ON DEMAND ACCESS TO BIG DATA THROUGH SEMANTIC TECHNOLOGIES

Peter Haase, Michael Schmidt
fluid Operations AG

fluid Operations (fluidOps)

Linked Data & Semantic Technologies

Information Workbench

Enterprise Cloud Computing

eCloudManager Product Suite

Software company founded Q1/2008 by team of serial entrepreneurs, privately held, VC funded

Headquarters in Walldorf / Germany, SAP Partner Port

Currently 45 employees

Named "**Cool Vendor**" by Gartner Mar 2010

Global **reseller agreement with EMC** focus large enterprise customers Apr 2010

NetApp Advantage Alliance Partner Oct 2010

Outline

- **Big Data Challenges: Beyond Volume**
- **Semantic Technologies for Big Data Challenges**
- **Linked Data as a Service**
- **On Demand Data Access in a Self-service Process**
- **FedX: Federated Query Processing over Linked (Big) Data**
- **Application Examples**

Big Data

"**Big data** consists of data sets that grow so large that they become awkward to work with using on-hand database management tools." (Wikipedia)

- *12 terabytes of Tweets created daily*
- *30 terabytes of telescope data each night*
- *350 billion meter readings*
- *…*

Optique Case Study: Statoil Exploration

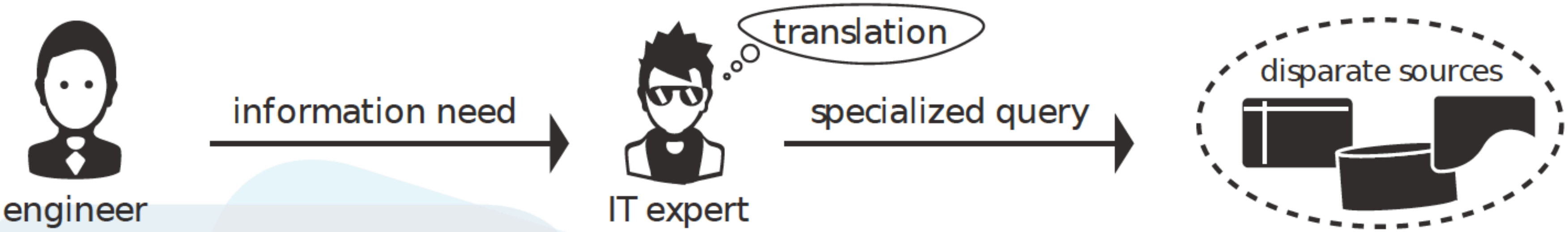
Experts in geology and geophysics develop stratigraphic models of unexplored areas.

- Based on production and exploration data from nearby locations
- Analytics on:
  - 1,000 TB of relational data
  - using diverse schemata
  - spread over 2,000 tables
  - spread over multiple individual data bases

- 900 experts in Statoil Exploration
- up to 4 days for new data access queries
- assistance from IT-experts required

©Øyvind Hagen, Statoil

Scalable End-user Access to Big Data



engineer

information need

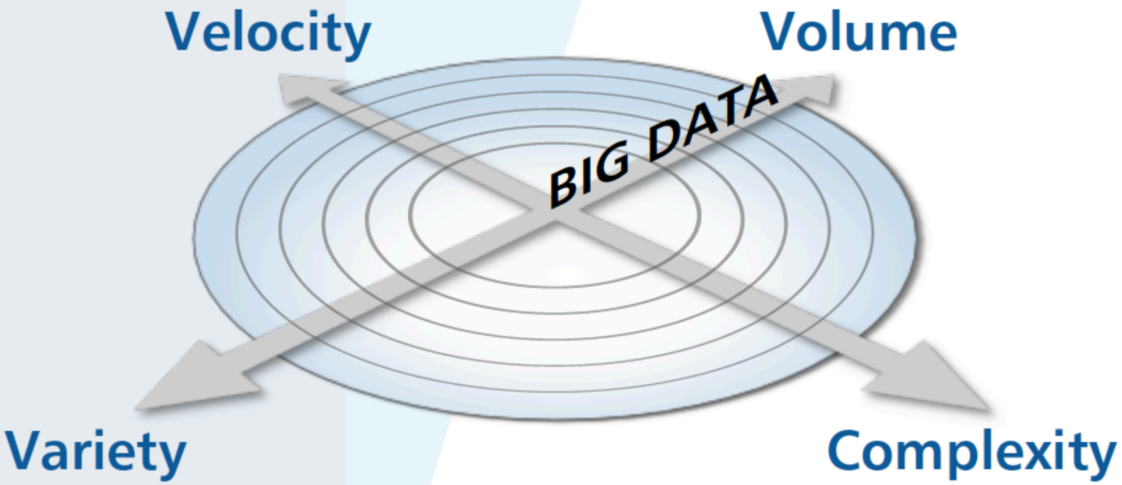
IT expert

translation

specialized query

disparate sources

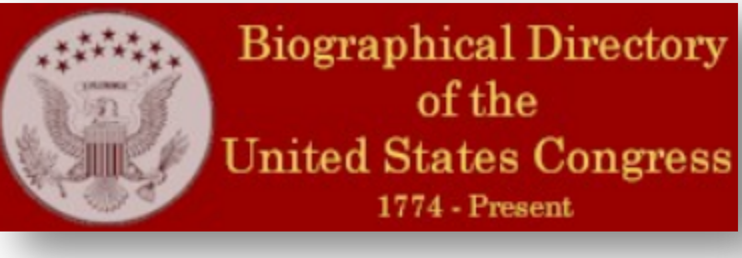
Up to **80%** of experts' time spent accessing Big Data



Life Sciences and Pharma Databases

prosite
SIB
GO
the Gene Ontology
ChEBI
UniProt
the universal protein resource
UniSTS
Integrating Markers and Maps
PubMed
www.p
Entrez Gene
RCSB PDB
PROTEIN DATA BANK
BIOCARTA
PubChem
KEGG
Kyoto Encyclopedia of Genes and Genomes
OBO
open biomedical ontologies
Pfam
UniParc
The UniProt Archive
MGI
MeSH
OMIM
HG
NC
KEGG COMPOUND
Taxonomy Browser
UniRef
Non-redundant Reference
100%>90%>50%
Entrez Nucleotide
ProDom
InterPro
Protein Archive
HomoloGene
Discover Homologs

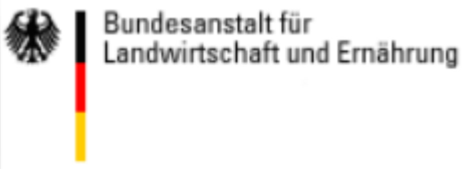
Wealth of Open Gov Data


Biographical Directory of the United States Congress
1774 - Present

European Commission
European Commission > Beneficiaries > Fts
Financial Transparency System

USASPENDING.GOV
Government spending at your fingertips

Welcome to
the European Parliament


Bundesanstalt für
Landwirtschaft und Ernährung

OFFICE OF
MANAGEMENT AND BUDGET

Unklarheiten.de
Politische Datenbank

Semantic Technologies for Horizontal Big Data

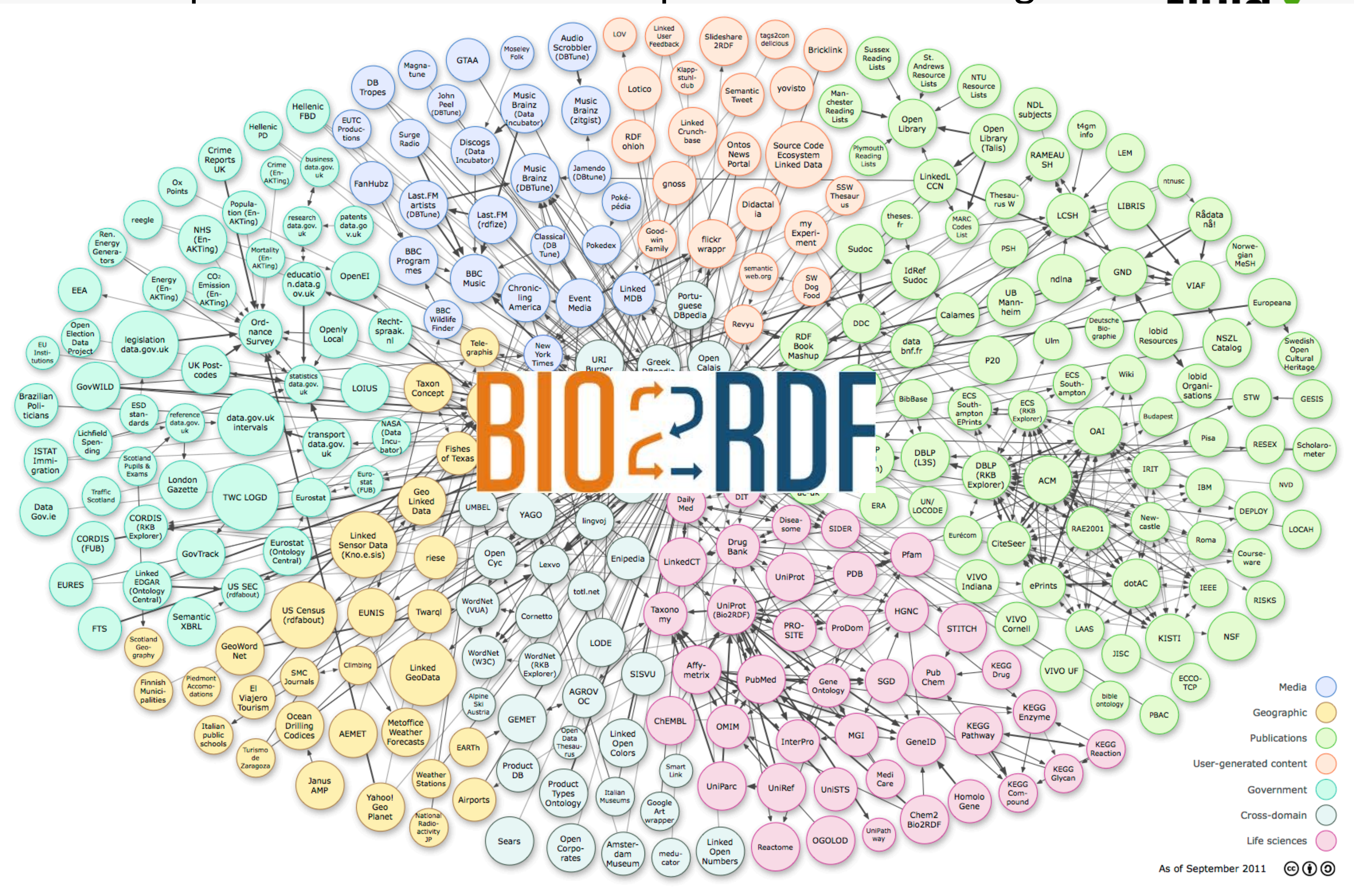
Linked Data

- Set of standards, principles for publishing, sharing and interrelating structured data: RDF as data model, SPARQL for querying
- Graph-based data model for achieving higher degree of variety
- Semantically interlink data scattered among different information spaces: from data silos to a **Web of Data**
- Linked Data as abstraction layer for virtualized data access across data spaces

Ontologies

- For **describing the semantics** of the data
- As **conceptual models** for end-user oriented access
- For the **integration** of heterogeneous sources
- For (light-weight) **reasoning**

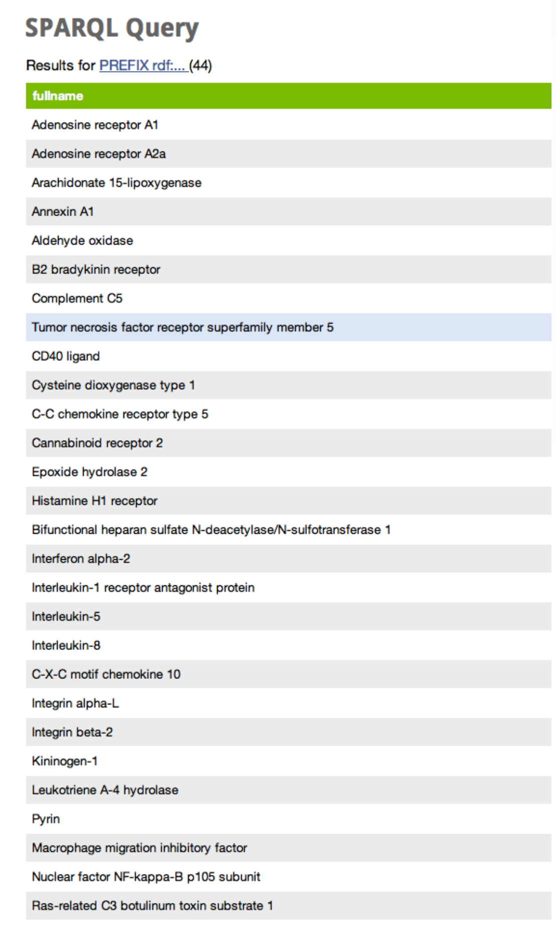
Linked Open Data Cloud: An Example of Horizontal Big Data



Sample SPARQL Query

Find all proteins that are linked to a curated molecular interaction, to inflammatory response and to a target of an existing drug

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX biopax2: <http://www.biopax.org/release/biopax-level2.owl#>
PREFIX uniprot: <http://purl.uniprot.org/core/>
PREFIX drugbank: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/>

