



# Industrial implementations of IKS Nuxeo

8 April 2011

Stefane Fermigier, Nuxeo



IKS is co-funded by the European Union and develops  
new technology for intelligent content management

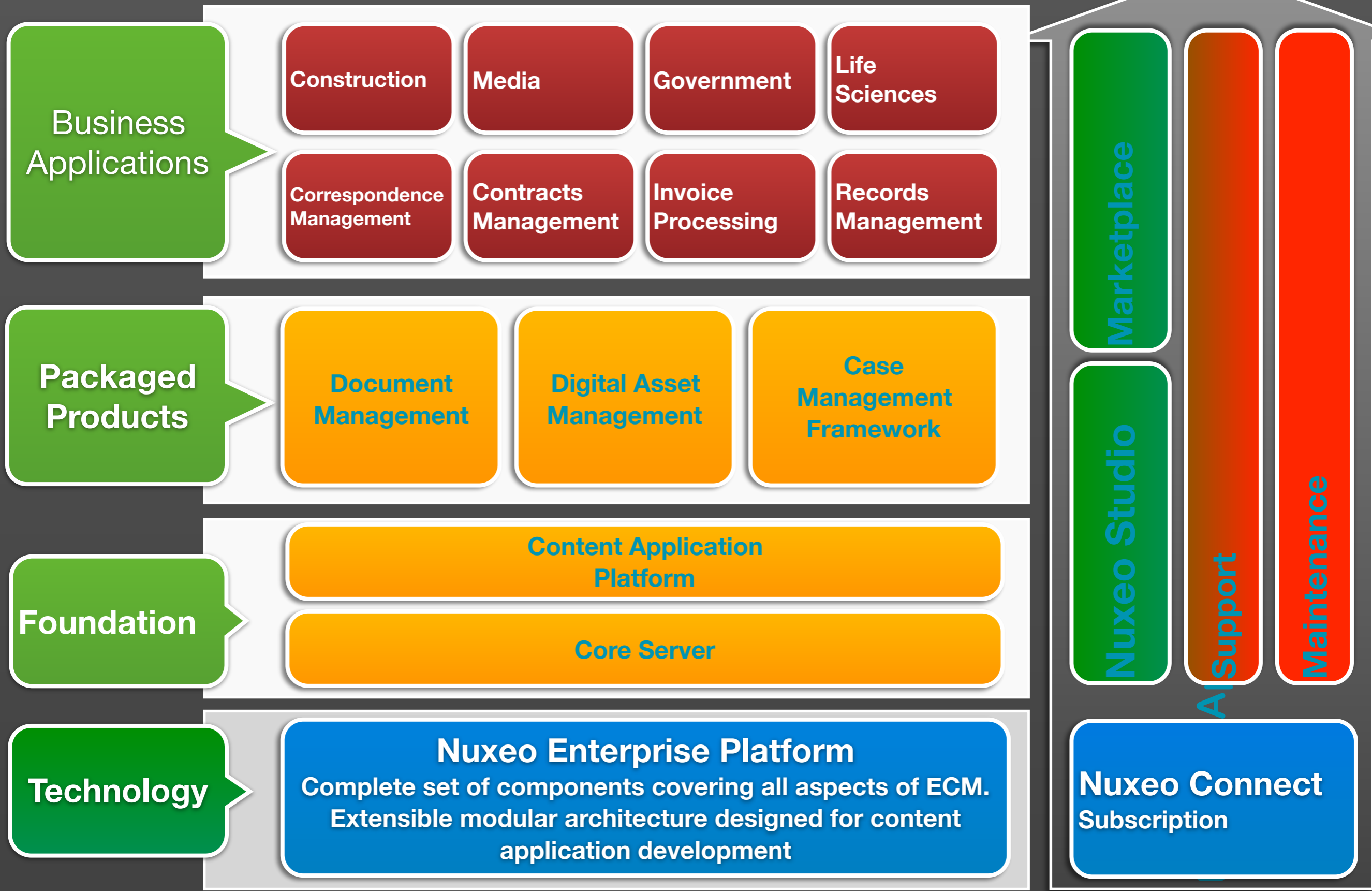


**Nuxeo**

- | **Founded in 2000, 50 employees**
- | **Located in Paris, Boston and SF**
- | **Open source stack rewritten in 2006, using open source Java technologies**
  - | **Large code base (400 kSLOC of Java), very modular and extensible (thanks in part to OSGi)**
- | **Main markets:**
  - | **Document Management (DMS) -> public administrations and large companies, alternative to Documentum or OpenText**
  - | **Repository / Platform -> custom / complex applications**
  - | **News management system -> press agencies**
  - | **Digital Asset Management (DAM) -> various**



# Nuxeo ECM - Our Approach



Business Applications

Construction

Media

Government

Life Sciences

Correspondence Management

Contracts Management

Invoice Processing

Records Management

Packaged Products

Document Management

Digital Asset Management

Case Management Framework

Foundation

Content Application Platform

Core Server

Technology

Nuxeo Enterprise Platform

Complete set of components covering all aspects of ECM. Extensible modular architecture designed for content application development

Marketplace

Nuxeo Studio

AI Support

Maintenance

Nuxeo Connect Subscription



## Our (initial) expectations

### | Improving DM and DAM

- | NLP (named entities, quotes, co-occurrences... extraction)
- | Pictures analysis
- | Document classification and clustering
- | Faceted semantic search

### | Correspondence / Case Management

- | Analyze and route incoming mails / cases depending on their content

