# Zentity 1.0:

## *Microsoft Research & the Scholarly Information Ecosystem*

**Alex D. Wade**

*Director, Scholarly Communication*

Microsoft External Research

Microsoft Corporation

Microsoft Research

# Microsoft External Research

- Organization within Microsoft Research that engages in strong **partnerships** with academia, industry and government to advance computer science, education, and research in fields that rely heavily upon advanced computing
- Initiatives that focus on the **research process** and its role in the innovation ecosystem, including support for open access, open tools, open technology, and interoperability
- Developers of **advanced technologies** and services to support every stage of the research process

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Mission

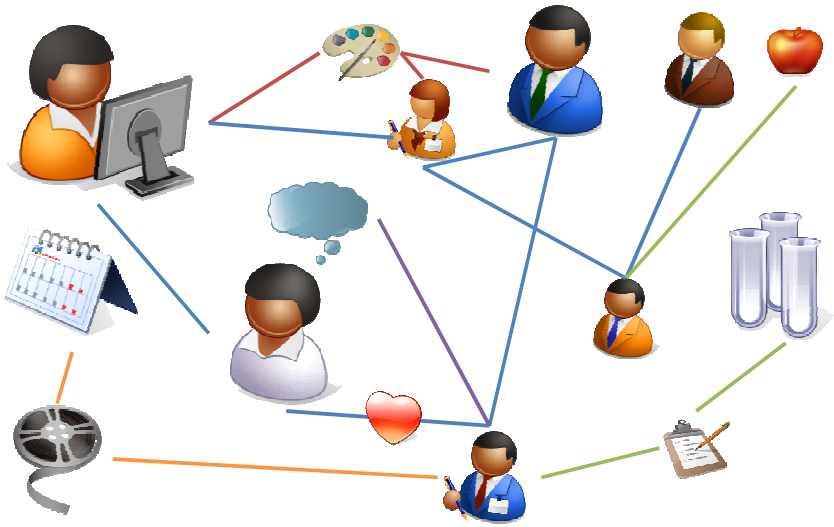
**Optimize and extend Microsoft software to meet the specific needs of the academic community**

**Our approach:**

*Conduct applied projects to enhance academic productivity by evolving Microsoft's scholarly communication offerings*

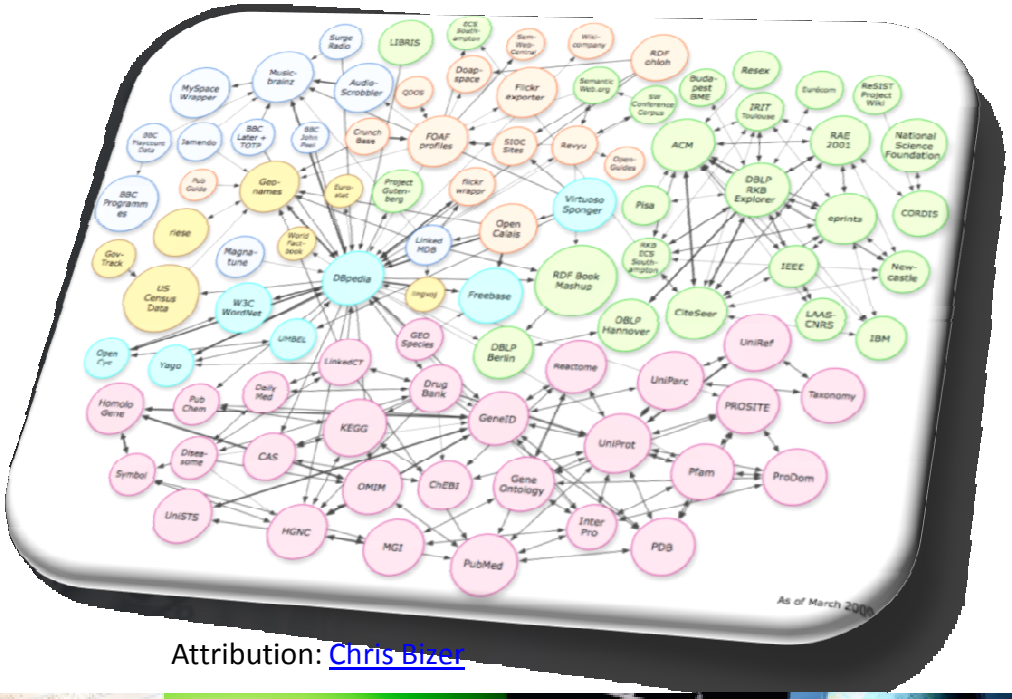
Microsoft External Research is uniquely positioned to drive this initiative across Microsoft

# A world where all data is linked …



- Data/information is inter-connected through machine-interpretable information (e.g. paper X **is about** star Y)
- Social networks are a special case of ‘data meshes’

- A knowledge ecosystem:
  - A rich user experience
  - An ecosystem of services
  - Semantic storage
  - Open, Collaborative, Interoperable, and Automatic

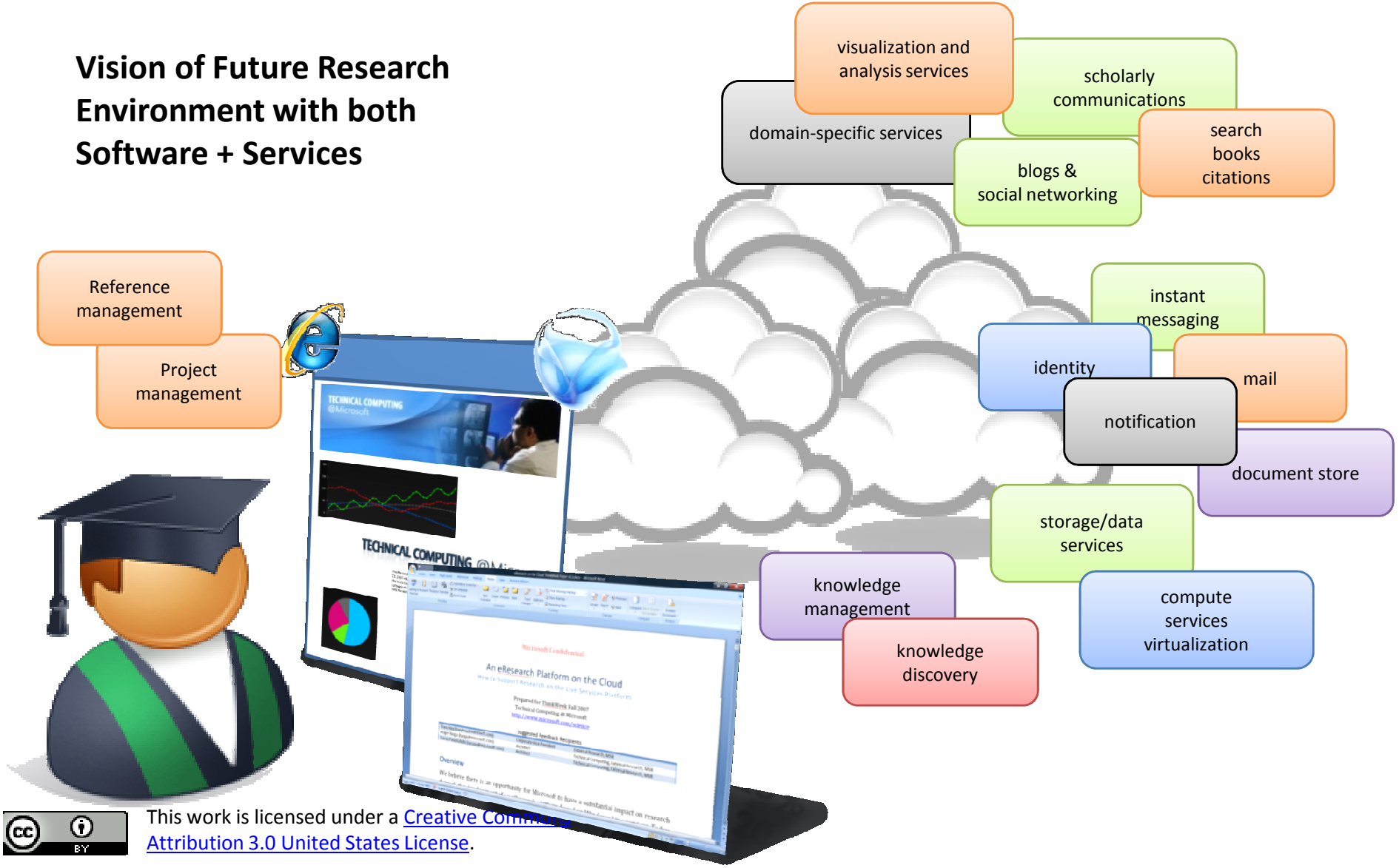



Attribution: Chris Bizer

This work is licensed under a Creative Commons Attribution 3.0 United States License.

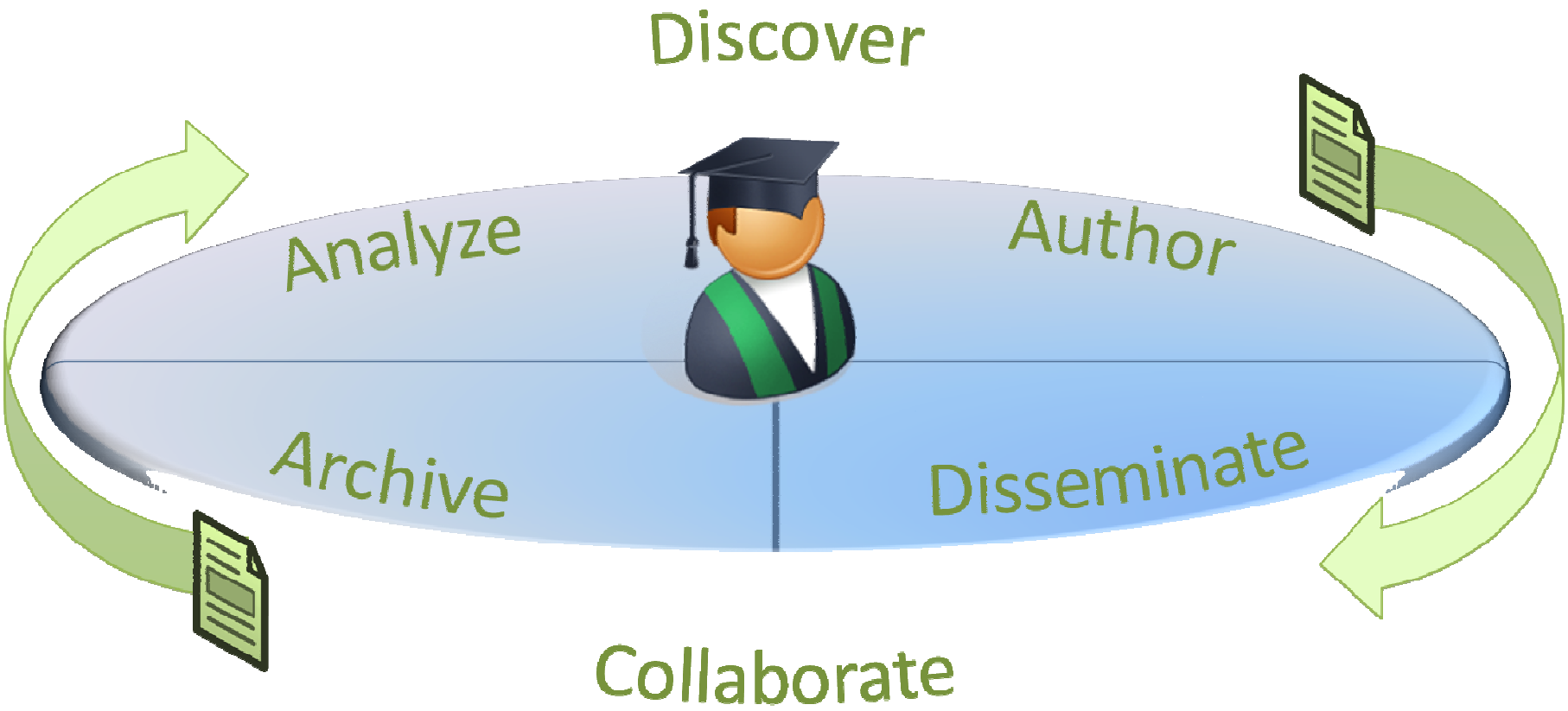
# …and stored/processed/analyzed in the Cloud

**Vision of Future Research Environment with both Software + Services**




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# The Scholarly Communication Lifecycle




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Our Goals

- **For Microsoft end-users**, making it easier to use our software for all aspects of their research process
- **For Microsoft developers**, demonstrating the toolset and showing how our platform can be extended

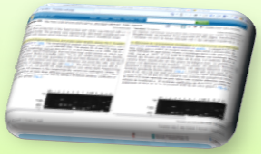
---

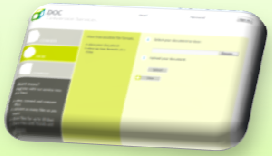
- **For non-Microsoft end-users**, working to ensure the ability to interoperate with our software across all phases of the research process, as necessary
- **For non-Microsoft developers**, enabling transparency to our efforts in this space and encouraging a dialogue

This work is licensed under a Creative Commons Attribution 3.0 United States License.