SELECT distinct ?fullname
WHERE {
   ?interaction biopax2:PARTICIPANTS ?participant .
   ?participant biopax2:PHYSICAL-ENTITY ?physicalEntity .
   ?physicalEntity skos:exactMatch ?protein .
   ?protein uniprot:classifiedWith <http://purl.uniprot.org/go/0006954>.
   ?protein uniprot:recommendedName ?name.
   ?name uniprot:fullName ?fullname .
   ?protein uniprot:mnemonic ?mnemonic .
   ?target drugbank:swissprotName ?mnemonic .
}
```

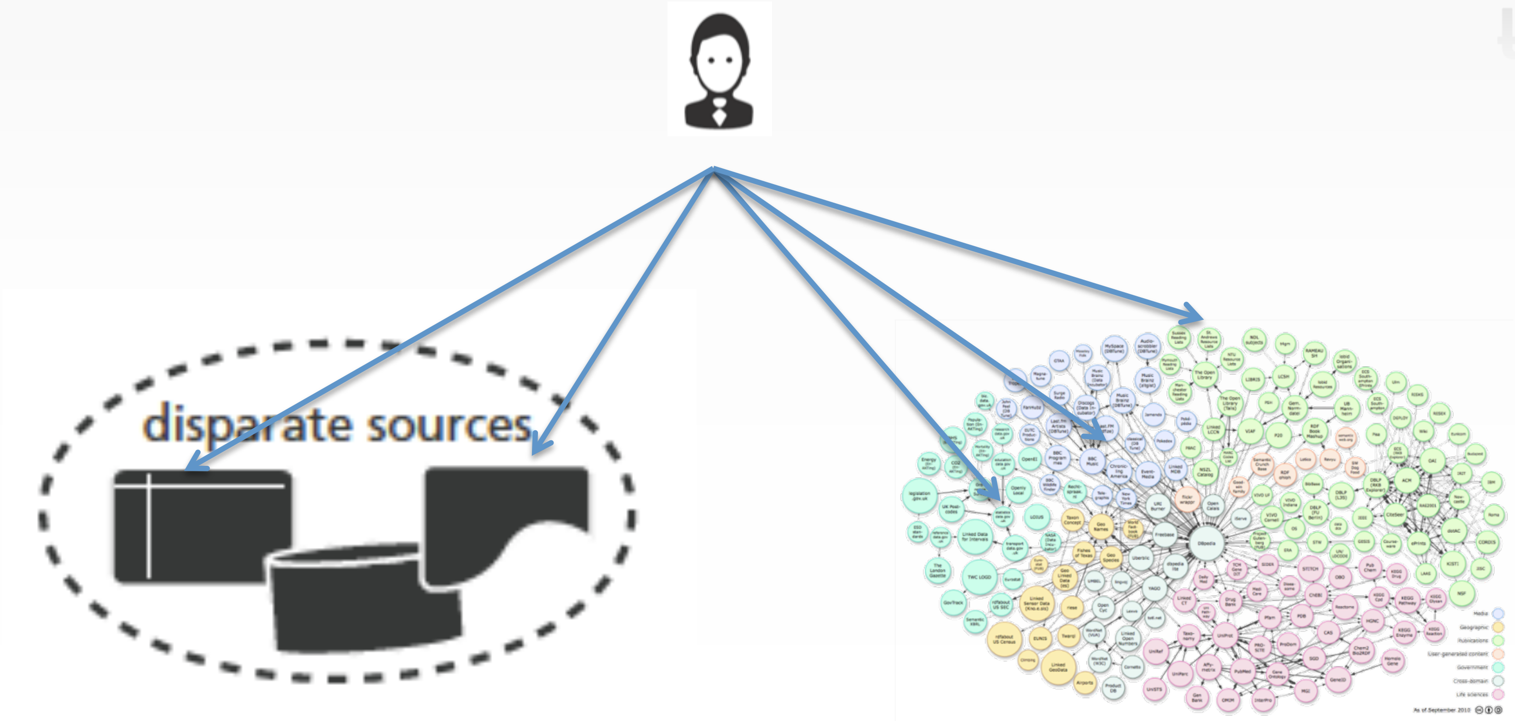


SPARQL Query

Results for PREFIX rdf:... (44)

fullname
Adenosine receptor A1
Adenosine receptor A2a
Arachidonate 15-lipoxygenase
Annexin A1
Aldehyde oxidase
B2 bradykinin receptor
Complement C5
Tumor necrosis factor receptor superfamily member 5
CD40 ligand
Cysteine dioxygenase type 1
C-C chemokine receptor type 5
Cannabinoid receptor 2
Epoxide hydrolase 2
Histamine H1 receptor
Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1
Interferon alpha-2
Interleukin-1 receptor antagonist protein
Interleukin-5
Interleukin-8
C-X-C motif chemokine 10
Integrin alpha-L
Integrin beta-2
Kininogen-1
Leukotriene A-4 hydrolase
Pyrin
Macrophage migration inhibitory factor
Nuclear factor NF-kappa-B p105 subunit
Ras-related C3 botulinum toxin substrate 1

On Demand Access to Big Data



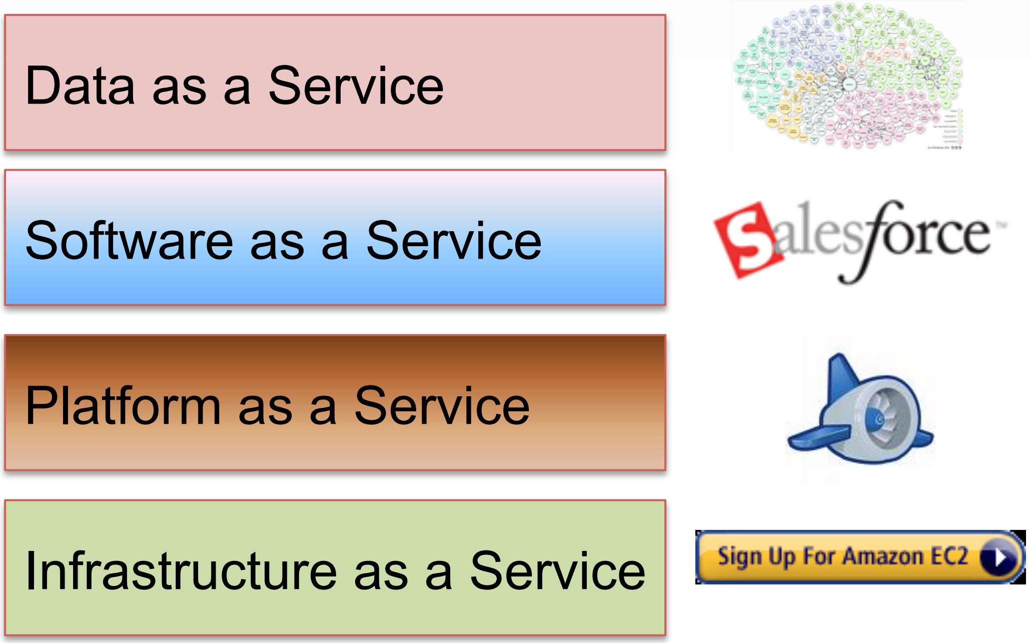
Enabling on demand data access

1. discovery of relevant data sources
2. automated integration and interlinking of sources, and
3. interactive exploration and ad hoc analysis of data

Everything as a Service

- **Abstract from** physical **implementation** details and **location** of resources
- **Regardless of** geographic or organizational **separation of provider and consumer**

- **“In the cloud”**
- **Web based**
- **Virtualized**
- **On-demand**
- **Self-service**
- **Scalable**
- **Pay as you go**



Data as a Service

Software as a Service

Platform as a Service

Infrastructure as a Service

Linked Data as a Service

> “Like all members of the "as a Service” family, DaaS is based on the concept that the product, data in this case, can be provided on demand to the user regardless of geographic or organizational separation of provider and consumer.”
>
> Source: Wikipedia

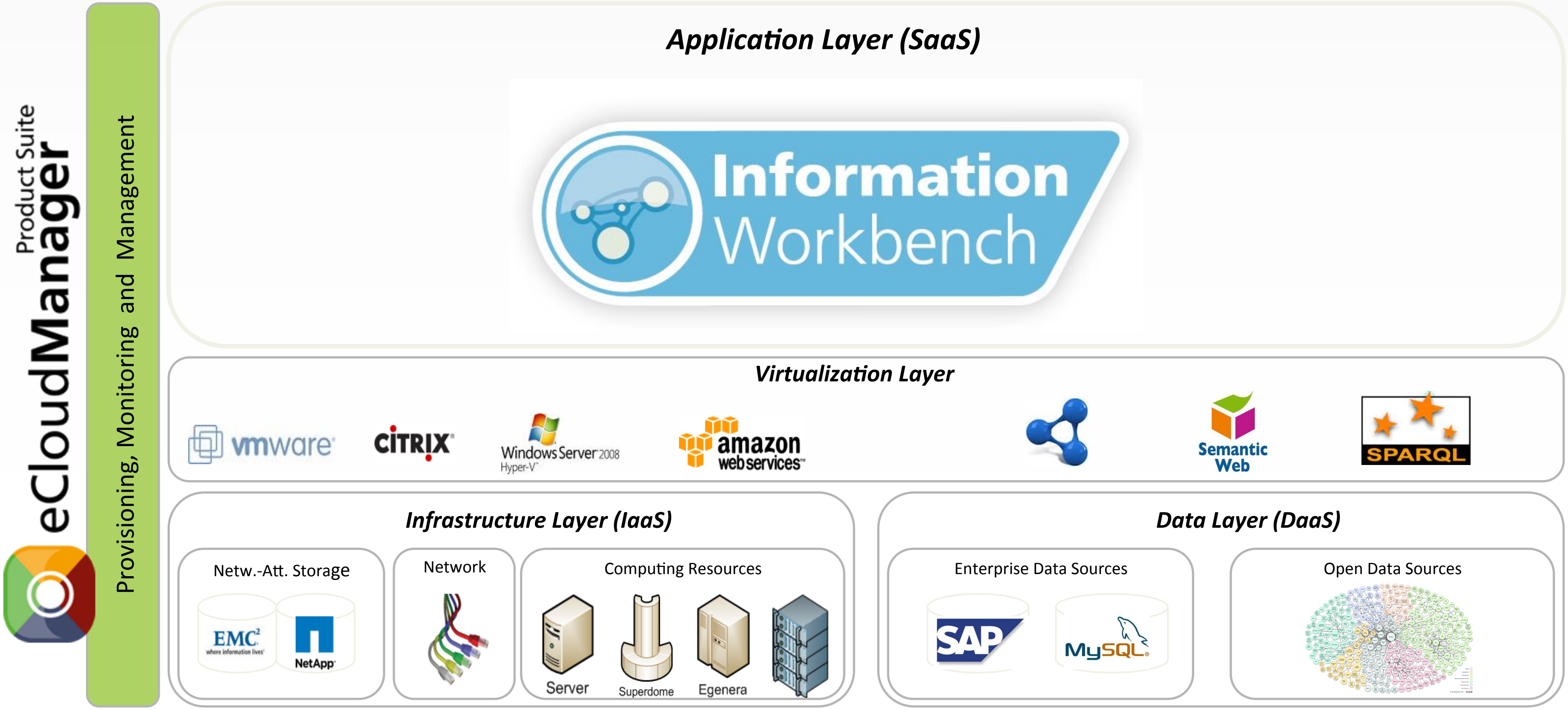
- **Data virtualization** supported by Linked Data principles
  1. Use **URIs** as names for things
  2. Use **HTTP** URIs so that people can look up those names.
  3. When someone looks up a URI, provide useful information, using the standards: **RDF, SPARQL**
  4. Include **links** to other URIs, to discover more things.

- Linked Data as abstraction layer for virtualized data access across data spaces
- Enables data portability across current data silos
- Platform independent data access

- Basis for enabling **automation** of **discovery**, **composition**, and **use of datasets**

Information Workbench:
Linked Data as a Service in a Cloud Platform Architecture

fluid Operations



Product Suite eCloudManager

Provisioning, Monitoring and Management

Application Layer (SaaS)

Virtualization Layer

Infrastructure Layer (IaaS)

Netw.-Att. Storage

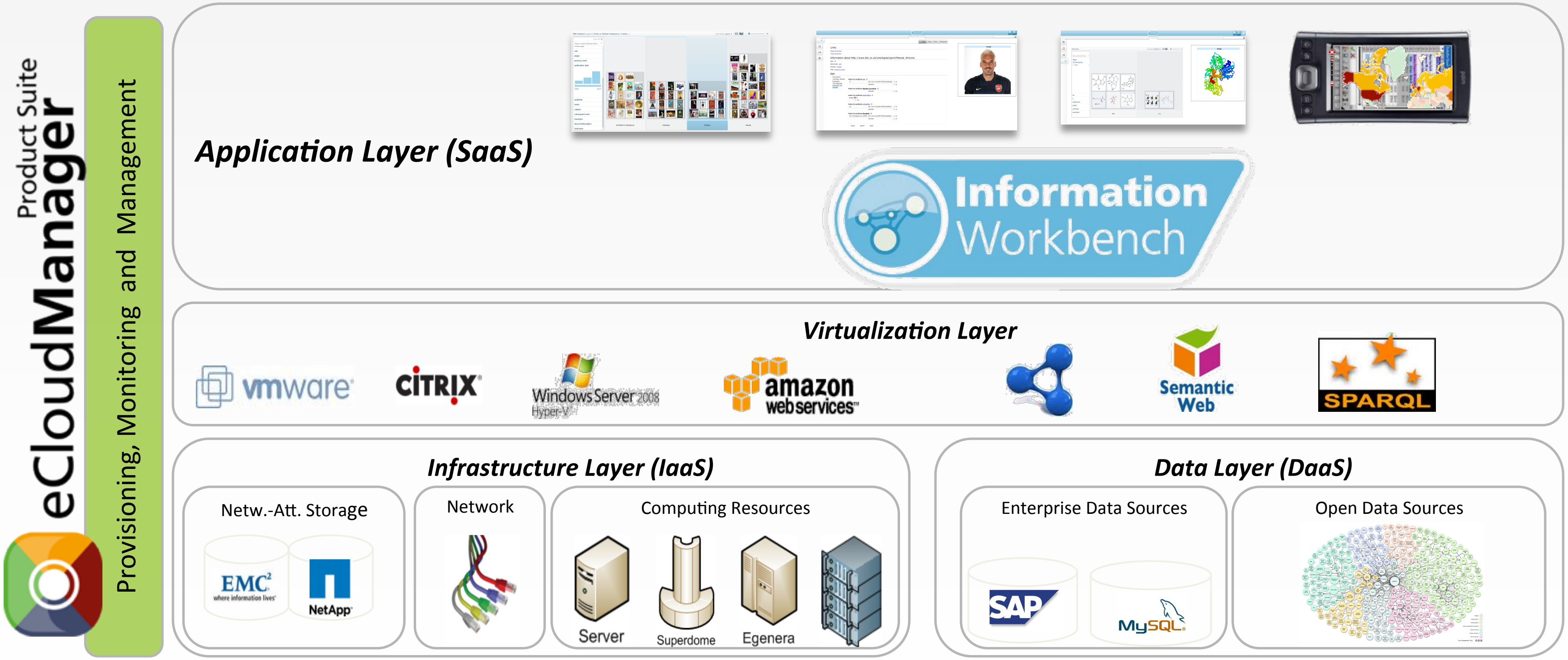
Network

Computing Resources

Data Layer (DaaS)

Enterprise Data Sources

Open Data Sources



Product Suite eCloudManager

Provisioning, Monitoring and Management

Application Layer (SaaS)

Virtualization Layer

Infrastructure Layer (IaaS)

Netw.-Att. Storage

Network

Computing Resources

Data Layer (DaaS)

Enterprise Data Sources

Open Data Sources

Self-service Deployment

- Self-service deployment of the *Information Workbench* in the cloud
