*Everything Self-Service:*
**Linked Data Applications with the Information Workbench**

Peter Haase, Michael Schmidt
fluid Operations AG

# Agenda

- **Linked Data in the Enterprise**

- **Challenges**

- **The Information Workbench Platform**

- **Solution Areas and Application Scenarios**

- **Conclusions**

# Increasing Popularity of Linked Data

- **Linked Data**
  - Set of standards, principles for publishing, sharing and interrelating structured knowledge
  - From data silos to a **Web of Data**
  - RDF as data model
  - SPARQL for querying
  - Ontologies to describe the semantics
- **Linked Open Data of Sep 2010**
  - 25 billion triples
  - 395 million RDF links
- **Covering various domains**
  - Media
  - Life Science
  - Geography
  - Publications
  - …
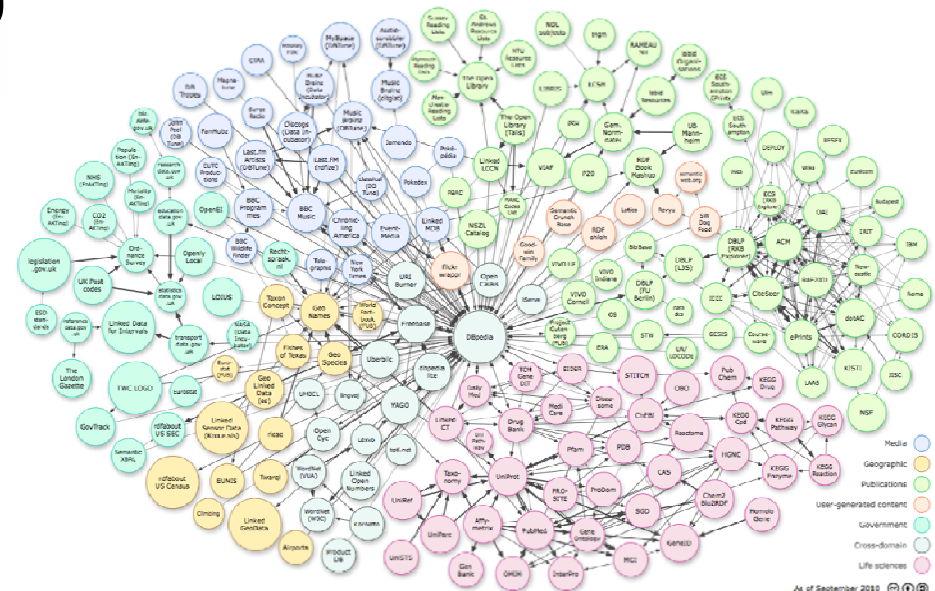
# Benefits of Linked Data in the Enterprise

- **Enterprise Data Integration**
  - Semantically integrate and interlink data scattered among different information systems
  - Linked Data standards as Lingua Franca
  - No global schema needed up-front, dynamic integration of new data sources

- **Openness and Accessibility of Enterprise Data**
  - Shared and agreed identifiers for data access inside the enterprise
  - Simplified publishing and sharing of data

- **Value Add by Linking to Linked Open Data**
  - Cross the chasm between enterprise-internal and public data
  - Enrichment and contextualization through interlinking

- **Innovative Applications through Semantic Technologies**
  - Leverage semantic technologies for improved search and presentation
  - Allow for addressing expressive information needs

