The Linked Data Value Chain Model: A Methodology for Information Integration and Orchestration

Daniel Hladky
Semantic Web Lab at HSE/W3C

28 November 2013
Background

Computer Science and Economics
SAP AG, iXOS (OpenText), Ontos

Research Interest

Linked (Open) Data for Government & Enterprises
NLP, Semantic Web, Business Impact of Linked Data

Activities

Researcher EU FP7 – GeoKnow (within Ontos), LOD Russia
KESW 2012 (Lecture Linked Enterprise Data), KESW’13 (Co-Chair)
W3C Russia office hosted by NRU HSE
Moderator Open Data at RIAN
Co-organising Skolkovo CREI proposal with KAIST, ICSI Berkley, EIS Bonn
The Linked Data Value Chain Model: A Methodology for Information Integration and Orchestration
2. Topic in the research community

References (excerpt)
- D. Wood – Linking Enterprise Data
- T. Heath, C. Bizer – Linked Data…
- B. Hu, G. Svensson – A case study of linked enterprise data
- FP. Servant – Linking Enterprise Data
- P. Frischmuth, S. Auer – Linked Data in Enterprises

Sources that talk about this are
- ISWC
- LDOW
- WWW
- Semantic-Web-Journal

EU funded projects

Organization Data (Intranet)

Public Data (External)

Semantics

Knowledge Base

Structured Data
- ERP, CRM, RDBMS

Unstructured Data
- Papers, Legacy documents
  E-mails, Blogs

Semi-Structured Data
- Wikis, Calendars, CMS

Linked Open Data

News & Social Media

Semi-Structured Data
- Wikis, Calendars, CMS

Unstructured Data
- Papers, Legacy documents
  E-mails, Blogs

Data Lift
4. Research Tasks and Objectives

1. Can we develop a simple and suitable model that captures the Linked Data paradigm in relation to the value chain?

2. Does the Linked Data model support the understanding of information integration using Linked Data technologies?

3. Can we define a model for measuring and quantifying the value of Linked Data information integration along the value chain?
Criteria 1 “Model”
A model to capture the Linked Data Value Chain.

Criteria 2 “Value”
Algorithm and method to measure the value by comparing the metrics from the manual and the tool supported process.

Criteria 3 “Prototype”
A software prototype that will demonstrate how the automation of the Linked Data stack can be orchestrated.
### 5. Importance

<table>
<thead>
<tr>
<th>Information Integration</th>
<th>ROI Valuation</th>
<th>LD Orchestration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Database Integration</strong></td>
<td><strong>Justifying Investment</strong></td>
<td><strong>Ease of Use</strong></td>
</tr>
<tr>
<td>- taxonomies, schemas</td>
<td>- Model to evaluate investment</td>
<td>- Simplifying process for non expert users</td>
</tr>
<tr>
<td><strong>Fusion and Linking</strong></td>
<td><strong>ROI Adoption</strong></td>
<td><strong>Workbench / Platform</strong></td>
</tr>
<tr>
<td>- Across heterogeneous data</td>
<td>- Extend existing models to the new Linked Data Paradigm</td>
<td>- Integration of tools</td>
</tr>
<tr>
<td><strong>Web Services</strong></td>
<td><strong>Flexibility</strong></td>
<td><strong>- Configuration and adoption to new needs</strong></td>
</tr>
<tr>
<td>- REST, SOA, API etc</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tools</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- MDM, BI, Portals</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. New Research Area

1. A New Model
   1. Linked Data Value Chain Model
   2. ROI algorithm to valuate the model

2. Linked Data Framework
   1. Orchestrating the LD Process
   2. Simplifying the information integration
   3. Proof of Concept to measure impact
6.1 Research Area – Linked Data Value Chain Model

**Types of Data**

- **Raw Data**
- **Linked Data**
- **Human-Readable Data**

**Linked Data Roles**

- Raw Data Provider
- Linked Data Provider
- Linked Data Application Provider
- End User

**Participating Entities**

- Entity
6.2 Linked Data Orchestration Framework

OntoQuad plays role of the central Storage for the collected triplified RDF data.
7. Experimental results
## 7.1 Experimental results – Task comparison (simplified)

<table>
<thead>
<tr>
<th>Task</th>
<th>Avg. time before</th>
<th>Avg. time with LD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. Comparison – ROI using the LDVCM
Criteria 1 “Model”
A new Linked Data Value Chain Model is developed and captures the Linked Data Paradigm for Information Integration.

Criteria 2 “Value”
The LD Value Chain ROI Algorithm is suitable to measure the impact and efficiency. A significant impact is shown using the Linked Data Orchestration Framework.

Criteria 3 “Prototype”
A Linked Data Orchestration Framework is developed and tested using the CRM example, LOD Russia and GeoKnow.
10. Research Summary

**Practical implementations**

**LOD Russia**
Ministry of Education and Science Russia, Contract № 07.524.11.4005 of October 20, 2011

**GeoKnow – Generator**
EU FP7 funded project, Grant Agreement No.318159

**DoW – Linked Data Orchestration Workbench**
Swiss CTI funded project

**Papers (excerpt)**


Q&A
Thank You