**Ecosystem of Tool/Services**

Peer-Review


Translation


Conversion

- Visualization
- Discovery
- Entity Extraction
- etc.

**Repositories**

arXiv.org

eprints

Zentity

Fedora Commons

DSPACE

**Rich User Experience**

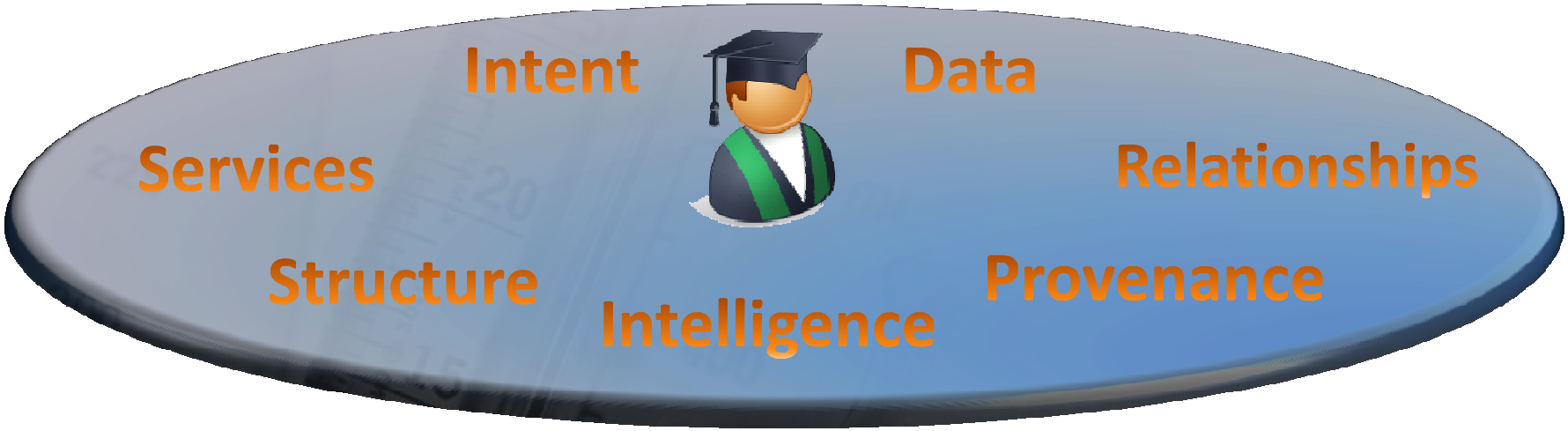
Authoring

Collaboration/VREs

- Search
- Desktop Tools
- ELNs
- etc.

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Rich User Experience




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Creative Commons Add-in for Office 2007

**Intent**: Insert Creative Commons licenses from within Office 2007

**Services**: Integrates with Creative Commons Web API to create new licenses

**Relationships**: license information stored as RDF XML within the document OOXML

This work is licensed under a Creative Commons Attribution 3.0 United States License.

Source code and binary:
http://ccaddin2007.codeplex.com

# Ontology Add-in for Word 2007

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

science commons

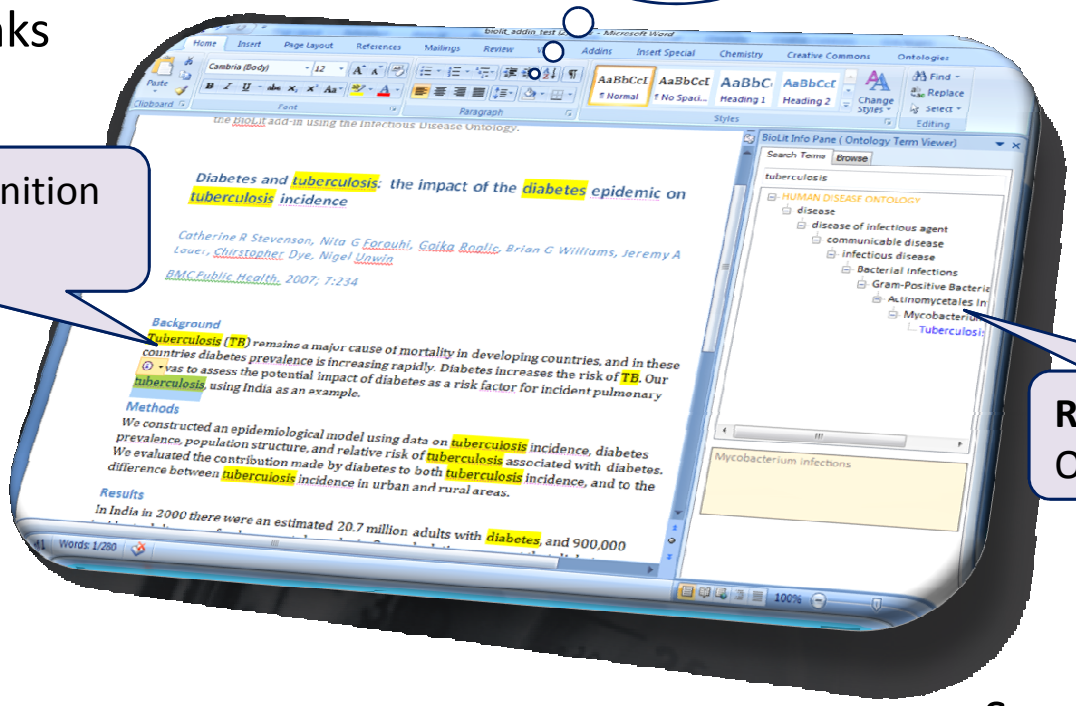
- John Wilbanks

**Services:** Ontology download web service

University of California San Diego

- Phil Bourne
- Lynn Fink

**Intent:** Term recognition & disambiguation




**Relationships:** Ontology browser

Source code and binary:
http://research.microsoft.com/ontology/

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Chem4Word - Chemistry Drawing in Word

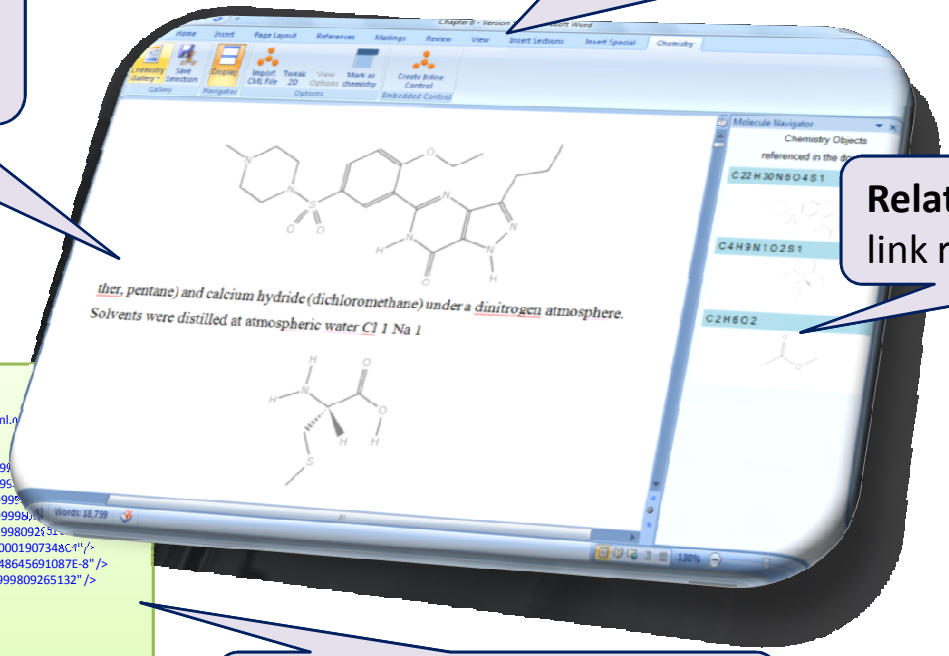

UNIVERSITY OF CAMBRIDGE

- Peter Murray-Rust
- Joe Townsend
- Jim Downing


**Author/edit 1D and 2D chemistry. Change chemical layout styles.**

**Intent:** Recognizes chemical dictionary and ontology terms

**Data:** Semantics stored in Chemistry Markup Language




**Relationships:** Navigate and link referenced chemistry

```
<?xml version="1.0" ?>
<cml version="3" convention="org-synth-report" xmlns="http://www.xml-cml.
 <molecule id="m1">
  <atomArray>
   <atom id="a1" elementType="C" x2="-2.9149999618530273" y2="0.769999
   <atom id="a2" elementType="C" x2="-1.5813208400249916" y2="1.539999
   <atom id="a3" elementType="O" x2="-0.24764171819695613" y2="0.769999
   <atom id="a4" elementType="O" x2="-1.5813208400249912" y2="3.079999980
   <atom id="a5" elementType="H" x2="-4.248679083681063" y2="1.539999980924
   <atom id="a6" elementType="H" x2="-2.914999961853028" y2="-0.770000019073486
   <atom id="a7" elementType="H" x2="-4.248679083681063" y2="-1.907348645691087E-8" />
   <atom id="a8" elementType="H" x2="1.0860374036310796" y2="1.5399999809265132" />
  </atomArray>
  <bondArray>
   <bond atomRefs2="a1 a2" order="1" />
   <bond atomRefs2="a2 a3" order="1" />
   <bond atomRefs2="a2 a4" order="2" />
   <bond atomRefs2="a1 a5" order="1" />
   <bond atomRefs2="a1 a6" order="1" />
   <bond atomRefs2="a1 a7" order="1" />
   <bond atomRefs2="a3 a8" order="1" />
  </bondArray>
 </molecule>
</cml>
```

**Intelligence:** Verifies validity of authored chemistry

Available soon:
http://research.microsoft.com/chem4word/


This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Article Authoring Add-in for Word 2007