# Example: Integrated Knowledge Management for the Life Sciences
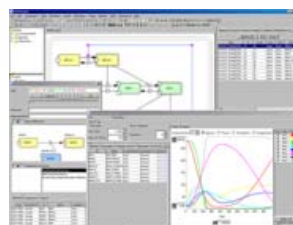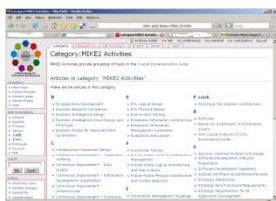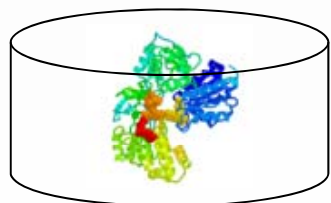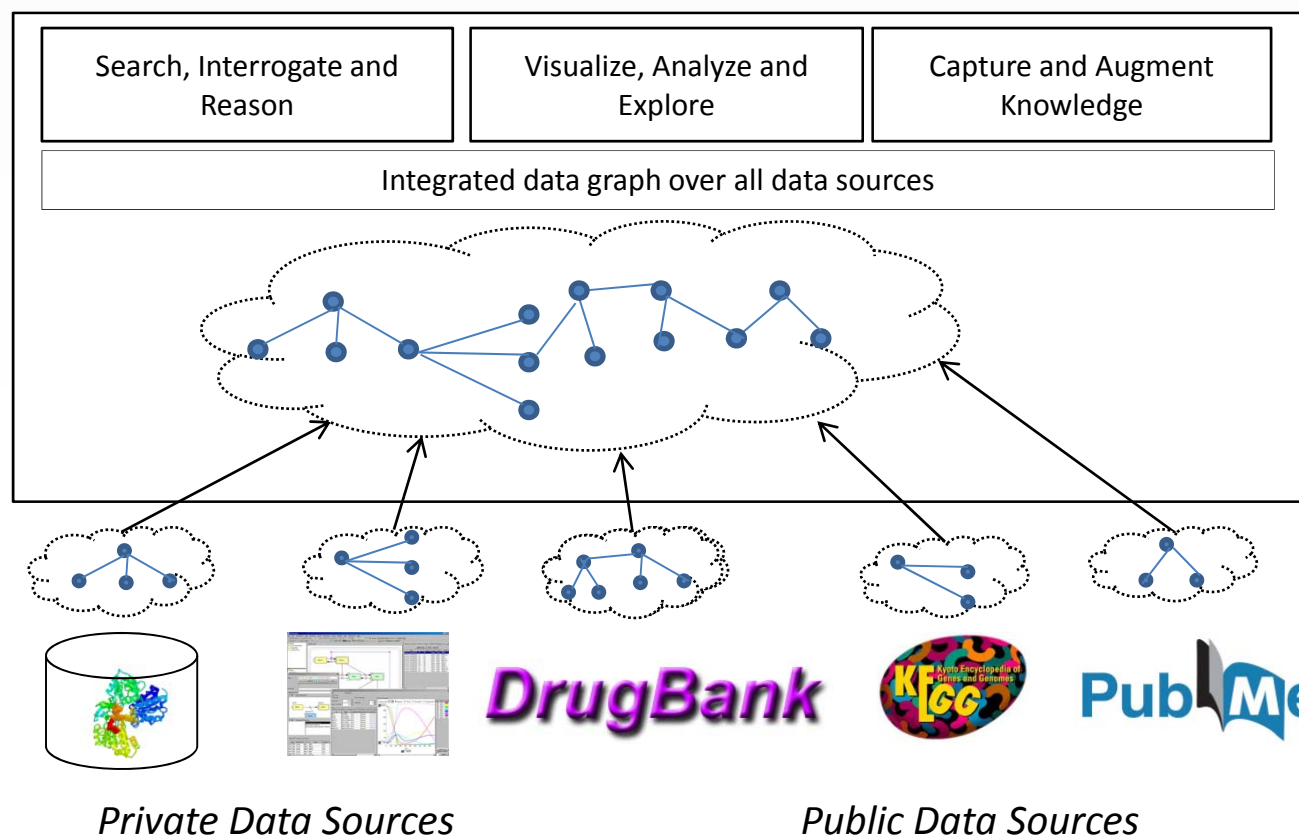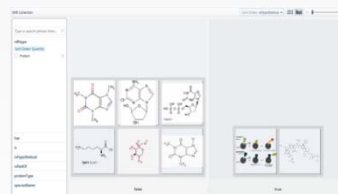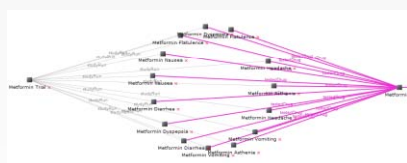


*Private Data Sources*

*Public Data Sources*

# Example: Integrated Knowledge Management for the Life Sciences



| Search, Interrogate and Reason | Visualize, Analyze and Explore | Capture and Augment Knowledge |
|---|---|---|

Integrated data graph over all data sources

*Private Data Sources*            *Public Data Sources*

# Challenges in Building Linked Data Applications

- **Discovery** and **automated integration** of relevant data sources

- **Heterogeneity** in various dimensions
  - Location, ownership of data (internal / external, open / closed)
  - Identifiers, structure, and vocabularies
  - Structured and unstructured data

- **Quality** of Linked Data
  - Various forms of imperfection (erroneous, incomplete, imprecise data)
  - Trustworthiness

- **End-user oriented** interfaces and interaction paradigms
  - Abstraction from technical details of Linked Data
  - Interfaces that operate on large amounts of data, flexible and dynamic schemas
  - Meaningful aggregation of the data
  - Support for expressive information needs, while retaining intuitive interfaces

# The Information Workbench

- **Platform** for Linked Data Application Development
  - Base functionality to build applications without any programming
  - SDK for easy extensions

- Covering the all aspects of **interacting with Linked Data**
  - Discovery of data sources
  - Integration of data sources
  - Visualization
  - Search and exploration
  - Advanced analytics
  - Collaborative authoring of data

- Targeted at
  - Linked Open Data
  - Linked Enterprise Data
  - Combinations thereof

