

Development of St. Petersburg's linked open data site using Information Workbench

Dmitry Mouromtsev, Vitaly Vlasov, Mikhail Galkin, Vitaly Knyazev and
Olga Parkhimovich
NRU ITMO, Russia



Outline

- Reasons why we do it
- Objectives and goals
- Methods and approaches
- Development process and tools
- Features
- Future work



Reasons

- Government has an opportunity to publish their data
- Society has a strong intention to control authorities' work



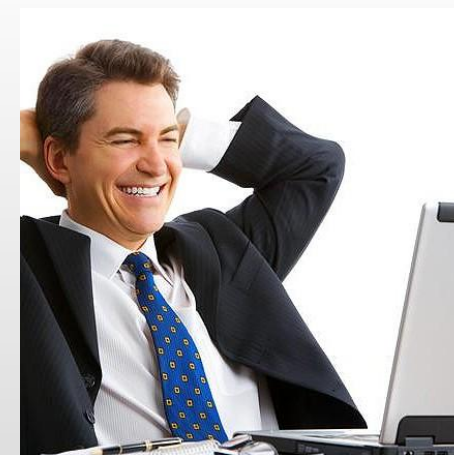
Reasons



Current level of Open
Government Data in Russia



Community



Reasons

- Government has an opportunity to publish their data
- Society has a strong intention to control authorities' work
- One of the best fields to follow the Linked Open Data (LOD) concept
- Semantic technologies may help greatly
- Successful examples of such portals in the USA, the UK and other countries

➤ LOD + Semantic Web = 



Project Scope

Main goal:

- Provide a powerful and flexible platform to publish open government data

Primary objectives:

- Full support of RDF, triple stores and available SPARQL endpoint
- Big Data
- Different sources of datasets
- Visualizing the data in convenient forms



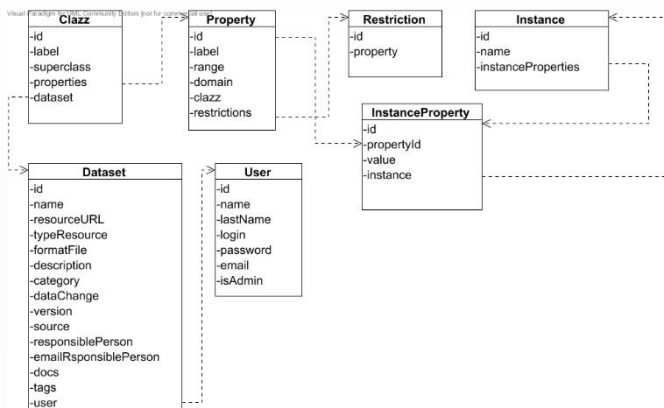
Approaches

1st approach – processing the budget of Saint Petersburg

- ~100 MB RDF data
- Google App Engine
- Trying out some algorithms to work with triples
- Manual visualizing



Approaches



Заголовок + Logo

[На главную](#) О проекте

Добро пожаловать!

Последние данные

- [Бюджет 2012](#)

Категории данных

- [Бюджет СПб](#)
- [Категория 2](#)
- [Категория 3](#)
- [Категория 4](#)
- [Категория 5](#)

Buttons: Загрузить данные, Построить таблицу, Построить график

Заголовок + Logo

[На главную](#) О проекте

Buttons: Загрузить данные, Построить таблицу, Построить график

Таблица

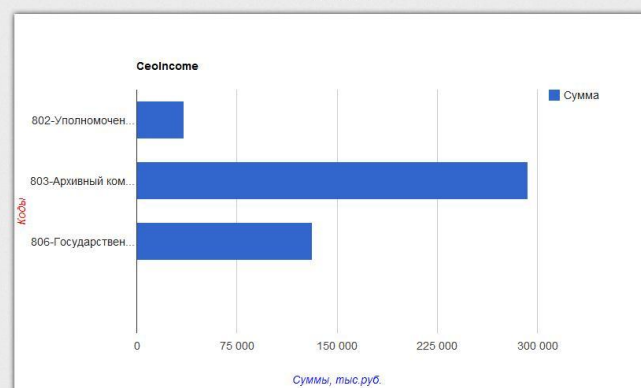
| Наименование | Код ОКУ | Код вида расходов | Код целевой статьи | Сумма |
|--------------------------------------------------------------------------------------------------------------|---------|-------------------|--------------------|----------|
| Содержание исполнительного органа государственной власти Санкт-Петербурга | | | | |
| Оплата труда и начисления на выплаты по оплате труда | 210 | 012 | 0010009 | 106164.9 |
| Заработная плата | 211 | 012 | 0010009 | 87272.1 |
| Начисления на выплаты по оплате труда | 213 | 012 | 0010009 | 18892.8 |
| Оплата работ, услуг | 220 | 012 | 0010009 | 20978.2 |
| Услуги связи | 221 | 012 | 0010009 | 540.0 |
| Транспортные услуги | 222 | 012 | 0010009 | 12490.6 |
| Коммунальные услуги | 223 | 012 | 0010009 | 963.2 |
| Работы, услуги по содержанию имущества | 225 | 012 | 0010009 | 4519.4 |
| Прочие работы, услуги | 226 | 012 | 0010009 | 2465.0 |
| Прочие расходы | 290 | 012 | 0010009 | 683.7 |
| Увеличение стоимости основных средств | 310 | 012 | 0010009 | 1933.0 |
| Увеличение стоимости материальных запасов | 340 | 012 | 0010009 | 1660.0 |
| Расходы на эксплуатацию и развитие программных продуктов автоматизированного ведения бюджетного учета | | | | |
| Оплата работ, услуг | 220 | 012 | 3300036 | 136.1 |
| Прочие работы, услуги | 226 | 012 | 3300036 | 136.1 |

