ontotext

Linked Data for the Enterprise: Opportunities and Challenges

Marin Dimitrov (Ontotext)

Semantic Days 2012, Stavanger

About Ontotext

- Provides software and expertise for creating, managing and exploiting semantic data
  - Founded in 2000
  - Offices in Bulgaria, USA and UK
- Major clients and industries
  - Media & Publishing (BBC, Press Association)
  - HCLS (AstraZeneca)
  - Cultural Heritage (The British Museum, The National Archives)
  - Defense and Homeland Security

Outline

- Semantic Technologies & Linked Data for the enterprise
- Success stories from Ontotext
- Challenges and lessons learned

SEMANTIC TECHNOLOGIES AND LINKED DATA FOR THE ENTERPRISE

Are Semantic Technologies Suitable for the Enterprise?

- Various standards related to different aspects of information management
- Recent work @ W3C focused on enterprise applications
  - SPARQL 1.1 extensions
  - RDB2RDF
  - HCLS
  - Provenance
  - Semantic Sensor Networks

Enterprise Information Management Challenges

- Many disparate data sources and data silos
- Many point-to-point interfaces
- Data sources with similar/inconsistent information
- Complex data integration processes inadequate for changing business requirements
- Most of the knowledge is hidden in texts
- Difficult to integrate structured data and text

Semantic Web and Linked Data Opportunities for the Enterprise

- Simplify the information integration processes
  - Flexible, easy to evolve data model
  - Bottom-up / incremental integration
  - Efficiently integrate structured and unstructured data
  - Reconciliation of duplicated information
- Provide an enterprise metadata layer
  - Unified metadata vocabulary for the enterprise
  - Align the legacy data silos
  - Improve the information sharing and reuse
  - Agile Master Data Management

Semantic Web and Linked Data Opportunities for the Enterprise (2)

- Discovery and enrichment of information
  - Interlink people, organisations, events, …
  - Enrich enterprise content with structured annotations
  - Discover implicit links and relationships
- Unified access to information within the enterprise
  - Simplified infrastructure based on open standards
- Information interchange across different enterprises
  - Easy publishing and consumption of Linked Data
- Augments existing IT assets and technologies
  - No need for disruptive replacement

Typical Enterprise Use Cases for Linked Data and Semantic Technologies

- Publish / consume Linked Data
  - Linked Data is *not* necessarily free data
  - Facilitate data interchange within the value chain
- Information integration within the enterprise
  - Integrated asset management / align data silos
  - Master Data Management
- Knowledge discovery and semantic search
  - Integrate structured and unstructured data
  - Semantic search

SUCCESS STORIES & LESSONS LEARNED

Success Stories

- Linked Life Data
  - Semantic warehouse integrating and interlinking 25+ public biomedical data sources
  - 1 billion biomedical entities described
  - Interactive discovery and exploration
- Mining relationships in Linked Data
  - Pharmaceutical in the US
  - Interactive relationship discovery between biomedical objects
  - Identify *explicit* relationships (from structured data) and implicit relationships (unstructured data)

