

## COMSODE FP7 Project - position paper for LAPSI workshop

### Encouraging open data usage by commercial developers

Miroslav Konecny (ADDSEN) and the COMSODE project team

#### COMSODE contribution to Open Data exploitation

The project COMSODE (Components Supporting the Open Data Exploitation) is an SME-driven RTD project aimed at progressing the capabilities in the Open Data re-use field. The concept is an answer to barriers still present in this young area: data published by various open data catalogues are poorly integrated; quality assessment, and cleansing are seldom addressed. Data consumers have to integrate the data before they can use them which increases significantly the costs of open data consumption and hinders open data usage and uptake, etc.

It is a project ambition to lay the foundations for a data integration platform based on Open Data which will allow the re-use of data not only between public bodies and end-users but also among public bodies themselves: Public bodies can exchange information by using the same infrastructure and tools as end-users which will decrease costs of exchanging the data and in most cases also enhance the quality and speed-up of the exchange. What is even more important, OpenData APIs can be used by integration projects between public bodies, again saving costs and enhancing the quality of the resulting solution. This in turns strengthens OpenData publishing, with end-users benefiting again – a self reinforcing loop.

#### Exploitable commercial results created by COMSODE:

(1) The main result of the project will be a **software product - Open Data Node (ODN) publication platform**. The exploitation of ODN already started in a Slovak IT project Single Portal for Public Administration. The operational version will be available in September 2015 with following functions:

- Centralized data governance system - internal harvesting, transformation of data, cataloguing;
- Local tool for institutions for open data publishing.

(2) **UnifiedViews** is the core component of ODN - an ETL framework with a native support for transforming RDF data. UnifiedViews allows users to define, execute, monitor, debug, schedule, and share data transformation tasks. UnifiedViews allows users to define and adjust data processing tasks (pipelines) using a GUI. The core components of every data processing task are data processing units (DPUs). DPUs can be exploited separately.

- UnifiedViews was originally developed as a student project at Charles University in Prague and now it is maintained by [Semantica.cz](http://Semantica.cz) (CZ), Semantic Web Company (AT), and EEA (SK). It is part of LOD2 stack.
- Semantic Web Company promotes UnifiedViews as the product. So the future exploitation of this product has very good foundation.
- DSE CUNI already used UnifiedViews in real world implementation in Czech Republic.
- Used for transformation of more than 30 datasets (some of them really large) into RDF.

(3) **Spinque Linked Data** is a new product that applies the Search by Strategy concept to the output of the Open Data Node. It is the result of extending the core software stack of Spinque with support for RDF data. Spinque Linked Data allows data publishers, as well as developers, to quickly create customized search engines over multiple datasets.

- An initial prototype demonstrates the approach with the open data of the Rijksmuseum Amsterdam. Information from DBpedia is integrated in the search strategies to improve the ranking of the artwork search, enable multilingual search and add recommendation.
- Spinque is already using this new product (currently in prototype state) with the members and users of the COMSODE project. For example, an application was created on top of open data from the Czech inspection authorities and the business registry to find inspections conducted at restaurants. Spinque plans to further exploit their new product to expand its current commercial offering, also besides the scope of COMSODE.

**(4) Methodologies for publishing open data** available (final version expected in 8 months) under convenient open licence (CC-BY 4.0) for free download from project website (deliverables section, later to be transformed to Showcase section). Methodologies will be free of charge for use. Licensing will empower 3rd parties to enhance the documentation and even publish and sell printed copies.

- COMSODE methodologies are available as public deliverables – please note the D5.1 – Methodology for publishing datasets as open data – [DOWNLOAD](#); [ANNEX 1](#) (Documentation of practices); [ANNEX2](#) (Methodology Master Spreadsheet)

**(5) New services** based on knowledge and experiences gained during COMSODE project

- Consulting – methodology, standards, strategy, plan, architecture
- SW development and integration services – based on ODN, UnifiedViews and Search solution (Implementation services - installation, configuration, integration, sizing, security issues, support, administration, hosting, etc.)
- Trainings based on methodologies and best practice from concrete implementations.

### Examples of commercial services from COMSODE countries using Open Data

- Czech republic:
  - [DATY.cz](#) - financial data of companies using public registries. Some reports and features are paid – Freemium business model.
  - [DATLOWE.cz](#) – product [Drug encyklopedia](#) - Advanced drug information for healthcare professionals and patients enabled with semantic technologies and Linked Data – Service business model.
  - Annual hackathon organized by the Otakar Motejl Fund "[Openig up data together](#)"
- Slovakia:
  - [Finstat.sk](#) – financial data of companies using public registries. Some reports and features are paid – Freemium business model.
  - [govdata.sk](#) is a product of [minio.sk](#) – customized data services for professional use
  - Others supported by hackathons and networking (Open Scraper Challenge, Restart Slovakia, Rubyslava, etc.)
- Italy:
  - [Dandelion.eu](#) – geospatial data - by leveraging Big Data & Semantic Web technologies are building a high-quality knowledge-graph accessible via simple APIs. Professional use is paid on monthly fee.

### Main barriers identified and topics for discussion at the Share-PSI workshop

- **Missing clear message for investors and entrepreneurs** – numbers, business models and good examples that convince to put efforts and investments into open data services.
  - How to promote best practice of commercial cases?
  - What business models can reveal the value of Open Data?
  - What market potential numbers are credible (when certainly talking about disruptive approach to information management!)?
- **Critical mass of quality open data is still not available** - It is not easy to find the right dataset. There are islands of dataset groups (catalogues, national servers) but cross-information is missing. COMSODE contributes by its activities to solve this problem.
  - How much Open Data is needed to create critical mass with commercial value?
  - How to interlink available datasets and enable their cross-use?
  - What quality is needed for commercial applications?

## Links

- COMSODE (Components Supporting the Open Data Exploitation - <http://www.comsode.eu>)
- Semantica - <http://semantica.cz/en/>
- COMSODE D5.1 – [http://www.comsode.eu/wp-content/uploads/D5.1-Methodology\\_for\\_publishing\\_datasets\\_as\\_open\\_data.pdf](http://www.comsode.eu/wp-content/uploads/D5.1-Methodology_for_publishing_datasets_as_open_data.pdf)
- ANNEX I – [http://www.comsode.eu/wp-content/uploads/Annex2\\_D5.1-Methodology\\_Master\\_Spreadsheet.xlsx](http://www.comsode.eu/wp-content/uploads/Annex2_D5.1-Methodology_Master_Spreadsheet.xlsx)
- ANNEX II - [http://www.comsode.eu/wp-content/uploads/Annex2\\_D5.1-Methodology\\_Master\\_Spreadsheet.xlsx](http://www.comsode.eu/wp-content/uploads/Annex2_D5.1-Methodology_Master_Spreadsheet.xlsx)
- DATY.CZ – [www.daty.cz](http://www.daty.cz)
- Drug encyklopedia - <http://datlowe.org/en/drug-encyclopedia.html>
- Otakar Motejl fund hackaton Opening data together- <http://www.otevrenadata.cz/soutez/>
- FINSTAT – [www.finstat.sk](http://www.finstat.sk)
- Portal GOVDATA – [www.govdata.sk](http://www.govdata.sk)
- MINIO – [www.minio.sk](http://www.minio.sk)
- Dandelion.eu – [www.dandelion.eu](http://www.dandelion.eu)