

# Using the Semantic Web



German Research  
Center for Artificial  
Intelligence GmbH



Information Society  
Technologies



Leo Sauermann

[leo.sauermann@dfki.de](mailto:leo.sauermann@dfki.de)

<http://www.dfki.de/~sauermann/>



# *God morgen!*

# Leo Sauermann

- Researcher on Semantic Desktop since 2002, Vienna
- At **DFKI** since 2004
- Currently at the **Nepomuk EU** project
- **Scientist**: workshops, lectures, students, diploma theses
- **Community**: open source, developer, blogger, SWEO W3C
- **Training**: workshops, consulting, talks

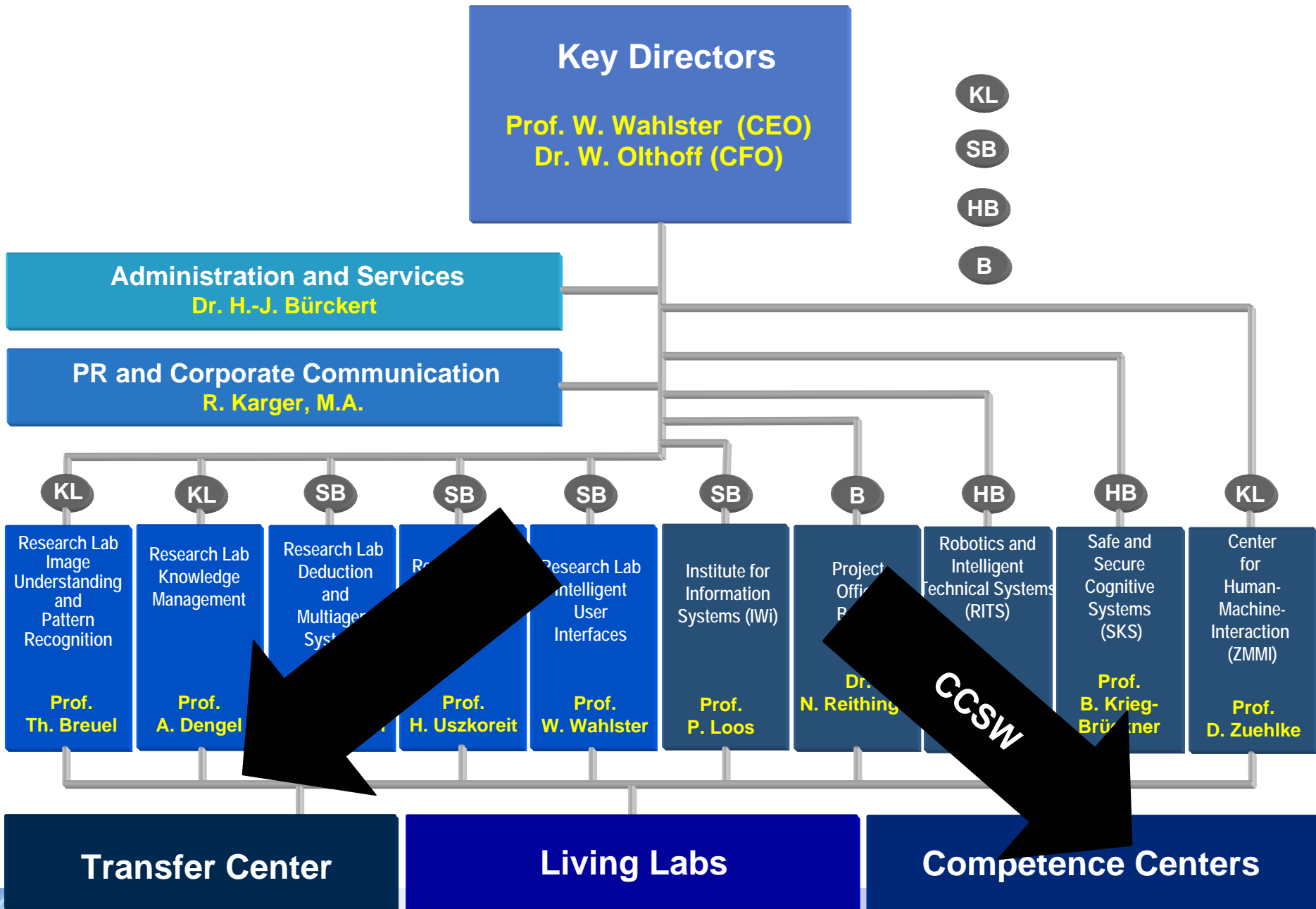


[leo.sauermann@dfki.de](mailto:leo.sauermann@dfki.de)

