical Analytics systems.[22] Eureka provides Web interfaces and tools for simplifying the task of extracting, transferring, and loading clinical data into the I2B2 clinical data warehouse. The advantage of this integration is that subject cohorts identified in I2B2 can be easily sent to WaveformECG for further analysis using a newly developed Send Patient Set plug-in that communicates with WaveformECG using JavaScript Object Notation (JSON). In the example in Figure 6, we use I2B2 to query data from the MIMIC-II database,[23] looking for all Asian subjects for whom there exists a measurement of cardiac output. The plug-in extracts subject IDs satisfying this particular query from I2B2 and sends them to WaveformECG as a JSON message. WaveformECG receives and processes the JSON, creating a folder displayed in the upload interface that corresponds to the I2B2 cohort name (Monitor-Asian@14:11:45[10-29-2015]). WaveformECG creates subfolders for each of the corresponding subjects, with their subject identifiers as the folder names. The user can then upload the waveform data for each of the subjects into their respective folders (left side of Figure 6). In the example, I2B2 returned three subjects who met this criteria, but only one of the subjects had corresponding waveform data. We then uploaded that subject's data to WaveformECG and processed those waveforms with multiple analysis algorithms.

Eureka lets users define a data source and specify cohort definitions. In the case of WaveformECG, we defined OpenTSDB as the data source. Through that definition, Eureka performed a RESTful call to OpenTSDB, searching for analysis results linked with files in the Eureka folder. Once found, analysis results along with their ECG lntology IDs are transferred to Eureka, where they are reorganized into a format acceptable for automatic loading into I2B2. A subset of those results can be seen in Figure 6 under the "EKG annotations" folder in the I2B2 "navigate terms" window.

WaveformEC G Case Study

Sudden cardiac death (SCD) accounts for 200,000 to 450,000 deaths in the US annually.[24] Current screening strategies fail to detect roughly 80 percent of those who die suddenly. The ideal screening method for increased risk should be simple, inexpensive, and reproducible in different settings so that it can be performed routinely in a physicians' offices, yet be both sensitive and specific. A recent study has shown that features computed from the 12-lead ECG known as the QRS score and QRS-T angle can be used to identify patients with fibrotic scars (determined using late-gadolinium enhancement magnetic resonance imaging) with 98 percent sensitivity and 51 percent specificity.[25] Motivated by these findings, we assisted in a large-scale screening of all ECGs obtained over a six-month period at two large hospital

systems. The challenges faced in this study were the large number of subjects and ECGs (~35,000) to be managed and analyzed, the use of different ECG instrumentation and thus different data formats at the two sites, and the fact that instrument vendors don't make either of the algorithms to be tested available in their systems. WaveformECG proved to be a powerful platform for supporting this study. The QRS score and QRS-T angle algorithms were implemented and deployed, making it possible for the research team to quickly select and analyze ECGs from different sites. The two ECG-based features were shown to be a useful initial method (a sensitivity of 70 percent and a specificity of 55 percent) for identifying those at risk of SCD in the population of patients having preserved left ventricular ejection fraction (LVEF > 35 percent).

Other physiological time-series data arise in many other healthcare applications. Blood pressure waveforms, peripheral capillary oxygen saturation, respiratory rate, and other physiological signals are measured from every patient in the modern hospital, particularly those in critical care settings. Currently in most hospitals, these data are "ephemeral," meaning they appear on the bedside monitor and then disappear. These data are among the most actionable in the hospital because they reflect the patient's moment-to-moment physiological functioning. Capturing these data and understanding how they can be used along with other data from the electronic health record to more precisely inform patient interventions has the potential to significantly improve healthcare outcomes. In future work, we will extend WaveformECG to serve as a general-purpose platform for working with other types physiological time-series data.

Acknowledgments

References

1. Surawicz B, et al. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part III: Intraventricular Conduction Disturbances. Circulation. Mar 17.2009 119:e235–240. [PubMed: 19228822]

2. Rautaharju PM, et al. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part IV: The ST Segment, T and U Waves, and the QT Interval. Circulation. Mar 17.2009 119:e241–250. [PubMed: 19228821]

3. Wagner GS, et al. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part VI: Acute Ischemia/Infarction. J Am College Cardiology. Mar 17.2009 53:1003–1011.