PubMed
National Library of Medicine NLM

Hindawi

**Services:** repository deposit via SWORD

arXiv.org

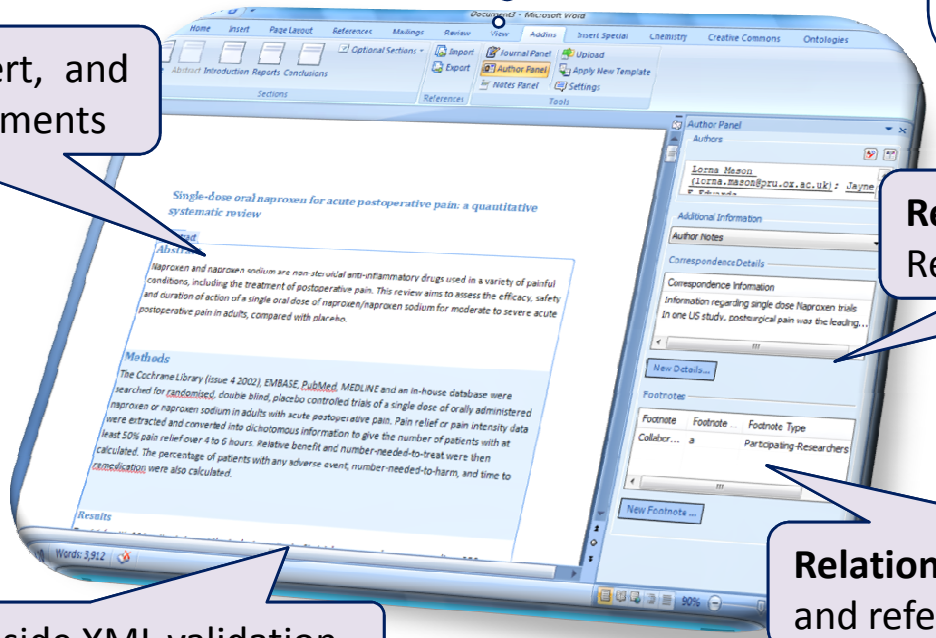
eprints

Zentity

Fedora Commons

DSPACE

**Structure:** Read, convert, and author NLM XML documents




**Relationships:** ORE Resource Map creation

**Relationships:** Citation lookup and reference management

**Structure:** Client-side XML validation

Binary (version 2.0):
http://research.microsoft.com/authoring/

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Research Information Centre

Personal site for each researcher and project site for each project

**Collaborative environment for researchers**

Federated search, co-search, annotations, tags, ratings, etc.

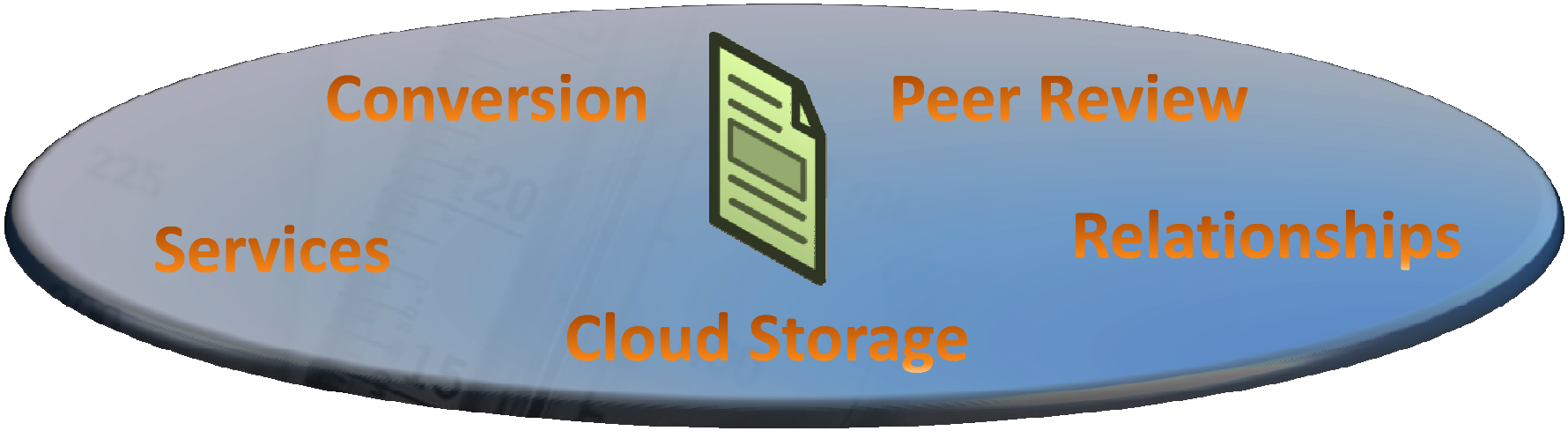
Social networking, real-time communication, blogs, wikis

Project site navigation and tool based on project lifecycle

Available soon:
http://research.microsoft.com/ric/

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Semantic Storage




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Zentity – *a Research Output Repository Platform*

Default web UI with CSS support and custom ASP.Net controls

Native support for RSS, OAI-PMH, OAI-ORE, AtomPub and SWORD

Flexible data model enables many scenarios and can be easily extended over time

**A semantic computing platform to store and expose relationships between digital assets**

Binary (version 1.0):
http://research.microsoft.com/zentity/

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Zentity – Goals

## Quick

- Easy to install
- 'Scholarly Works' data model
  - Authors, Papers, Data, Videos, Code, Lectures, Books, etc.
- Default Web UI

## Extensible

- UI Toolkit
- Intuitive programming experience
- Extensible Data Model (entities, relationships)
- RDFs for new data models

## Interoperable

- BibTeX Import
- RSS/Atom Syndication
- METS support
- OAI-PMH Provider
- OAI-ORE
- Simple Search API
- Atom Publishing Protocol
- SWORD

## Free & Open

- Freely available
- Based on open standards
- SQL Server and Developer tools available via Dreamspark

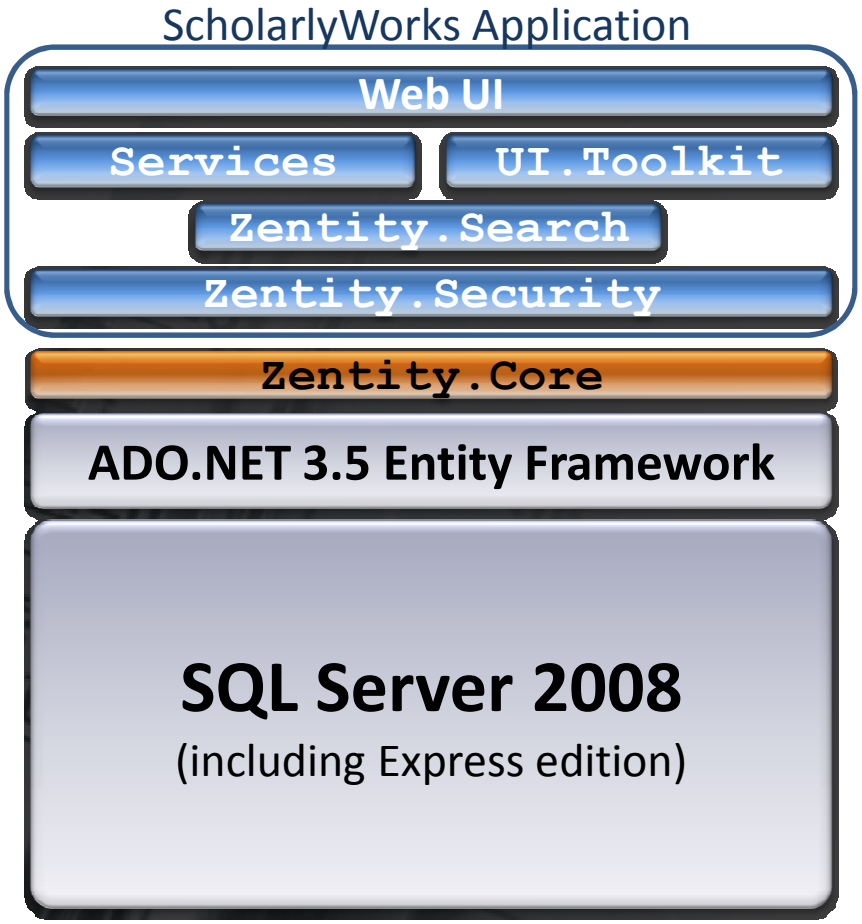
This work is licensed under a Creative Commons Attribution 3.0 United States License.

# System Requirements

- **Supported Processor Architectures**
  - x86 and x64.
- **Supported Operating Systems**
  - Microsoft Windows Server 2008 (x86 and x64)
  - Microsoft Windows Vista SP1 (x86 and x64)
- **Installation Requirements**
  - Microsoft .Net Framework 3.5
  - Supported Microsoft SQL Server
    - Microsoft SQL Server 2008 Enterprise Edition
    - Microsoft SQL Express 2008 **with Advanced Services**
- **User and Configuration Requirements**
  - Site Admin privileges are granted to the user installing Zentity
  - The selected Microsoft SQL Server instance must have “Windows Authentication” enabled.
  - User running the installer must have ‘database creation’ permissions on the Microsoft SQL Server instance.

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Application Stack

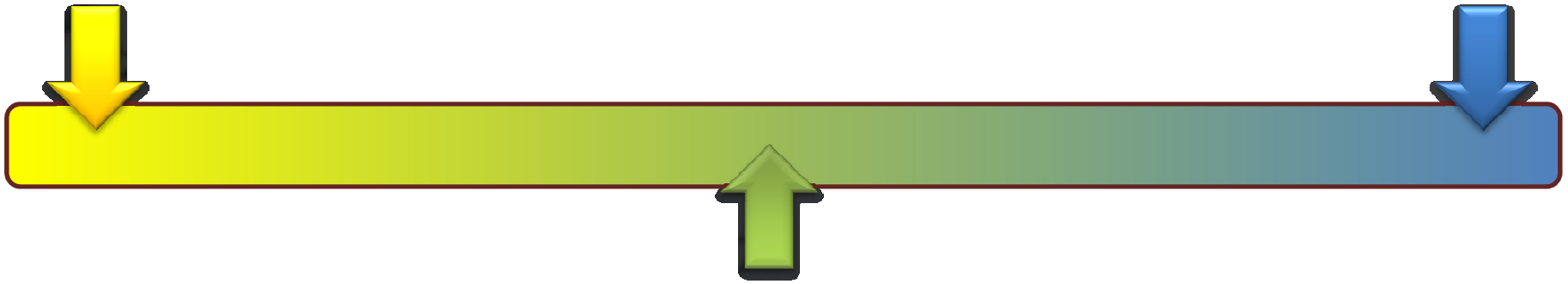



This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Zentity - Store

**Triple stores**
-Evolution friendly
-Poor performance
-No need to model everything in advance
-Semantic interpretation at the application level