Заголовок + Logo

[На главную](#) О проекте

Buttons: Загрузить данные, Построить таблицу, Построить график

Графики



Своincome

802-Уполномочен...

803-Архивный ком...

806-Государствен...

Суммы, тыс.руб.

Legend: Сумма



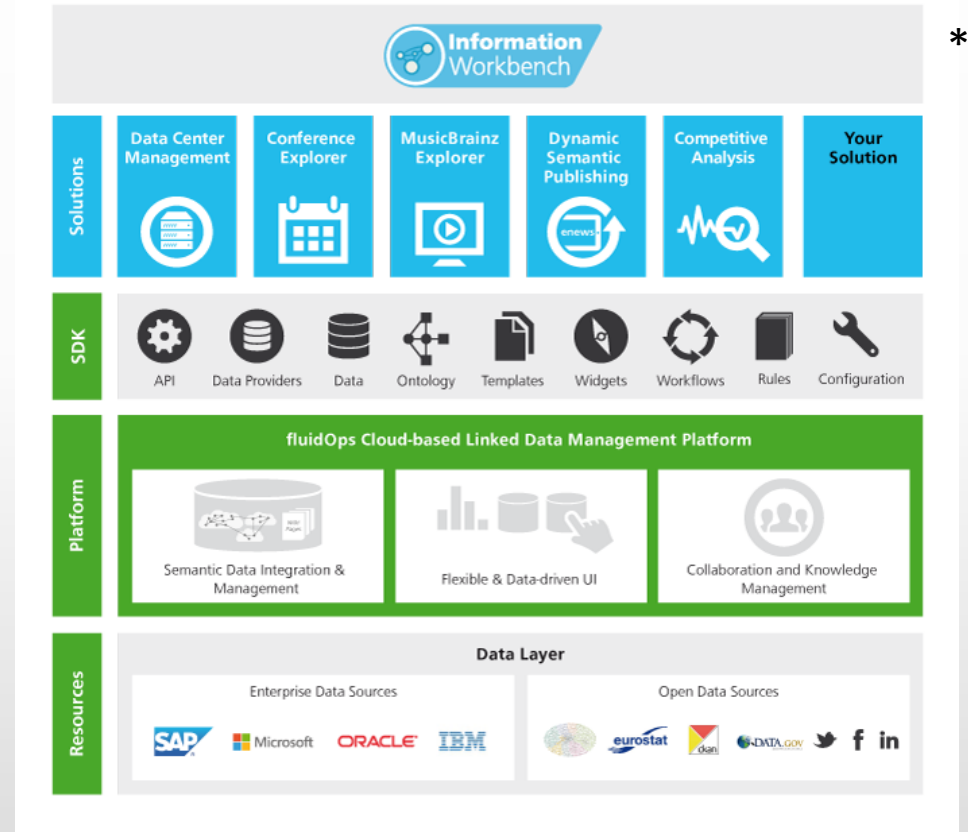
Approaches

2nd approach is based on The Information Workbench by fluid Operations.

(<http://www.fluidops.com/>)

What is the Information Workbench?

- A powerful tool that supports a large amount of SW formats (RDF, N3, Turtle, TriG, etc)
- Has built-in Sesame store, SPARQL endpoint and visualizing methods
- Satisfies the requirement about different sources of data by introduction of the Data Provider concept.



* source: <http://www.fluidops.com/information-workbench/>

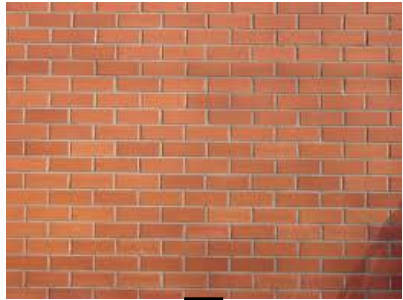
Approaches

What is the Data Provider ?