# Information Workbench Architecture
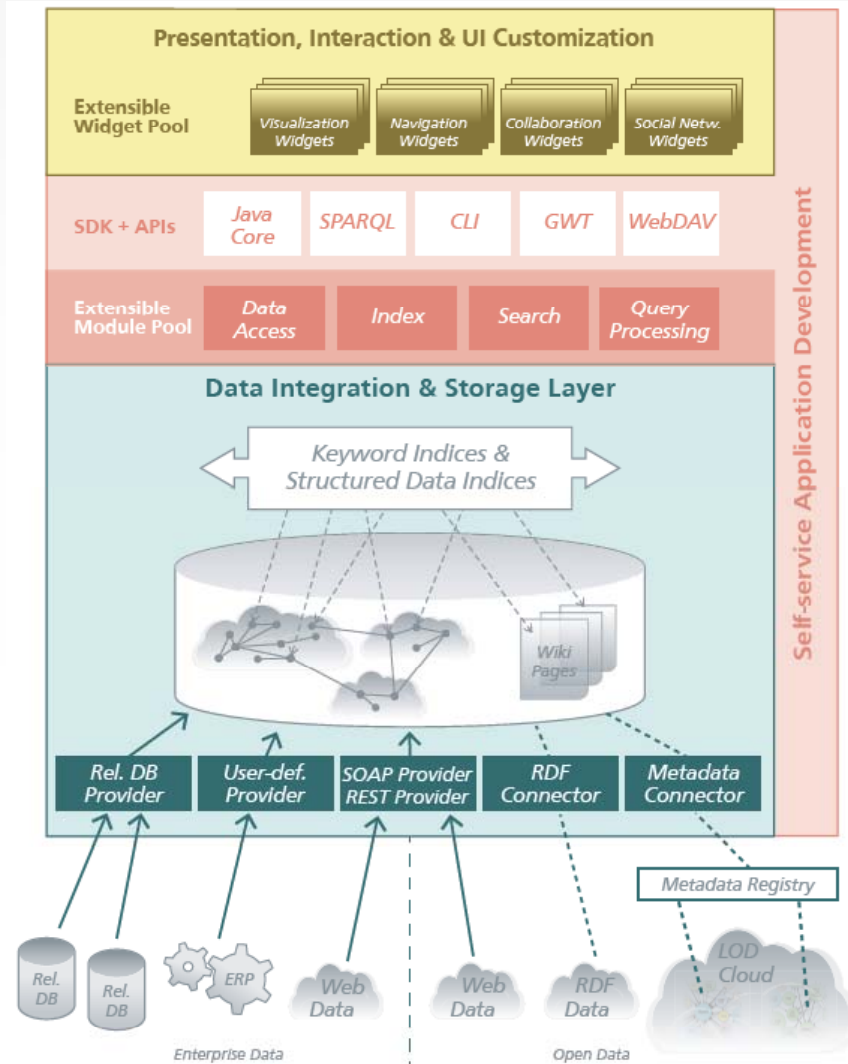


- **Extensible, widget-based UI**
  - Resource-centric presentation
  - Living UI, which exploits semantics of underlying data
  - Large collection of predefined widgets, easily extendable

- **Search and information access**
  - Coexistence of structured and unstructured data
  - Different search paradigms (keyword and faceted search, semantic query completion)

- **Data integration through providers**
  - Convert data from a data source into the RDF data format
  - Customizable, easily extensible
  - Use of public LOD registries (CKAN)
  - Support for Google Refine

# Self-Service Application Development

*Everything Self-service!*

1. ## Linked Open **Data Discovery**

   - Access to data registered at global registries, e.g. ckan.org, data.gov, …
   - Visually explore data sets and identify those relevant for your application
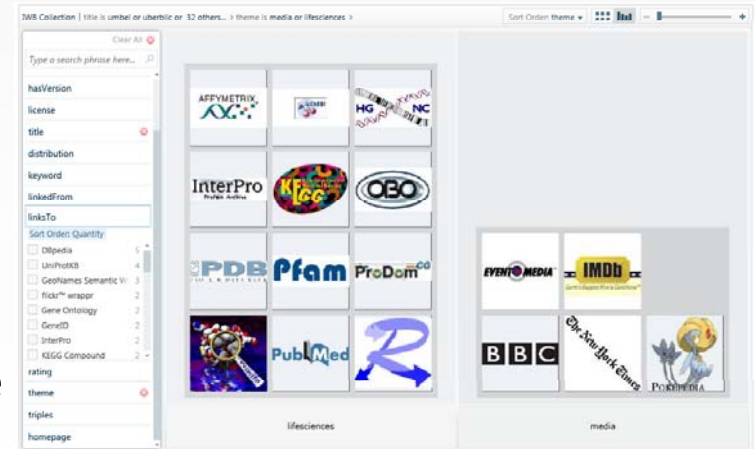
2. ## **Data Integration**

   - Integrate the data sets with the click of your mouse
   - Add providers for internal and external legacy data
   - Interlink different data sets

