Applying Visual Analytics to the Global Legal Entity Identifier System to Enhance Financial Transparency

Victoria L. Lemieux, Peter Phillips, Harkaran Singh Bajwa, and Carol Li
The University of British Columbia

Conference on Data Standards, Information and Financial Stability, Loughborough University, April 11-12, 2014
Acknowledgements

The financial support provided by the Natural Science and Engineering Research Council of Canada and The Boeing Corporation is gratefully acknowledged, along with additional support from Tableau Software.
Motivation

- Desire to support efforts to increase global financial transparency
- Focused on the GLEI system
- Realized that, even though LEI data is open and available:
  - It has gaps
  - It is semantically heterogeneous
  - It is overwhelming
The Challenge

How to help people make sense of LEI data
Visual Analytics – A Potential Solution

Definition

“Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.”

Use technology for human analysis & decision-making:

• Synthesize information and deriving insight from massive, dynamic, ambiguous and often conflicting data.
• Provide timely, defensible, and understandable assessments.
• Effectively communicating assessments for action.
What Can You Recognize?
Visual Analytics: From this....
And this...
To This . . .
Support for Sensemaking

Van Wijk, 2005
Phase One – Initial Explorations and Sensemaking

• Harvested and integrated CICI Utility and WM Datenservice data

• Created visualizations using Tableau and Gephi, two common tools that can be used for visual analysis
Phase One

CICI Utility Records by City and Legal Form
Phase One

Distribution of CICI Utility Records (US)  Distribution of CICI Utility Records (EUR)
Phase One

Hierarchical structure of registered LEIs, Jan. 2013
Phase Two – Enabling Public Exploration

- Objective was to build a platform to enable public exploration and sensemaking of LEI data to support a taxonomy of analytic tasks.

TASK TAXONOMY

(a) What are the cities in which the most registrations occur?
(b) What are the cities with the highest number of registrations?
(c) Which LOUs have the highest number of registrations?
(d) What is the pattern of registrations over time by country and by city and by registration authority?
(e) What is the proportion of legal forms being registered by country, by city and by registration authority?
(f) What companies have been registered?
(g) What is the global reach of these companies?
(h) Where are the companies registered?
(i) How are the companies related to one another?
(j) What is the volume of swaps trades per legal entity?
Phase Two – A Tower of Babel

• But first . . . Data harvesting and integration!

• To permit visualization of LEI data, we have integrated data from seven different Local Operating Units (LOUs) to date:
  – the DTCC/SWIFT CICI Utility (US)
  – WM Datenservice (Germany)
  – Irish Stock Exchange
  – Institut National de la Statistique et des Etudes Economiques (France)
  – Irish Stock Exchange
  – Takasbank (Turkey)
  – London Stock Exchange (UK)
  – Dutch Chamber of Commerce.
Phase Two – A Tower of Babel (aka Semantic Heterogeneity)

- Each of LEI source has its own XML schema to store the LEIs for its region/country.

- Challenges:
  - level of standardization of field names among different LOUs is relatively low

![Bar chart showing frequency distribution of usage instances of a given term](image-url)
Phase Two – A Tower of Babel

• Further Challenges:
  – lack of clear definition of field names
  – problematic data organizing structure
  – Ambiguity of meaning
  – Overly complex XML schemas
### Phase Two – Integrated Schema & Data Dictionary

<table>
<thead>
<tr>
<th>Term</th>
<th>DatServ</th>
<th>France</th>
<th>Ireland</th>
<th>Turkey</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>EII</td>
</tr>
<tr>
<td>Business Registry No.</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>OtherIdentifier Type</td>
<td>LEIFR</td>
<td>OtherIdentifier Type</td>
</tr>
<tr>
<td>City</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>Corporate Legal Address</td>
<td>LEIFR</td>
<td>Registered Address</td>
</tr>
<tr>
<td>Country</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>Two sets, only Registered Address is empty if Maybe can be extracted from LegalForm</td>
<td>LEIFR</td>
<td>Registered Country</td>
</tr>
<tr>
<td>Entity Profile</td>
<td>DTCC</td>
<td>see City</td>
<td>see City</td>
<td>see City</td>
<td>see City</td>
</tr>
<tr>
<td>Entity_status</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>Event/PortalDateTime</td>
<td>LEIFR</td>
<td>EntityStatusCode</td>
</tr>
<tr>
<td>Expiry Date</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>Expiry Reason</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>Fund Manager Name</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>FundManager/Name</td>
<td>EII</td>
<td>FundManager</td>
</tr>
<tr>
<td>Headquarters Addresses</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>OtherAddress Type= <strong>HEAL</strong></td>
<td>LEIFR</td>
<td>HeadquartersAddress</td>
</tr>
<tr>
<td>Last Update Date</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>LastUpdateDate</td>
<td>LEIFR</td>
<td>LEIRecord_LastUpd</td>
</tr>
<tr>
<td>Latitude</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>Legal Form</td>
<td>DTCC</td>
<td>see City</td>
<td>see City</td>
<td>see City</td>
<td>see City</td>
</tr>
<tr>
<td>Legal Form Effective Date</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
</tr>
<tr>
<td>Legal Name</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>LEI</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
</tr>
<tr>
<td>Longitude</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
<td>LEIFR</td>
</tr>
<tr>
<td>Maintenance State</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>CertificationState</td>
<td>LEIFR</td>
<td>LEIFR</td>
</tr>
<tr>
<td>Next Maintenance Date</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>nextCertificationDate / CER</td>
<td>LEIFR</td>
<td>LEIFR</td>
</tr>
<tr>
<td>Other Name</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>Portal Date Time</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>InternalSystemDateTime / AIS</td>
<td>LEIFR</td>
<td>PortalDateTime/ASSIGNED</td>
</tr>
<tr>
<td>PostCode</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>see City</td>
<td>LEIFR</td>
<td>LegalFormAddressPostcode/ZIP</td>
</tr>
<tr>
<td>Record State</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>ProLoiStatus</td>
<td>LEIFR</td>
<td>LEIStatus</td>
</tr>
<tr>
<td>Redirected lei</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>Related LEI</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>Sector</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
<tr>
<td>State</td>
<td>DTCC</td>
<td>LEIFR</td>
<td>EII</td>
<td>EII</td>
<td>EII</td>
</tr>
</tbody>
</table>
Phase Two - Visualization

www.leiwatch.org
Sensemaking – A worked Example
Phase Two – Evaluation (In Progress)

• Running psychometric experiments to test efficacy of visualizations in relation to analytic tasks

• Evaluation techniques drawn from the field of Human-Computer Interaction (HCI)
  – Participants asked to complete tasks
  – Participants prepare pre and post-test domain knowledge map
  – Post task semi-structured questions to gain additional insight

• Metrics include time to completion, completion rates, and success rate of task completion (correctness of answer) and level of insight (based on) a procedure for scoring the domain knowledge maps.

• Data analysis in progress
Future Directions

Global Legal Entity Identifier Watch Site Development

• Continue to integrate LEI data
• Integrate new data (e.g., from SDRs) to support exploration and sensemaking
• Improve existing visualizations and website based on evaluation
• Create new visualizations e.g., network visualizations

Theory Development

• What are the mechanisms by which visual analytics supports sensemaking and insight generation?
• Can participatory visual analysis of LEIs and other financial data support financial stability monitoring that harnesses the wisdom of crowds?