<http://www.dfki.de/~sauermann/>



# The Structure of DFKI



# DFKI is based in Kaiserslautern, Saarbrücken, Bremen and Berlin



**Saarbrücken Site**

DFKI GmbH  
Campus D32  
Stuhlsatzenhausweg 3  
D-66123 Saarbrücken

Tel.: +49 681 302 5252  
Fax.: +49 681 302 5341  
email: [info@dfki.de](mailto:info@dfki.de)  
[www.dfki.de](http://www.dfki.de)



**Kaiserslautern Site**

DFKI GmbH  
Trippstadter Straße 122  
D-67663 Kaiserslautern

Tel.: +49 631 20575 0  
Fax.: +49 631 20575 503  
email: [info@dfki.de](mailto:info@dfki.de)  
[www.dfki.de](http://www.dfki.de)



**Bremen Labs**

DFKI GmbH  
Robert-Hooke-Straße 5  
D-28359 Bremen

Tel.: +49 421 218 64100  
Fax.: +49 421 218 64150  
email: [info@dfki.de](mailto:info@dfki.de)  
[www.dfki.de](http://www.dfki.de)



**Project Office Berlin**

DFKI GmbH  
Rosenstraße 2  
D-10178 Berlin

Tel.: +49 30 243102 101  
Fax.: +49 30 243102 22  
email: [info@dfki.de](mailto:info@dfki.de)  
[www.dfki.de](http://www.dfki.de)

**in a changing world...**



**You are happy driving a steam wheeler.  
but what when others start driving cars... ?**

# So why do people use Semantic Web?

- I did some research on SWEO case studies



**W3C Semantic Web Use Cases and Case Studies**  
 Case Study: Prioritization of Biological Targets for Drug Discovery  
 Lutz Dreyer, D. Eddy  
 October 2007  
*Sloxy*



Figure 1: The current version of the Target Assessment Tool (TargetAssessmentTool) is a Java application that can be used to identify a set of targets for a given drug. The graph shows the relationships between the targets and the drug. The graph is a network of nodes and edges. The nodes represent the targets and the edges represent the relationships between them.

**Drug Target Assessment Tool**  
 The Target Assessment Tool is a Java application that can be used to identify a set of targets for a given drug. The graph shows the relationships between the targets and the drug. The graph is a network of nodes and edges. The nodes represent the targets and the edges represent the relationships between them.

- Key Benefits of Using Semantic Web Technology**
- The ability to integrate diverse source of data
  - More flexibility in being able to incorporate additional unanticipated data sets
  - The ability to be able to perform targeted queries or to be able to navigate through the relationships in the data
  - The capability of viewing all data relating to entities of interest, even if the information available for each entity differs

## Key Benefits of Using Semantic Web Technology

- The ability to integrate diverse source of data
- More flexibility in being able to incorporate additional unanticipated data sets
- The ability to be able to perform targeted queries or to be able to navigate through the relationships in the data
- The capability of viewing all data relating to entities of interest, even if the information available for each entity differs

•Significantly enhance and facilitate archive access  
 •Navigation and Discovery of new, potentially interesting facts «hidden» in the repository  
 •Highly efficient music archive, combining multi-channel access with a fully automated ordering and production flow  
 •enhanced metadata representation, including multiple file formats (not only music files with full metadata, but including pictures, links, interviews and many other resources) helping journalists to faster produce better trailers and talk-throughs  
 •Ease of integration across multiple archives and resources in the nearby future.

•Use of familiar, local terminology  
 •Support for unanticipated modeling extensions  
 •High degree of automation  
 •High-fidelity integration and mapping with external systems and terminologies  
 •Support for accurate answering of expressive queries

•For the City of Zaragoza the benefits include  
 •Providing a better service to its citizens, and thus improving the image of public administration to citizens  
 •Stimulate e-Government by providing easier access to relevant information  
 •Reduced cost by lowering the load on the call center and physical offices  
 •On the way to "citizen self-service": not only information retrieval but also semi-automatic service execution

•For citizens the benefits include  
 •Easy interaction, either through keywords or natural language  
 •High precision, without loosing recall  
 •Concrete answers instead of long lists of documents  
 •Ability to suggest related services (serendipity)  
 •Perceived as a search engine which "understands" the citizen

•An integrated, coordinated and richly-interconnected repository of eleven cultural heritages of its home region  
 •Transferring cultural knowledge in an economic and scalable way to society  
 •Providing a unique point of access for all people interested in cultural information of Catalonia  
 •The ontology guarantees interoperability between different applications, allowing easy addition of new ones  
 •Possibility to export knowledge and applications to different regions  
 •Since the ontology is based on existing standards, easy interpretation and federation is possible with other repositories.

