caGrid, caBIG, CVRG and NCIBI

Joel Saltz MD, PhD
Director Center for
Comprehensive Informatics
Biomedical Middleware: caGrid

caGrid Components

- Security (GAARDS)
- Language (metadata, ontologies)
- Semantic/Federated query
- Workflow
- Grid Service Graphical Development Toolkit (Introduce)
- DICOM, IHE compatibility
- Advertisement and Discovery

*All Services Register with the Index Service*
Translation: Same ideas, different words
**Vocabulary/Ontology**

<table>
<thead>
<tr>
<th>Prostate Adenocarcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identifiers:</strong></td>
</tr>
<tr>
<td>name</td>
</tr>
<tr>
<td>code</td>
</tr>
<tr>
<td><strong>Relationships to other concepts:</strong></td>
</tr>
<tr>
<td>Disease_Has_Abnormal_Cell</td>
</tr>
<tr>
<td>Disease_Has_Associated_Anatomic_Site</td>
</tr>
<tr>
<td>Disease_Has_Associated_Anatomic_Site</td>
</tr>
<tr>
<td>Disease_Has_Normal_Cell_Origin</td>
</tr>
<tr>
<td>Disease_Has_Normal_Tissue_Origin</td>
</tr>
<tr>
<td>Disease_Has_Primary_Anatomic_Site</td>
</tr>
<tr>
<td><strong>Information about this concept:</strong></td>
</tr>
<tr>
<td>Preferred_Name</td>
</tr>
<tr>
<td>Semantic_Type</td>
</tr>
<tr>
<td>Unified Medical Language System Concept Identifier</td>
</tr>
<tr>
<td>DEFINITION</td>
</tr>
<tr>
<td>Synonym with source data</td>
</tr>
<tr>
<td>Synonym with source data</td>
</tr>
<tr>
<td>Synonym with source data</td>
</tr>
</tbody>
</table>
Interoperability

- Registered metadata
- Ontology concept codes used to annotate models
- XML schemas that define data structures also registered
- Thus both data semantics AND data structures are registered. That is how we achieve (relative) interoperability.
Will Treatment work and if not, why not?

Motivating Example: Avastin and Glioblastoma in RTOG-0825

Treatment: Radiation therapy and Avastin (anti angiogenesis)

Predict and Explain: Genetic, gene expression, microRNA, Pathology, Imaging

RT, imaging, Pathology markup/annotations
For the sake of quality control, reproducibility and data sharing, results of RT, imaging, Pathology observations and analyses need to be described in a well defined manner.

Finding: mass
Mass ID: 1
Margins: spiculated
Length: 2.3cm
Width: 1.2cm
Cavitary: Y
Calcified: N
Spatial relationships: Abuts pleural surface; invades aorta
Digital Pathology

Multiheaded Microscope

caMicroscope
In Silico Center Collaboration: Distinguish (and maybe redefine) astrocytic, oligodendroglial and oligoastrocytic tumors using TCGA and Rembrandt. Important since treatment and outcome differ.

- Link nuclear shape, texture to biological and clinical behavior.
- How is nuclear shape, texture related to gene expression category defined by clustering analysis of Rembrandt data sets?
- Relate nuclear morphometry and gene expression to neuroimaging features (Vasari feature set).
- Genetic and gene expression correlates of high resolution nuclear morphometry and relation to MR features using Rembrandt and TCGA datasets.
Annotation and Markup of Pathology Data needs Human/Algorithm Cooperation

Astrocytoma vs Oligodendroglioma
- TCGA finds genetic, gene expression overlap
- Pathologists have also long seen overlap
- Relationship between Pathology, Molecular, Radiology
- *Relationship to Outcome, treatment response*
What you find depends on where you look

- GBM gene expression patterns will be influenced by necrosis, degree and type of angiogenesis
- Systems biology hypotheses being evaluated – ideal context for NCIBI collaboration
- Degree and pattern of necrosis/angiogenesis varies within a given tumor so *molecular analyses need to be interpreted in the context of what was sampled*
Use of randomly selected sample to determine whether the earth is wet or dry …
The Cardiovascular Research Grid (PI Rai Winslow -- JHU, OSU, UCSD)
The D. W. Reynolds Cardiovascular Clinical Research Center

- Who should receive ICDs?
- Large patient cohort (~1,200) at high risk for sudden cardiac death
- All have CAD, LV dysfunction, received ICD placement
- Multi-scale data from each patient
- Patients with appropriate ICD firings are defined as high risk, patients without as low risk
- Challenge – discover biomarkers that are predictive of high risk
- Test biomarkers on novel (~500) patient population
Common Application Services

- caGrid Service: OpenClinica Model
  - OpenClinica Data
- caGrid Service: WFDB Metadata Model
  - Physionet WFDB
- caGrid Service: HL7 aECG Model
  - ECG XML
- caGrid Service: JHU SNP Model
  - Taqman Data
- caGrid Service: ProteinDB Model
  - ProteinDB Data

- caGrid Service: Physionet Algorithms
- caGrid Service: Berger QT Algorithm
- caGrid Service: DICOM/AIM Model
  - MRI/CT Images

caBIG™ cancer Biomedical informatics Grid™
an initiative of the National Cancer Institute
Links to NCIBI

- Employ caGrid tools to wrap NCIBI tools
- Leverage NCIBI tools to explore systems biology hypotheses
- Leverage Pathology and Radiology analysis, annotation tools and middleware for joint integrative studies joint with NCIBI
Final caGrid/CVRG Example: CTSA/RCMI
Treatment Resistant Hypertension (Gary Gibbons)

- Leverage caGrid, i2b2, CVRG, CTSA TRIAD infrastructure
- Links to NCIBI via i2b2, caGrid connectivity
- Accrual from RCMI community clinics
- SNP, ECG, Echo, vascular studies
Thank you