- Pay-per-use
- Scalability on demand

Data Discovery

- On demand access to private and public data sources
- Dynamic Discovery

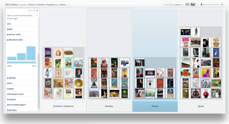
Data Integration & Federation

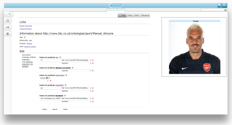
- Virtualized data access
- Dynamic integration & federation of data sources

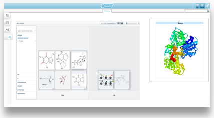
Self-service UI & Analytics

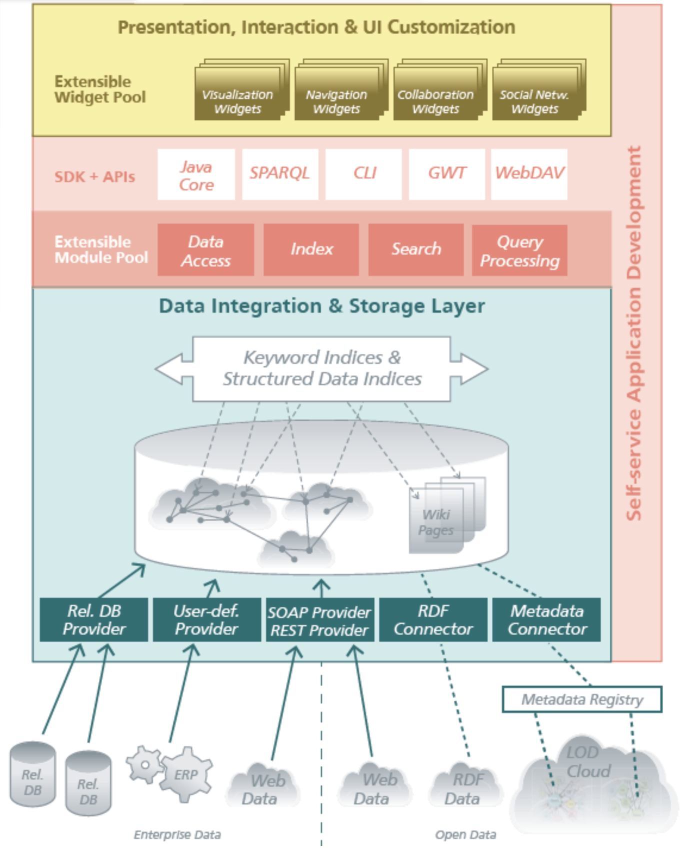
- Living UI, composed from semantics-aware widgets
- Ad hoc data exploration, visualization, analytics

Information Workbench - Linked Data Platform









Presentation, Interaction & UI Customization

Extensible Widget Pool

Visualization Widgets

Navigation Widgets

Collaboration Widgets

Social Netw. Widgets

SDK + APIs

Java Core

SPARQL

CLI

GWT

WebDAV

Extensible Module Pool

Data Access

Index

Search

Query Processing

Data Integration & Storage Layer

Keyword Indices & Structured Data Indices

Wiki Pages

Rel. DB Provider

User-def. Provider

SOAP Provider REST Provider

RDF Connector

Metadata Connector

Metadata Registry

Self-service Application Development

Rel. DB

Rel. DB

ERP

Web Data

Web Data

RDF Data

LOD Cloud

Enterprise Data

Open Data

Semantic Web Data

Information Workbench:

- Semantics- & **Linked Data-based integration** of private and public data sources
- **Intelligent Data Access and Analytics**
  - Visual Exploration
  - Semantic Search
  - Dashboarding and Reporting
- **Collaboration** and knowledge management platform
  - Wiki-based curation & authoring of data
  - Collaborative workflows

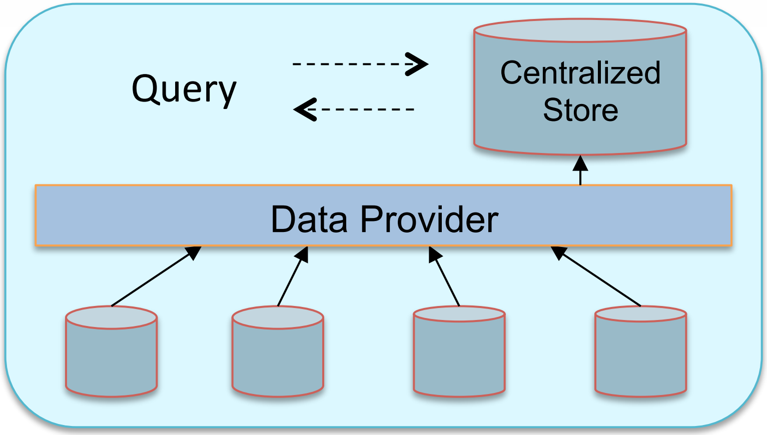
Linked Data Integration Approaches

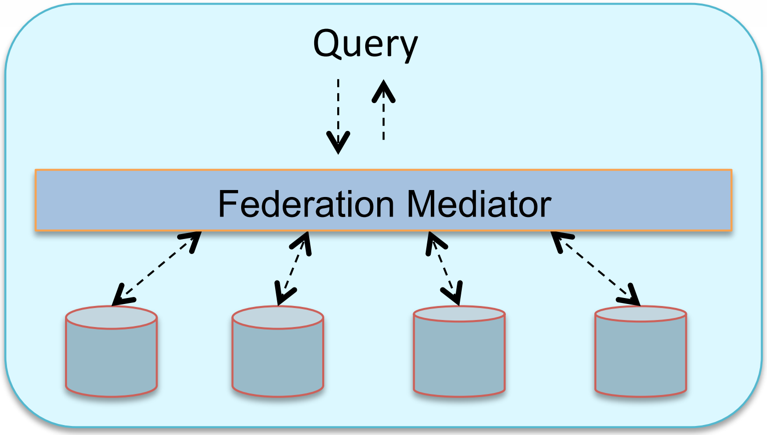
Centralized Integration

- Following a data warehousing approach
- Data providers periodically gather data from sources and lift it to semantic data formats
- Graph-based data format enables pay-as-you-go integration of legacy data sources
- Information Workbench comes with predefined providers for various formats and data sources (Spreadsheets, XML, …)

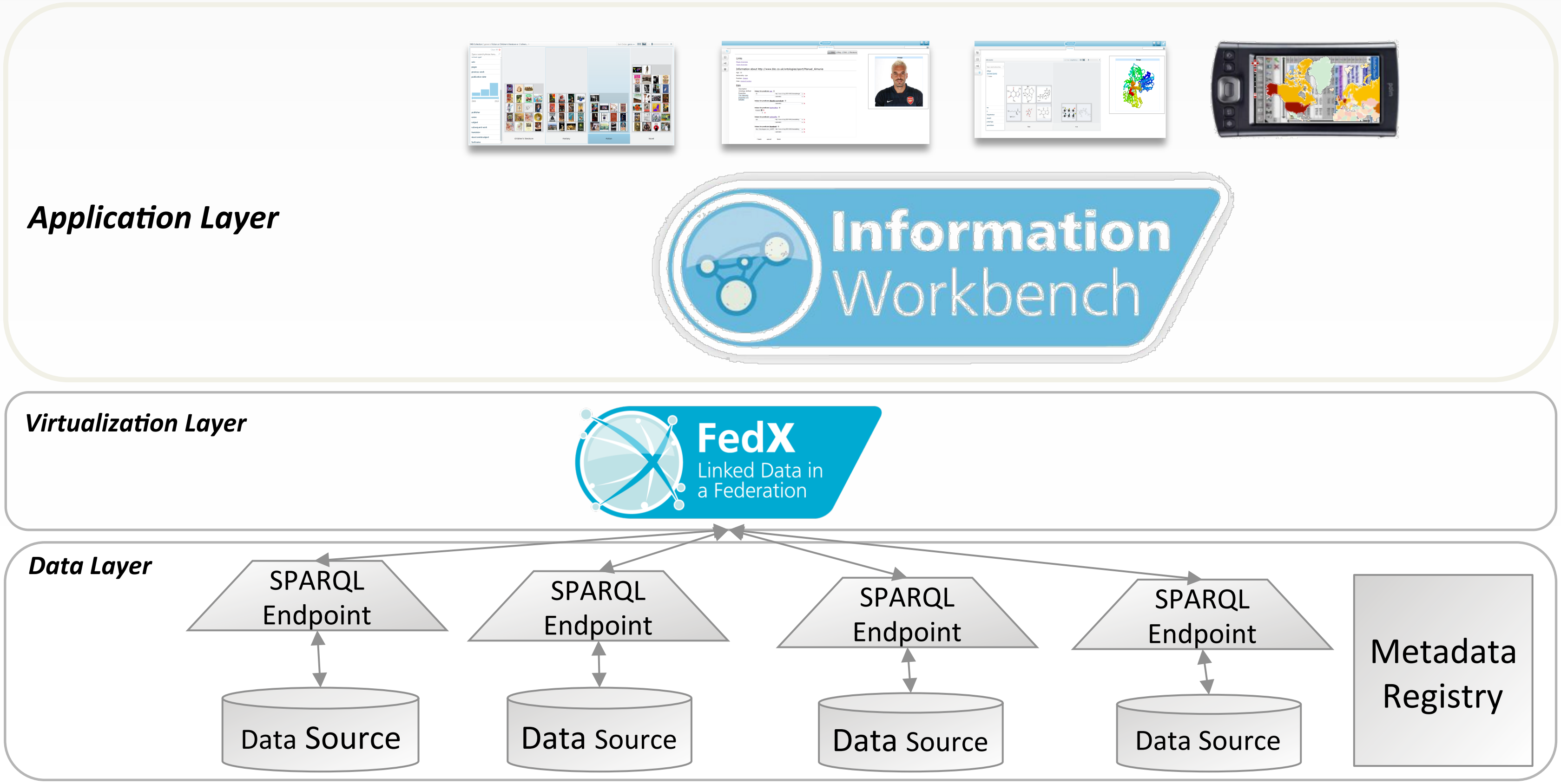
Virtualized Integration

- Autonomous, distributed data sources linked through a federation layer
- No central integration required
- Data sources can be added ad hoc, on demand
- Federation mediator for query processing (routing sub queries to relevant sources)





Enabling Data Composition & Integration: *Federation of Virtualized Data Sources*



See also: ***FedX: Optimization Techniques for Federated Query Processing on Linked Data (ISWC2011)***

FedX Query Processor

- Efficient SPARQL query processing over multiple distributed sources
- Data sources are known and accessible as SPARQL endpoints
- FedX is designed to be fully compatible with SPARQL 1.0
  - Complementary approach to SPARQL 1.1, where sources are specified in the query
  - Many of the optimization techniques naturally carry over to SPARQL 1.1 query processing
- Querying requires no a-priori knowledge about data sources
  - No local preprocessing of the data sources required
  - No need for pre-computed statistics
  - On-demand federation setup

-> enables ad hoc queries against arbitrary SPARQL endpoint federations

Federated Query Processing

Example scenario

- DBpedia and New York Times collections
  - DBpedia as structured knowledge base
  - New York Times as a news provider

➔ **Query both data collections in an integrated way**

Example query might look as follows:

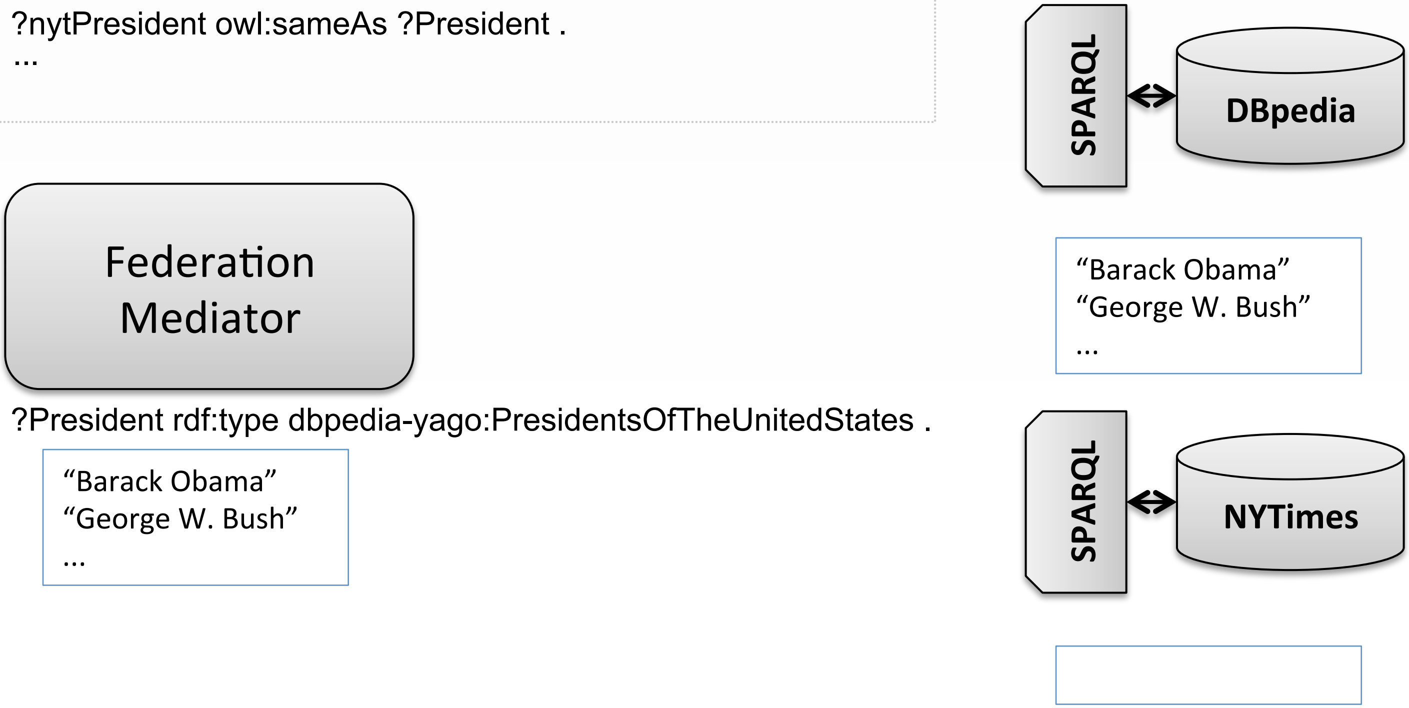
- Find US presidents and associated news articles