3. ## **Customization** of the Frontend

   - Declaratively customize the user interface using a selection of widgets
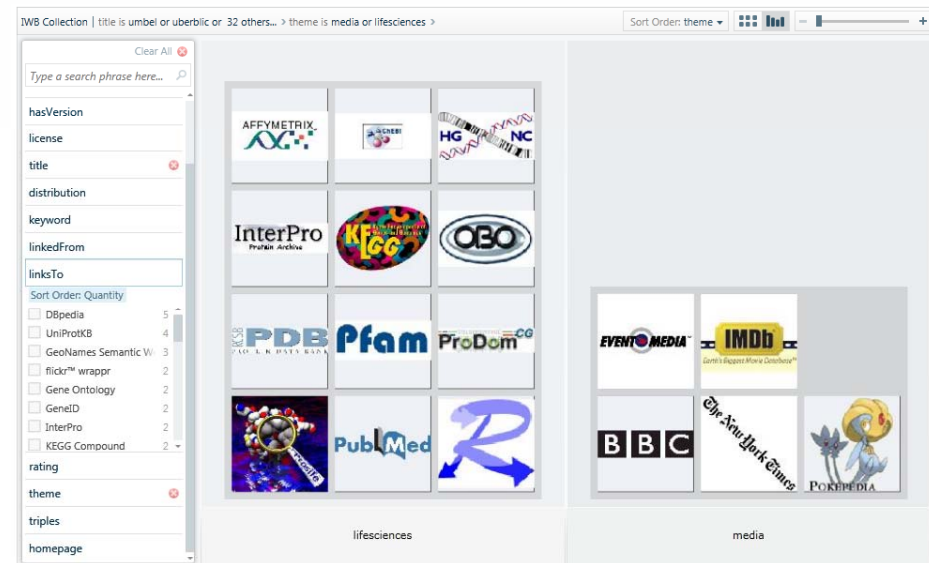
4. ## **Extending the Platform** with own Components

   - Use APIs and SDKs to implement your own widgets, providers, etc.

# Discovery of Data Sets

- Access to data registered at global registries, e.g. ckan.org, data.gov, …

- Based on metadata vocabularies (voID, DCAT)

- Sort/filter data sets by topic, license, size and many more facets to identify relevant data

- Visually explore data sets

# Data Integration

- Integrate relevant Linked Data sources with the click of a button

- Support for different distributions of the data
  - RDF dumps
  - SPARQL endpoints
  - URI lookup

- Add providers for internal and external legacy data sources

- Interlink different data sets
  - Exploit existing links and mappings between data sources
  - Use third party mappings
  - User-generated mappings
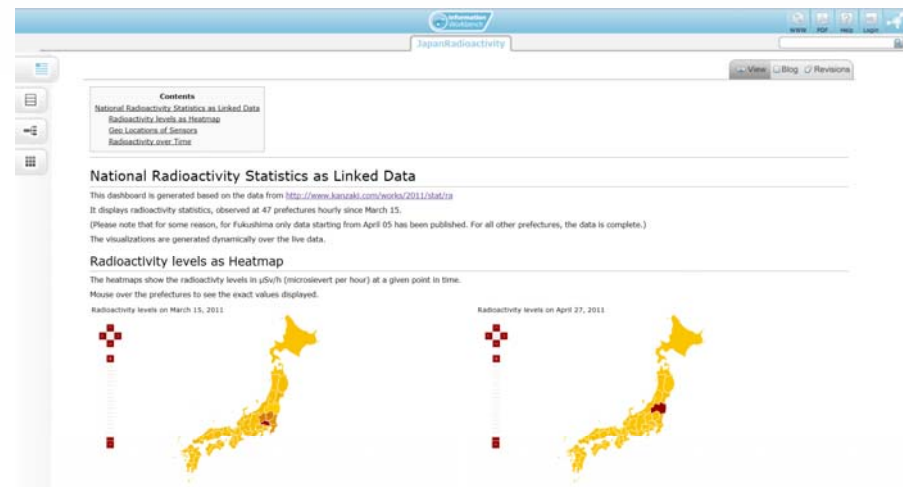  - Integrate components for automated link/mapping discovery

# Semantic Wiki + Widgets as Self-service Linked Data Frontend

- **Semantic Wiki** for **collaborative authoring** and **linking** of unstructured and structured semantic data
- **Declarative specification** of the UI based on available pool of widgets and simple wiki-based syntax
- Widgets have direct access to the database
- Embed dynamic data, visualizations, forms, etc.
- Type-based template mechanism
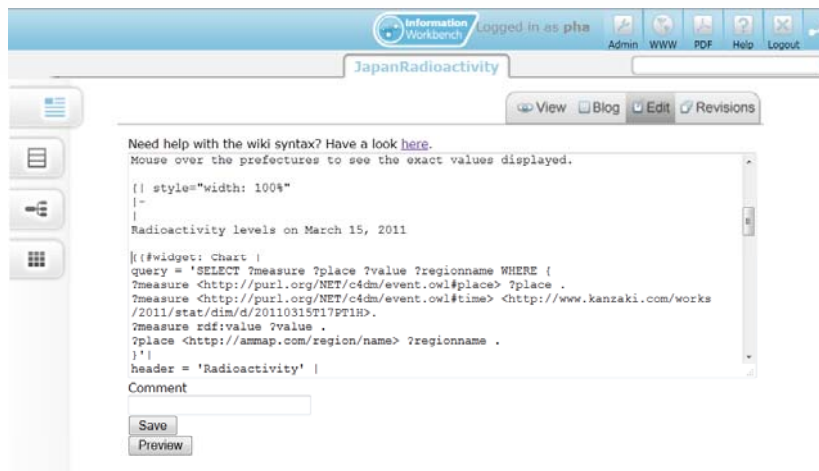


