Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

Transforming Scholarly Communication

Overview of Recent Projects from Microsoft Research

Lee Dirks
Director, Education & Scholarly Communication
Microsoft Research | Connections

Dr. Camilo A. Acosta-Márquez
Director Centro de Robótica e Informática
Universidad Jorge Tadeo Lozano

Transforming Scholarly Communications

Overview of Recent Projects from Microsoft Research

Lee Dirks
Director, Education
& Scholarly Communication
Microsoft Research | Connections

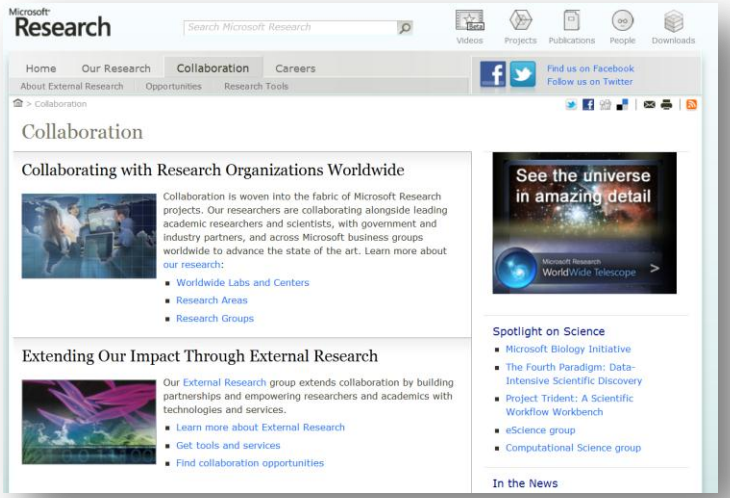
Dr. Camilo A. Acosta-Márquez
Director Centro de Robótica e Informática
Universidad Jorge Tadeo Lozano

Microsoft Research Connections

Microsoft Research | **Connections**

Outreach. Collaboration. Innovation.

- Division within Microsoft Research focused on **partnerships** between academia, industry and government to advance computer science, education, and research in fields that rely heavily upon advanced computing
- Supporting **groundbreaking research** to help advance human potential and the wellbeing of our planet
- Developing **advanced technologies and services** to support every stage of the research process
- Microsoft Research Connections is **committed to interoperability** and to providing open access, open tools, and open technology