```
SELECT ?President ?Party ?TopicPage WHERE {
   ?President rdf:type dbpedia-yago:PresidentsOfTheUnitedStates .
   ?nytPresident owl:sameAs ?President .
   ?President dbpedia:party ?Party .
   ?nytPresident nytimes:topicPage ?TopicPage .
}
```

Federated Query Processing

Example:

```
SELECT ?President ?Party ?TopicPage WHERE {
   ?President rdf:type dbpedia-yago:PresidentsOfTheUnitedStates .
   ?nytPresident owl:sameAs ?President .
   ...
}
```



SPARQL

DBpedia

“Barack Obama”
“George W. Bush”
…

Federation Mediator

?President rdf:type dbpedia-yago:PresidentsOfTheUnitedStates .

“Barack Obama”
“George W. Bush”
…

SPARQL

NYTimes

Federated Query Processing

Example

```
SELECT ?President ?Party ?TopicPage WHERE {
   ?President rdf:type dbpedia-yago:PresidentsOfTheUnitedStates .
   ?nytPresident owl:sameAs ?President .
   ...
}
```

Federation Mediator

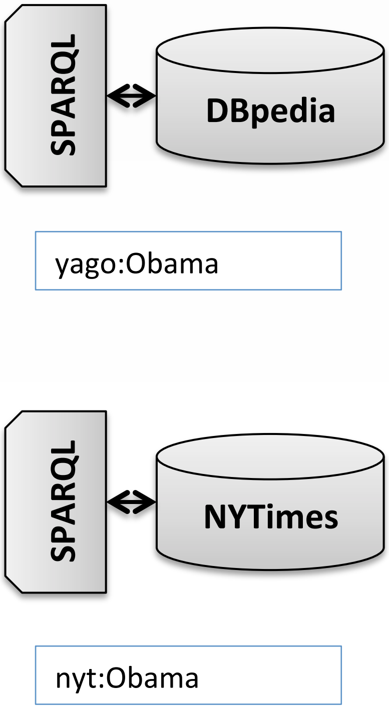
?nytPresident owl:sameAs “Barack Obama” .

Input:

“Barack Obama”
“George W. Bush”
…

Output:

“Barack Obama”, yago:Obama
“Barack Obama”, nyt:Obama



Federated Query Processing

Example

```
SELECT ?President ?Party ?TopicPage WHERE {
   ?President rdf:type dbpedia-yago:PresidentsOfTheUnitedStates .

   ...
}
```

Federation Mediator

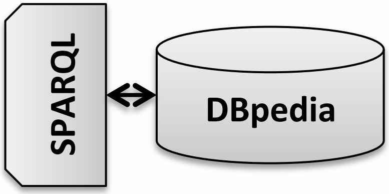
?nytPresident owl:sameAs "George W. Bush" .

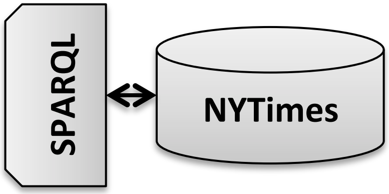
Input:

"Barack Obama"
"George W. Bush"
...

Output:

"Barack Obama", yago:Obama
"Barack Obama", nyt:Obama
"George W. Bush", nyt:Bush

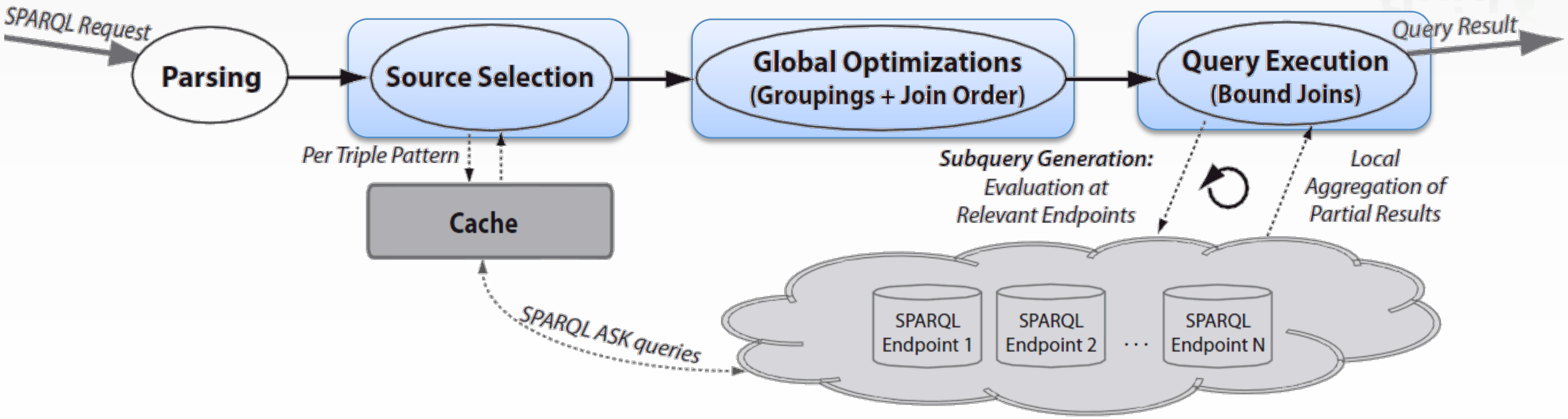




nyt:Bush

... and so on for the other intermediate mappings and triple patterns ...

FedX Federated Query Processing



1.) Involve only relevant sources in the evaluation

Problem: Subqueries are sent to all sources, although potentially irrelevant

2.) Compute joins close to the data

Problem: All joins are executed locally in a nested loop fashion

3.) Reduce remote communication

Problem: Nested loop join causes many remote requests

Optimization Techniques

1.) Involve only relevant sources in the evaluation

Problem: Subqueries are sent to all sources, although potentially irrelevant

Optimization Approach: *Improved Source Selection*

Idea: Annotate Triple patterns with relevant sources

- Identify sources that can contribute information for a particular triple pattern
- Done via SPARQL ASK requests in conjunction with a local cache

  After a warm-up period the cache knows the capabilities of the data sources

  ➔ During source selection remote requests can be avoided

Optimization Techniques

2.) Compute joins close to the data

Problem: All joins are executed locally in a nested loop fashion

Optimization Approach: *Exclusive Groups*

Idea: Group triple patterns with the same single relevant source

- Evaluation in a single (remote) subquery
- Push join to the relevant endpoint

Optimization Techniques

Example: Source Selection + Exclusive Groups

```
SELECT ?President ?Party ?TopicPage WHERE {
 ?President rdf:type dbpedia-yago:PresidentsOfTheUnitedStates .
 ?President dbpedia:party ?Party .
 ?nytPresident owl:sameAs ?President .
 ?nytPresident nytimes:topicPage ?TopicPage .
}
```

Source Selection

@ DBpedia
@ DBpedia
@ DBpedia, NYTimes
@ NYTimes

Exclusive Group

- → **Avoid sending subqueries to sources that are not relevant**
- → **Delegate joins to the endpoint by forming exclusive groups (i.e. executing the respective patterns in a single subquery)**

Optimization Techniques

3.) Reduce remote communication

Problem: Nested loop join causes many remote requests

Optimization Approach: *Improve Join Order*

Idea: Iteratively determine the join order based on count-heuristic

- Count free variables of triple patterns and groups
- Consider "resolved" variable mappings from earlier iteration

Optimization Approach: *Bound Joins*

Idea: Compute joins in a block nested loop fashion

- Reduce the number of requests by "vectored" evaluation of a set of input bindings
- Renaming and Post-Processing technique for SPARQL 1.0
- In SPARQL 1.1: send multiple bindings using new BIND clause

Optimization Techniques

Example: Bound Joins