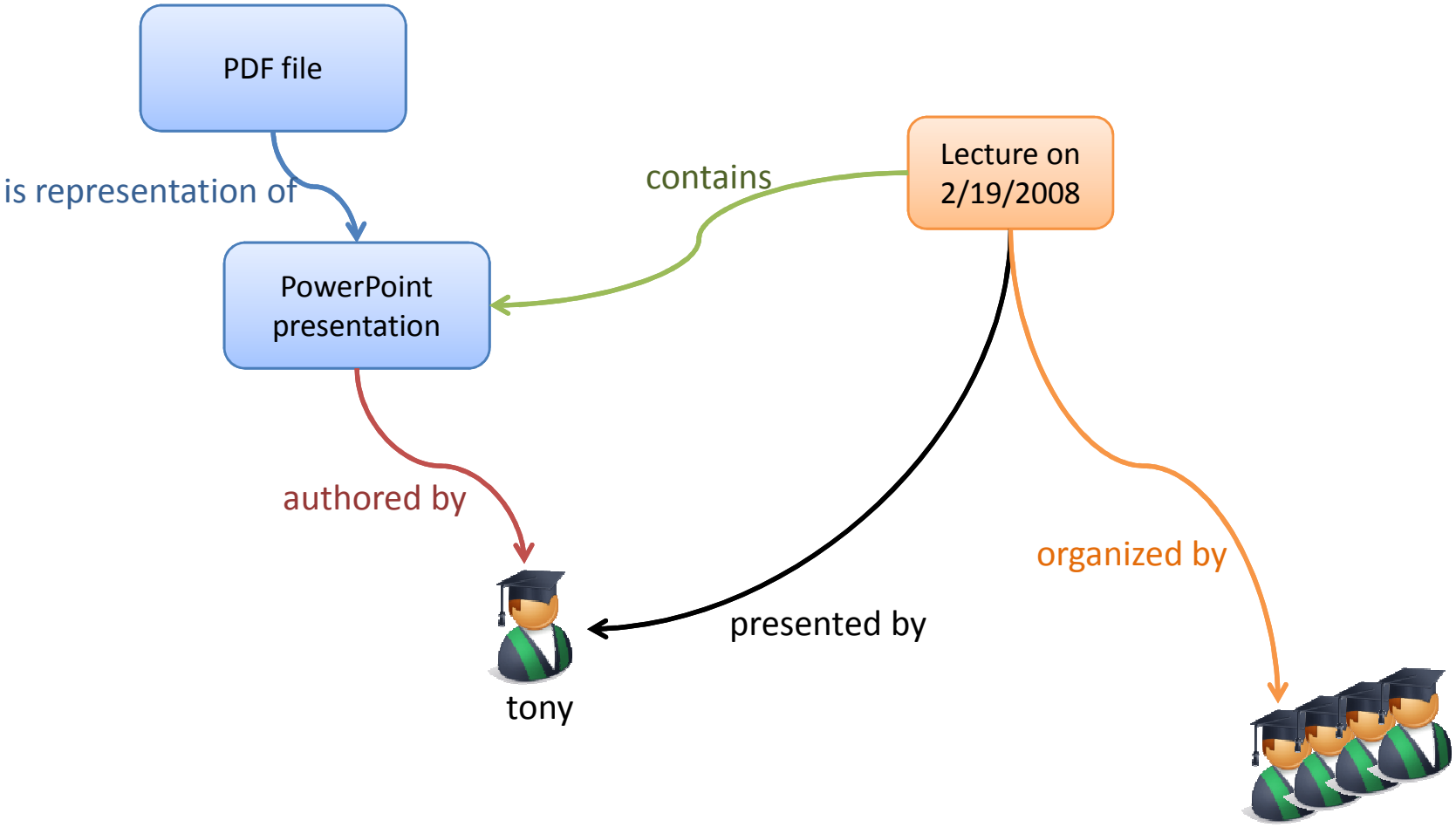
**Relational schema**
-Evolution not so easy
-Great opportunities for optimization
-Model everything in advance



**Zentity Store**
-Maintain a balance
-Try to model the frequently used entities in our app domain
-Try to capture the frequently used relationships
-Allow for extensibility (Relationships, Properties)

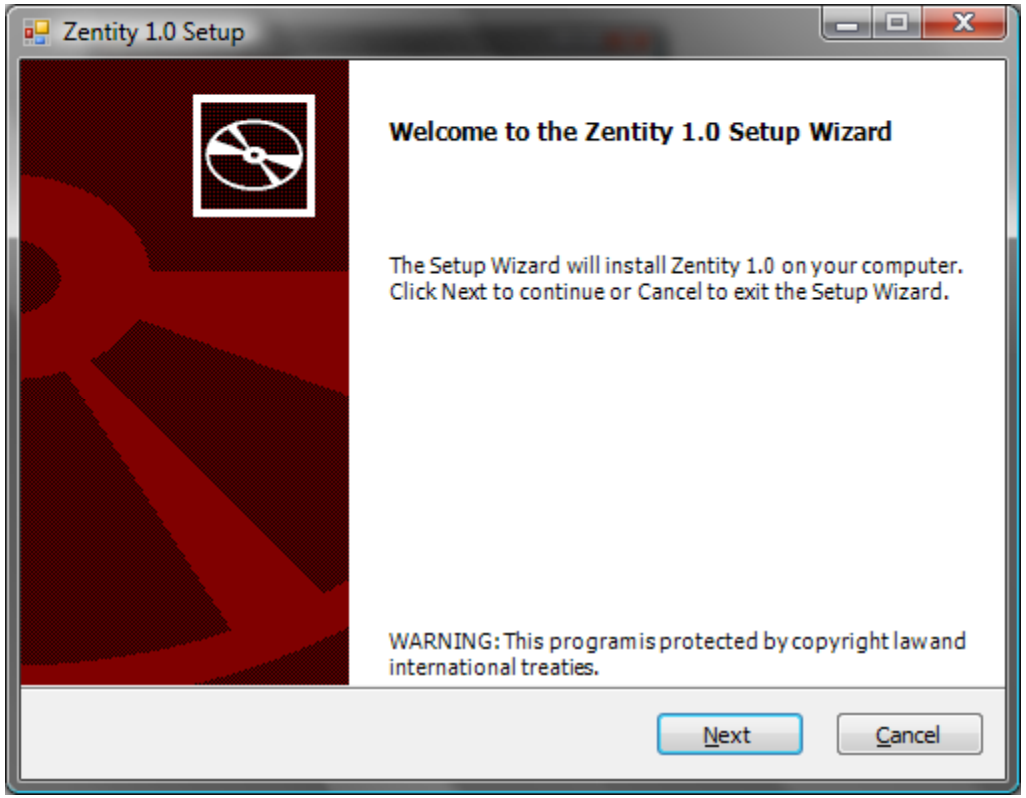
This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Research Output Repository Platform

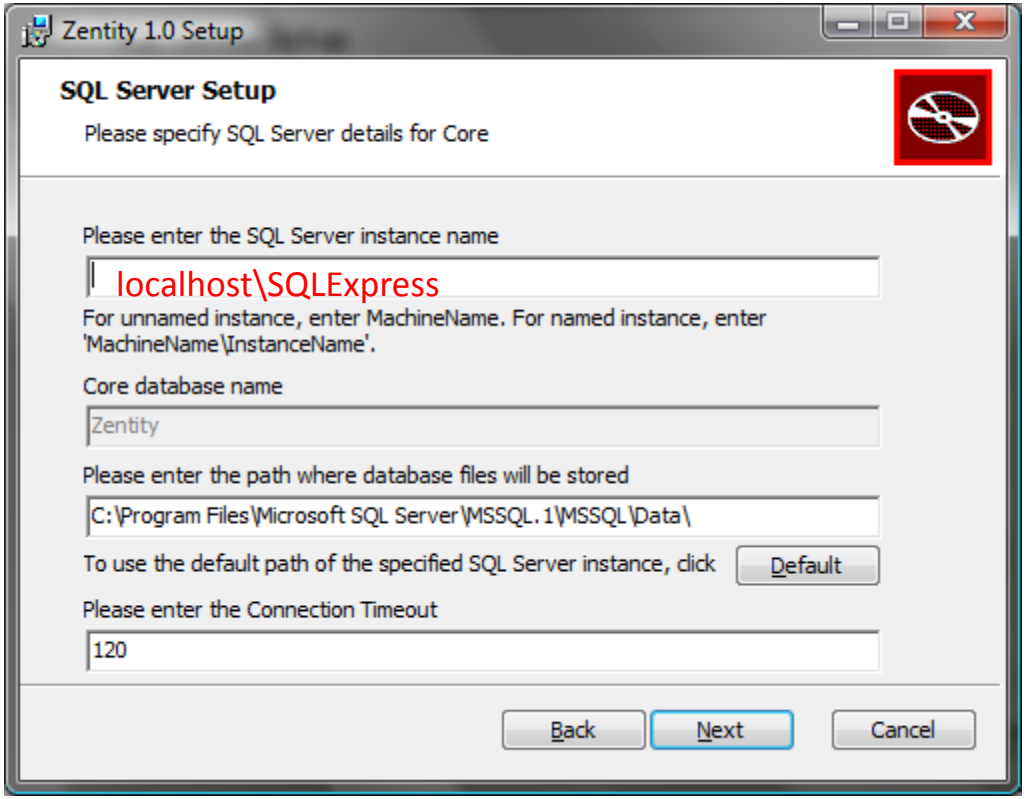



This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Installation

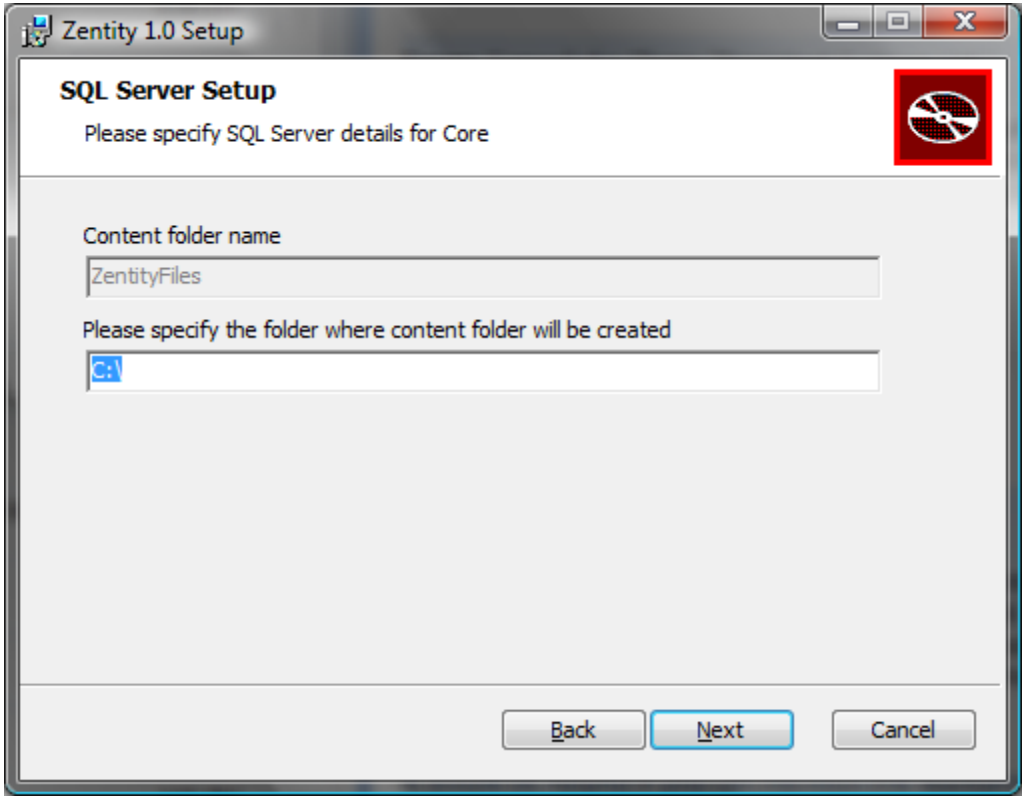



This work is licensed under a Creative Commons Attribution 3.0 United States License.