### | Collaboration and KM

- | Promote emergence, capture and sharing of knowledge within an organization by combining social and semantic technologies





## What I wrote for last year's review

- | **“The semantic part is “easy” (using state of the art libraries or algorithms) ...”**
  - | **Actually I was wrong, lots of work has been going on on the FISE / Stanbol project**
- | **“... the hardest part is actual integration (UI, user feedback, changing internal data structures when needed), even with a modular, service-oriented and event-driven architecture such as ours”**
  - | **But I will show you some progress on this part today**



## This year's focus

- | We worked (quite a bit) on the “easy” part (FISE then Stanbol)
- | Later in 2010, we've conducted a first integration of FISE / Stanbol / IKS Alpha Stack into Nuxeo DM
  - | Focus on named entities suggestion / management
  - | Leverage Linked Open Data
- | Iterative approach



## From Tags to Entities

- | **Nuxeo already knows how to classify documents by tag or Dublin Core metadata (e.g. DC Subject)**
- | **So what's wrong with tags and metadata?**
  - | **Several terms for same meaning**
    - | Ex: "NYC", "New York", "New York City"
  - | **Several meanings for same terms**
    - | Need context to remove any ambiguity

Navigation sidebar with icons for Home, Search, Add, Refresh, and Logout.

- Default domain
  - Sections
  - Templates
  - Workspaces
  - Entities

WorkList | Clipboard








No document in clipboard






> Entities




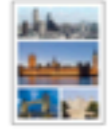




## Entities

Search [ ]

People (81134457860387652)      Organizations (95967456873405)      Places (39856734586)