```
SELECT ?President ?Party ?TopicPage WHERE {
   ?President rdf:type dbpedia:PresidentsOfTheUnitedStates .
   ?President dbpedia:party ?Party .
   ?nytPresident owl:sameAs ?President .
   ?nytPresident nytimes:topicPage ?TopicPage .
}
```

Assume that the following intermediate results have been computed as input for the last triple pattern

Block Input

"Barack Obama"
"George W. Bush"
…

Before (NLJ)

SELECT ?TopicPage WHERE { "Barack Obama" nytimes:topicPage ?TopicPage }
SELECT ?TopicPage WHERE { "George W. Bush" nytimes:topicPage ?TopicPage }
…

Now: Evaluation in a single remote request using a SPARQL UNION construct + local post processing (SPARQL 1.0)

Experimental Setting I:
Comparison with State-of-the-art Systems

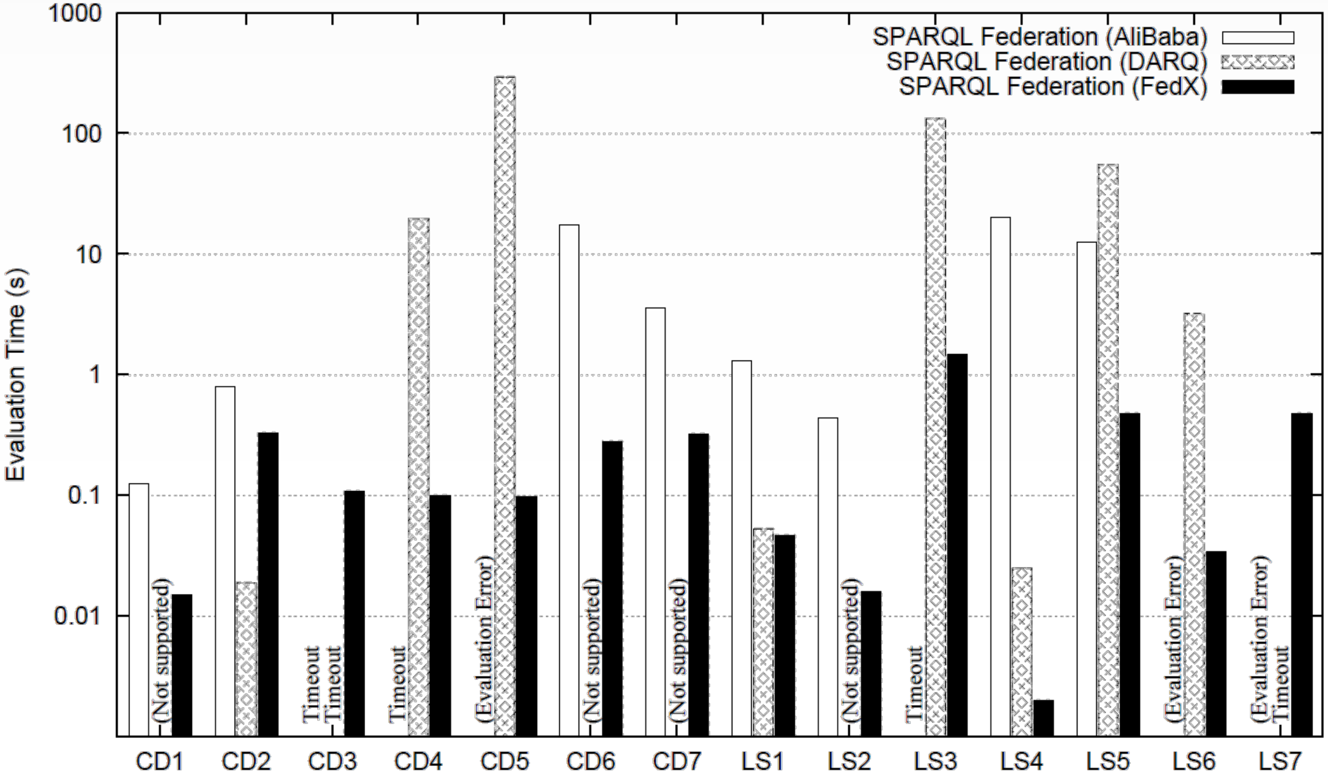
Evaluation based on FedBench benchmark suite

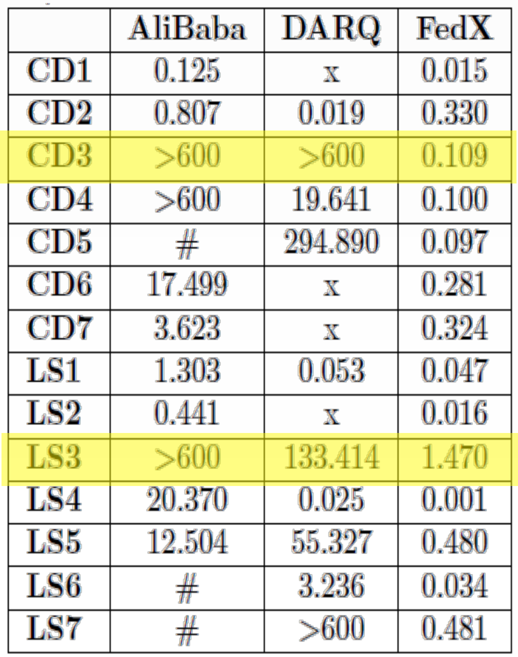
- 14 queries from the *Cross Domain* (CD) and *Life Science* (LS) collections
- Executed over real-world data from the Linked Open Data cloud
  - CD scenario: 6 data sources containing about 150M triples
  - LS scenario: 4 data sources containing about 50M triples
- Queries vary in complexity, size, structure, and sources involved
- Comparison with **AliBaba** and **DARQ** systems for federated query processing:

Benchmark environment

- Local copies of the SPARQL endpoints to ensure reproducibility and reliability of the service
- Set up on a HP Proliant 2GHz 4Core, 32GB RAM
- 20GB RAM for server (federation mediator)
- Use the infrastructure/execution framework provided by FedBench

Setting I: Evaluation Results

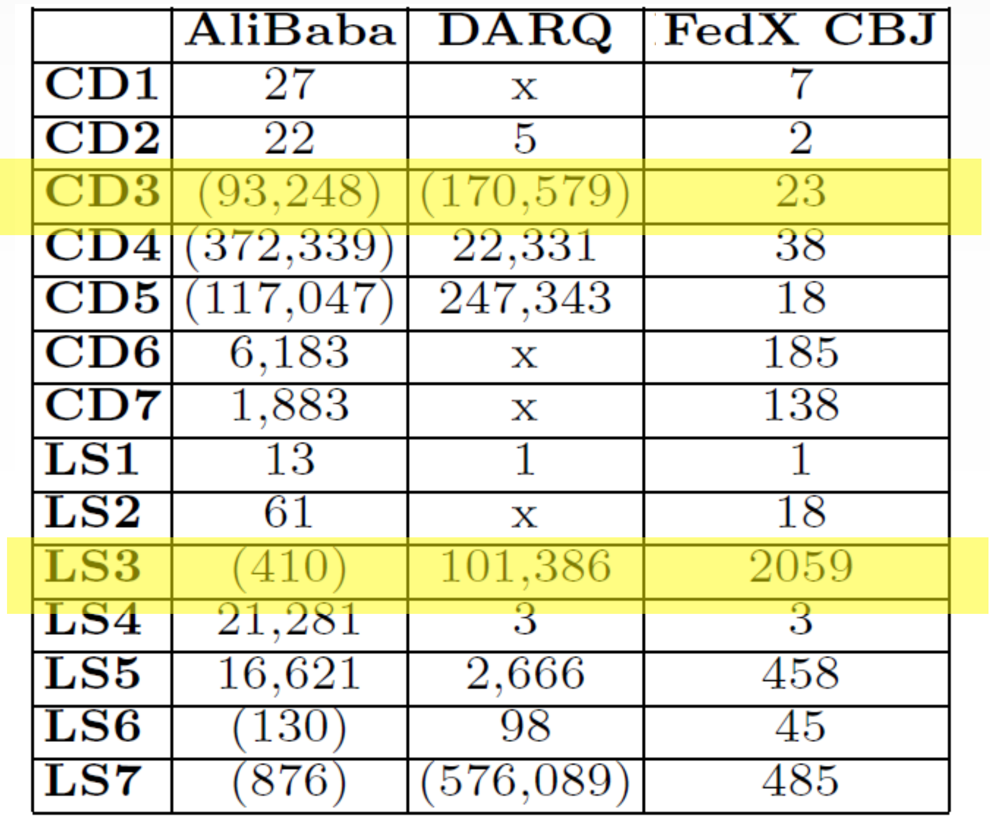




	AliBaba	DARQ	FedX
CD1	0.125	x	0.015
CD2	0.807	0.019	0.330
CD3	>600	>600	0.109
CD4	>600	19.641	0.100
CD5	#	294.890	0.097
CD6	17.499	x	0.281
CD7	3.623	x	0.324
LS1	1.303	0.053	0.047
LS2	0.441	x	0.016
LS3	>600	133.414	1.470
LS4	20.370	0.025	0.001
LS5	12.504	55.327	0.480
LS6	#	3.236	0.034
LS7	#	>600	0.481

Evaluation times of Cross Domain (CD) and Life Science (LS) queries

Setting I: Evaluation Results



	AliBaba	DARQ	FedX CBJ
CD1	27	x	7
CD2	22	5	2
CD3	(93,248)	(170,579)	23
CD4	(372,339)	22,331	38
CD5	(117,047)	247,343	18
CD6	6,183	x	185
CD7	1,883	x	138
LS1	13	1	1
LS2	61	x	18
LS3	(410)	101,386	2059
LS4	21,281	3	3
LS5	16,621	2,666	458
LS6	(130)	98	45
LS7	(876)	(576,089)	485

Number of requests sent to the endpoints

Runtimes
AliBaba: >600s
DARQ: >600s
FedX: 0.109s

Runtimes
AliBaba: >600s
DARQ: 133s
FedX: 1.4s

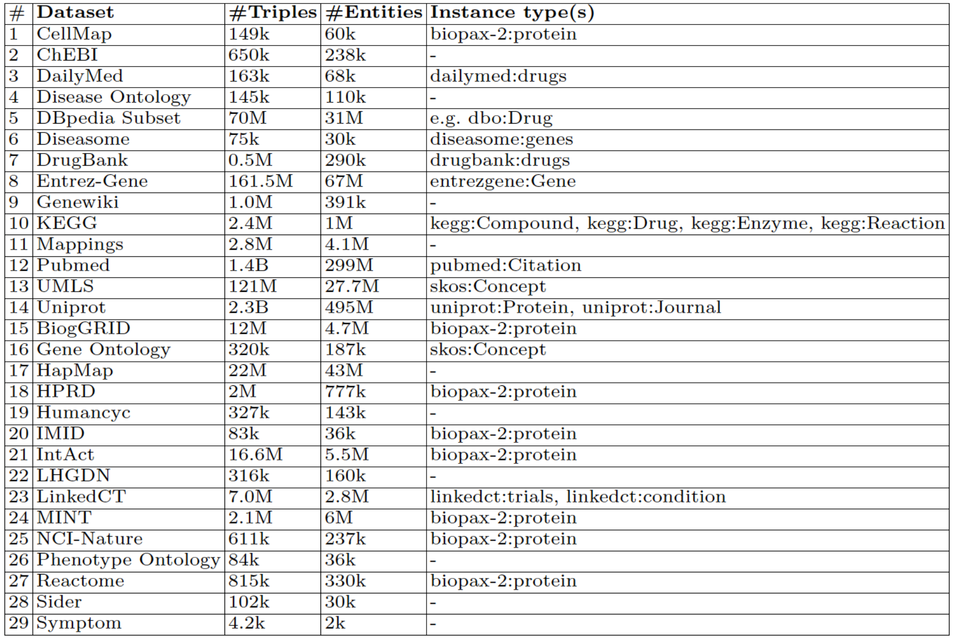
Experimental Setting II: Evaluating a Large-scale Federation

Evaluation based on Bio2RDF federation

- Queries from FedBench *Life Science* (LS) collections + *Linked Life Data* project (LLD)
- Executed over a large scale federation:
  - 29 SPARQL endpoints, varying from tens of thousands to billions of triples
  - Total sum of RDF triples: 4B+

Benchmark environment

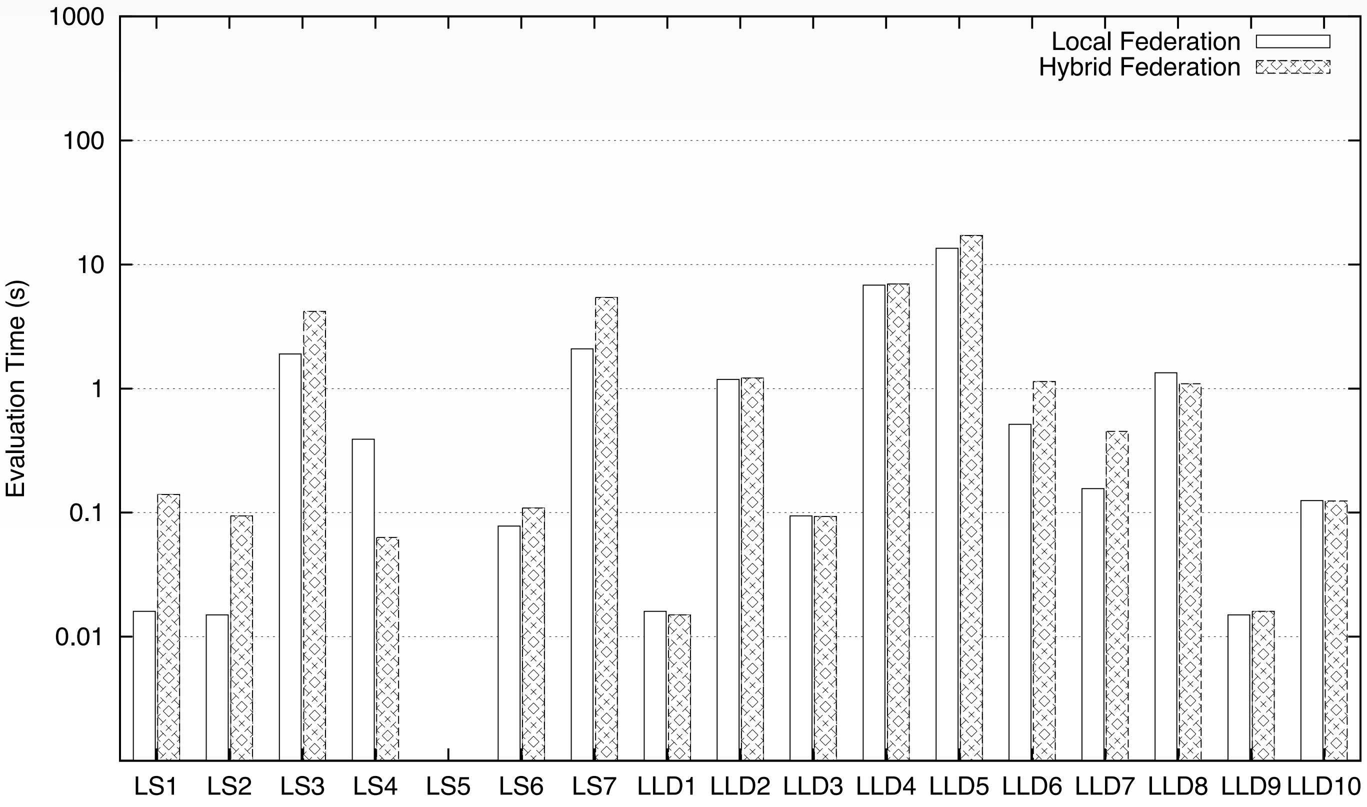
- Local Setting
  - SPARQL endpoints hosted on 2 machines in local data center
  - Total of 10 CPUs with 2-3GHz
  - Total of 84GB RAM, fast storage
- Hybrid Setting
  - Outsourcing of three SPARQL endpoints to AWS cloud (m2.2xlarge instance)
  - Remaining instances distributed on local infrastructure described before)



#	Dataset	#Triples	#Entities	Instance type(s)
1	CellMap	149k	60k	biopax-2:protein
2	ChEBI	650k	238k	-
3	DailyMed	163k	68k	dailymed:drugs
4	Disease Ontology	145k	110k	-
5	DBpedia Subset	70M	31M	e.g. dbo:Drug
6	Diseasome	75k	30k	diseasome:genes
7	DrugBank	0.5M	290k	drugbank:drugs
8	Entrez-Gene	161.5M	67M	entrezgene:Gene
9	Genewiki	1.0M	391k	-
10	KEGG	2.4M	1M	kegg:Compound, kegg:Drug, kegg:Enzyme, kegg:Reaction
11	Mappings	2.8M	4.1M	-
12	Pubmed	1.4B	299M	pubmed:Citation
13	UMLS	121M	27.7M	skos:Concept
14	Uniprot	2.3B	495M	uniprot:Protein, uniprot:Journal
15	BiogGRID	12M	4.7M	biopax-2:protein
16	Gene Ontology	320k	187k	skos:Concept
17	HapMap	22M	43M	-
18	HPRD	2M	777k	biopax-2:protein
19	Humancyc	327k	143k	-
20	IMID	83k	36k	biopax-2:protein
21	IntAct	16.6M	5.5M	biopax-2:protein
22	LHGDN	316k	160k	-
23	LinkedCT	7.0M	2.8M	linkedct:trials, linkedct:condition
24	MINT	2.1M	6M	biopax-2:protein
25	NCI-Nature	611k	237k	biopax-2:protein
26	Phenotype Ontology	84k	36k	-
27	Reactome	815k	330k	biopax-2:protein
28	Sider	102k	30k	-
29	Symptom	4.2k	2k	-

Table 1: Lifescience datasets used for federation scenario: 29 datasets/4B+ RDF triples

Setting II: Performance



Comparison of query evaluation time in local and hybrid federation

Key Findings

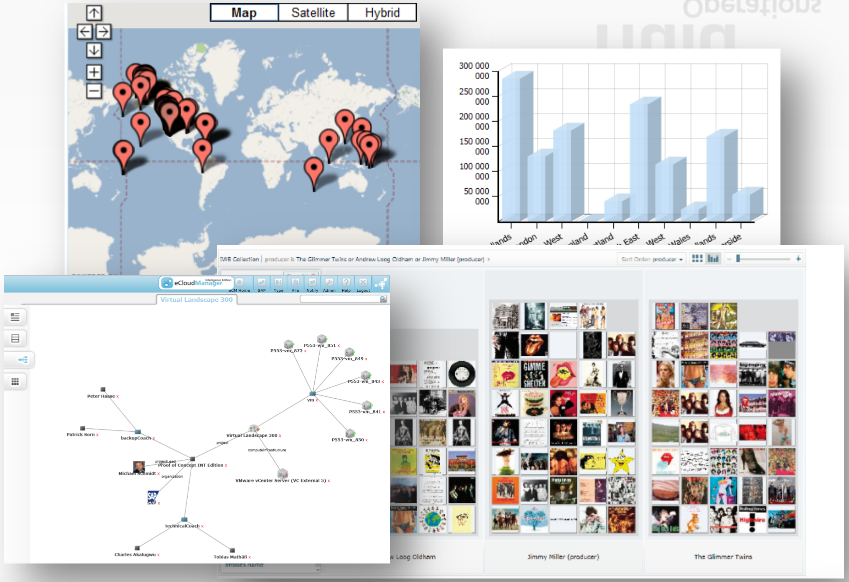
1. Both local and hybrid setting exhibit practical evaluation times

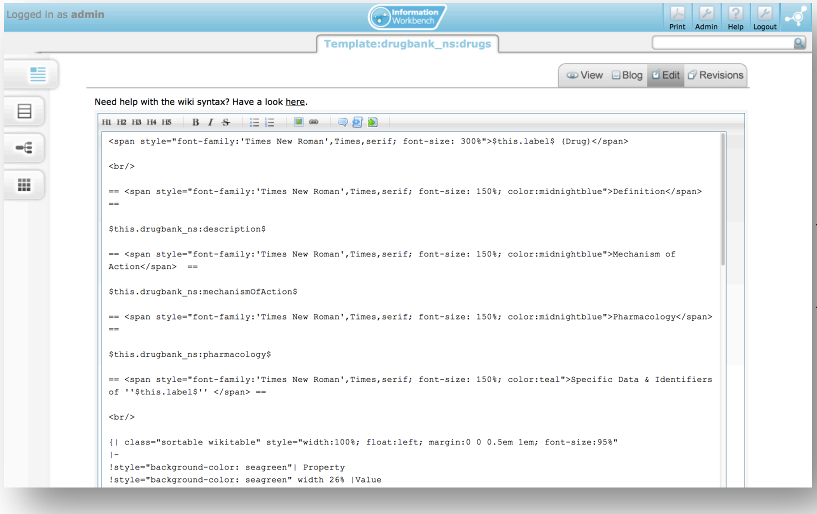
2. Typically low overhead in hybrid federation setup due to increased communication costs

3. For some queries, hybrid federation even outperforms local federation due to better load distribution

Enabling On Demand Use:
Self-service Linked Data Frontend

- Ontology-driven template mechanism
- **Declarative specification** of the UI based on available pool of widgets and declarative wiki-based syntax
- Widgets **have direct access to the DB**
- Ad hoc data exploration, visualization, analytics, dashboards, …

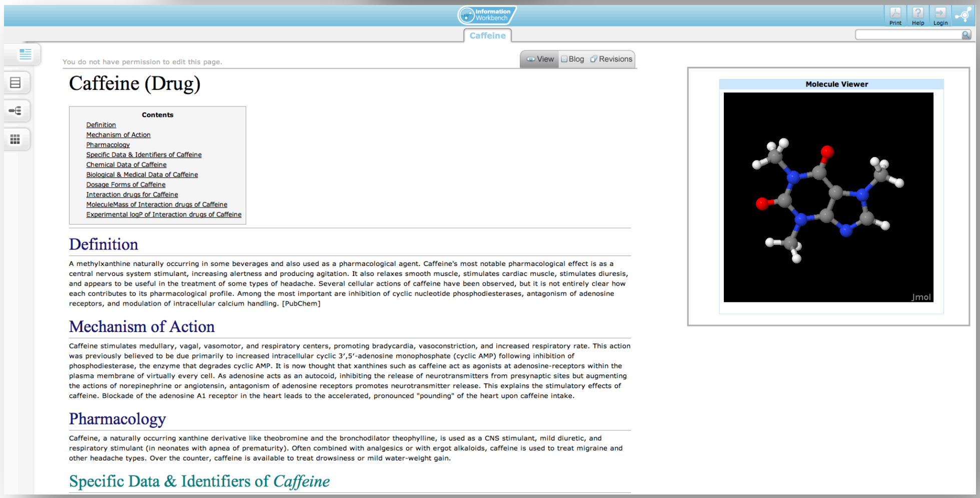




Need help with the wiki syntax? Have a look here.