•Web access to resources enabled by URI-based naming  
 •Non-ambiguous inference via OWL's formal semantics  
 •Dynamic schema evolution enabled by Semantic Web triple format  
 •Stable knowledge exchange via RDF(S) and OWL  
 •Numerous choices for the management of Semantic Web knowledge

•Enhanced search and navigation of a complex corporate Web site  
 •Improved visualization capabilities to guide users to data of interest  
 •Improved ability to personalize the Web site  
 •Enabled quick and simple implementation  
 •Stored data in RDF and enabled queries using SPARQL

•For the end user  
 •Users can find what they are looking for and trust what they find  
 •Users will be warned if they are browsing to a site that has made fraudulent claims  
 •For business  
 •Sites using Content Labels will get highlighted in search results. Depending on the user settings they may be the ONLY results displayed

•Lightweight information integration infrastructure  
 •The rapid integration of loosely coupled infrastructure  
 •Ability to share and reuse information across disparate data sources and geographies  
 •A consistent data model to build upon  
 •The ability to reuse a powerful user interface to data

•The ability to integrate diverse source of data  
 •More flexibility in being able to incorporate additional unanticipated data sets  
 •The ability to be able to perform targeted queries or to be able to navigate through the relationships in the data  
 •The capability of viewing all data relating to entities of interest, even if the information available for each entity differs

•Better innovation management leading to faster time-to-market  
 •Differentiation from competition  
 •Talent management  
 •Leverage full intellectual capital for improvements of business and operations  
 •Employee satisfaction  
 •Documented, shared understanding of relevant vocabulary for business

•By providing semantic metadata, the content provider facilitates the discovery of their content to Vodafone users.  
 •Vodafone are able to infer content relationships from disparate providers, and

to produce common-themed pages  
 Users can search more precisely via extended search forms, which allow search based on predicates used in the RDF ("song title", etc.)  
 Vodafone can filter pages and search results based on the user's content preferences, device and connection speed)

Distributed collaboration and interoperably disparate and heterogeneous data can be exchanged, integrated and utilized seamlessly and dynamically between remote systems.  
 Dynamic adaptability new requirements, data, functions and tasks can be introduced to the system without major rewrite or reprogramming to address novel situations or new tasks.  
 Multidisciplinary reuse of information: existing data in the system can be reprocessed to address unprecedented use cases.  
 Human-computer interaction systems interact intelligently with human users and more effectively, intuitively and easily.

Exposing of legacy data through a semantic layer so that it can be more easily re-used and recombined.  
 Linking data across database boundaries so as to enable more intuitive query, search, and navigation without the awareness of the boundaries.  
 The ontology serves as the control vocabulary to make semantic suggestions such as synonyms, related concepts to facilitate query and search.  
 Reasoning capability such as sub-classing, transitive property can then be implemented at the semantic layer to increase the query expressiveness so as to retrieve more concise answers.  
 Allows for more advanced data analysis and integrative knowledge discovery based on the huge web of data.

One major advantage of an ontology-based system over a traditional system is its ability to perform concept-based searches. The FNA portal allows searches for specific concepts; for example, a search for "child" or "children" will give same results though there are two different localizations of the same concept, that is one is the plural of the other. The portal also allows multilingual searches: "children" or "enfants" will return the same results.

Semantic Web technology, as now used by Sun Microsystems' swoRDFish Metadata Initiative:  
 Improves web content quality and consistency  
 Provides intelligence for auto-assembly of web content (dynamic content generation)  
 Enables auto-assembly of related web content based on predefined relationships within swoRDFish's metadata  
 Enables creation and management of multiple taxonomies, which allows groups the capability to organize content in different ways depending on their target customers  
 Provides storage of multiple product names (aliases), which allows consumers of web content to search using variations on a product's name, spellings of that name, code names, or acronyms  
 Provides one unique identifier (swoRDFish ID) for each product that can be used to integrate relevant product data across Sun (for example, to publish web content, for internal data-to-data applications, or to embed in firmware to enable identification of customer services for reporting, audits, and notification)  
 Provides Return on Investment (ROI) factors such as resourcing benefits (through automation of web content publishing, significantly reduces human touch points from the development team), improved content time-to-market and quality, reduction of redundant work in other projects  
 Future applications of Semantic Web Technology through the Sun swoRDFish Metadata Initiative could include:  
 Defining and implementing more automated data maintenance procedures for swoRDFish, reducing administrative costs  
 Managing relationships between products and product parts (what works with what)  
 Extending the capability to auto-assemble related web content  
 Expanding the use of swoRDFish IDs to help unify service offerings with more relevant reporting and notifications  
 Further exploring swoRDFish as a data gateway for various types of data across Sun  
 Decentralizing ontology management for added flexibility and improved data quality  
 Internationalized application and localized data  
 Tighter integration with product and technology usage (as products are more integrated with the web, this can provide a global key, integrating with connecting swoRDFish to extend the reach of swoRDFish metadata)