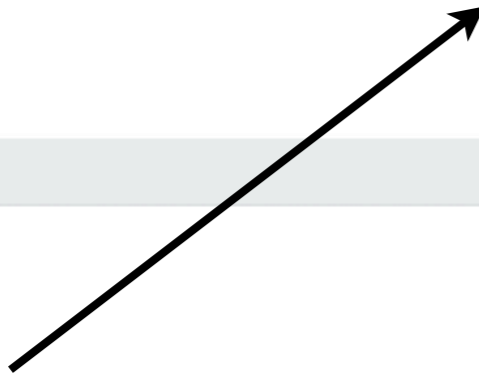
-  **Jimi Hendrix**  
James Marshall "Jimi" Hendrix (born Johnny Allen Hendrix; November 27, 1942 - September 18, 1970) was an American guitarist, singe[...]
-  **Elmore James**  
Elmore James (January 27, 1918 - May 24, 1963) was an American blues guitarist, singer, song writer and band leader. He was known [...]
-  **Howlin' Wolf**  
Chester Arthur Burnett (June 10, 1910 - January 10, 1976), better known as Howlin' Wolf, was an influential American blues singer,[...]
-  **Albert King**  
Albert King (April 25, 1923 - December 21, 1992) was an American blues guitarist and singer.
-  **B.B. King**  
Riley B. King (born September 16, 1925), known by the stage name B.B. King, is an American blues guitarist and singer-songwriter a[...]
-  **Muddy Waters**  
McKinley Morganfield (April 4, 1913 - April 30, 1983), known as Muddy Waters, was an American blues musician, generally considered[...]
-  **Steve Cropper**  
Steven Lee Cropper (born October 21, 1941 in Willow Springs, Missouri), also known as Steve "The Colonel" Cropper is an American g[...]

-  **American Red Cross**  
The American Red Cross (also known as the American National Red Cross) is a humanitarian organization that provides emergency ass[...]
-  **Apple Inc.**  
Apple Inc. is an American multinational corporation that designs and manufactures consumer electronics, computer software, and com[...]
- Guardian Media Group**  
Not to be confused with the Guardian Enterprise Group, an American media company. Guardian Media Group plc (often referred to as [...]
-  **International Red Cross and Red Crescent Movement**  
The International Red Cross and Red Crescent Movement is an international humanitarian movement with approximately 97 million volu[...]
- O'Reilly Media**  
O'Reilly Media (formerly O'Reilly & Associates) is an American media company established by Tim O'Reilly that publishes books and [...]
- Stanford University**  
The Leland Stanford Junior University, commonly referred to as Stanford University or Stanford, is a private research university [...]
-  **Starbucks**  
Starbucks Corporation is an international coffee and coffeehouse chain based in Seattle, Washington, United States. Starbucks is [...]
-  **United Nations**

-  **Aylesbury railway station**  
Aylesbury railway station (also called Aylesbury Town railway station) is a railway station in Aylesbury, Buckinghamshire, England[...]
-  **Europe**  
Europe is, by convention, one of the world's seven continents. Comprising the westernmost peninsula of Eurasia, Europe is generall[...]
-  **Isle of Wight**  
The Isle of Wight is a county and the largest island of England, located in the English Channel, 3-5 miles (5-8 km) off the south [...]
-  **London**  
London is the capital of England and the United Kingdom. It is the UK's largest and most populous metropolitan area and the larges[...]
-  **Monterey Pop**
-  **United States**  
The United States of America (commonly referred to as the United States, the U.S. , the USA, or America) is a federal constitution[...]
-  **West Hollywood, California**  
West Hollywood, a city of Los Angeles County, California, was incorporated on November 29, 1984. The latest residential population[...]
-  **Woodstock, New York**  
Woodstock is a town in Ulster County, New York, United States. The population was 6,241 at the 2000 census. The Town of Woodstock [...]