```
<span style="font-family:'Times New Roman',Times,serif; font-size: 300%">$this.label$ (Drug)</span>

<br/>

== <span style="font-family:'Times New Roman',Times,serif; font-size: 150%; color:midnightblue">Definition</span>
==

$this.drugbank_ns:description$

== <span style="font-family:'Times New Roman',Times,serif; font-size: 150%; color:midnightblue">Mechanism of
Action</span>  ==

$this.drugbank_ns:mechanismOfAction$

== <span style="font-family:'Times New Roman',Times,serif; font-size: 150%; color:midnightblue">Pharmacology</span>
==

$this.drugbank_ns:pharmacology$

== <span style="font-family:'Times New Roman',Times,serif; font-size: 150%; color:teal">Specific Data & Identifiers
of ''$this.label$'' </span> ==

<br/>

{| class="sortable wikitable" style="width:100%; float:left; margin:0 0 0.5em 1em; font-size:95%"
|-
!style="background-color: seagreen"| Property
!style="background-color: seagreen" width 26% |Value
```

Wiki Page in Edit Mode …



You do not have permission to edit this page.

Caffeine (Drug)

Contents

Definition

A methylxanthine naturally occurring in some beverages and also used as a pharmacological agent. Caffeine's most notable pharmacological effect is as a central nervous system stimulant, increasing alertness and producing agitation. It also relaxes smooth muscle, stimulates cardiac muscle, stimulates diuresis, and appears to be useful in the treatment of some types of headache. Several cellular actions of caffeine have been observed, but it is not entirely clear how each contributes to its pharmacological profile. Among the most important are inhibition of cyclic nucleotide phosphodiesterases, antagonism of adenosine receptors, and modulation of intracellular calcium handling. [PubChem]

Mechanism of Action

Caffeine stimulates medullary, vagal, vasomotor, and respiratory centers, promoting bradycardia, vasoconstriction, and increased respiratory rate. This action was previously believed to be due primarily to increased intracellular cyclic 3′,5′-adenosine monophosphate (cyclic AMP) following inhibition of phosphodiesterase, the enzyme that degrades cyclic AMP. It is now thought that xanthines such as caffeine act as agonists at adenosine-receptors within the plasma membrane of virtually every cell. As adenosine acts as an autocoid, inhibiting the release of neurotransmitters from presynaptic sites but augmenting the actions of norepinephrine or angiotensin, antagonism of adenosine receptors promotes neurotransmitter release. This explains the stimulatory effects of caffeine. Blockade of the adenosine A1 receptor in the heart leads to the accelerated, pronounced "pounding" of the heart upon caffeine intake.

Pharmacology

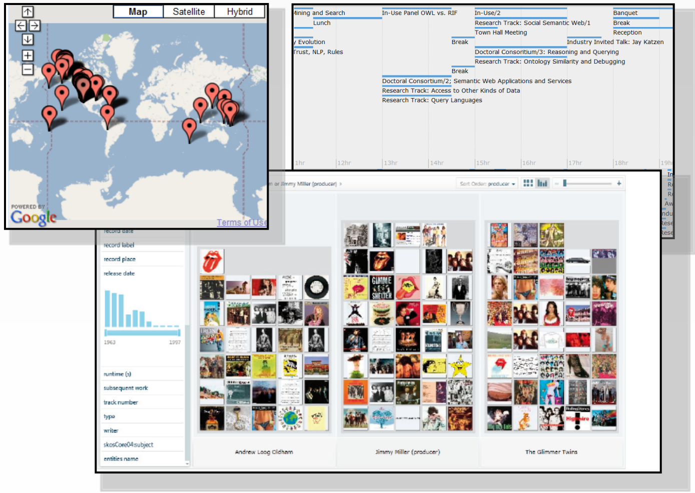
Caffeine, a naturally occurring xanthine derivative like theobromine and the bronchodilator theophylline, is used as a CNS stimulant, mild diuretic, and respiratory stimulant (in neonates with apnea of prematurity). Often combined with analgesics or with ergot alkaloids, caffeine is used to treat migraine and other headache types. Over the counter, caffeine is available to treat drowsiness or mild water-weight gain.

Specific Data & Identifiers of *Caffeine*

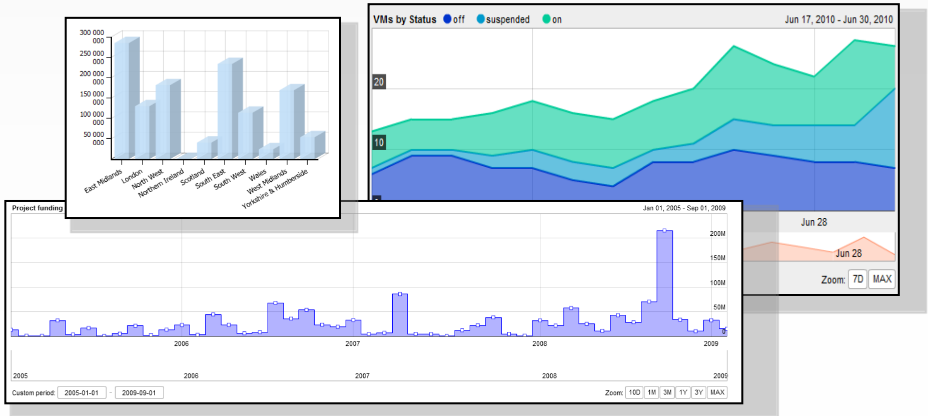
… and Displayed Result Page

Rich Pool of Available Widgets for Interacting with the Integrated Data

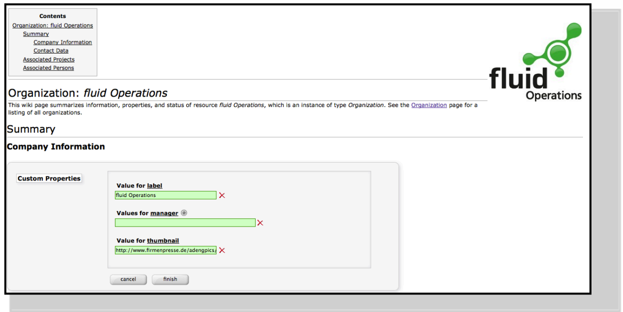
Visualization and Exploration



Analytics and Reporting



Authoring and Content Creation



Contents
Organization: fluid Operations
Summary
Company Information
Contact Data
Associated Projects
Associated Persons

Organization: *fluid Operations*

This wiki page summarizes information, properties, and status of resource *fluid Operations*, which is an instance of type *Organization*. See the Organization page for a listing of all organizations.

Summary

Company Information

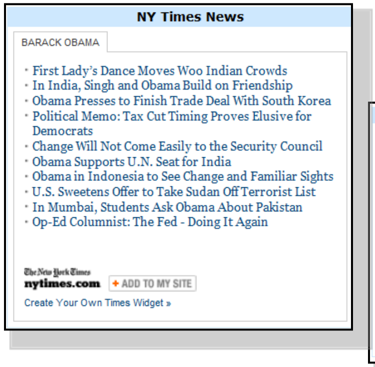
Custom Properties

Value for label: fluid Operations ✕

Values for manager ⊕: ✕

Value for thumbnail: http://www.firmenpresse.de/adengpics/ ✕

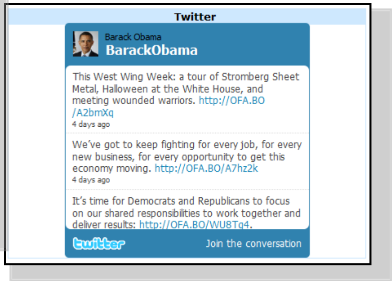
Mashups with Social Media



NY Times News

BARACK OBAMA

- First Lady's Dance Moves Woo Indian Crowds
- In India, Singh and Obama Build on Friendship
- Obama Presses to Finish Trade Deal With South Korea
- Political Memo: Tax Cut Timing Proves Elusive for Democrats
- Change Will Not Come Easily to the Security Council
- Obama Supports U.N. Seat for India
- Obama in Indonesia to See Change and Familiar Sights
- U.S. Sweetens Offer to Take Sudan Off Terrorist List
- In Mumbai, Students Ask Obama About Pakistan
- Op-Ed Columnist: The Fed - Doing It Again



Twitter

Barack Obama
BarackObama

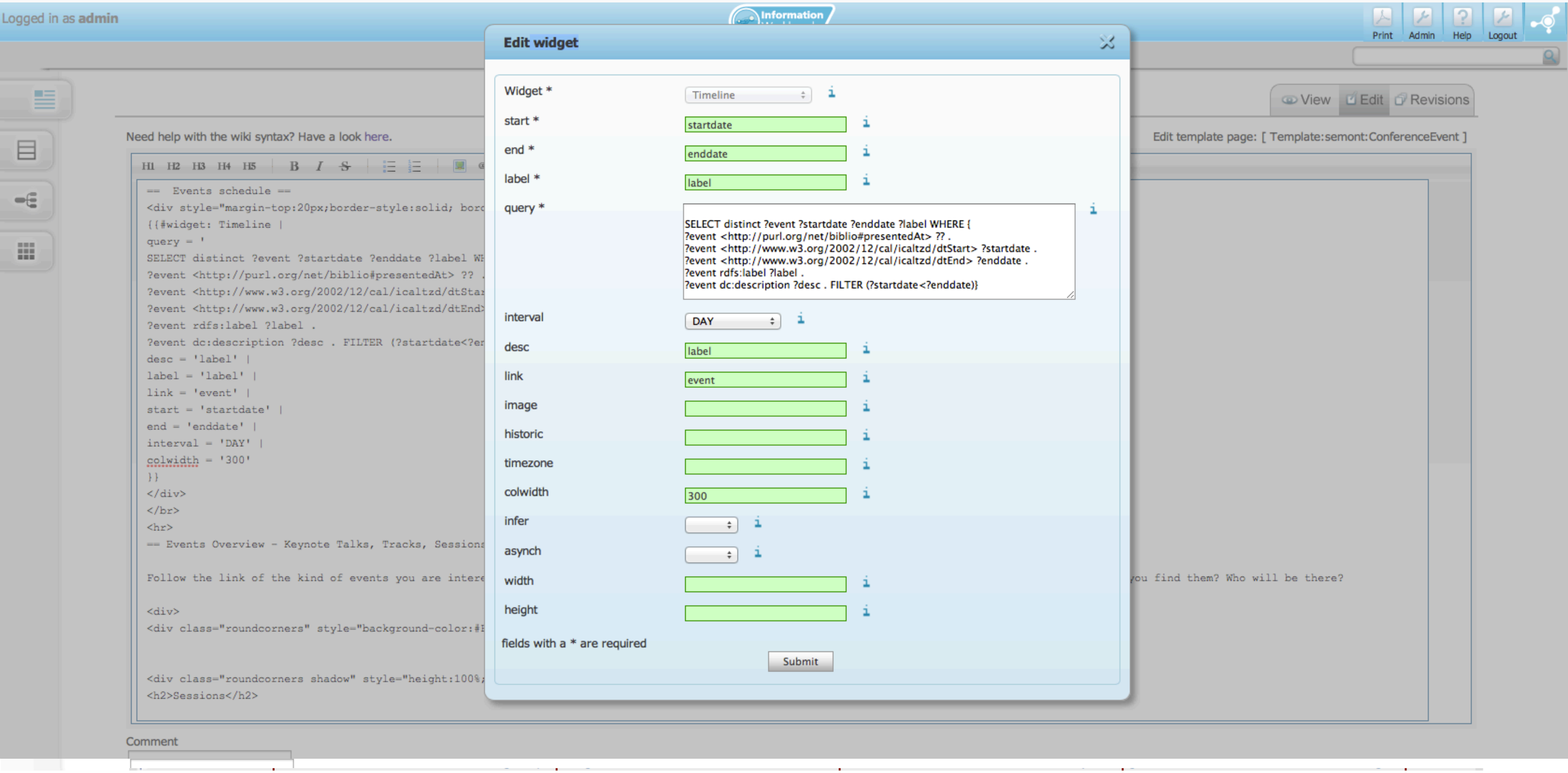
This West Wing Week: a tour of Stromberg Sheet Metal, Halloween at the White House, and meeting wounded warriors. http://OFA.BO/A2bmXq
4 days ago

We've got to keep fighting for every job, for every new business, for every opportunity to get this economy moving. http://OFA.BO/A7hz2k
4 days ago

It's time for Democrats and Republicans to focus on our shared responsibilities to work together and deliver results: http://OFA.BO/WU8Tq4.

Widgets can be integrated into Semantic Wiki pages using an intuitive, declarative syntax.

Widget-based Visualization and Query Construction



Need help with the wiki syntax? Have a look here.

Edit template page: [Template:semont:ConferenceEvent]