4. Hancock EW, et al. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part V: Electrocardiogram Changes Associated with Cardiac Chamber Hypertrophy. Circulation. Mar 17.2009 119:e251–261. [PubMed: 19228820]

5. Winslow R, et al. The CardioVascular Research Grid (CVRG) Project. Proc AMIA Summit on Translational Bioinformatics. 2011:77–81.

6. Musen MA, et al. BioPortal: Ontologies and Data Resources with the Click of a Mouse. Proc Am Medical Informatics Assoc Ann Symp. 2008:1223–1224.

7. Murphy SN, et al. Serving the Enterprise and Beyond with Informatics for Integrating Biology and the Bedside (I2B2). J Am Medical Informatics Assoc. 2010; 17(2):124–130.

8. Bild DE, et al. Multi-ethnic Study of Atherosclerosis: Objectives and Design. Am J Epidemiology. Nov 1.2002 156:871–881.

9. Lynch EB, et al. Cardiovascular Disease Risk Factor Knowledge in Young Adults and 10-Year Change in Risk Factors: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. Am J Epidemiology. Dec 15.2006 164:1171–1179.

10. Cheng A, et al. Protein Biomarkers Identify Patients Unlikely to Benefit from Primary Prevention Implantable Cardioverter Defibrillators: Findings from the Prospective Observational Study of Implantable Cardioverter Defibrillators (PROSEICD). Circulation: Arrhythmia and Electrophysiology. 2014; 7(12):1084–1091. [PubMed: 25273351]

11. Yuan, JX. Liferay Portal Systems Development. Packt Publishing; 2012.

12. Ananthakrishnan, R., et al. Globus Nexus: An Identity, Profile, and Group Management Platform for Science Gateways and Other Collaborative Science Applications. Proc. Int' l Conf. Cluster Computing; 2013; p. 1-3.

13. Sigoure, B. OpenTSDB: The Distributed, Scalable Time Series Database. Proc. Open Source Convention; 2010; http://opentsdb.net/misc/opentsdb-oscon.pdf

14. Taylor RC. An Overview of the Hadoop/MapReduce/HBase Framework and Its Current Applications in Bioinformatics. BMC Bioinformatics. 2010; 11:S1.

15. Pautasso, C. RESTful Web Services: Principles, Patterns, Emerging Technologies. Springer; 2014. p. 31-51.

16. Douglas, K., Douglas, S. PostgreSQL: A Comprehensive Guide to Building, Programming, and Administering PostgreSQL Databases. SAMS publishing; 2003.

17. Moody GB, Mark RG, Goldberger AL. PhysioNet: A Web-Based Resource for the Study of Physiologic Signals. IEEE Eng Medicine and Biology Magazine. 2001; 20(3):70–75.

18. Jayasinghe, D., Azeez, A. Apache Axis2 Web Services. Packt Publishing; 2011.

19. Moody, GB., Mark, RG., Goldberger, AL. PhysioNet: Physiologic Signals, Time Series and Related Open Source Software for Basic, Clinical, and Applied Research. Proc. Conf. IEEE Eng. Medicine and Biology Soc; 2011. p. 8327-8330.

20. Chesnokov Y, Nerukh D, Glen R. Individually Adaptable Automatic QT Detector. Computers in Cardiology. 2006; 33:337–341.

21. Strauss DG, et al. Screening Entire Health System ECG Databases to Identify Patients at Increased Risk of Death. Circulation: Arrhythmia and Electrophysiology. 2013; 6(12):1156–1162. [PubMed: 24122522]

22. Post A, et al. Semantic ETL into I2B2 with Eureka! AMIA Summit Translational Science Proc. 2013:203–207.

23. Saeed M, et al. Multiparameter Intelligent Monitoring in Intensive Care II: A Public-Access Intensive Care Unit Database. Critical Care Medicine. 2011; 39(5):952–960. [PubMed: 21283005]