http://research.microsoft.com/collaboration/

Microsoft®

Research Connections

Community and Geographic Outreach

Core Computer Science

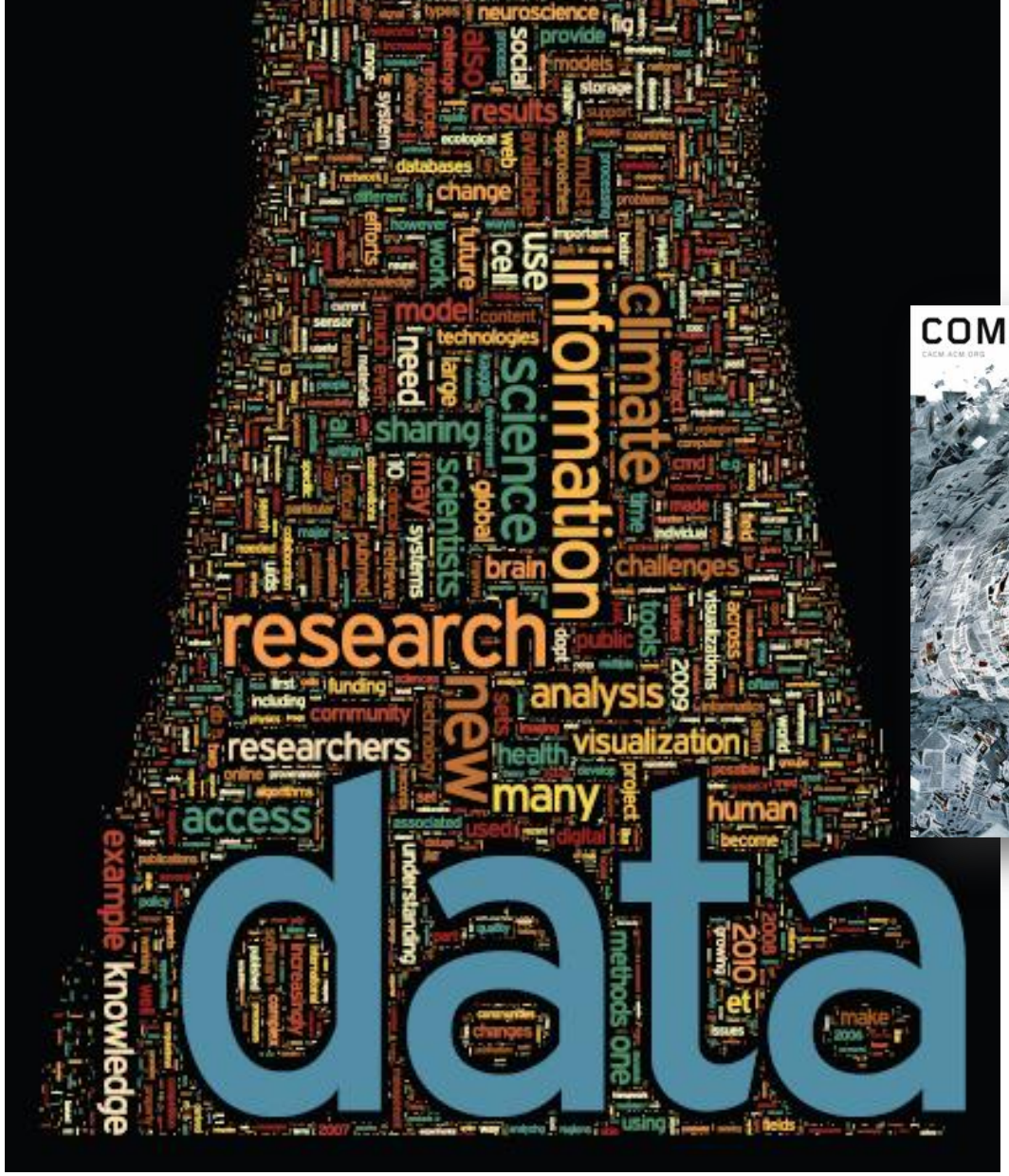
Earth, Energy and Environment

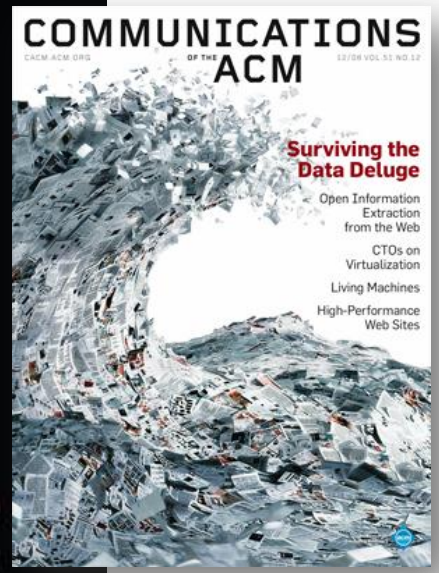
Education & Scholarly Communication

Health & Wellbeing

Natural User Interface

Advanced Research Tools and Services

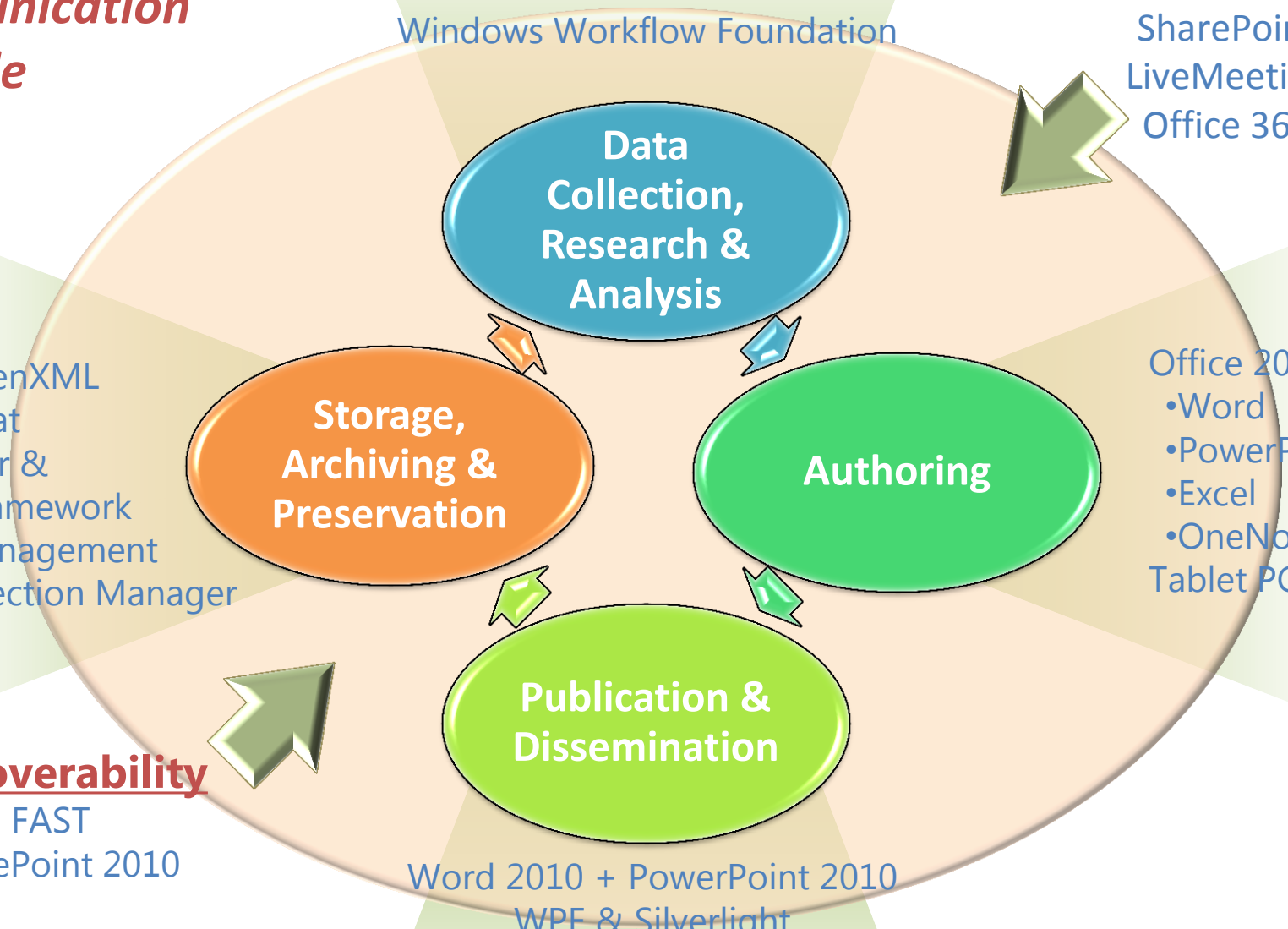

data
research
information
science
climate
analysis
visualization
researchers
access
many
human
challenges
sharing
model
results
neuroscience
provide
models
storage
databases
change
technologies
need
scientists
brain
public
tools
community
funding
health
new
example
knowledge
understanding
methods
one
2010
2009
2008
2007
make
using
fields
used
digital
become


COMMUNICATIONS OF THE ACM
CACM.ACM.ORG
12/08 VOL.51 NO.12
Surviving the Data Deluge
Open Information Extraction from the Web
CTOs on Virtualization
Living Machines
High-Performance Web Sites

nature
BIG DATA
SCIENCE IN THE PETABYTE ERA

The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

The Scholarly Communication Lifecycle



Excel 2010
Windows Server HPC
Windows Workflow Foundation

Collaboration
SharePoint
LiveMeeting
Office 365

Office OpenXML
XPS Format
SQL Server &
Entity Framework
Rights Management
Data Protection Manager

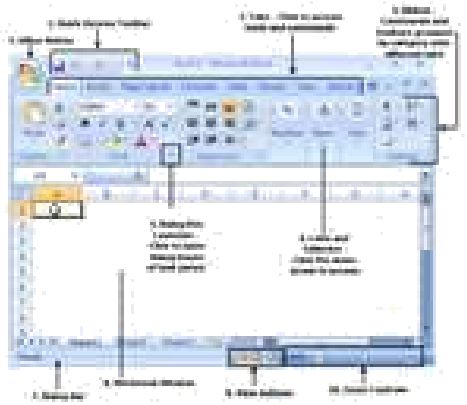
Office 2010:
- Word
- PowerPoint
- Excel
- OneNote

Tablet PC/UMPC

Discoverability
FAST
SharePoint 2010

Word 2010 + PowerPoint 2010
WPF & Silverlight
"Sea Dragon" / "PhotoSynth" / "Deep Zoom"

Data Curation Add-in for Microsoft Excel



University of California
CDL
