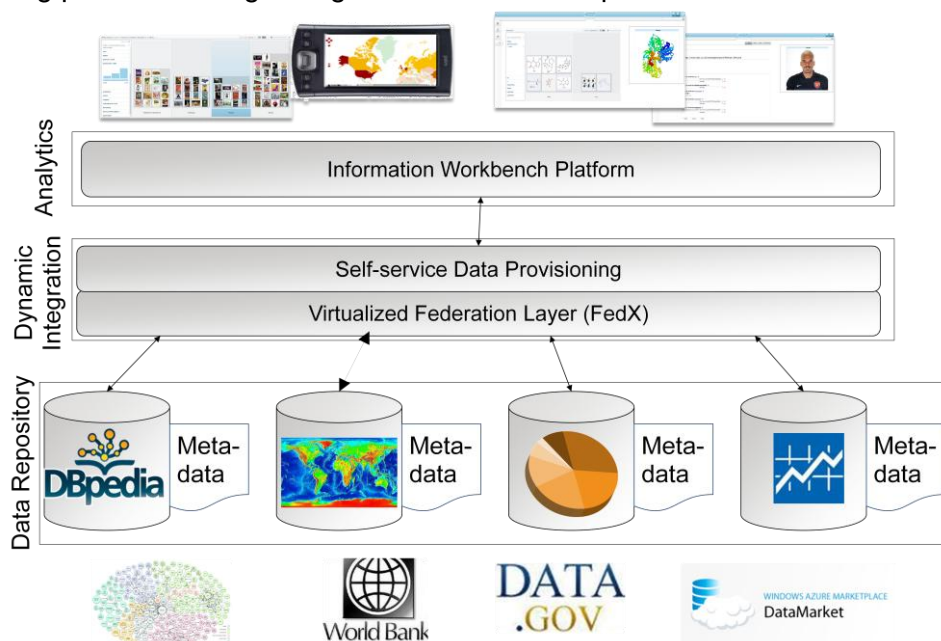


Bachelorprojekt für 2011/2012
 Fachgebiet Informationssysteme
 Prof. Dr. Felix Naumann

A Cloud Platform for On-Demand Access to Open Data

In the last years the concept of **Open Data** has emerged as a game-changing development in a number of countries. Examples include the Linking Open Data Initiative (LOD), governmental initiatives such as data.gov, data.gov.uk, publicdata.eu, or open data initiatives of international organizations such as data.worldbank.org. Open Data can improve public services, create innovation, and lead to greater transparency. In particular, making public sector data openly available to analytics solutions for quick and effective decision making is more than ever one of the core enablers of corporate growth, productivity, and a sustainable competitive advantage. While so far great progress has been made on opening data and publishing it, mechanisms that support low-cost, target-oriented, and **on-demand use of open data** are still missing.

To leverage the benefits of open data, fluid Operations is developing a cloud platform that will simplify and considerably accelerate the entire data utilization process, namely support in **data discovery, self-service deployment** of data sources, **scalable data storage, data integration, data curation**, as well as **analytics** and custom **application development** on top of the data. Implementing the **Data-as-a-Service (DaaS)** paradigm, the platform will facilitate data access to support ad hoc, dynamic information needs. It will allow organizations to (i) search and acquire openly available data from public Linked Data sources, (ii) to structure and prepare this data according to user requirements; and (iii) to integrate and aggregate it with local or private data for on-demand analytical applications. On the frontend side, we provide an extensible set of tools for exploring, browsing, searching and analyzing public data originating from different data providers.



Project Description

In the context of this cloud platform, the goal of the project is to create a repository of open data sources. Concrete challenges and tasks include:

- The **population of the repository** from public sources, including data from the Linked Open Data initiative, publicdata.eu, the Azure datamarket, Worldbank, etc.
- The extraction and **generation of metadata** about data sets (such as schema, coverage, topics) to support the search and discovery of data sets
- The **generation of statistics** about data sets to support efficient query processing
- **Link discovery** across data sources to enable search and queries across multiple data sets
- **Data cleansing** to deal with imperfections in the data, especially when integrating heterogeneous sources
- **Versioning and change management** of datasets to deal with the aspects of data dynamics

In addition to the creation of the repository itself, tasks will include the development of the frontend to support end-user oriented interaction with the data, including:

- **Search** for data sets, **visual exploration** of the data and metadata
- Mash-Ups and **widgets** to visualize and analyze the data
- Interface for user-friendly **structured queries** across multiple, federated data sources

The value of the platform will be demonstrated in concrete application scenarios.

Organization

A group of 5-8 students may participate in the project. Organization and tasks will be determined by the project participants in close cooperation with fluid Operations, following an agile development process. The programming language will be Java. The project will be carried out at the Hasso Plattner Institute in Potsdam under the supervision of Prof. Naumann, Johannes Lorey, and Christoph Böhm.

The expected start date is 1.10.2011.

fluid Operations AG

fluid Operations is a startup software company based in Walldorf, Germany. fluidOps specializes in topics centred on Cloud Computing and Semantic Technologies and has invented a number of new technologies for cloud infrastructures that enable a true Enterprise Compute Cloud. All resources of an adaptive, cloud-enabled data centre can be set up, monitored, and maintained from a single, unified, and intuitive management console, and new instances of services or applications can be created at the click of a button. These technologies are available in fluid Operations' main product, the eCloudManager Suite, which manages the entire cloud stack from hardware to software. At the interface of cloud computing and semantic technologies, fluidOps develops the Information Workbench, a platform for the management of linked data and semantic applications in the cloud.

fluid Operations AG: <http://www.fluidops.com/>

Information Workbench: <http://www.fluidops.com/information+workbench>

FedX: <http://www.fluidops.com/FedX>