**Wiki Page in Edit Mode …**

**… and Displayed Result Page**

# Information Workbench Self Service Platform



- Create your own instance of the Information Workbench at the click of a button

- Select from pre-defined configurations for certain application domains

- Provision in private or public cloud environments (such as Amazon EC2)

# Live Demo: ISWC 2011 Explorer
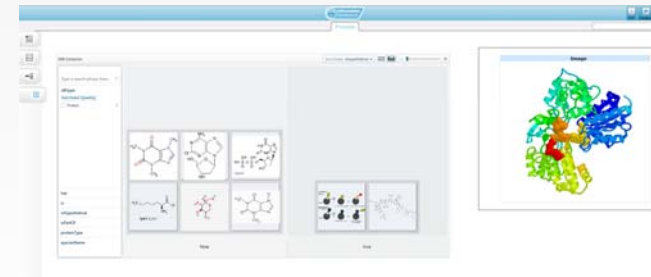
Events

Topics

People

Publications

Places

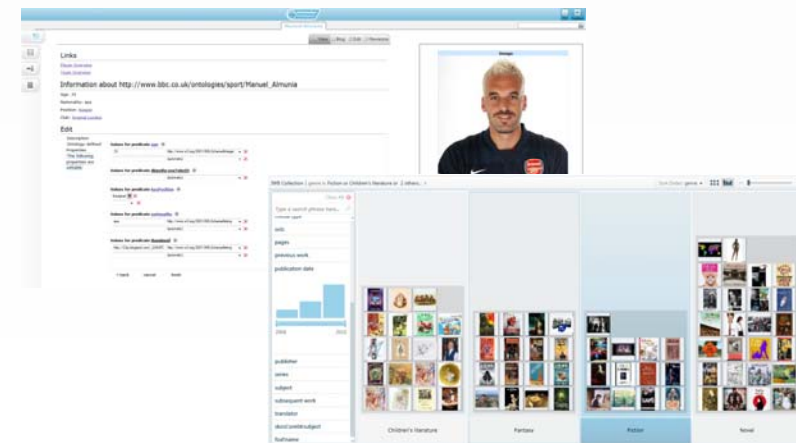http://data.semanticweb.org/conference/iswc/2011
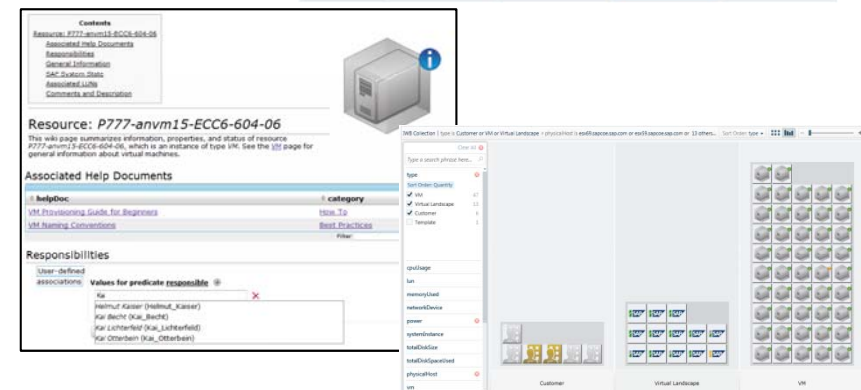
# Information Workbench Application Areas

**Knowledge Management in the Life Sciences**

**Digital Libraries, Media and Content Management**

**Intelligent Data Center Management**

# Intelligent Data Center Management

## BEFORE

- Business data not interlinked with technical data

- CXOs struggle to have an integrated view on the resources employed in the data center

## NOW

- Semantic, resource-centric view on data: link business data with data center resources and interrogate heterogeneous resources in a unified way