24. Goldberger JJ, et al. American Heart Association/American College of Cardiology Foundation/ Heart Rhythm Society Scientific Statement on Noninvasive Risk Stratification Techniques for Identifying Patients at Risk for Sudden Cardiac Death: A Scientific Statement from the American Heart Association Council on Clinical Cardiology Committee on Electrocardiography and Arrhythmias and Council on Epidemiology and Prevention. Circulation. Sep 30.2008 118:1497–1518. [PubMed: 18833586]

25. Strauss DG, et al. ECG Quantification of Myocardial Scar in Cardiomyopathy Patients with or without Conduction Defects: Correlation with Cardiac Magnetic Resonance and Arrhythmogenesis. Circulation: Arrhythmia and Electrophysiology. 2008; 1(12):327–336. [PubMed: 19808427]

Biographies

Raimond L. Winslow is the Raj and Neera Singh Professor of Biomedical Engineering and director of the Institute for Computational Medicine at Johns Hopkins University. His research interests include the use of computational modeling to understand the molecular mechanisms of cardiac arrhythmias and sudden death, as well as the development of

informatics technologies that provide researchers secure, seamless access to cardiovascular research study data and analysis tools. Winslow is principal investigator of the CardioVascular Research Grid Project and holds joints appointments in the departments of Electrical and Computer Engineering, Computer Science, and the Division of Health Care Information Sciences at Johns Hopkins University. He's a Fellow of the American Heart Association, the Biomedical Engineering Society, and the American Institute for Medical and Biological Engineers. Contact him at rwinslow@jhu.edu.

Stephen Granite is the director of database and software development of the Institute for Computational Medicine at Johns Hopkins University. He's also the program manager for the CardioVascular Research Grid Project. Granite has an MS in computer science with a focus in bioinformatics and an MS in business administration with a focus in competitive intelligence, both from Johns Hopkins University. Contact him at sgranite@jhu.edu.

Christian Jurado is a software engineer in the Institute for Computational Medicine at Johns Hopkins University. He's also lead developer of WaveformECG for the CardioVascular Research Grid Project. Jurado has a BS in computer science, specializing in Java Web development and Liferay. Contact him at cjurado2@jhu.edu.
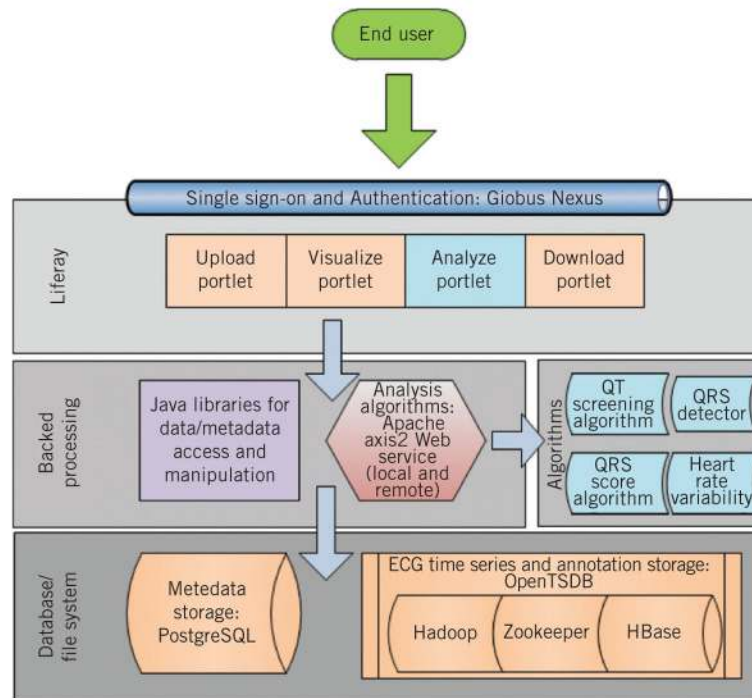
Figure 1.
Platform architecture. Users authenticate to WaveformECG via Globus Nexus to upload, visualize, analyze, and download ECG data, analysis results, and annotations. To accomplish this, WaveformECG makes use of Java libraries and Web services that provide access to data, metadata, analysis results, annotations, and data analysis algorithms.
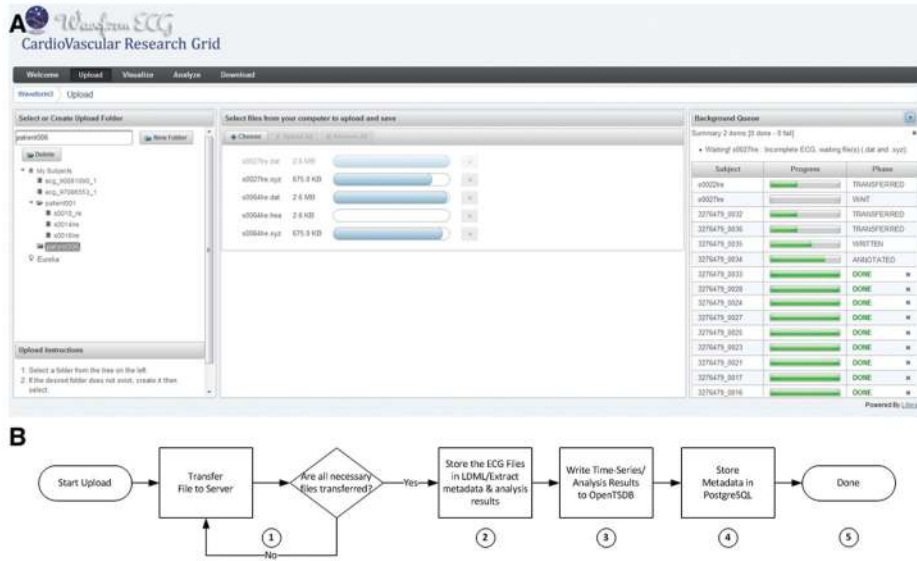
Figure 2.
Upload portal. (a) The listing on the left shows that the user has created the patient006 folder, into which data will be uploaded. Datasets under "my subjects" are owned by the user. Folders group datasets by subjects, and progress bars next to the file names in the center of the screen show progress on the upload of each file to the server. The background queue on the right provides users with a real-time update of progress on dataset processing. (b) The upload processing flow consists of five parts: server upload ("wait"); storage in LDML and parsing file data ("transferred"); transfer of time-series data and analysis results to OpenTSDB ("written"); transfer of metadata to PostgreSQL ("annotated"); and completion ("done").