•Better infomation management leading to faster time-to-market  
 •Differentiation from competition  
 •Talent management  
 •Leverage full intellectual capital for improvements of business and operations  
 •Employee satisfaction  
 •Documented, shared understanding of relevant vocabulary for business

•By providing semantic metadata, the content provider facilitates the discovery of their content to Vodafone users.  
 •Vodafone are able to infer content relationships from disparate providers, and

17 case studies

Key Benefits

89 Key Benefits



# The Key Benefits of Semantic Web

<b>profit</b>	customer satisfaction, shareholder value, user work support	21
<b>data integration across sources</b>	A consistent data model to build upon. integrate content from different organizations /providers/ departments, disparate data sources, legacy data	13
<b>expressive queries</b>	i.e. sparql	11
<b>taxonomy multilingual</b>	software uses the terminology of the user, vocabulary, synonyms, multilinguality	9
accessible for users	generate website make data accessible to users	8
navigation, related items, <b>serendipity</b>	discover hidden information, navigate accross data boundaries, "what works with what"	8
extensible metadata	incorporate unanticipated data sets, new requirements do not change code, dynamic adaptability	5



# The Key Benefits of Semantic Web

export knowledge	existing data can be repurposed, unprecedented use cases, conserve culture standards	4
<b>semantics</b>	<b>inference owl reasoning formal</b>	4
simple	cheap implementation lightweight	2
natu		2
pers		2
auto		2
expla		1
improve content quality	consistency	1
<b>web of linked open data</b>		1
<b>unique id</b>		1

**this will grow**





# Open linked data

~ 2.000.000.000 facts

~ 680.000 links

- <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>



... more pe

Google Social API + FO

unique id

March 13, 2008

## The Yahoo! Search Open Ecosystem

In the coming weeks, we'll be releasing more detailed specifications that will describe **our support of semantic web standards**. Initially, we plan to support a number of [microformats](#), including [hCard](#), [hCalendar](#), [hReview](#), [hAtom](#), and [XFN](#). Yahoo! Search will work with the web community to evolve the vocabulary framework for embedding structured data. **For starters, we plan to support vocabulary components from [Dublin Core](#), [Creative Commons](#), [FOAF](#), [GeoRSS](#), [MediaRSS](#), and others based on feedback. And, we will support [RDFa](#) and [eRDF](#) markup to embed these into existing HTML pages.** Finally, we are announcing support for the [OpenSearch](#) specification, with extensions for structured queries to deep web data sources.

<http://www.ysearchblog.com/archives/000527.html>

**applications  
from SWEO case studies  
and foafs**



[www.flickr.com/photos/wonderferret/1447024668](http://www.flickr.com/photos/wonderferret/1447024668)

4

# Vodafone live!

## Problem:

- Vodafone is one of the largest mobile carriers
- **Content** for portal *Vodafone live!*
- Multiple **content** providers

## Solution

- *Vodafone live!* publishes RDF vocabulary
- Content providers conform
- Content metadata is delivered in RDF

## Project:

- **Content** ringtones, games, pictures
- **Metadata** phone platforms, ratings adult|violence, availability
- exact documentation, validation of XML
- key is involvement of content providers

Courtesy of Kevin Smith, Vodafone Group R & D ([SWEQ Use Case](#))

# Vodafone live!

## vocabularies

- Dublin Core
- PRISM for embargos and availability
- Internet Content Ratings Association (ICRA) for violence, erotic, gambling
- extensions
  
- user enters profile at portal, appropriate content delivered
- significant increase in content download after the introduction