- A program
- A module of the IWB
- Connects to some source (local, online) and processes the data to the database according to its purpose
- Updates the data automatically and regularly
- Versatile processing methods - up to the developer



Approaches



Data Provider



Data Provider



Implementation

- Separate committees  large number of web-sites with data
- Roads repairs dataset – for a prototype

http://gov.spb.ru/gov/otrasl/tr_infr_kom/tekobjekt/tek_rem/



Initially, we have a table:

| Street Name | Roadway area (thousands m^2) | Pedestrian zone area (thousands m^2) | Total (thousands m^2) |
|-------------|------------------------------------|--------------------------------------------|-----------------------------|
| Name1 | X1 | Y1 | X1+Y1 |
| Name2 | X2 | Y2 | X2+Y2 |
| ... | ... | ... | ... |



Implementation

Processing to triples:

URI = URL

Street Name → uri:Name

Roadway area → uri:Roadway

Pedestrian zone area → uri:Pedestrian

Total area → uri:Total

| Subject | Predicate | Object |
|-----------|----------------|----------|
| uri:name1 | rdf:type | uri:Name |
| uri:name1 | rdfs:label | "Name1" |
| uri:name1 | uri:Roadway | X1 |
| uri:name1 | uri:Pedestrian | Y1 |
| uri:name1 | uri:Total | X1+Y1 |

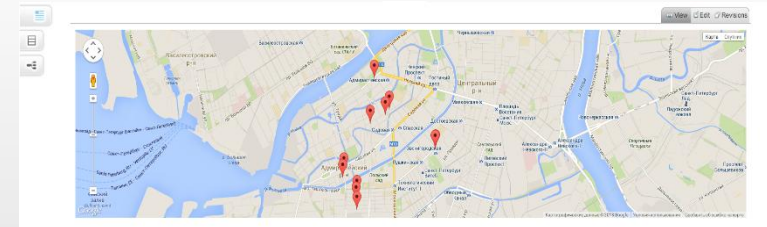
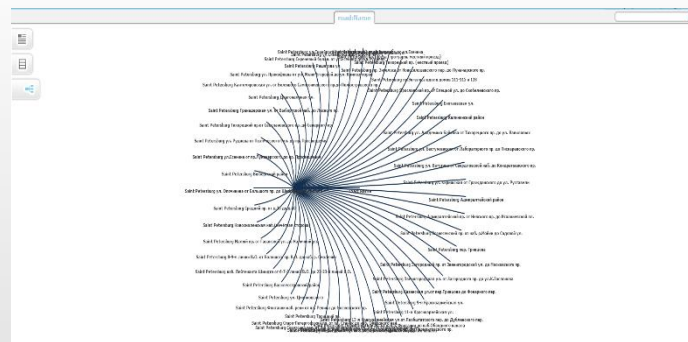
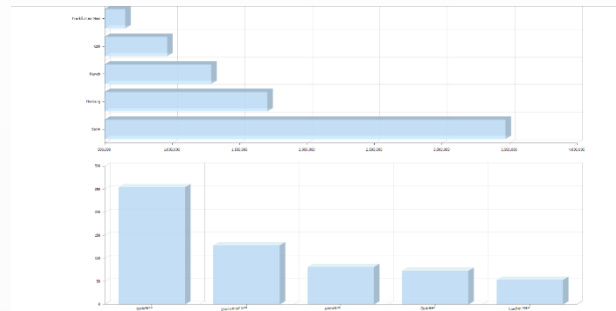
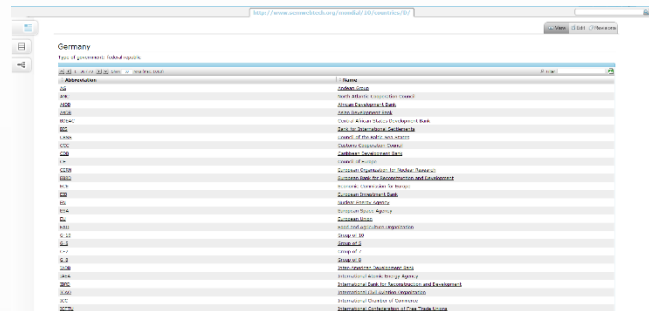


Features

- Transforming text data to Linked Data
- An ontology is constructed automatically
- In case of roads dataset – binding street names to a map
- Data is accessible via SPARQL endpoint
- Huge opportunities for improving
 - Number of injuries
 - The most dangerous roads
 - Reliability of newly repaired roads



Features



Future work

Extensive way

More Data Providers for government web-sites since everyone is doing front-end differently



Bad

Intensive way

A universal algorithm.

- Specify URL and a title of the data we need
- Algorithm crawls through a web page, finds the data and processes it to the datastore



YES!

Future work

- A universal algorithm
- More datasets
- **MORE** datasets
- Work with authorities to obtain the data that is already suitable for machine processing
- Emphasis on practical usage of Linked Data



Thank you!

<http://ailab.ifmo.ru/>

mikhgalkin@gmail.com

olya.parkhimovich@gmail.com

