Ontology-Based Annotation of Biomedical Time Series Data

Rai Winslow, Steve Granite
The Institute for Computational Medicine
Johns Hopkins University
Motivation For This Project

• Cardiovascular disease
  – Mortality ~1 million /year, ~70 million live with CV disease

• ECG most most frequently collected biomedical time series data type in CV research and in medical practice

• No ontology, data model, or tool for describing, annotating, and sharing ECG data

• Creating the above resources will facilitate analysis of ECG data, and will support many large-scale clinical studies
Project Aims

• Develop a comprehensive ECG ontology
• develop an ECG data model and data service
• Extend the capabilities of the ECG data management and analysis portal being developed as part of the CardioVascular Research Grid (CVRG – next slide) project so that users may:
  – Upload, query, and retrieve data from an ECG data service;
  – scroll through ECG waveforms, view the annotations generated by ECG analysis algorithms, edit these annotations (correct algorithm mistakes), and create new annotations labeling temporal features of the ECG waveform (...missed by the algorithms) and the data derived from ECG waveform analysis;
  – store annotated ECGs and derived data back into the data service as part of the multi-scale data federation;
• Enable further development of the ontology by using the features of BioPortal 2.0 that allow users to comment on or add new terms. Put in place a process for adjudicating these new terms.
Context – The CardioVascular Research Grid (CVRG)

• NHLBI funded effort to develop a national infrastructure for managing, sharing, and analyzing a broad range of cardiac data

• Test-bed project – the Reynolds study of predictors of Sudden Cardiac Death in the setting of LV dysfunction and coronary artery disease

• Reynolds project is ideal for understanding how to re-use and extend middleware developed in other grid projects (BIRN, caGRID) to meet needs of the CV community
The Driving Biomedical Project
The Reynolds Study

• Large subject cohort (~1,200) at high risk for sudden cardiac death
• All have:
  – CAD
  – LV dysfunction
  – received ICD placement
• Multi-scale data from each patient
• Challenge – discover biomarkers predictive of high risk and need for ICD placement
• Test biomarkers on novel (currently ~500) subject population
ECG Data and Analytic Services

Norav ECG Data (RDT)

Physionet ECG Data (WFDB)

HL7 aECG XML

Repository (Raw and Analyzed Data)

Berger Algorithms

Physionet Algorithms

Meta-data XML

WFDB
Demo of Existing CVRG ECG Portlets

https://portal.cvrgrid.org:8443/web/guest/home
ECG Ontology

• ECG data collection protocols
  • Number of leads, lead name, lead placement
  • ECG acquisition instrumentation
    • deskside, ICD, pacemaker, Holter monitor
  • Software version #, raw data format

• Signal attributes
  • Duration, Filtering, sampling, # bits, signal processing

• Data analysis algorithms
  • Name features computed from ECG (large number)
  • Name specific algorithms used to compute ECG features (large number)
  • Name reference describing the algorithm(s)
ECG Visualization Portlet
Automated ECG Annotation

- Data passed to ECG analysis services on upload
- ECG properties are computed, for example
  - Q, R, S, T wave peaks
  - QRS width, RR & QT intervals
  - QT and RR interval variability
  - ST segment elevation
  - many other properties
- Computed features are displayed
  - meaning of each term
  - Reference to algorithm used to compute each feature
- Users can scroll through the annotated data
  - “electronic” rather than paper ECG chart recording
Challenges


- Bioset 9000 instrumentation
- 12 lead ECG
- Digital caliper electrodes
- ECG computed from leads II, V2, V5
- Hannover ECG Analysis Software Version V3.22-12
- Software specific definition of QT interval
- QT interval correlated with genotype is corrected for heart rate, age, and sex of subject
Challenges (cont.)


\[
QTc_{-RAS} = \overline{QT} - \lambda_1 \cdot (\overline{RR} - 1000) - \lambda_2 \cdot (A - 60) - 4.58
\]

QTc_RAS is Average QT interval corrected for heart rate R, age A, and sex S

\(\overline{QT}\) is mean QT interval, \(\overline{RR}\) is mean RR interval

\(\lambda_1\) and \(\lambda_2\) are sex dependent factors

** Definition of computed ECG features is algorithm dependent and currently there is no unique name for each quantity
CVRG Project 2 Aim 2D

- EP data is another type of time-series data important in cardiovascular research
- No ontology/data model for disseminating EP data, no culture of even sharing this data.
- Develop an ontology and data model
  - Protege
  - Single-channel electrophysiological experiments and data
    - voltage-clamp
    - current clamp
    - fluorescence imaging
  - Multi-channel fluorescence imaging
EP Ontology Layout

EPStudy

Creator

Biomaterial

Procedure
  - EPPprocedure
  - EPTechnique

Observation
  - EPObservation
  - EPData
Integrating Ontologies

- National Center for Biological Ontologies
  - BioPortal
    - Foundational Model of Anatomy (FMA)
    - Unit Ontology (UO)
    - Gene Ontology (GO)
    - Chemical Entities of Biological Interest (ChEBI)
    - NCBI Taxonomy Ontology
EPprocedure

Technique
- Voltage Clamp
- Current Clamp
- Fluorescence Imaging

Patch Clamp
- Cell Attached Patch Clamp
- Whole Cell Patch
- Inside Out Patch Clamp
- Single Electrode Voltage Clamp
- Two Electrode Voltage Clamp

Sharp Electrode Voltage Clamp
EP Ontology Testing

- Creation is an iterative process
- Describe experimental methods section
- Web-based submission form for novel EP related terms (in future)
- Version 1 ready for submission
- Use this experience in development of the ECG ontology
Manage the Ontology

• Deposit the ontology at the NCBO and manage using NCBO tools (BioPortal)

• First level of review
  • Expert review panel to triage terms
  • Members - Ary Goldberger, Ron Berger, Gordon Tomaselli
  • Add members as needed over time

• Second level of review
  • User feedback capabilities of BioPortal 2.0

• Third level of review
  • Annual meetings of the CVRG ECG Working Group