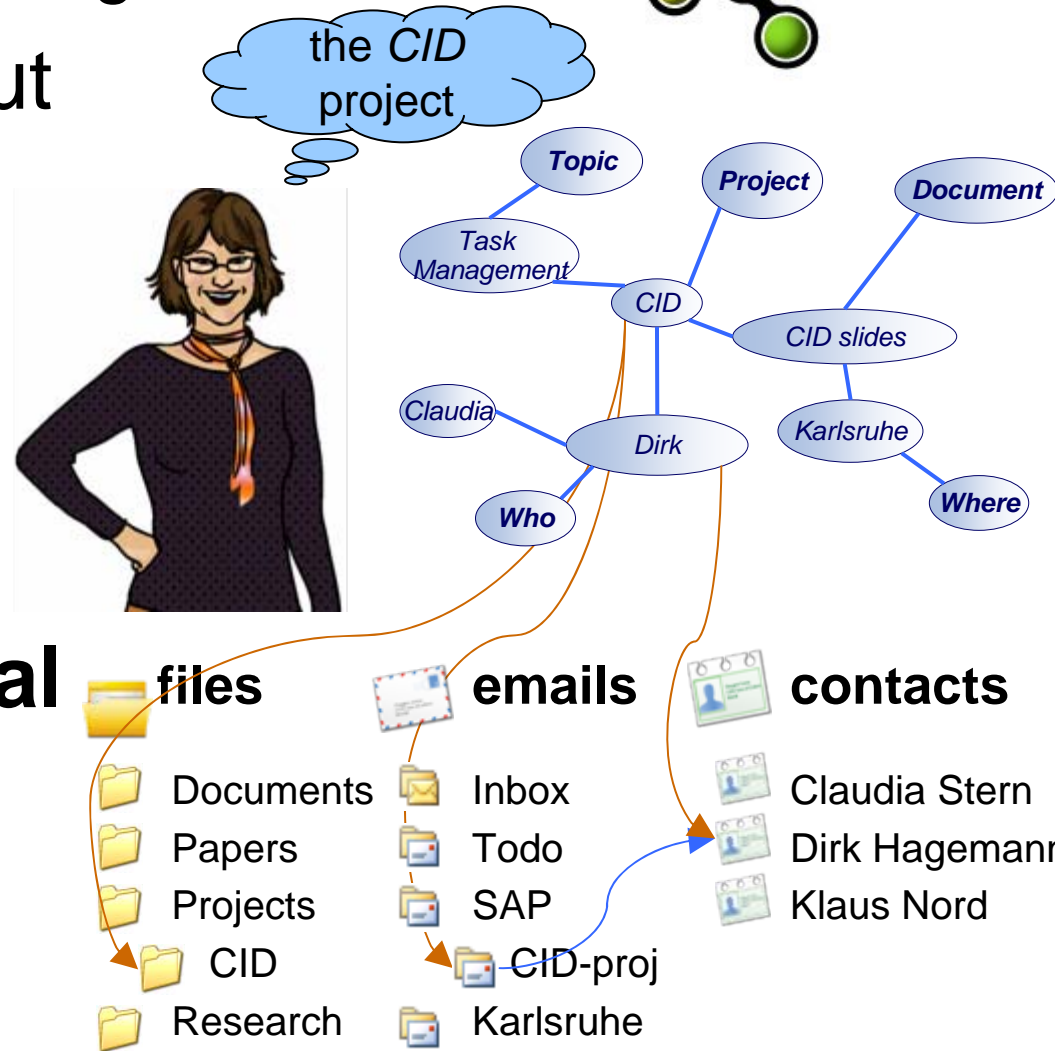




3

# Semantic Desktop: Problem of Personal Information Management

- Problem: **Data** about **single ideas** are stored in **several applications** and **without context**
- People have **mental models** of their environment