```
== Events schedule ==
<div style="margin-top:20px;border-style:solid; bord
{{#widget: Timeline |
query = '
SELECT distinct ?event ?startdate ?enddate ?label WH
?event <http://purl.org/net/biblio#presentedAt> ?? .
?event <http://www.w3.org/2002/12/cal/icaltzd/dtStar
?event <http://www.w3.org/2002/12/cal/icaltzd/dtEnd>
?event rdfs:label ?label .
?event dc:description ?desc . FILTER (?startdate<?en
desc = 'label' |
label = 'label' |
link = 'event' |
start = 'startdate' |
end = 'enddate' |
interval = 'DAY' |
colwidth = '300'
}}
</div>
</br>
<hr>
== Events Overview - Keynote Talks, Tracks, Sessions

Follow the link of the kind of events you are intere
ou find them? Who will be there?

<div>
<div class="roundcorners" style="background-color:#E


<div class="roundcorners shadow" style="height:100%;
<h2>Sessions</h2>
```

Edit widget

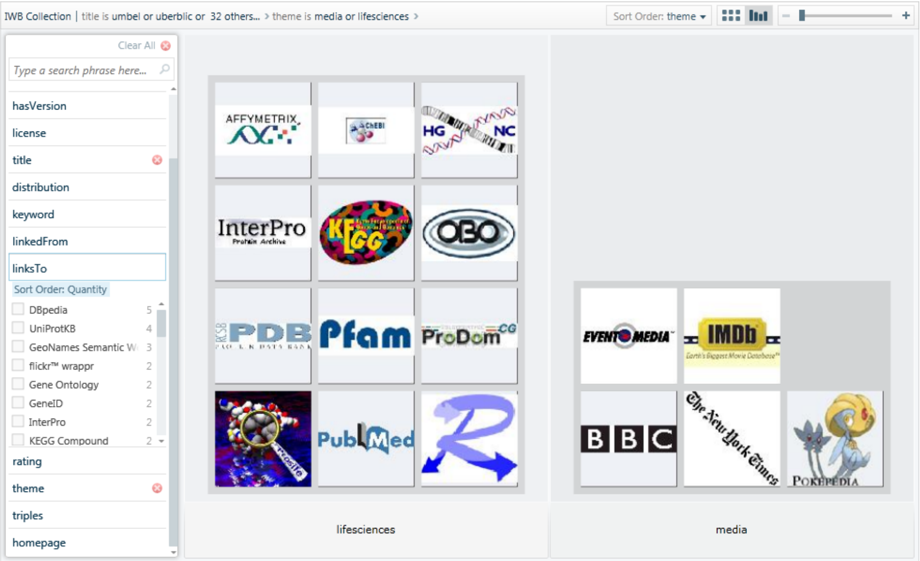
Widget *	Timeline
start *	startdate
end *	enddate
label *	label
query *	SELECT distinct ?event ?startdate ?enddate ?label WHERE { ?event <http://purl.org/net/biblio#presentedAt> ?? . ?event <http://www.w3.org/2002/12/cal/icaltzd/dtStart> ?startdate . ?event <http://www.w3.org/2002/12/cal/icaltzd/dtEnd> ?enddate . ?event rdfs:label ?label . ?event dc:description ?desc . FILTER (?startdate<?enddate)}
interval	DAY
desc	label
link	event
image	
historic	
timezone	
colwidth	300
infer	
asynch	
width	
height	

fields with a * are required

Submit

Enabling On Demand Data Discovery: *Metadata about Data Sets*

- Metadata about data sources essential for dynamic discovery
- Based on metadata vocabularies (VoID, DCAT)
- Access to data registered at global registries, e.g. ckan.org, data.gov, …
- Sort/filter data sets by topic, license, size and many more facets to identify relevant data
- Visually explore data sets



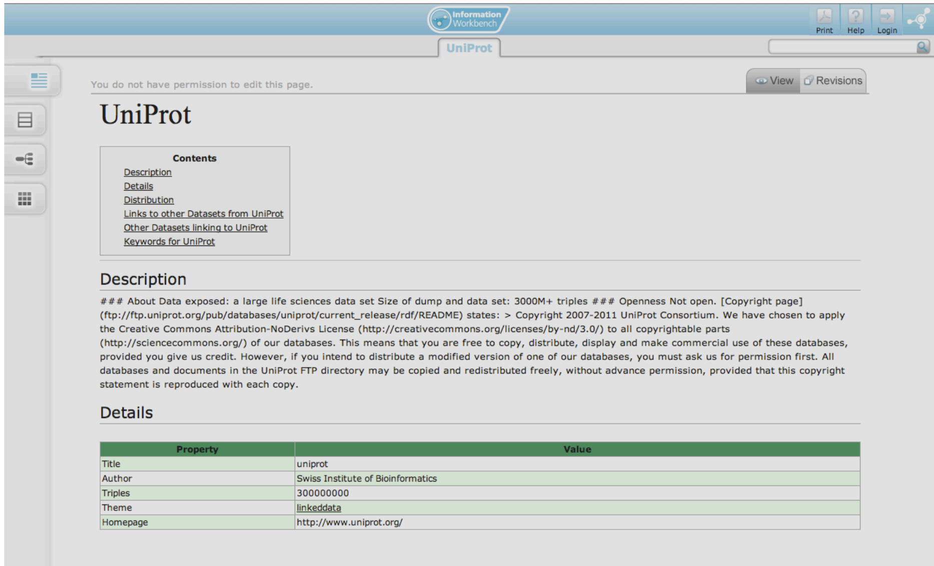
IWB Collection | title is umbel or uberblic or 32 others… › theme is media or lifesciences ›

Sort Order: theme

Type a search phrase here…

Sort Order: Quantity

□ DBpedia	5
□ UniProtKB	4
□ GeoNames Semantic W	3
□ flickr™ wrappr	2
□ Gene Ontology	2
□ GeneID	2
□ InterPro	2
□ KEGG Compound	2



You do not have permission to edit this page.

UniProt

Contents

Description
Details
Distribution
Links to other Datasets from UniProt
Other Datasets linking to UniProt
Keywords for UniProt

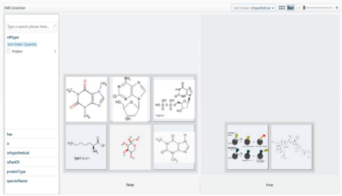
Description

About Data exposed: a large life sciences data set Size of dump and data set: 3000M+ triples ### Openness Not open. [Copyright page] (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/rdf/README) states: > Copyright 2007-2011 UniProt Consortium. We have chosen to apply the Creative Commons Attribution-NoDerivs License (http://creativecommons.org/licenses/by-nd/3.0/) to all copyrightable parts (http://sciencecommons.org/) of our databases. This means that you are free to copy, distribute, display and make commercial use of these databases, provided you give us credit. However, if you intend to distribute a modified version of one of our databases, you must ask us for permission first. All databases and documents in the UniProt FTP directory may be copied and redistributed freely, without advance permission, provided that this copyright statement is reproduced with each copy.

Details

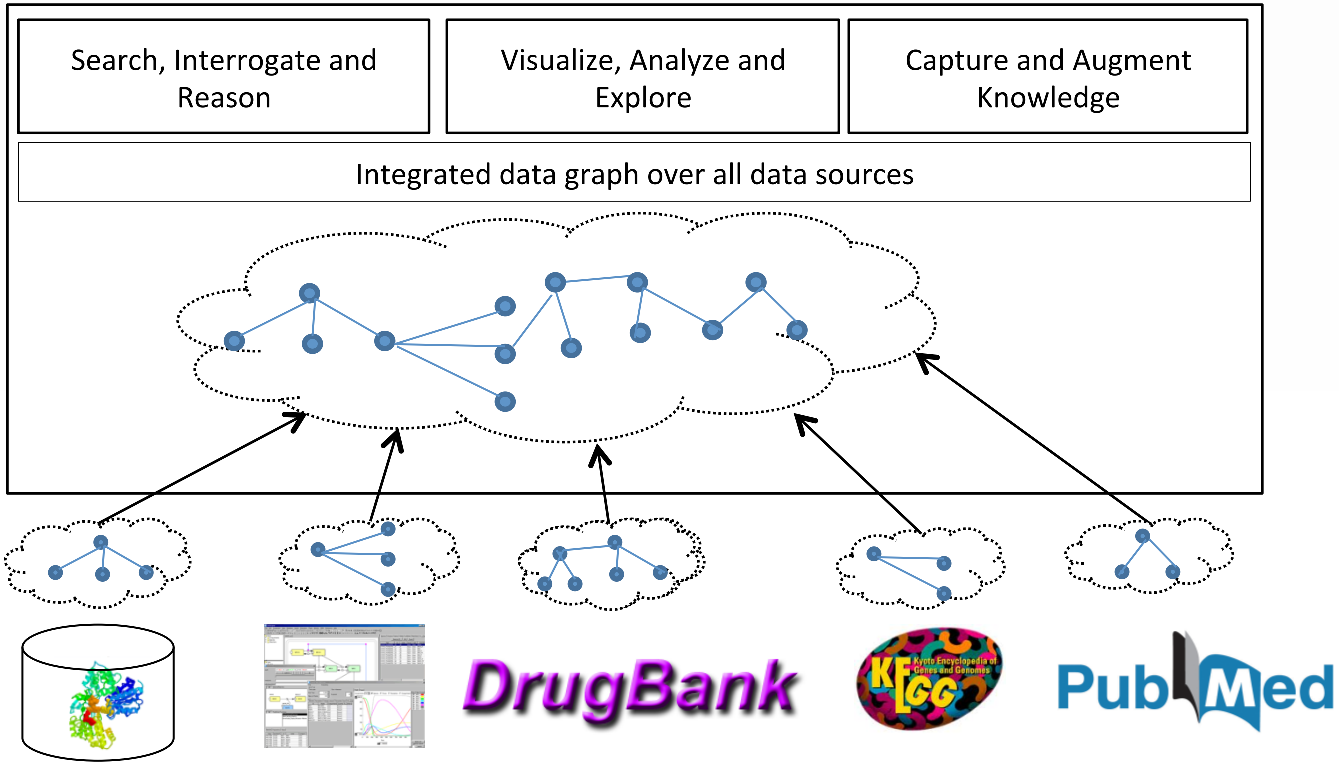
Property	Value
Title	uniprot
Author	Swiss Institute of Bioinformatics
Triples	300000000
Theme	linkeddata
Homepage	http://www.uniprot.org/

Example: Linked Data in Pharma



Caffeine (Drug)

Mechanism of Action


Search, Interrogate and Reason
Visualize, Analyze and Explore
Capture and Augment Knowledge
Integrated data graph over all data sources
DrugBank
KEGG
Kyoto Encyclopedia of Genes and Genomes
PubMed

Private Data Sources

Public Data Sources

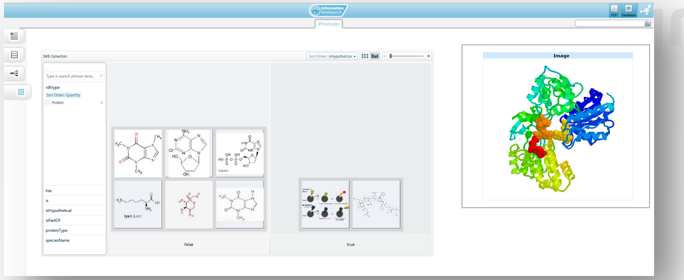
Main Use Cases

- Integrate data from company-internal data silos
- Augment company-internal data with Linked Open Data
- Collaborative knowledge management
- Support of internal processes (drug development)

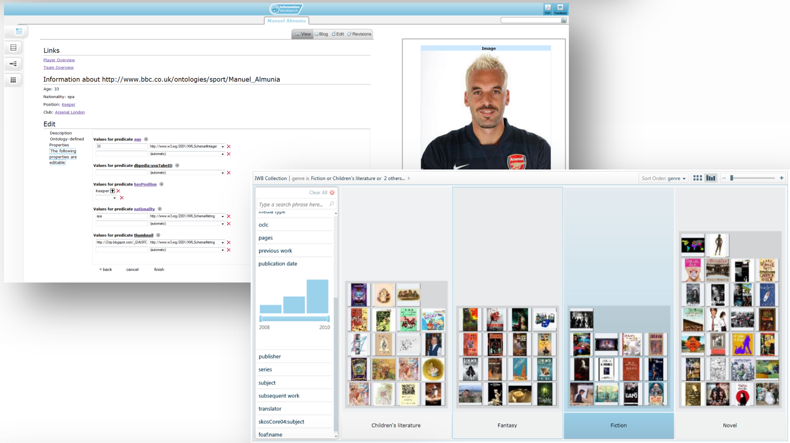
Information Workbench – Linked Data as a Service

Application Areas

Knowledge Management in the Life Sciences



Digital Libraries, Media and Content Management



Links

Player Overview

Team Overview

Information about http://www.bbc.co.uk/ontologies/sport/Manuel_Almunia

Age: 33

Nationality: spa

Position: Keeper

Club: Arsenal London

Edit

Description

Ontology-defined Properties

The following properties are editable

Values for predicate age

33 http://www.w3.org/2001/XMLSchema#integer

(automatic)

Values for predicate dbpedia:youTubeID

(automatic)

Values for predicate hasPosition

Keeper

Values for predicate nationality

spa http://www.w3.org/2001/XMLSchema#string

(automatic)

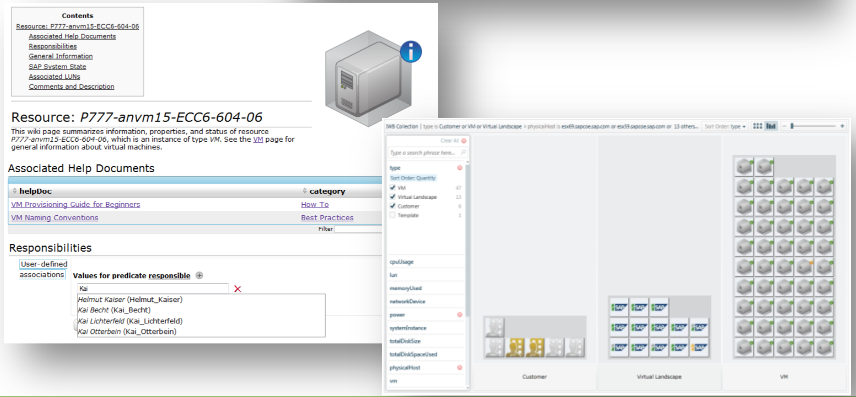
Values for predicate thumbnail

http://2.bp.blogspot.com/_QVb5FC http://www.w3.org/2001/XMLSchema#string

(automatic)

< back cancel finish

Intelligent Data Center Management



