The Onco-i2b2 project: integrating biobank information and clinical data to support translational research in oncology

Riccardo Bellazzi
University of Pavia, Italy

Daniele SEGAGNI, Valentina TIBOLLO, Leonardo PERINATI Alberto ZAMBELLI, Silvia PRIORI
Fondazione Salvatore Maugeri, Pavia, Italy
Arianna DAGLIATI
IUSS Pavia, Italy
Clinical Bioinformatics

Knowledge discovery

Knowledge to practice
**I2b2** a project sponsored by the NIH - open source distribution – Partners (Boston)
i2b2 Cell: Canonical Hive Unit

HTTP XML
(minimum: RESTful, others like SOAP optional)

Programmatic Access

By Shawn Murphy
http://services.i2b2.org/webclient/
Clinical Bioinformatics – the i2b2 Pavia projects

Biobanks

DW / clinical research chart

Intelligent query

EMR

Discharge letters

Research data-bases
BIOINFORMATICS METHODOLOGY AND TECHNOLOGY TO INTEGRATE CLINICAL AND BIOLOGICAL KNOWLEDGE SUPPORTING ONCOLOGY TRANSITIONAL RESEARCH
Architecture overview

Clinical patient management

Data -> HIS -> CRC

Anonymized data

Match IDs, i2b2

Anonymized samples

Biobank

Researcher

Patient

Samples

Laboratory
I2b2-Pavia populating the datawarehouse

Data warehouse

Anatomical pathology unit

HIS Data

Domain Ontology

SNOMED

NLP System

Reports

Anagraghical and clinical data

Diagnosis, Morphology

Ontology Mapped Clinical Data
ETL is performed relying on KETTLE, a tool developed within the Pentaho project.

Development of complex ETL transformations based on a cancer-specific ontology that:
- combines atomic information and
- creates a well defined medical observation.
Counts and results

• 6,561 patients

• 235 biospecimens
  • collected from November 2010

• 960 concepts
  • divided into demographic data – diagnosis – clinical measurements – histological records – therapies and Biobank samples

• 45,743 visits and 91,857 observations
Adding statistical functionalities
Kaplan Meier Stat

Drop a Patient Set into the input boxes below, select one event and then click the "run Kaplan Meier" button to retrieve information about the analysis.

Patient Set: 61-70 y-Brugada@10:50:32 [11-9-2010] [Daniele] [PATIENTSET_380]

Events:
- Syncope
- Cardiac Arrest
- Appropriate Discharge

Run Kaplan Meier Analysis

Data description:

<table>
<thead>
<tr>
<th>Group</th>
<th>Patient ID</th>
<th>Time to Event</th>
<th>Event</th>
<th>Survival Rate</th>
<th>StdErr</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>52</td>
<td>71</td>
<td>1</td>
<td>0.9859</td>
<td>0.0140</td>
<td>0.9569</td>
<td>0.9598</td>
</tr>
<tr>
<td>Female</td>
<td>41</td>
<td>130</td>
<td>1</td>
<td>0.923</td>
<td>0.077</td>
<td>0.874</td>
<td>0.974</td>
</tr>
<tr>
<td>Female</td>
<td>46</td>
<td>128</td>
<td>1</td>
<td>0.9046</td>
<td>0.0163</td>
<td>0.8674</td>
<td>0.9437</td>
</tr>
<tr>
<td>Female</td>
<td>48</td>
<td>128</td>
<td>1</td>
<td>0.9761</td>
<td>0.0132</td>
<td>0.9515</td>
<td>0.9915</td>
</tr>
<tr>
<td>Female</td>
<td>51</td>
<td>125</td>
<td>1</td>
<td>0.9631</td>
<td>0.0152</td>
<td>0.9398</td>
<td>0.9898</td>
</tr>
<tr>
<td>Female</td>
<td>54</td>
<td>121</td>
<td>1</td>
<td>0.9611</td>
<td>0.0171</td>
<td>0.9262</td>
<td>0.9922</td>
</tr>
<tr>
<td>Female</td>
<td>55</td>
<td>119</td>
<td>1</td>
<td>0.9530</td>
<td>0.0187</td>
<td>0.9170</td>
<td>0.9870</td>
</tr>
<tr>
<td>Female</td>
<td>58</td>
<td>85</td>
<td>1</td>
<td>0.9430</td>
<td>0.0210</td>
<td>0.9026</td>
<td>0.9826</td>
</tr>
<tr>
<td>Female</td>
<td>59</td>
<td>60</td>
<td>1</td>
<td>0.9312</td>
<td>0.0239</td>
<td>0.8856</td>
<td>0.9656</td>
</tr>
<tr>
<td>Female</td>
<td>61</td>
<td>56</td>
<td>1</td>
<td>0.9146</td>
<td>0.0265</td>
<td>0.8601</td>
<td>0.9604</td>
</tr>
<tr>
<td>Female</td>
<td>63</td>
<td>36</td>
<td>1</td>
<td>0.8852</td>
<td>0.0375</td>
<td>0.8167</td>
<td>0.9137</td>
</tr>
<tr>
<td>Female</td>
<td>64</td>
<td>27</td>
<td>1</td>
<td>0.8562</td>
<td>0.0484</td>
<td>0.7664</td>
<td>0.7664</td>
</tr>
</tbody>
</table>

Log-rank test:

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>73</td>
<td>1</td>
<td>4.445</td>
</tr>
<tr>
<td>Male</td>
<td>130</td>
<td>11</td>
<td>7.555</td>
</tr>
</tbody>
</table>

Chi^2 = 4.247 on 1 degrees of freedom, p = 0.039
Project workplan

Phase 1 (Dec 2010- March 2011)

Training on i2b2

Phase 2 (Feb 2011- June 2011)

Functional specifications

Phase 3 (June 2011 - May 2012 )

Implementation

Expected results: implementation of ONCO-i2b2 at FSM
Acknowledgements

BMI Labs “Mario Stefanelli” (http://bioinfo.unipv.it)

Harvard Medical School

i2b2
Informatics for Integrating Biology & the Bedside

Daniele SEGAGNI, Valentina TIBOLLO
Arianna DAGLIATI, Leonardo PERINATI
Alberto ZAMBELLI, Silvia PRIORI

ItalBioNet

IUSS Sapere Aude

Regione Lombardia

IRCCS Fondazione S. Maugeri