- User-defined dashboards, queries, historical data management for analytics and reporting purposes
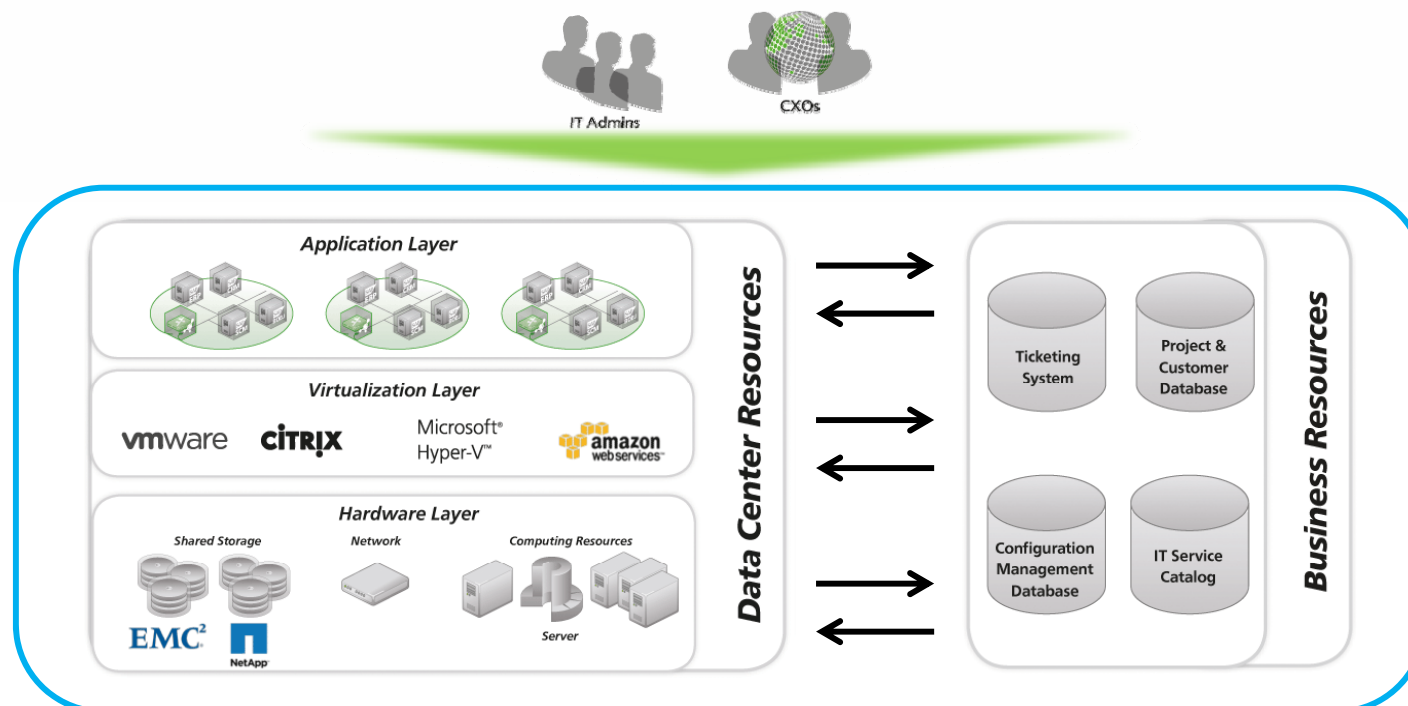
# Integrated View on the Data Center
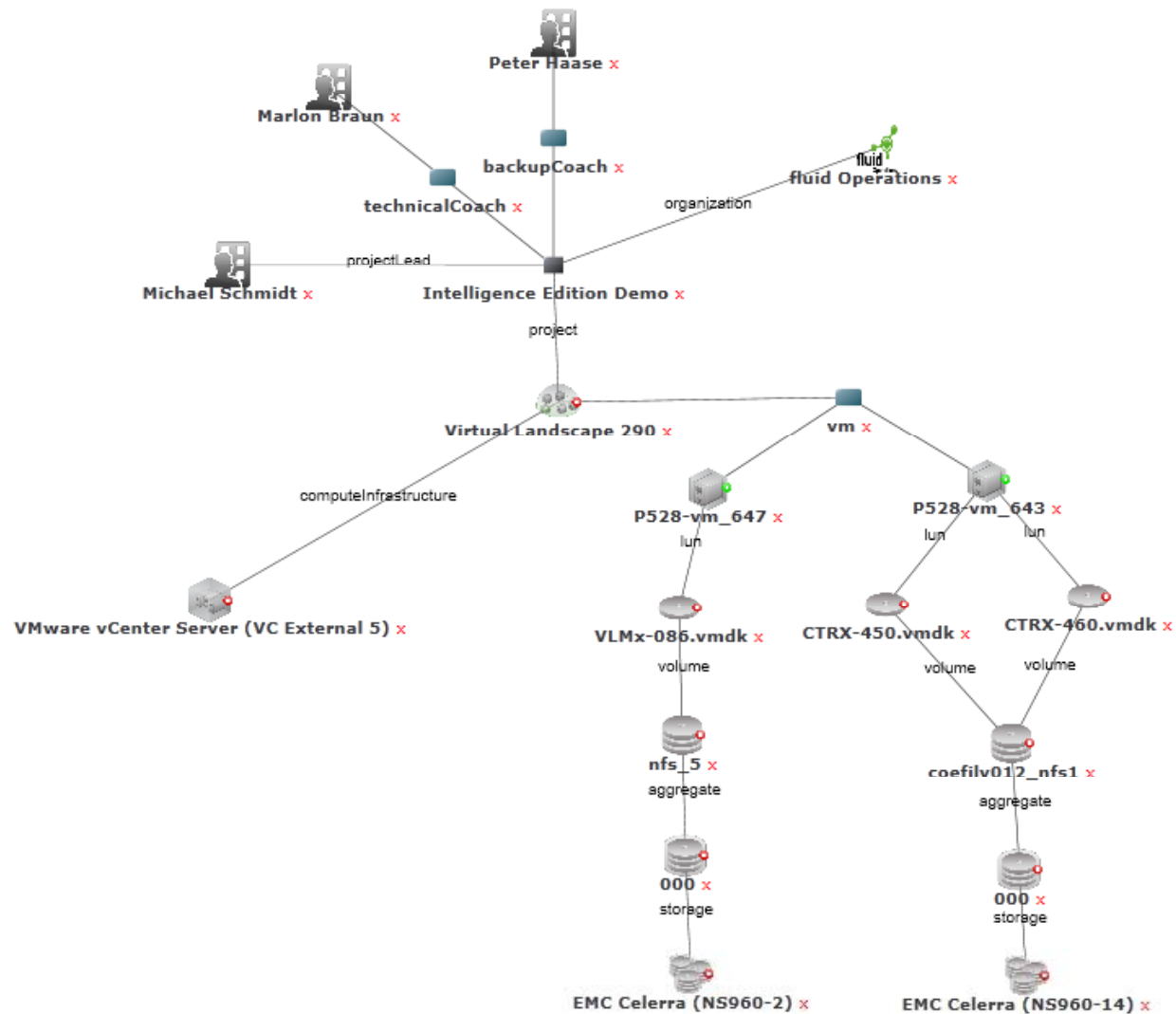
# Link Business Data To Data Center Resources