Figure 3.
Analysis portal. (a) Three datasets with different formats were selected for processing by multiple algorithms. The background queue shows progress in data processing: two datasets have each been processed using eight algorithms, while the third has completed processing by seven. (b) In the analysis process, data and algorithm selection follow a step-by-step workflow.
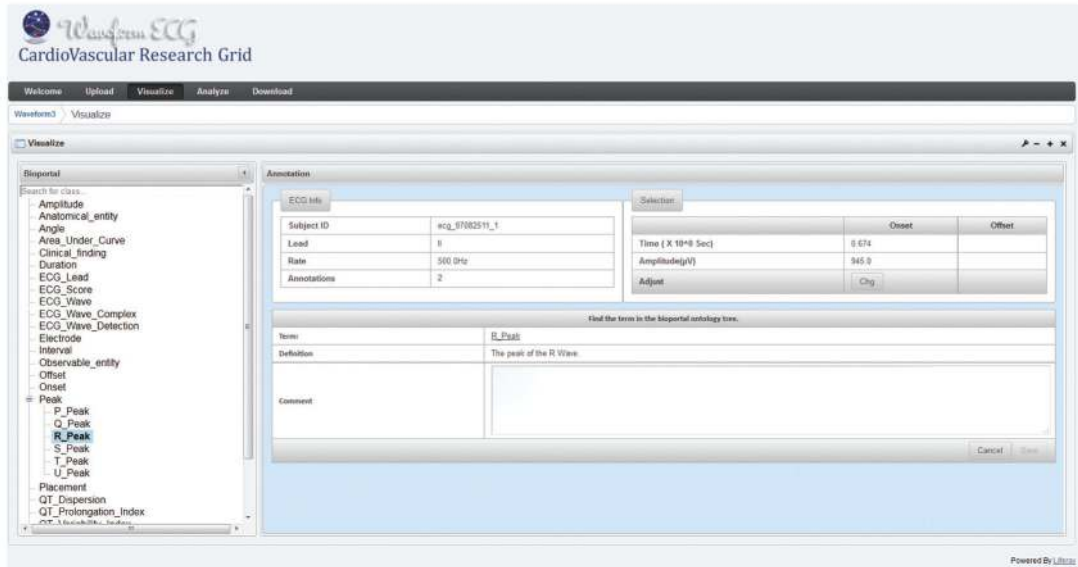
Figure 4.
Multilead visualization interface. In the multilead display for 4 of the 15 leads from a GE Marquette Universal System for Electrocardiography (MUSE) XML upload, the vertical bar in 3 of the graphs represents the cursor location in the first graph. The bars move with the cursor and change focus as the cursor changes graphs.