# Washington is...



Entities > Entities

## Entities

Search

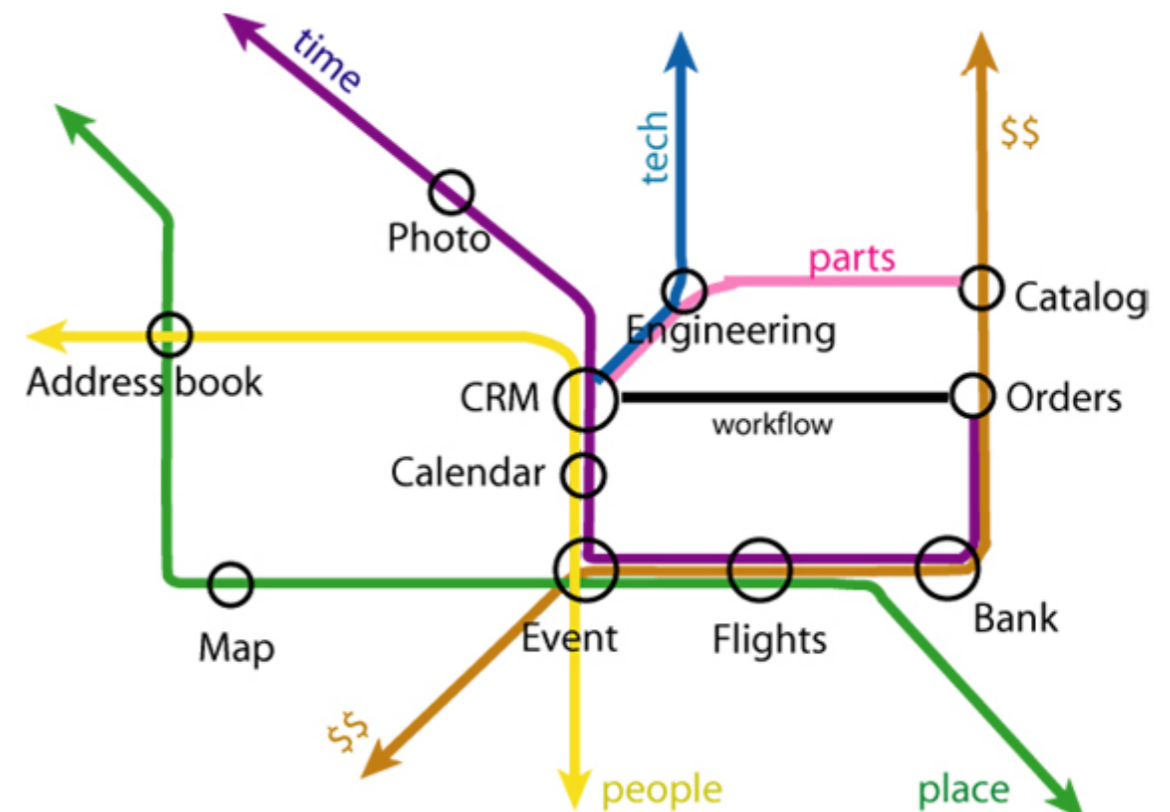
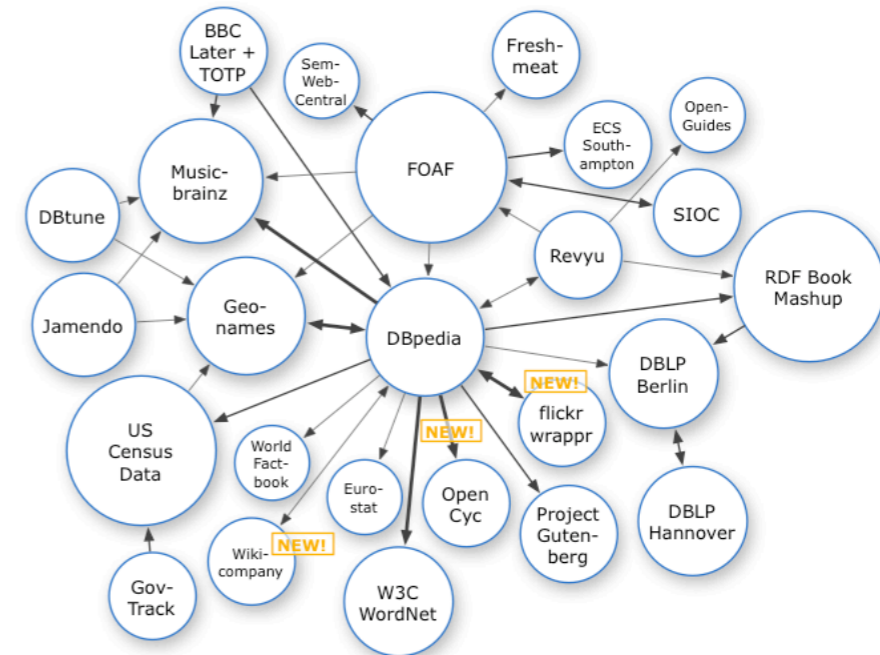
People (1)	Organizations (3)	Places (2)
 <p><b>George Washington</b> George Washington (February 22, 1732 – December 14, 1799) served as the first President of the United States from 1789 to 1797 and[...]</p>	 <p><b>George Washington University</b> The George Washington University (GW, GWU, or George Washington) is a private, coeducational comprehensive research university loc[...]</p>  <p><b>Washington Redskins</b> The Washington Redskins are a professional American football team based in the Washington, D.C. area. The team plays at FedExField[...]</p>  <p><b>Washington University in St. Louis</b> Washington University in St. Louis is a nonsectarian, private research university located in suburban St. Louis, Missouri. Founded[...]</p>	 <p><b>Washington</b> Washington is a state in the Pacific Northwest region of the United States. Washington was carved out of the western part of Washi[...]</p>  <p><b>Washington, D.C.</b> Washington, D.C. , formally the District of Columbia and commonly referred to as Washington, the District, or simply D.C. , is the[...]</p>