- E.g.: link your customers to their Virtual Landscapes using semantic annotations; visually explore the connections between the business information and the data center resources on-demand
- **Use Case: Root Cause Analysis and Error Handling**
  - ✓ Identify which customer's systems and applications are affected when an error on the storage level occurs
  - ✓ Determine where errors on the application level are coming from
  - ✓ Relate events to each other
  - ✓ Document and compare solutions for events
  - ✓ → allows fast reaction for error handling and ensures SLA enforcement



**Semantic Link in Wiki Page**
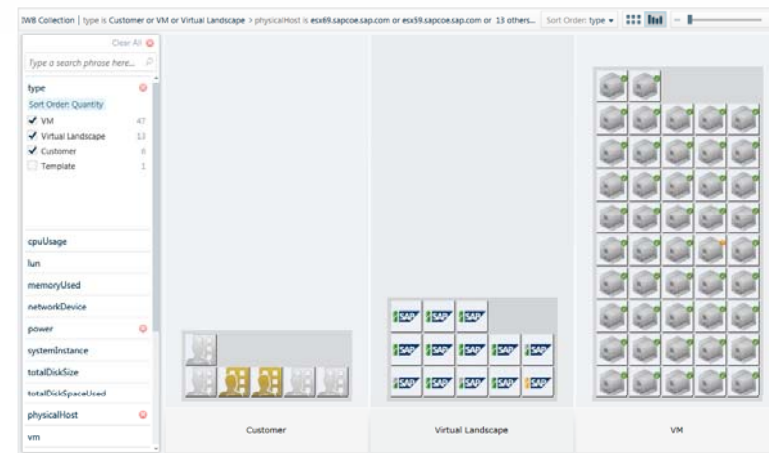
**Visual Data Exploration**

# Share Knowledge Within Your Company

**Collaborative Acquisition and Augmentation of Knowledge with Semantic Wiki Technology**

- **Use Case: Technical Documentation and Responsibility Management**
  - ✓ Use Wiki to collaboratively maintain technical documentation and best practices
  - ✓ Interlink hardware resources with documentation in a central place
  - ✓ Assign responsibilities directly to technical resource
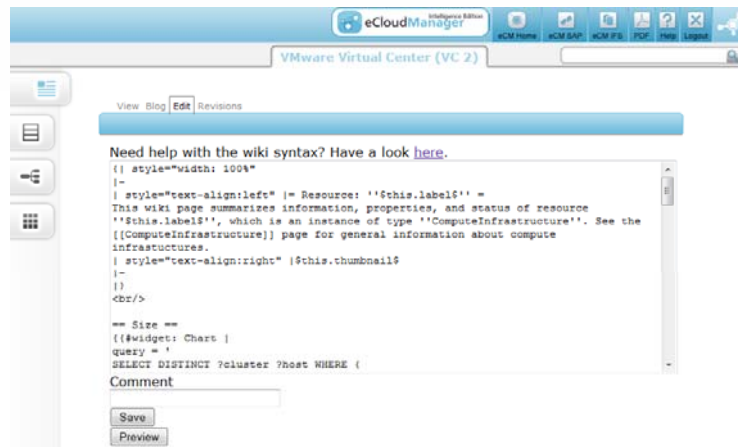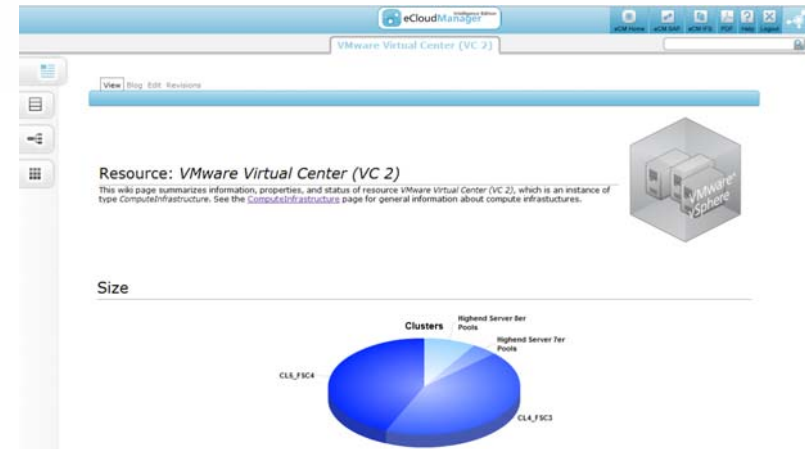