# NEPOMUK KDE 4.0

~ 1-8 million users

~ 1k developers

integrate the desktop data



11. January 2008: release v4.0



?! use RDF  



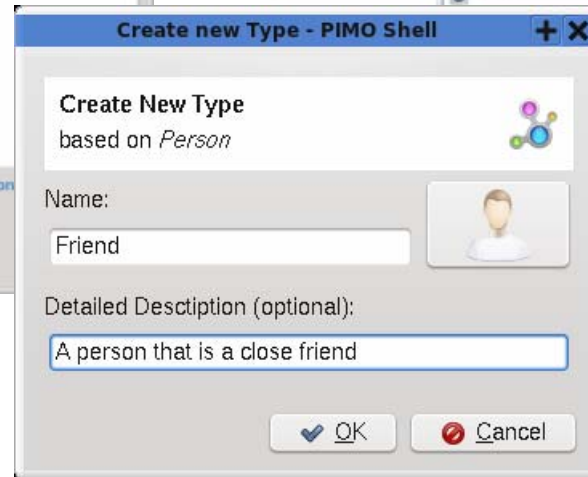
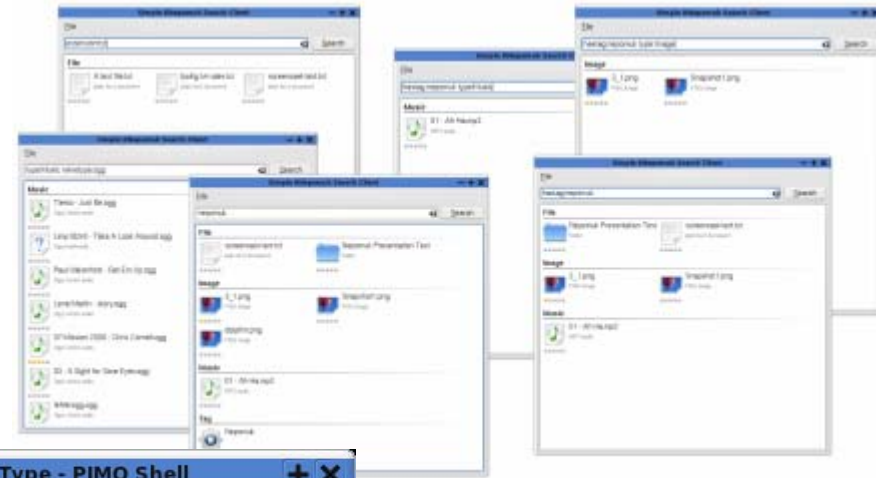
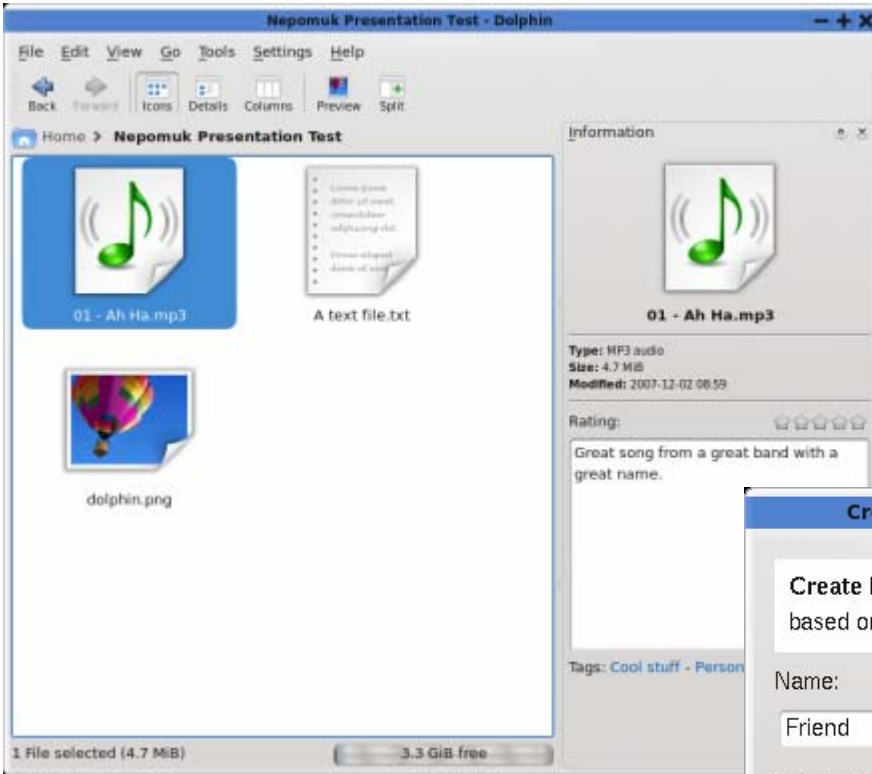
- convert all data to RDF, index it
- allow user to annotate with tags, ratings, etc
- allow applications to annotate

# NEPOMUK KDE 4.0



- Soprano indexing, Dolphin file explorer

Soprano



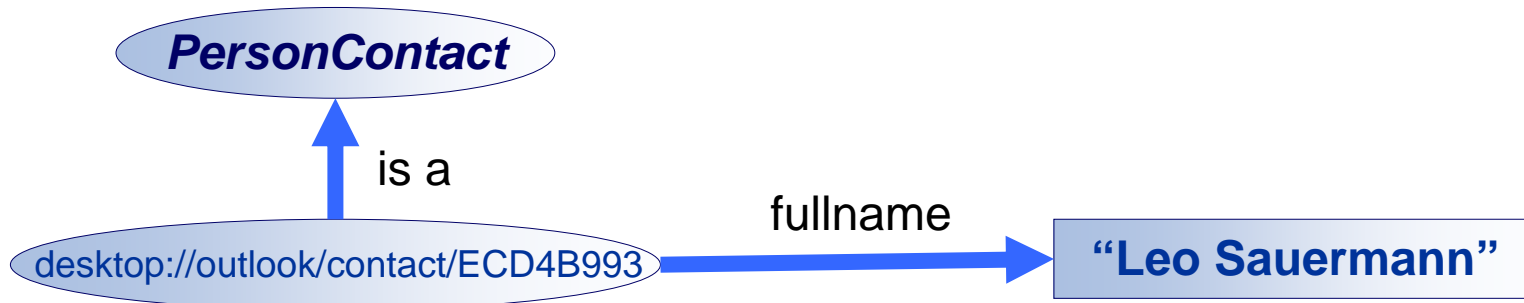
<http://nepomuk.kde.org/>

# Convert Leo from Address-Book to RDF



Microsoft Outlook

Name...	Leo Sauermann
Position:	Scientist
Eirma:	DFKI
Speichern unter:	Sauermann, Leo



```
<rdf:RDF
```

```
  xmlns:nco="http://www.semanticdesktop.org/ontologies/2007/03/22/nco#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
```

```
<nco:PersonContact rdf:about="desktop://outlook/contact/ECD4B993">
```

```
  <nco:fullname>Leo Sauermann</nco:fullname>
```

```
</nco:PersonContact >
```

```
</rdf:RDF>
```

done with <http://aperture.sourceforge.net>

2



since ~2000 RDF is used by plugins to  
configure how they extend the mozilla platform