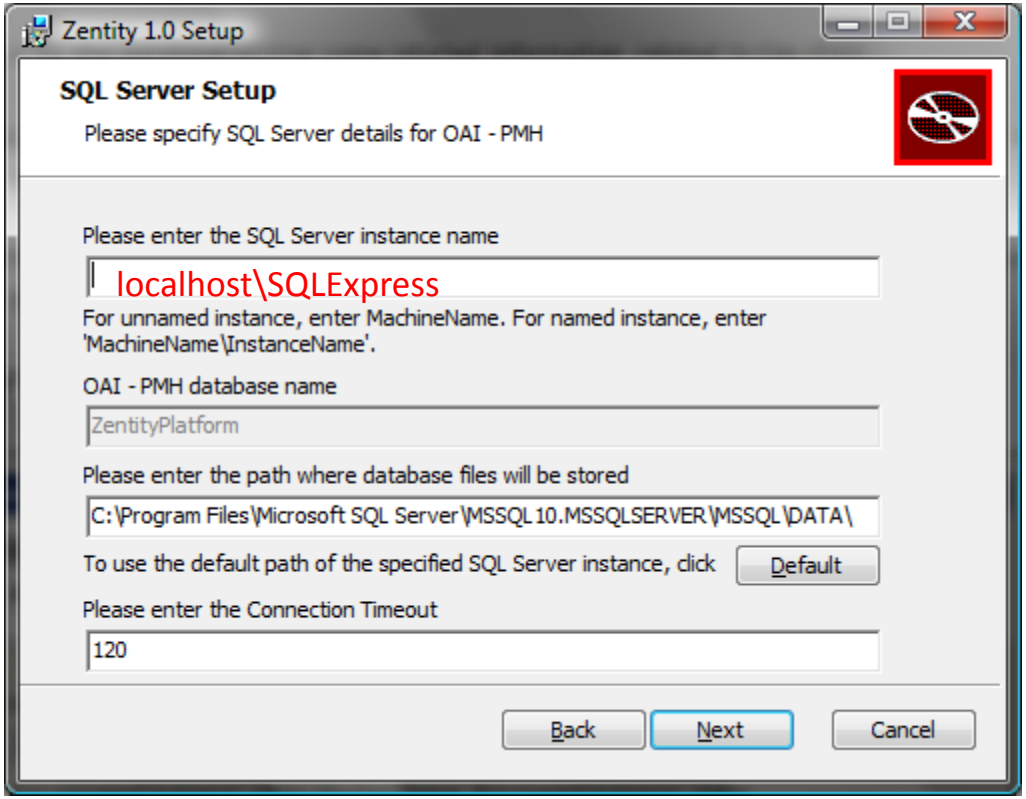

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# FILESTREAM File Location



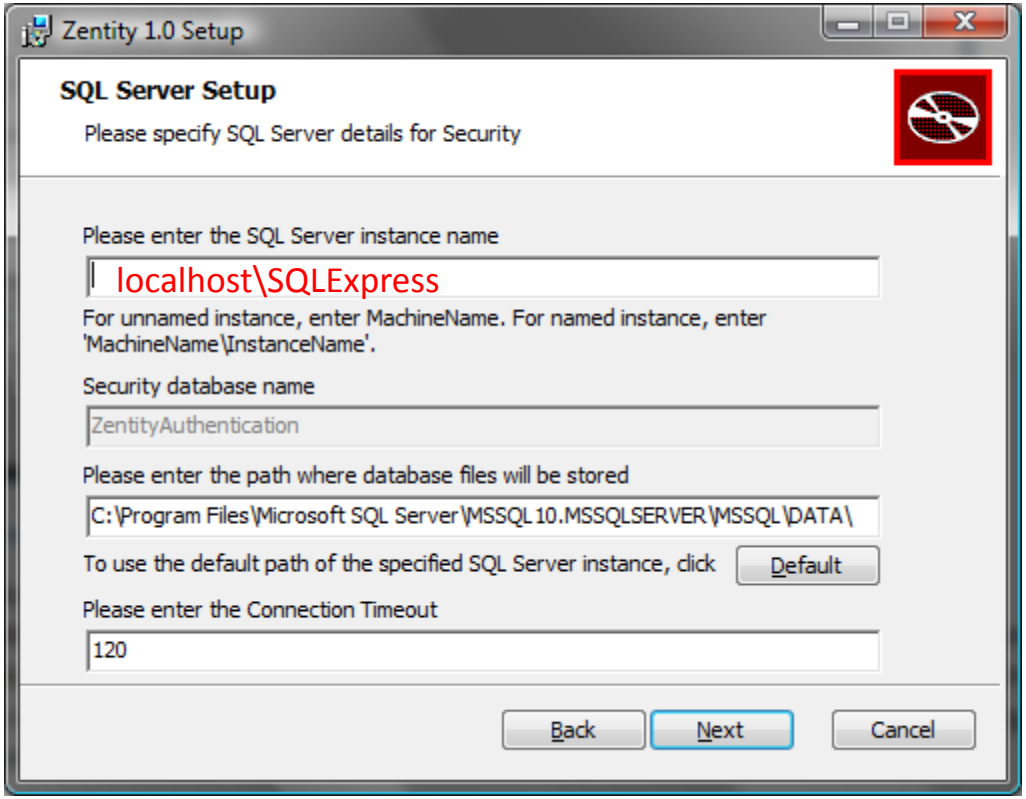

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# OAI-PMH database



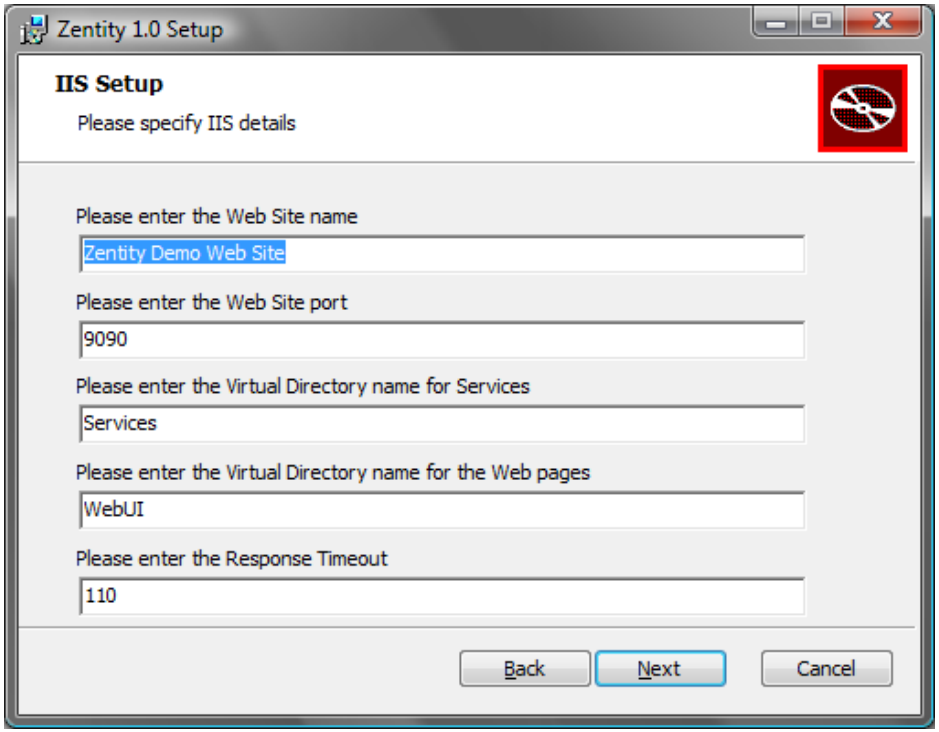

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Users/Groups Database



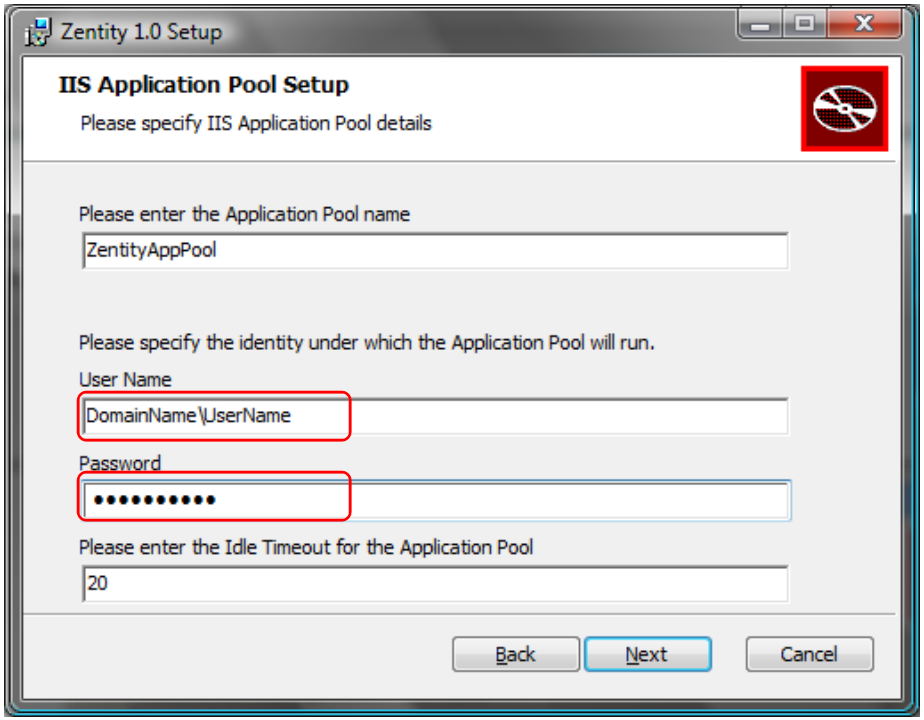

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Configure IIS

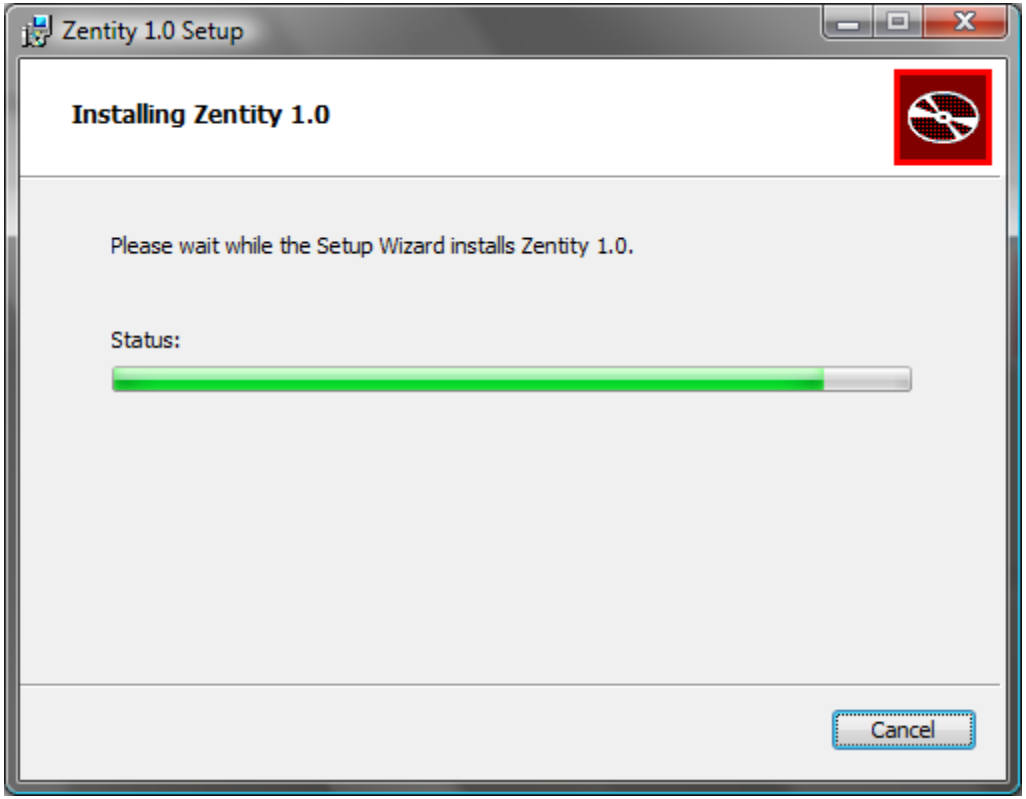



This work is licensed under a Creative Commons Attribution 3.0 United States License.

