Enabling Discovery in High-Risk Plaque using Semantic Web Approaches

C-SHALS 2009
Cambridge MA
February 25-27, 2009

© 2006 BG MEDICINE
What is HRP?

- “HRP” stands for “High-Risk Plaque”
- HRP is an authoritative industry-funded precompetitive activity to open novel markets for anti-atherosclerosis treatments.
- It involves all key stakeholders:
  - FDA
  - Payor
  - NHLBI/European Public Sector
  - Academia/thought leaders, and
  - Industry
    - Therapeutics
    - Imaging
    - In-vitro diagnostics
HRP Initiative Activities

- Circulating Biomarker Pilot Study (Duke—CATHGEN)
- BioImage Study
- Circulating Biomarker Study (Copenhagen City Hospital—CGPS, CCHS)
- Plaque Biology Study (U.Maastricht—CARIM)
- Health Economic Study
- Regulatory Initiative
Objective:
- To identify imaging and/or circulating biomarkers that predict 3-year cardiovascular events
- Determine predictive value of biological and/or imaging markers for near-term (1-3-year) outcomes
BiolImage Study Design

BiolImage is a payor-based observational study

Enrollment (n=7300)
Humana members ages 55-80 (men) or 60-80 (women), without known CV disease or serious medical conditions

Follow up (~3 years)
Ongoing event monitoring and survey every 6 months

Primary Endpoint: Major CV events (n= ~600)
Associate blood biomarkers with imaging data

- Physical measurements
  - Blood pressure, height, weight, EKG, ABI
- Blood samples
  - DNA, RNA, plasma, serum
- Cardiac CT and ultrasound (n=6000)
- For individuals found to be high risk:
  - CT angiogram of coronary vessels
  - MRI of carotids and abdominal aorta
  - FDG -PET
Enabling BioImage Data Mining (POC)

• Semantic Web methodologies can be applied to HRP datasets

• These approaches allow for:
  – Sharing the data across all HRP companies in a fully interoperable manner
  – Annotating the data with existing public biological knowledge
  – Enabling faceted browsing to slice & dice data displays for focused biological questions

• Adopting this approach in a pre-competitive environment will:
  – Enable easier use of image and molecular data, including associating interpretations and annotations
  – Internally jump-start efforts in this emerging standard
  – Demonstrate the utility to the larger community
Implementing a Semantic Web Approach

• Convert data format to RDF
  – **Improves interoperability**
    • Agnostic to the source
    • Easy to adopt with standards
  – **Interlinks different types of data & annotated resources**

• Generate an interface using MIT/CSAIL standard

  *Exhibit*
  – **Determine appropriate filters and lenses needed ask focused questions**
    • Scatter plots, tables, timelines, maps

• Web site hosted at BGM
  – **Underlying data files, interface available to member companies**
Lessons Learned from the POC Study

• Existing BioImage Data maps completely to Semantic Web format (RDF)

• Large data sets can be viewed and analyzed via enabled browsers, in any combination

• Configurable views and statistical lenses

• Metadata can be attached to Image data

• Still need to enhance access for CT and MRI images
BioImage for the Semantic Web (BISM)

Goals

• Design and Implement a full-scale Semantic Web Solution for BioImage Data: integration and views

• Enable complex queries and inferencing on the complete data-set

• Platform for complex analysis, dashboarding, and reporting

• Support annotations of data subsets

• Enable direct integration with public and private molecular and disease knowledge
BiolImage Data Model
Detailed Data Model
Phase II

• Develop Database Wrappers around BioImage to link to Semantic Web structures

• Lightweight web-service for enabling browser-based data viewing and analysis (perform SPARQL queries)

• Integration of clinical codes standard via Semantic Web (URI) identifiers
Additional Possibilities
Phase II

- Standardization of HRPI data with any internal data sets (genes, drugs, symptoms, etc)
- Interface with additional analytic tools
- Combining data views with other studies even if only partially compatible
- Ability to associate with various knowledge sources, e.g., Cardiovascular