+ If no results match, you can [create your entity](#).



# Tagging with Entities

- | Linking with LOD brings:
  - | Global namespace / universal meaning context
  - | Interoperability across domains
  - | Interoperability across applications



# Movie Time!

**AQUA TEEN  
HUNGER FORCE MOVIE  
YEAR OF THE DOG**

**YEAR OF THE DOG  
BLACK BOOK  
IN THE LAND OF WOMEN**



## Next steps for entity management

- | **Integration of the Entity Hub to ensure fast and reliable access to LOD**
  - | **Note: most commercial NER vendors lack this part**
- | **Current architecture can already be used to easily plug in commercial vendors tools**
- | **Multilingual models**
- | **New entity types**
  - | **Generic: commercial products, artworks...**
  - | **Specific: proteins, genes, drugs...**



## More engines

- | **Text summarization**
- | **Topic classification**
- | **Picture analysis**
  - | **Face detection / recognition**
  - | **Clustering**
- | **Faceted semantic search**



## Business perspectives

- | **Semantic features to be provided as add-ons to existing products**
- | **Demo / beta version of the plugin already available, but limited spontaneous adoption**
- | **A few interested customers or prospects so far, hopefully first projects could start in 2011**
- | **Potential partnerships with several semantic technology vendors (NER, sem. search) could bring significant business**



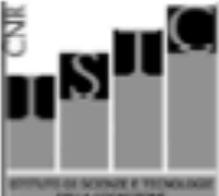








*Come in*

WE'RE

OPEN



# The IKS Consortium

<p>Project Lead and Coordination <b>Salzburg Research</b></p> <p>salzburg research</p>	<p><b>Wernher Behrendt</b> Salzburg Research Forschungsgesellschaft m.b.H. Jakob Haringer Straße 5/3   5020 Salzburg, Austria T +43.662.2288-409   F +43.662.2288-222 wernher.behrendt@salzburgresearch.at <a href="http://www.salzburgresearch.at">www.salzburgresearch.at</a></p>
<p>Deutsches Forschungsinstitut für Künstliche Intelligenz (DFKI)</p> 	<p>Universität St. Gallen Institute of Technology Management</p> 
<p>Consiglio Nationale delle Ricerche (CNR)</p> 	<p>Software Quality Lab Universität Paderborn</p>  
<p>Software Research and Development Consultancy Ltd (SRDC)</p> 	<p>Hochschule Furtwangen</p> 
<p>Nuxeo Sa.</p>	<p>Alkacon Software GmbH</p> 
<p>TXT Polymedia</p> 	<p>Pisano Holding GmbH</p> 
<p>Nemein Oy</p> 	<p>Day Software AG</p> 