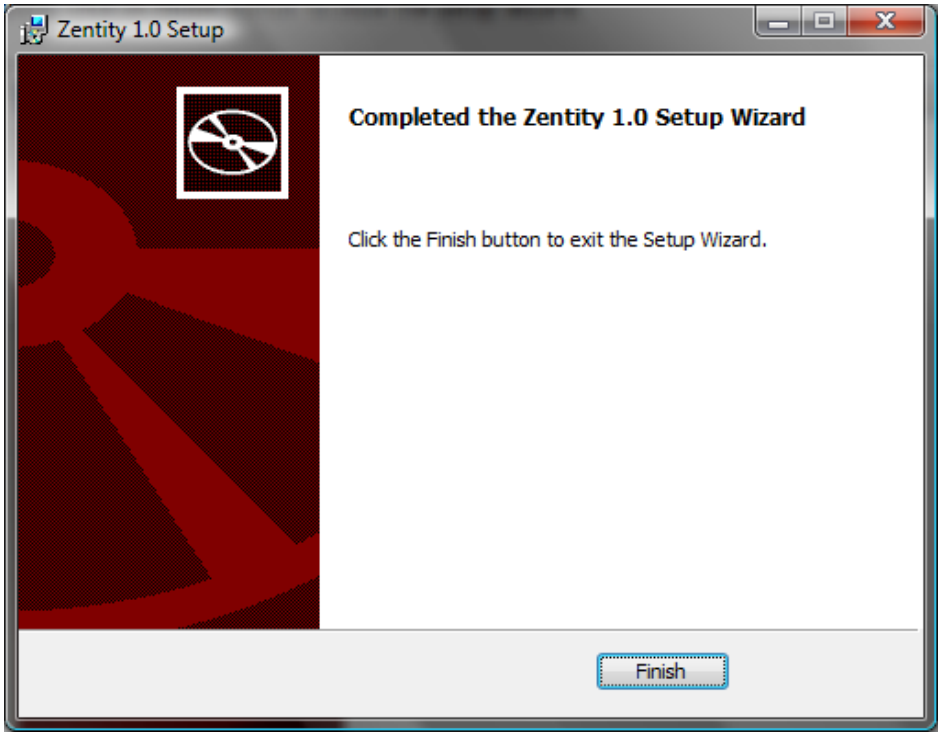
# IIS App Pool




This work is licensed under a Creative Commons Attribution 3.0 United States License.




This work is licensed under a Creative Commons Attribution 3.0 United States License.




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Zentity

John Campbell III
Profile | Articles | Help

## Browse

Browse the repository with pre-defined views. These views provide basic filter criteria to navigate through resources.

- By Year
- By Catagories
- By Type
- By Author
- Recently Added Resources
- Top Authors

## Search

Search the resources in the repository. You can search using Structured Query Syntax.

Search Zentity... GO

e.g. "Quantum Universe" Author:Tony

## Help

Use this help to learn about Zentity's many features and developer samples.

- How-to Guides
- Add-ins
- Developer References
- Zentity Forums
- About Zentity

## Manage

Add, Edit, and Delete resources from the Zentity repository.

- Resources
- Users and Roles
- Permissions
- Database Configurations

*Administrator privilages are required

## Services

Use the Open Standards compliant services provided by Zentity to access data directly in the repository database.

- RSS
- OAI-PMH
- SWORD
- AtomPub
- OAI-ORE
- RDFS

## Tags

audience audio become blogs biology century chapter computer culture dialog dissemination earth external floral form geometric internet mass media molecules newspapers papers people personal plato press print public radio research science sphere technology television today understand wade video

# Zentity

John Campbell III
Profile | Articles | Help

## Home

## Search

Go

E.g. Quantum Universe
Author:Tony

## Browse

By Year

- Year
  - 2009

By Category
By Resource Type
By Author
Recently Added Resources
Top Authors

## Manage

Add Resource
Add Tag
Edit/ Delete Resources
Edit/ Delete Tags
Manage Categories

Filters

### Categories

Department

- Department
  - Physics
  - Chemistry
  - Mathematics

### Resource Types

- ScholarlyWork
  - Lecture
  - ScholarlyWorkContaine
  - Experiment
  - Publication
  - Tutorial
- PersonalCommunication
- ElectronicSource
- Media

### Authors

A-E | F-J | K-O | P-T | U-Z | Others

david GO

David Nichols(David A
David Arthur(David A
David Arthur(David A
David Johnson(David
David Shmoys(David
David Wilson(David E
David Barber(David E
David Bargeron(Davi
David Bargeron(Davi

Prev Next

**SortBy :** Title / Date Added

**Title:** A brief report of the SIGPLAN/... (JournalArticle) **Date Added:** 5/11/2009 5:56:23 PM
**Authors:** Michael Schroeder **Files:**
**Description:** The goal of this meeting was to bring together researchers from the area of programming languages with those from the area of operating systems for the purpose of developing insight into the relations...

**Title:** Exploiting Semantic Query Cont... (Proceedings) **Date Added:** 5/11/2009 5:48:57 PM
**Authors:** S. Cucerzan, Z. Zhuang **Files:**
**Description:** One challenge for relevance ranking in Web search is underspecified queries. For such queries, top-ranked documents may contain information irrelevant to the search goal of the user; some newly-create...

**Title:** Evaluating fuzzy clustering fo... (Proceedings) **Date Added:** 5/11/2009 5:47:47 PM
**Authors:** M.E.S. Mendes, L. Sacks **Files:**
**Description:**

**Title:** Essential Pages (TechnicalReport) **Date Added:** 5/11/2009 5:47:30 PM
**Authors:** Darko Kirovski, Cherian Mathew, ... **Files:** Essential Pages
**Description:** Results to Web search queries are ranked using heuristics that typically analyze the global link topology

# Search

- Basic Search
- Search Filters
- Advanced Query Syntax (AQS)
  - Field Support
- Advanced Search

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Syndication

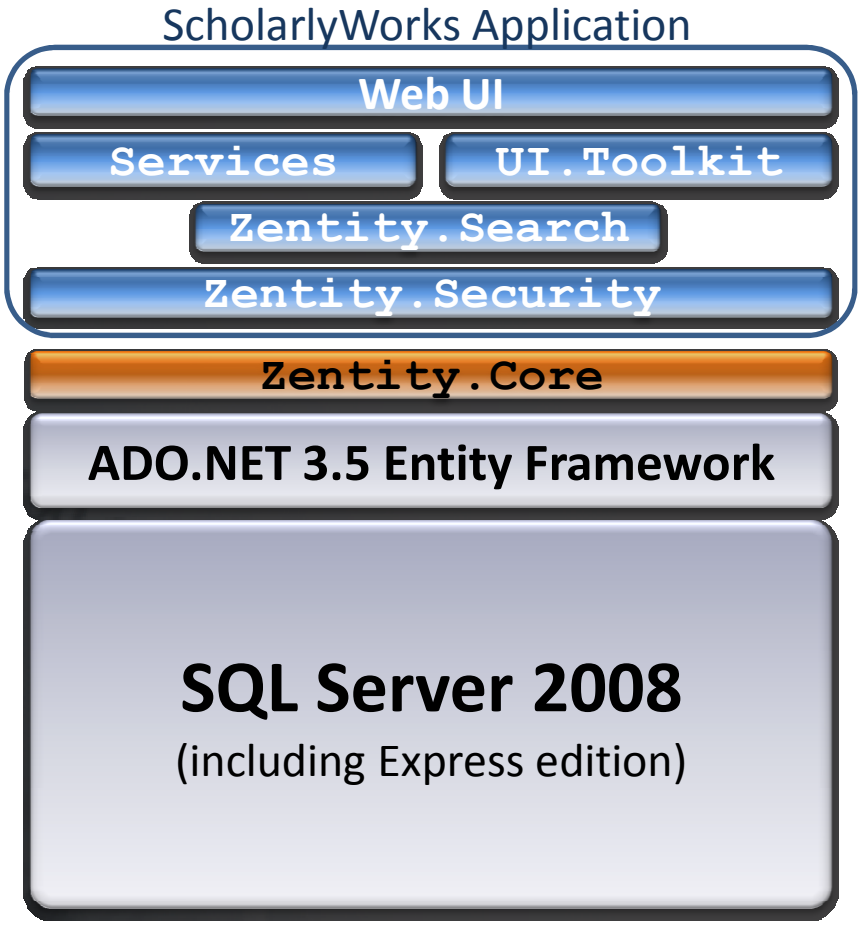
- http://localhost:9090/Syndication/Syndication.ashx?resourcetype: book author:(tony hey)

- Any search or browse view can also be exposed as RSS feed

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Extensibility

- Web UI & UI Toolkit
  - CSS
  - ASP.NET Controls
- Services
- Search
- Security
- Data Model




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Museum Scenario (Zentity.Core)