[http://developer.mozilla.org/en/docs/Deploying\\_a\\_Plugin\\_as\\_an\\_Extension](http://developer.mozilla.org/en/docs/Deploying_a_Plugin_as_an_Extension)

```
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:em="http://www.mozilla.org/2004/em-rdf#">
  <Description about="urn:mozilla:install-manifest">
    <em:id>RhapsodyPlayerEngine@rhapsody.com</em:id>
    <em:name>Rhapsody Player Engine</em:name>
    <em:version>1.0.0.487</em:version>
    <em:targetApplication>
      <Description>
        <em:id>{ec8030f7-c20a-464f-9b0e-13a3a9e97384}</em:id>
        <em:minVersion>1.5</em:minVersion>
        <em:maxVersion>1.5.0.*</em:maxVersion>
      </Description>
    </em:targetApplication>
  </Description>
</RDF>
```



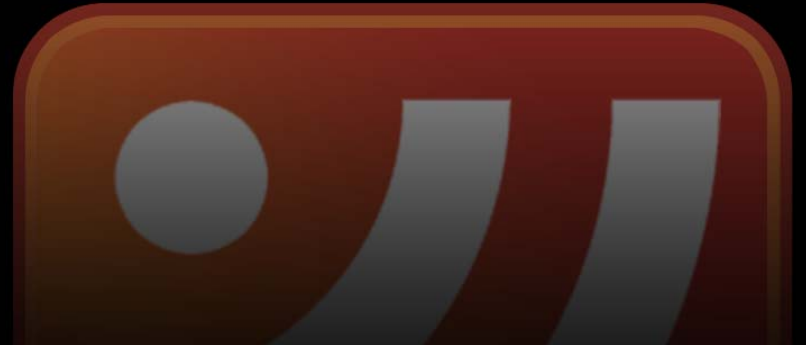
**Mozilla?**  
**Semantic Web has been**  
**in use for so long?**





# 1

# 1999 RSS



# RDF Site Summary

Many applications

Changes

- Version 1: <http://.../1>
- Version 2: <http://.../2>

Index of pages

- Page 1: <http://.../1>
- Page 2: <http://.../2>

- list of resources
- title
- author
- date
- content



# key learnings from W3C SWEO case studies

## Vodafone

- Use an **open, flexible model**
- Research **existing vocabularies and reuse industry metadata standards** where possible.
- Provide full **documentation and a schema (RDF or XML) for validation and IDE support.**
- Ensure appropriate **stakeholder support**

## Sun and Zepheira

- Start with, and maintain, an investment in **data integrity and process automation**
- Provide flexible and extensible administration tools
- Ensure appropriate **stakeholder involvement** and ownership
- Consider program name selection – choose a name that is simple to remember and spell, is not similar to other program or project names within that organization, and does not require explanation
- Produce an accurate definition and wide communication of the technology's success metrics

**make your own semantic  
web sites**



[www.flickr.com/photos/wonderferret/1447024668](http://www.flickr.com/photos/wonderferret/1447024668)

# Its easy!

## “Planning a Semantic Web site”

*Prepare your site for structured data*

Rob Crowther

04 Mar 2008

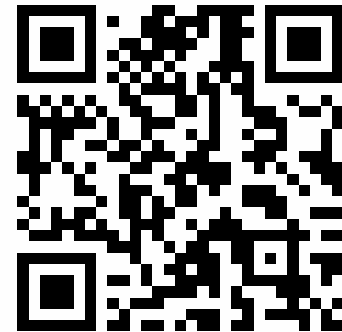
[http://www.ibm.com/developerworks/  
xml/library/x-plansemantic/index.html](http://www.ibm.com/developerworks/xml/library/x-plansemantic/index.html)



look on del.icio.us for  
“SWEO tutorial”

- Contact us to get connected to researchers within DFKI
  - Ontology creation and maintenance
  - Support for annotating and analysing data
  - Research on demand
  - Consulting, teaching, training

<http://semanticweb.dfki.de>



Michael Sintek  
michael.sintek@dfki.de



Paul Buitelaar  
paulb@dfki.de

+ Leo



mash it up!