**Wiki Page in Edit Mode …**                                    **… and Displayed Result Page**

# Analytics and Reporting

**Embed dynamic, user-defined charts directly into Semantic Wiki pages**

- E.g.: create tabular summaries of your data and existing connections; specify user-defined charts and dashboards; generate reports based on historical data

**Use case**: **Performance Monitoring and Capacity Planning**

- Monitor the performance & usage of your infrastructure over time using historical data
- Forecast when new infrastructure resources need to be ordered
- Analysis of the impact of new hardware options on utilization rates

**Use case: Cost and Demand Forecasting**

- Keep track of infrastructure costs for each customer / project
- Determine what infrastructure resources will be needed when and for which project
- Compare various infrastructure options in terms of cost
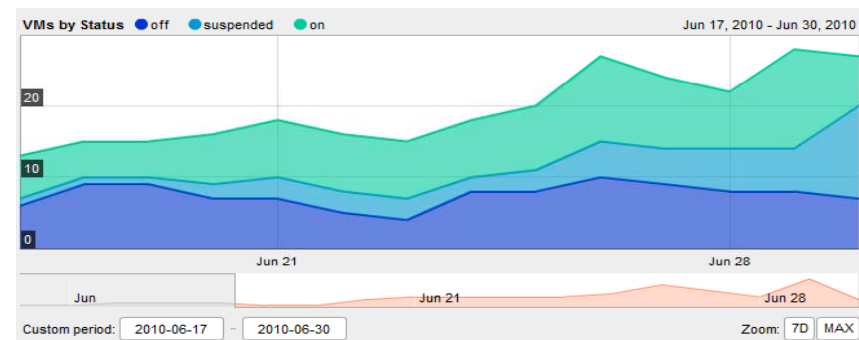
**Example: Employed VMs over Time grouped by Power Status**

# Summary

- Enormous **potential** for **Linked Data in the enterprise**

- Information Workbench as **platform** for implementing **Linked Data solutions**

- Addressing all aspects of **interacting with Linked Data**
  - Data Integration
  - Collaborative Knowledge Management and Authoring
  - Intelligent Data Access and Analytics

- "**Self-service**"-style application development
  - Linked Open Data discovery
  - Automated data integration
  - Customization of the frontend
  - Extensible via SDK

Visit our exhibitor booth
or arrange a private live demo!
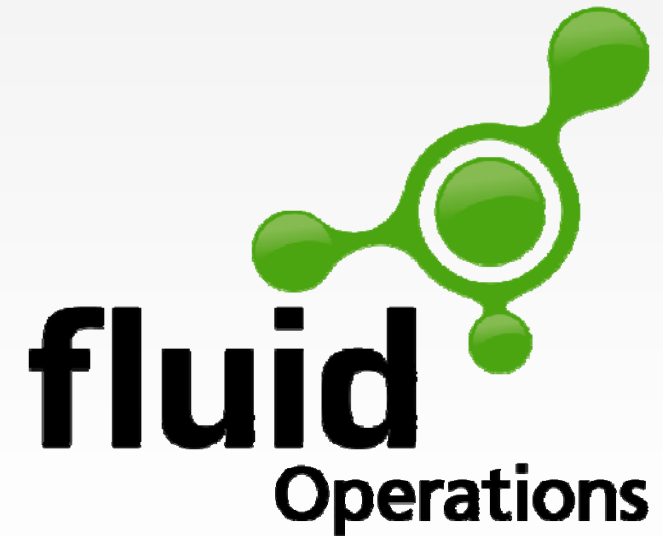
## CONTACT:

**fluid Operations**
**Altrottstr. 31**
**Walldorf, Germany**

**Email: sales@fluidOps.com**
**website: www.fluidOps.com**
**Tel.: +49 6227 3846-527**