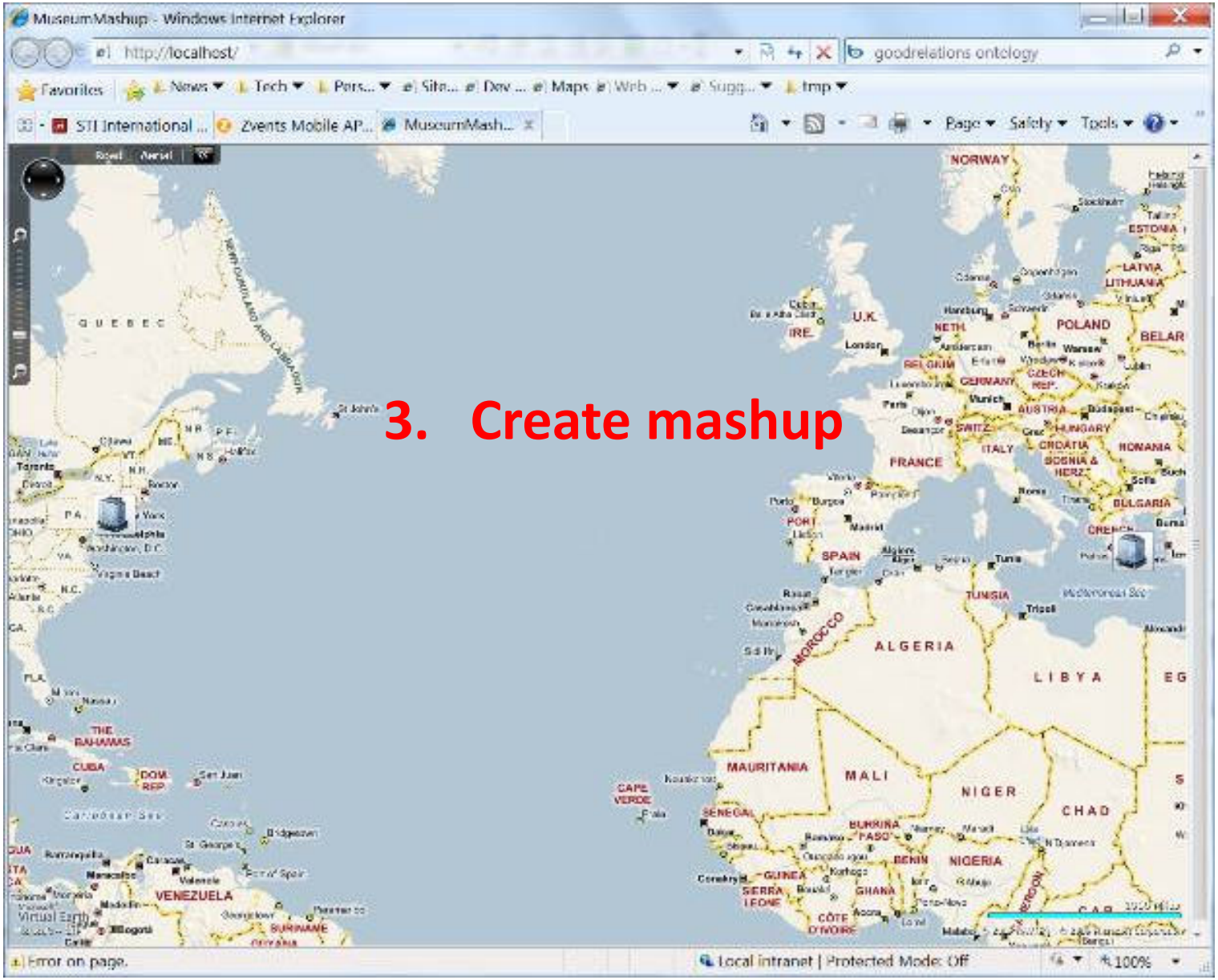
```
<?xml version="1.0" ?>
<!DOCTYPE rdf:RDF (View Source for full doctype...)>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:m="http://museum.example.org" xml:base="http://museum.example.org">
    <!-- Museum -->
    <rdfs:Class rdf:about="Museum" />
  + <rdf:Property rdf:about="Longtitude">
  + <rdf:Property rdf:about="Latitude">
  + <rdf:Property rdf:about="Name">
  + <rdf:Property rdf:about="City">
  + <rdf:Property rdf:about="Exhibitions">
  + <rdf:Property rdf:about="Collections">
  + <rdf:Property rdf:about="CollectionsOnLoan">
    <!-- Exhibition -->
    <rdfs:Class rdf:about="Exhibition" />
  + <rdf:Property rdf:about="Title">
  + <rdf:Property rdf:about="FromDate">
  + <rdf:Property rdf:about="ToDate">
  + <rdf:Property rdf:about="ShowCollections">
    <!-- Collection -->
    <rdfs:Class rdf:about="Collection" />
  + <rdf:Property rdf:about="Title">
  + <rdf:Property rdf:about="Artifacts">
    <!-- Artifact -->
    <rdfs:Class rdf:about="Artifact" />
  + <rdf:Property rdf:about="Title">
  + <rdf:Property rdf:about="Year">
  + <rdf:Property rdf:about="Image">
  + <rdf:Property rdf:about="Creator">
    <!-- Artist -->
    <rdfs:Class rdf:about="Artist" />
  + <rdf:Property rdf:about="FirstName">
  + <rdf:Property rdf:about="LastName">
  + <rdf:Property rdf:about="Born">
  + <rdf:Property rdf:about="Died">
```

1. **Define domain schema in RDFS**

# Museum Scenario (Zentity.Core)

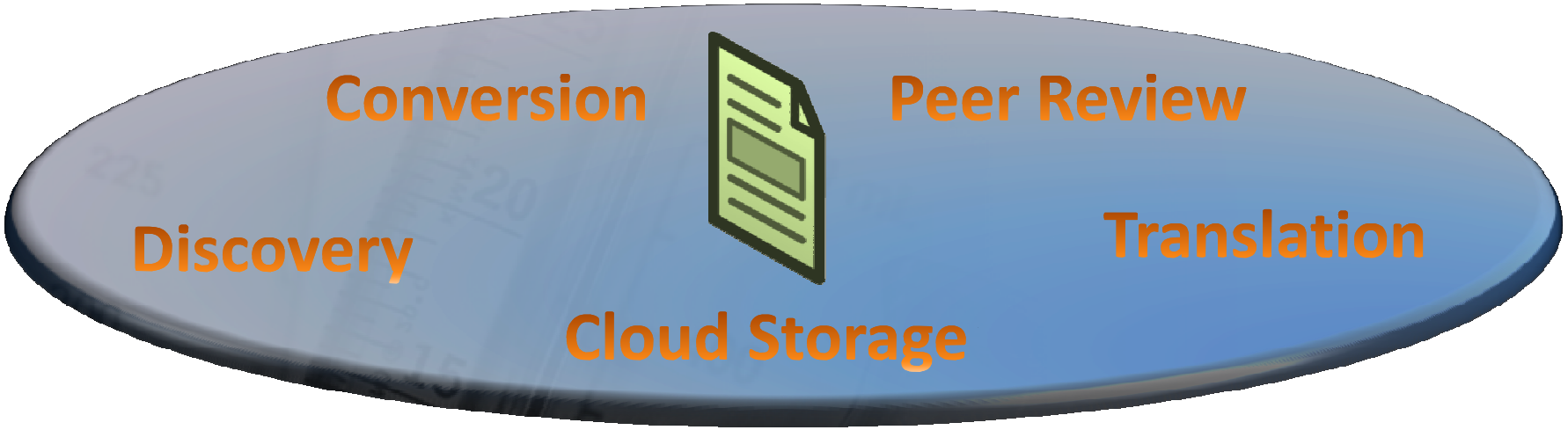
```
<?xml version="1.0" encoding="utf-8" ?>
- <Data>
  - <Museum>
      <Name>Athens National Museum</Name>
      <City>Athens</City>
      <Longtitude>23.71673</Longtitude>
      <Latitude>37.98236</Latitude>
      <Exhibition>The treasures of Greece</Exhibition>
      <Collection>Ancient Greece</Collection>
      <CollectionOnLoan />
    </Museum>
  - <Museum>
      <Name>Museum of Modern Art</Name>
      <City>New York City</City>
      <Longtitude>-73.97764</Longtitude>
      <Latitude>40.76132</Latitude>
      <Exhibition />
      <Collection />
      <CollectionOnLoan />
    </Museum>
  - <Collection>
      <Title>Ancient Greece</Title>
    - <Artifact>
        <Title>Statue of Dias</Title>
        <Year>2000 BC</Year>
        <Artist />
        <Image />
        <Artist>1</Artist>
      </Artifact>
    - <Artifact>
        <Title>Statue of Artemis</Title>
        <Year>1000 BC</Year>
        <Artist>1</Artist>
```

## 2. Import data




This work is licensed under a Creative Commons Attribution 3.0 United States License.

# An Ecosystem of Services




This work is licensed under a Creative Commons Attribution 3.0 United States License.

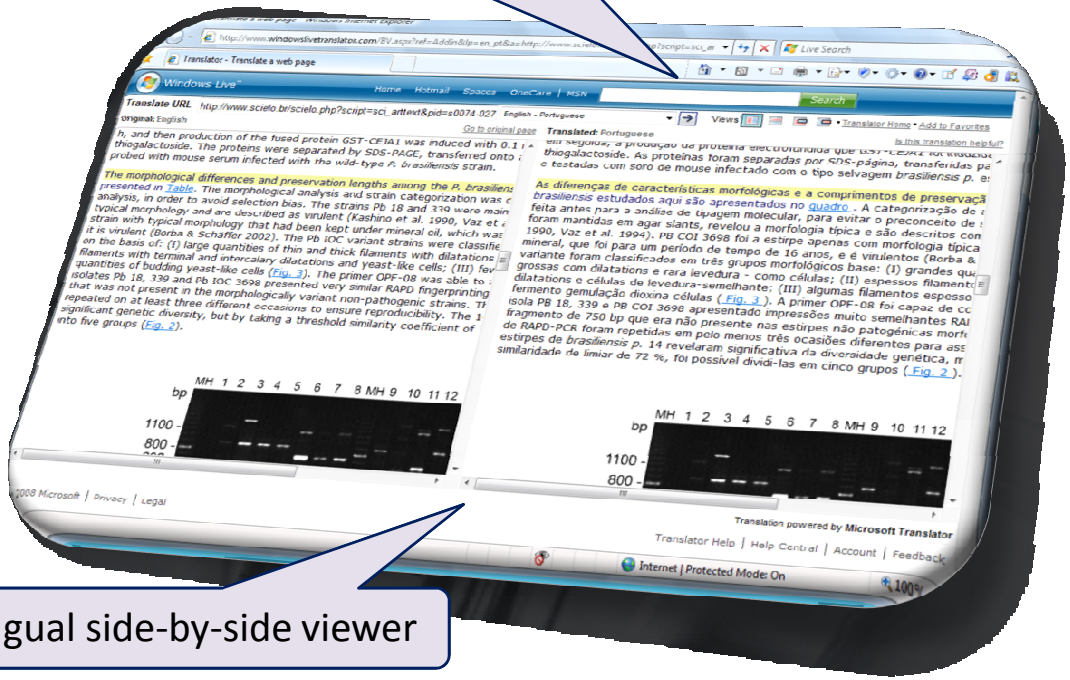
# Microsoft Translator

Microsoft® Translator

SciELO Brazil

JAIRO Japanese Institutional Repositories Online

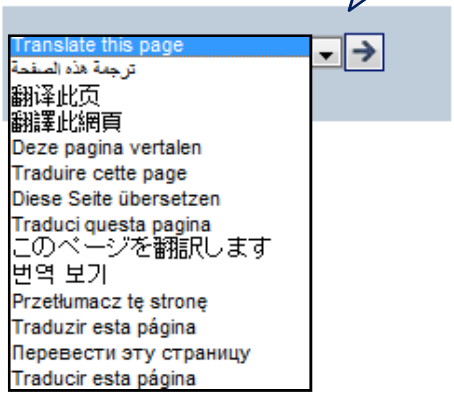
Query-time translation



Bilingual side-by-side viewer

Embeddable widget

**Your viewers will see this**




http://www.microsofttranslator.com/AddIn.aspx

http://www.microsofttranslator.com/dev/ajax/

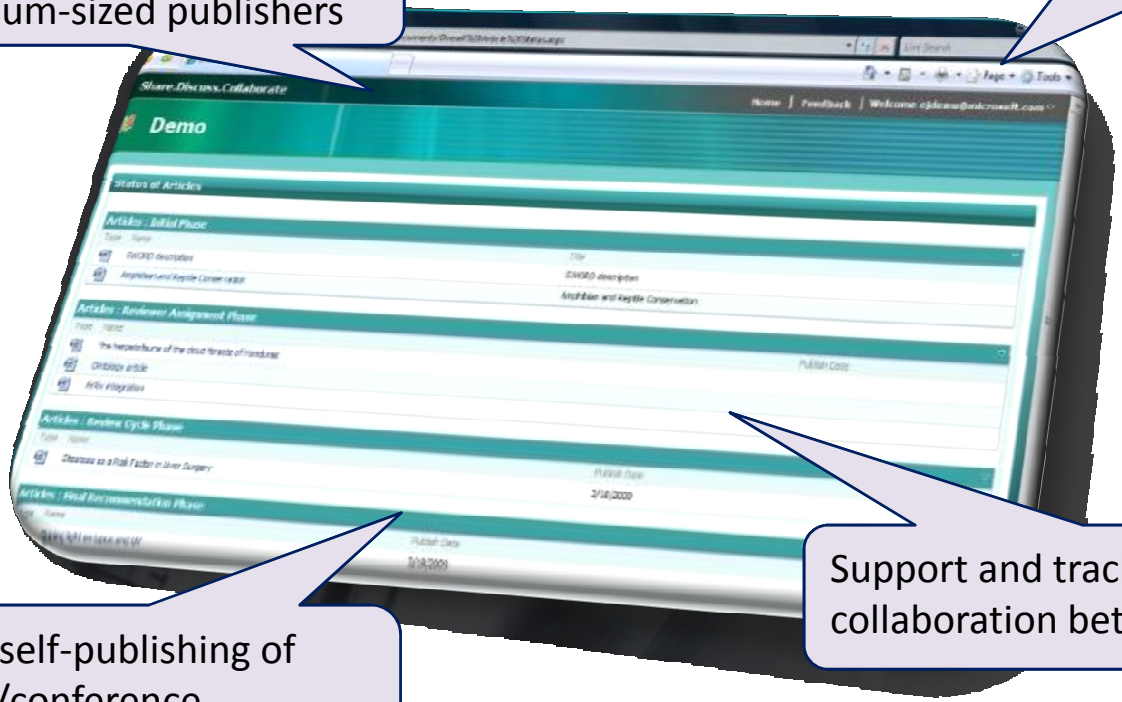
This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Microsoft Electronic Journals Service