(a)

(b)

Figure 5.
Visualization. (a) For lead II in a 12-lead ECG in Philips format, the table under "Analysis Results" displays the results of automated data processing by the Philips system used to collect this ECG. In the waveform graph, A denotes a QT interval annotation, with the yellow bar representing the interval itself. This annotation was made manually. The 1 denotes an R peak annotation, also made manually. All interval and point annotations are listed below the graph. (b) In the manual annotation interface, the R peak is highlighted and the information in the center shows the definition returned for that term selection. In

addition, there are details about the ECG and the point at which the annotation was made. To create these displays, the visualize interface initiates a RESTful call to OpenTSDB to retrieve the first 2.5 seconds of time-series data associated with all the leads in the file. Dygraphs, an open source JavaScript charting library, generates each of the graphs displayed.
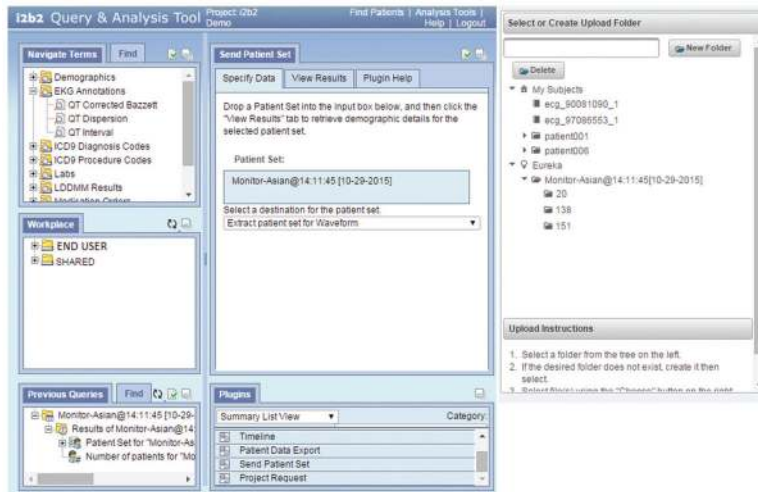
Figure 6.
Integration with a clinical data warehouse. A split screen shows information sent from I2B2 (left) to WaveformECG. In the expanded EKG annotations folder, three analysis results can be returned from WaveformECG to I2B2.

Table 1

Algorithm listing for the analyze interface.

Algorithm name	Developer	Purpose/dependencies
Rdsamp	Physionet	Converts the Waveform Database (WFDB) ECG format to human-readable format
Sigamp	Physionet	Measure signal amplitudes of a WFDB record
sqrs/sqrs2csv	Physionet	Detects onset and offset times of the QRS complex in single leads; second implementation produces output in CSV format
wqrs/wqrs2csv	Physionet	Detects onset and offset times of the QRS complex in single leads using the length transform; second implementation produces output in CSV format
ihr (sqrs & wqrs implementation)	Physionet	Computes successive RR intervals (instantaneous heart rate); requires input from sqrs or wqrs
pNNx (sqrs & wqrs implementation)	Physionet	Calculates time domain measures of heart rate variability; requires input from sqrs or wqrs
QT Screening	Yuri Chesnokov and colleagues[19]	Detects successive QT intervals based upon high- or low-pass filtering of ECG waveforms; works with data in WFDB format
QRS-Score	David Strauss and colleagues[20]	Produces the Strauss-Selvester QRS arrhythmia risk score based on certain criteria derived from GE MUSE analysis