Contents

Resource: P777-anvm15-ECC6-604-06

- Associated Help Documents
- Responsibilities
- General Information
- SAP System State
- Associated LUNs
- Comments and Description

Resource: *P777-anvm15-ECC6-604-06*

This wiki page summarizes information, properties, and status of resource *P777-anvm15-ECC6-604-06*, which is an instance of type *VM*. See the VM page for general information about virtual machines.

Associated Help Documents

helpDoc	category
VM Provisioning Guide for Beginners	How To
VM Naming Conventions	Best Practices

Filter

Responsibilities

User-defined associations

Values for predicate responsible

Kai

Helmut Kaiser (Helmut_Kaiser)

Kai Becht (Kai_Becht)

Kai Lichterfeld (Kai_Lichterfeld)

Kai Otterbein (Kai_Otterbein)

Example: A Cloud Portal for Access to Open Data with the Information Workbench

Goal

- Collect meta data from global data markets (LOD Cloud, WorldBank, Eurostat, CKAN, …)
- Allow integrated search and ad hoc integration of data sources from different repositories
- Link data with private/internal data sources, if desired
- Support semi-automated linking between data sets
- Provide visualization, exploration, and analytics functionality on top of integrated data sources

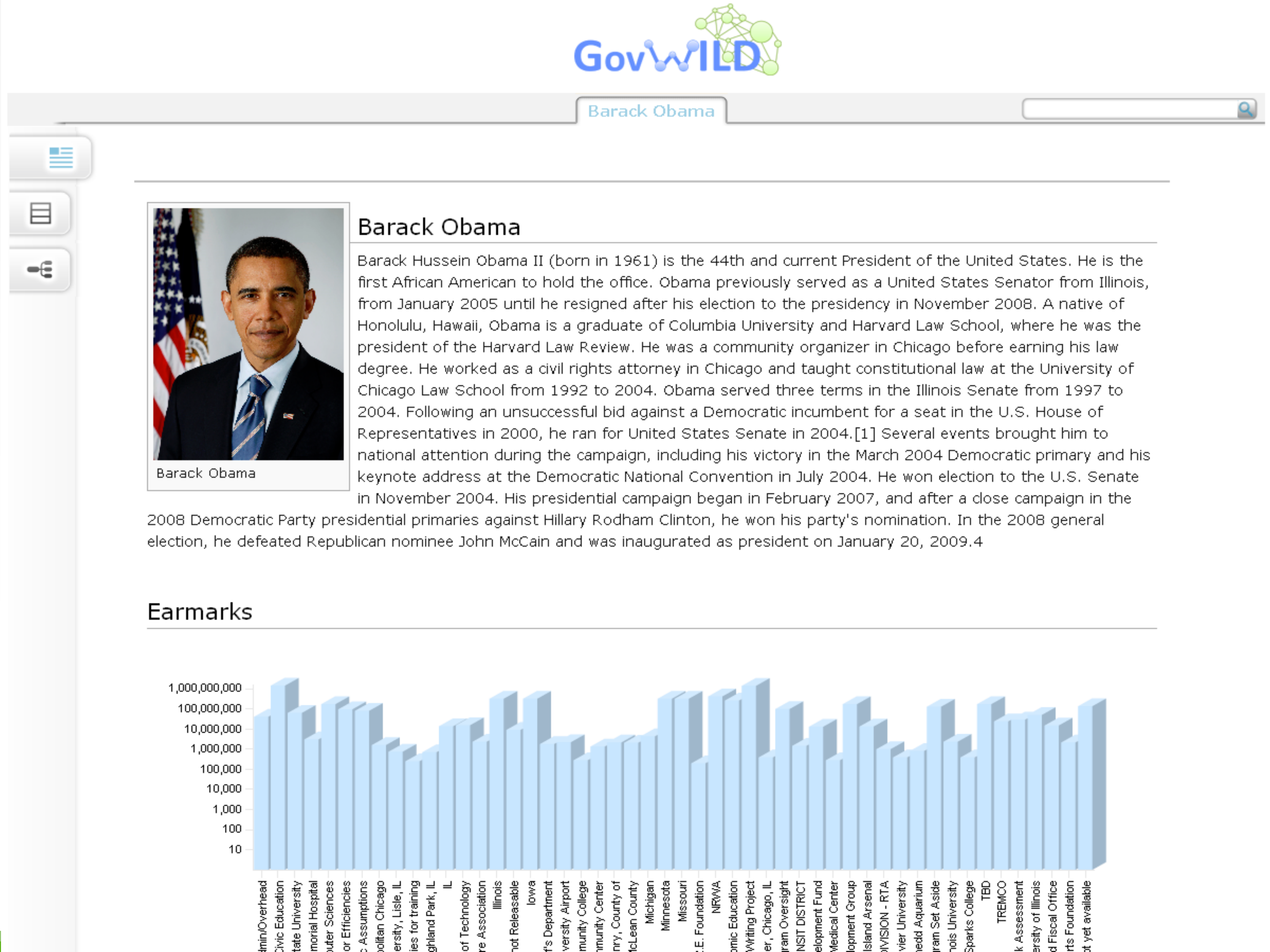
… using the *fluid Operations* Technology Stack

eCloudManager Product Suite

Information Workbench

Realization

- Project with the Hasso Plattner Institute (Potsdam, Germany)
- Create local repository containing data market metadata
- Use self-service technology to make services publicly available + Information Workbench for analytics



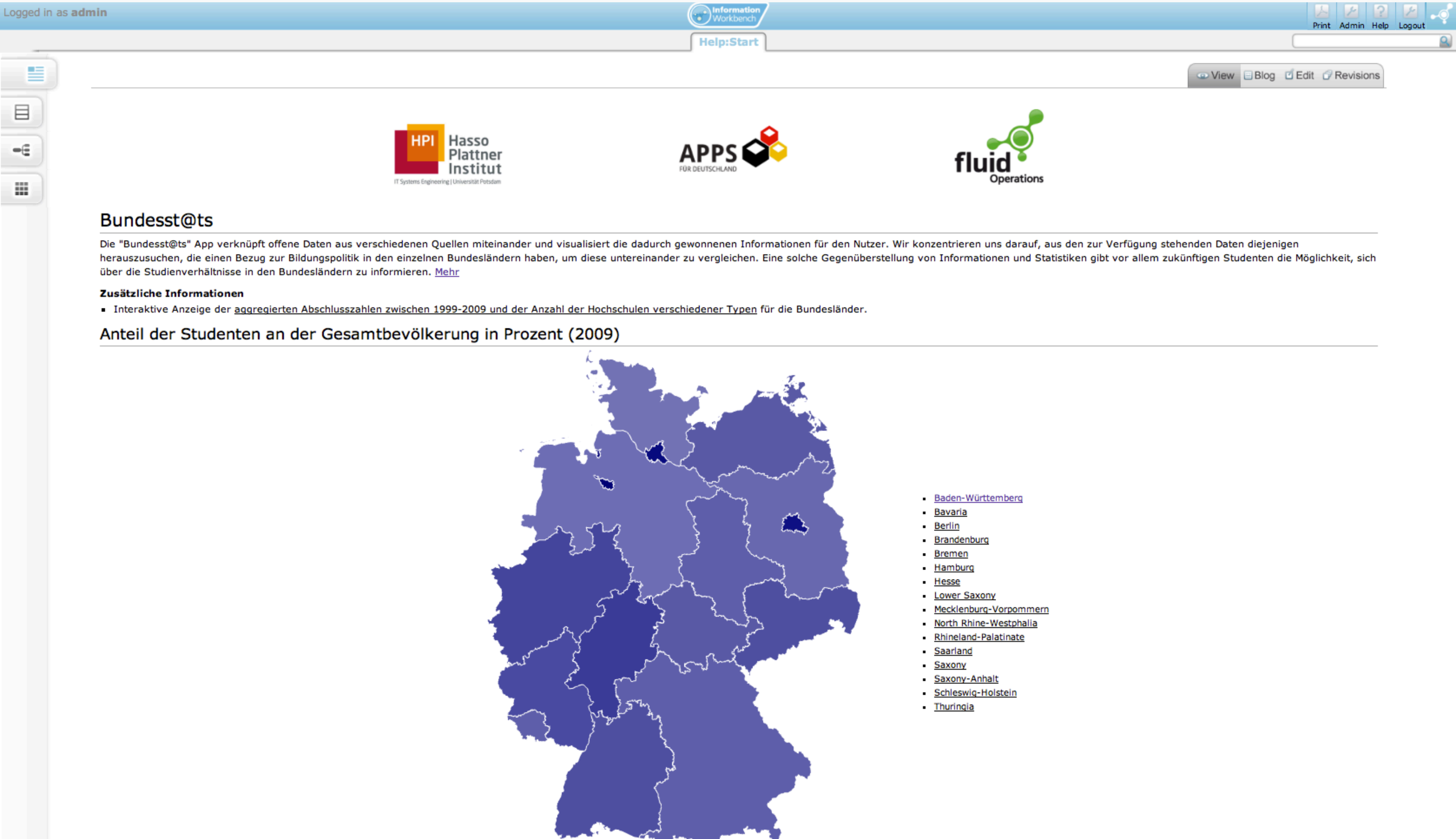
Barack Obama

Barack Obama

Barack Hussein Obama II (born in 1961) is the 44th and current President of the United States. He is the first African American to hold the office. Obama previously served as a United States Senator from Illinois, from January 2005 until he resigned after his election to the presidency in November 2008. A native of Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he was the president of the Harvard Law Review. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney in Chicago and taught constitutional law at the University of Chicago Law School from 1992 to 2004. Obama served three terms in the Illinois Senate from 1997 to 2004. Following an unsuccessful bid against a Democratic incumbent for a seat in the U.S. House of Representatives in 2000, he ran for United States Senate in 2004.[1] Several events brought him to national attention during the campaign, including his victory in the March 2004 Democratic primary and his keynote address at the Democratic National Convention in July 2004. He won election to the U.S. Senate in November 2004. His presidential campaign began in February 2007, and after a close campaign in the 2008 Democratic Party presidential primaries against Hillary Rodham Clinton, he won his party's nomination. In the 2008 general election, he defeated Republican nominee John McCain and was inaugurated as president on January 20, 2009.4

Earmarks

Example: Bundesst@ts



Bundesst@ts

Die "Bundesst@ts" App verknüpft offene Daten aus verschiedenen Quellen miteinander und visualisiert die dadurch gewonnenen Informationen für den Nutzer. Wir konzentrieren uns darauf, aus den zur Verfügung stehenden Daten diejenigen herauszusuchen, die einen Bezug zur Bildungspolitik in den einzelnen Bundesländern haben, um diese untereinander zu vergleichen. Eine solche Gegenüberstellung von Informationen und Statistiken gibt vor allem zukünftigen Studenten die Möglichkeit, sich über die Studienverhältnisse in den Bundesländern zu informieren. Mehr

Zusätzliche Informationen

- Interaktive Anzeige der aggregierten Abschlusszahlen zwischen 1999-2009 und der Anzahl der Hochschulen verschiedener Typen für die Bundesländer.

Anteil der Studenten an der Gesamtbevölkerung in Prozent (2009)

- Baden-Württemberg
- Bavaria
- Berlin
- Brandenburg
- Bremen
- Hamburg
- Hesse
- Lower Saxony
- Mecklenburg-Vorpommern
- North Rhine-Westphalia
- Rhineland-Palatinate
- Saarland
- Saxony
- Saxony-Anhalt
- Schleswig-Holstein
- Thuringia

Summary

- Big Data means more than volume and vertical scale

- Semantic Technologies for Big Data management
  - Linked Data as adequate data model
  - Ontologies as conceptual models to access big data
  - Integration of diverse, heterogeneous data sources

- Information Workbench: enabling on demand data access
  1. Discovery of relevant data sources
  2. Automated integration and interlinking of sources, and
  3. Interactive exploration and ad hoc analysis of data

- FedX: Federation over Linked Data sources
  1. Virtualized integration with horizontal and vertical scalability
  2. Experiments over Bio2RDF (5 billion triples in 29 databases)
  3. Outlook Optique: Possible combination with OBDA

CONTACT:

fluid Operations
Altrottstr. 31
Walldorf, Germany

Email: peter.haase@fluidops.com
website: www.fluidops.com
Tel.: +49 6227 3846-527