## A Hosted Offering for the Scholarly Community

Targeted at scholarly societies and small to medium-sized publishers

**Hosted editorial and peer review management tool**




Support and tracks online collaboration between authors

Simplifies self-publishing of workshop/conference proceedings and small journals

Alpha version available at:

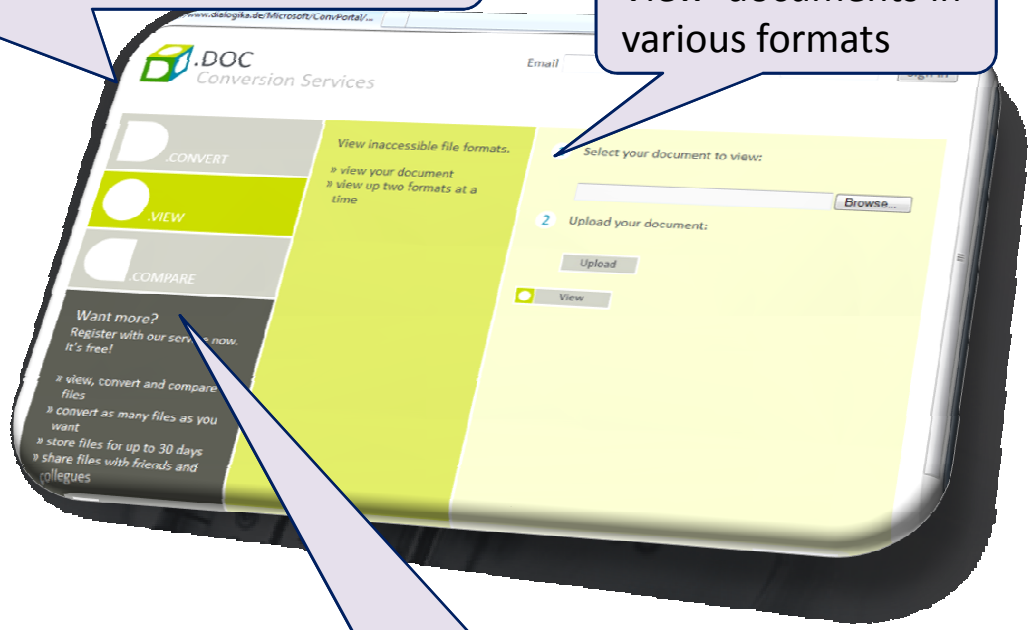
http://research.microsoft.com/ejournal/

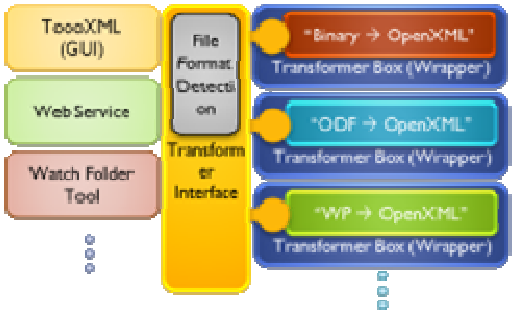
This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Document Conversion Service

Convert to and from Word, ODF, Word Perfect , RichText, and UOF

View documents in various formats




This work is licensed under a Creative Commons Attribution 3.0 United States License.

http://odf-converter.sourceforge.net/

# Open Document Standards

## May 18th announcement

- New project seeks to eliminate Open XML confusion and build interoperability
- Microsoft working with the Fraunhofer Institute for Open Communication Systems FOKUS in Berlin
- Building a document format test library and validation tool
- Tools will ease the effective exchange of data and improve the long-term benefits for data archiving

- At the Document Interoperability Initiative (DII) global forum in London, release of a number of products to support interoperable files:

- Open XML Document Viewer v1.0, a plug-in for the Opera browser to help users access documents via the web or across mobile devices
- The Apache POI 3.5 software development kit, includes a Java API to access information in the Open XML Format.
- The Open XML-ODF translator, has support for .XLS and .PPT file formats, improved ability to translate between ODF and Open XML formats.

This work is licensed under a Creative Commons Attribution 3.0 United States License.

# oreChem – *The Chemical Semantic Web*

PENNSTATE

- Lee Giles
- Karl Mueller
- Prasenjit Mitra

INDIANA UNIVERSITY

- Geoffrey Fox

Cornell University

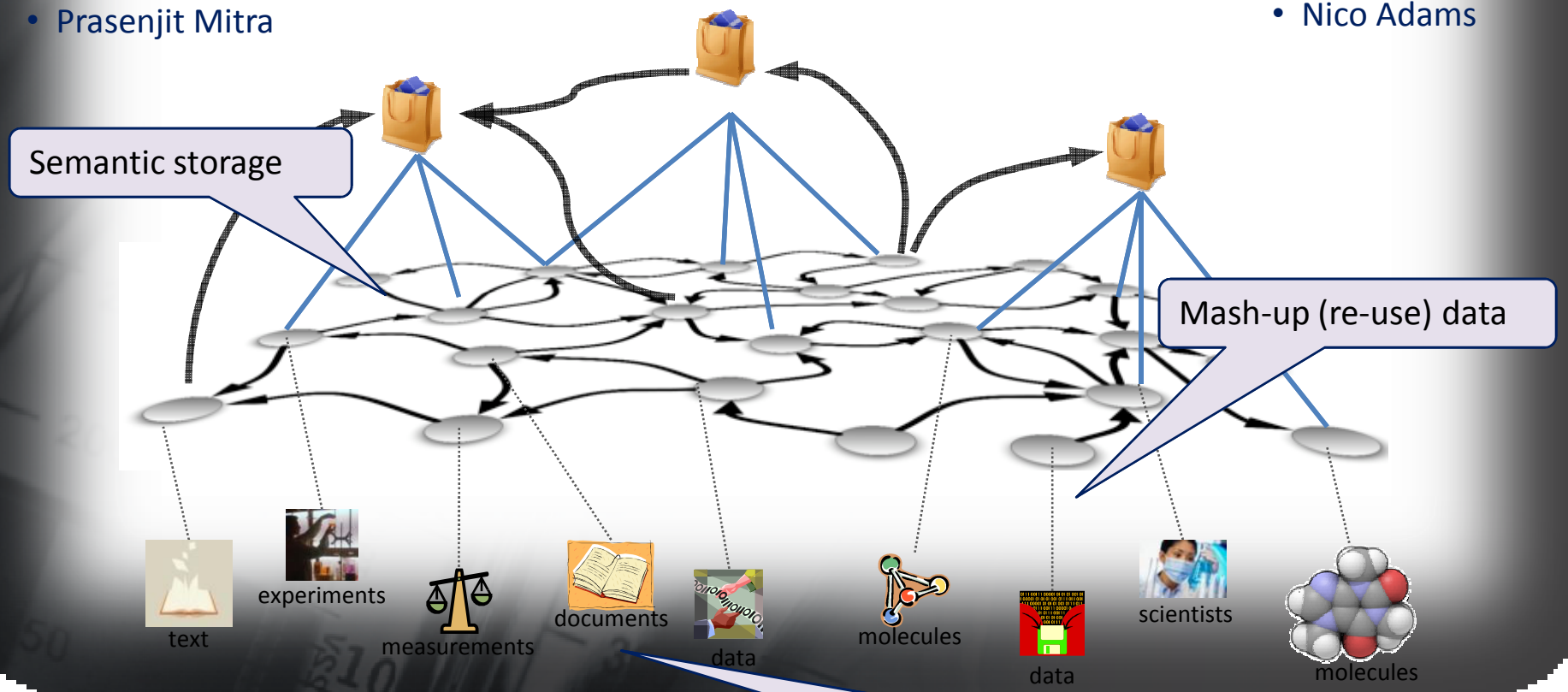
- Carl Lagoze

UNIVERSITY OF Southampton School of Chemistry

- Jeremy Frey
- Simon Coles

UNIVERSITY OF CAMBRIDGE

- Peter Murray-Rust
- Jim Downing
- Nico Adams

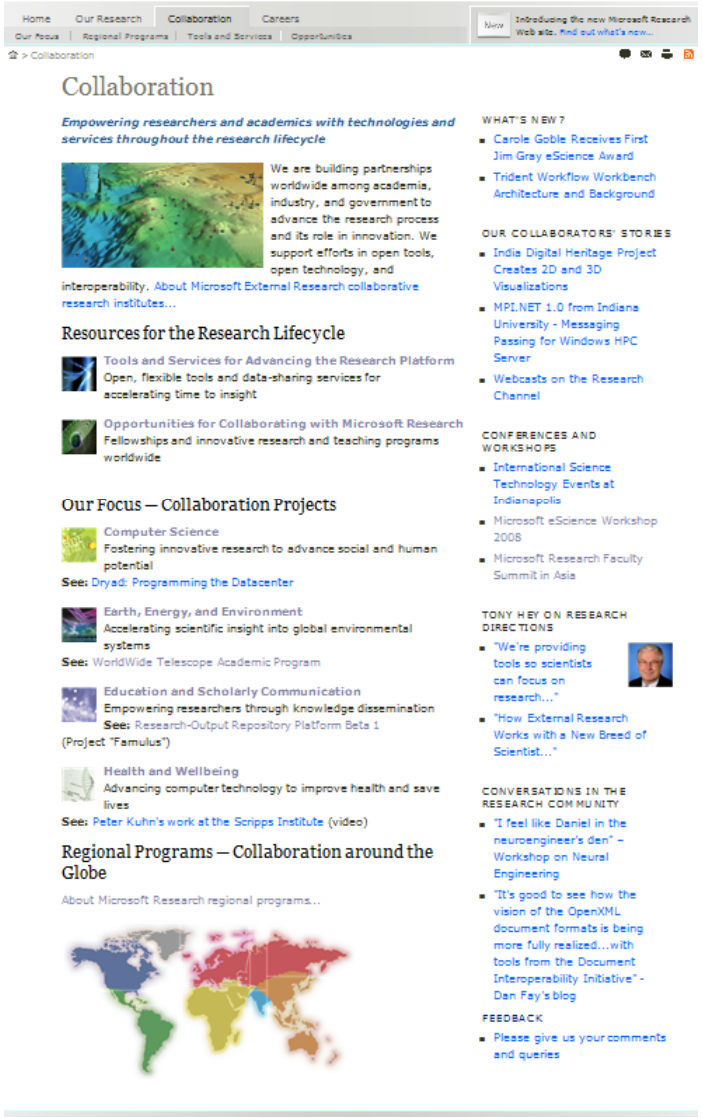



This work is licensed under a Creative Commons Attribution 3.0 United States License.

# Further Information and Resources

## http://research.microsoft.com

- The site contains access and downloads of relevant tools and resources for the worldwide academic research community. A small set of examples include:
  - **Research Output Repository**: building blocks, tools, and services for developers who are tasked with creating and maintaining an organization's repository ecosystem. http://research.microsoft.com/zentity
  - **Tools and Services for Research Collaboration**: http://www.microsoft.com/scholarlycomm/




This work is licensed under a Creative Commons Attribution 3.0 United States License.