</SemanticWeb>  
</rdf:RDF>

Leo Sauermann

[leo.sauermann@dfki.de](mailto:leo.sauermann@dfki.de)

<http://www.dfki.de/~sauermann/>

Thanks to Jesus, Ingrid Brunner-Sauermann, Heiko Maus, Sven Schwarz, Ansgar Bernardi, Andreas Dengel, Frank Osterfeld, Dominik Heim, Man Luo, Jeen Broekstra, Giovanni Tummarello, Michael Zeltner, Stephan Baumann, Gunnar A Grimnes, Ludger Van Elst, Harald Holz, Stefan Decker, Malte Kiesel, und alle anderen auf #swig



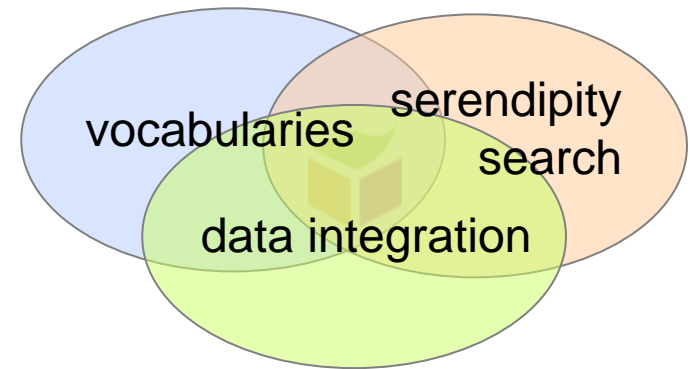
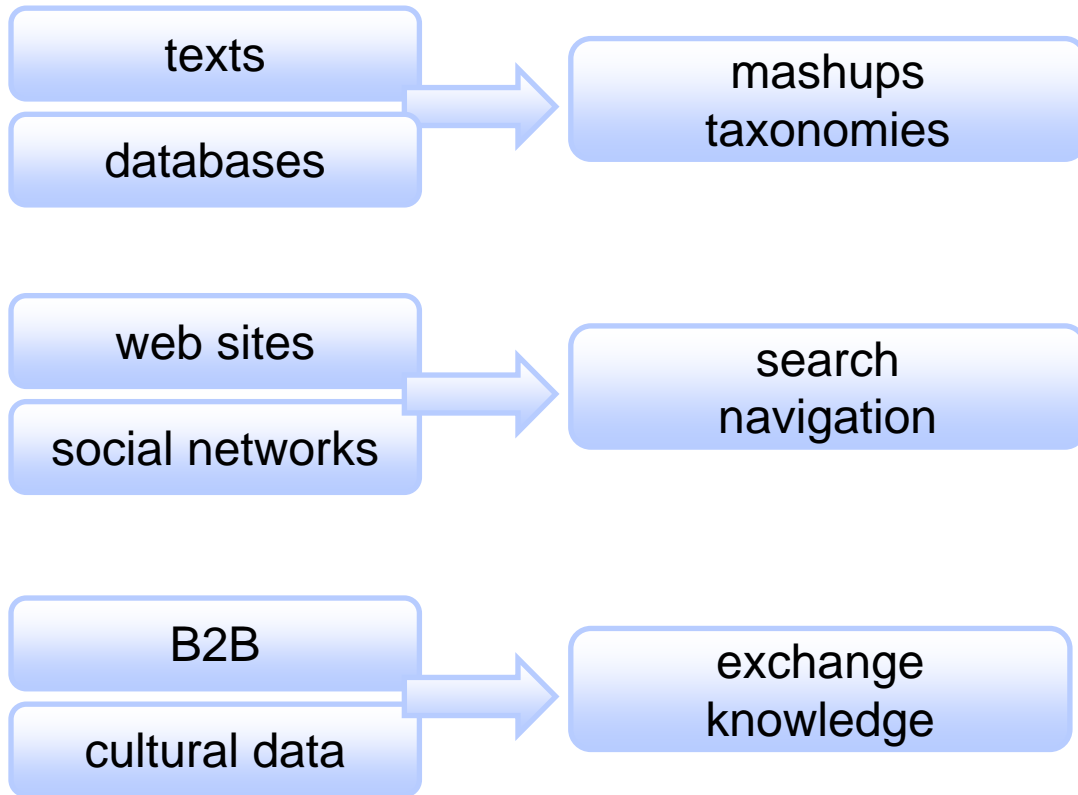


# Semantic Web is part of a changing world



<http://www.flickr.com/photos/pavel1998/1876595585/>

# Business Models



The New York Times and the Los Angeles Times tried charging for some online content, then abandoned the

<http://www.zemanta.com/>

Similar thing as Reuters did  
launch January 2008 with  
OpenCalais.com  
but with pictures, links, tags  
and users

#### Tags

Business Printing Publishing and Printing Brokers  
Los Angeles Times Computers Products Utilities

[huffingtonpost.com](http://huffingtonpost.com)

(link)

2 months ago

How a Startup  
Inspired HP's Print  
2.0 Strategy