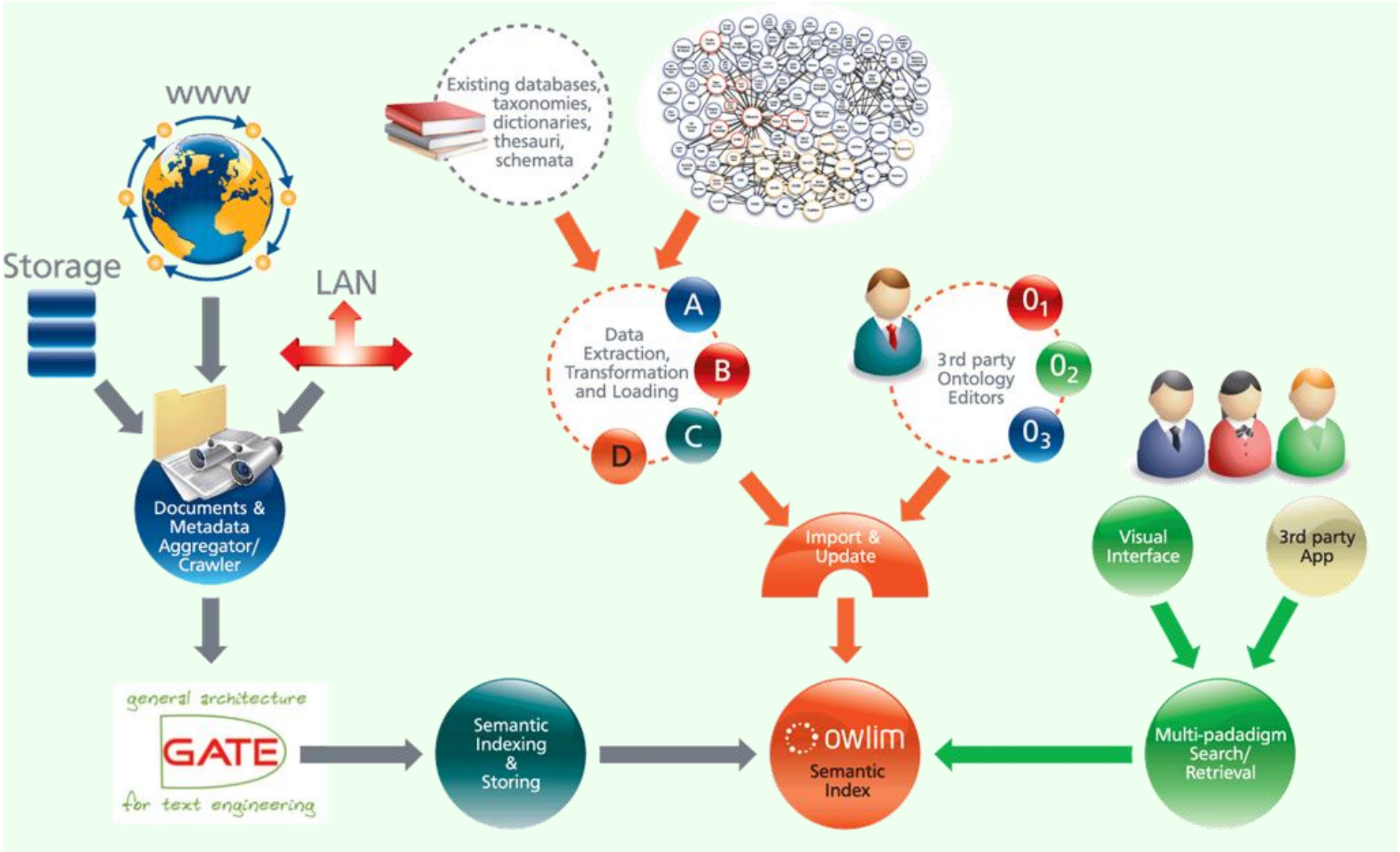
Success Stories (2)

- Assets enrichment
  - Media organisations in the UK and the USA
  - Annotate content repositories and interlink extracted concepts
  - Publish and consume Linked Data
  - Semantic search
- Dynamic Semantic Publishing
  - BBC World Cup 2010 and London Olympics 2012 websites
  - Use Linked Data to automate dynamic publishing of content

Success Stories (3)

- Semantic Knowledge Base & Semantic Search
  - UK Government web archive, also US government agencies
  - Enrich and Interlink 160M documents with Linked Data (LOD)
  - Semantic search over the interlinked data

Semantic Information Integration



Challenges & Lessons Learned

- Domain ontologies may need to be developed for the particular use case
- Tradeoff between ontological expressivity and query performance
- Semantic databases catching up with enterprise data warehouse systems (features and performance)
- Fine-grained provenance is quite challenging
- Data quality problems with many LOD sets
- Curation of text mining results is important

Challenges & Lessons Learned (2)

- Data-driven UX is important
- conventions for URIs of generated RDF data are important
- Multiple proof-of-concept projects to demonstrate the feasibility of the technology
- Enterprise IT teams still getting up to speed with Semantic Technologies
- Semantic Technologies infrastructure stack changing rapidly, technology still maturing

Semantic Technologies from Ontotext for Enterprise Information Management

- Scalable RDF database (OWLIM)
  - Standard compliance; full-text and geo-spatial extensions
  - Cluster set up for fail-over and scalability
  - Scalable inference & retraction
- Semantic annotation platforms (KIM and SBT)
  - Ontology based text mining
  - Multi-paradigm semantic search
  - Domain specific extensions
- Linked Data integration and reconciliation
  - Integrate and align diverse data sources

THANK YOU!