California Digital Library

DataONE

- **Microsoft Research, in partnership with California Digital Library's Curation Center**
  - Collaboration with **Trisha Cruse & John Kunze**
  - Part of the **DataONE** (an NSF DataNet Project)
- **Proposed functionality *under consideration*:**
  - **Support for versioning**, so that revision history and the original raw data can be easily protected and recovered,
  - **Standardized date/time stamps** so that researchers can easily determine when the data were created and last updated.
  - **A "workbook builder"** allowing researchers to select from globally shared standardized layouts for capturing data,
  - **Ability to export metadata in a standard format** (e.g., a DataCite citation or an EML document that describes the dataset(s) in a workbook) so that researchers can readily share their data,
  - **Ability to select from a globally shared vocabulary of terms for data descriptions** (e.g., column names), and as needed to add new terms to the globally shared vocabulary, to enable wide collaboration between researchers
  - **Ability to import term descriptions from the shared vocabulary and annotate them locally** to refine their definitions as used in the dataset,
  - **"Speed bumps" to discourage use of macros and customizations** that would impede interoperation of data imported from Excel into other applications, and
  - **Ability to deposit data and metadata directly into a data archive** to enable compliance with funding agency requirements to preserve and publish research data.

Project "Chem4Word"– Chemical Drawing in Word

Semantic chemistry for students and publishers

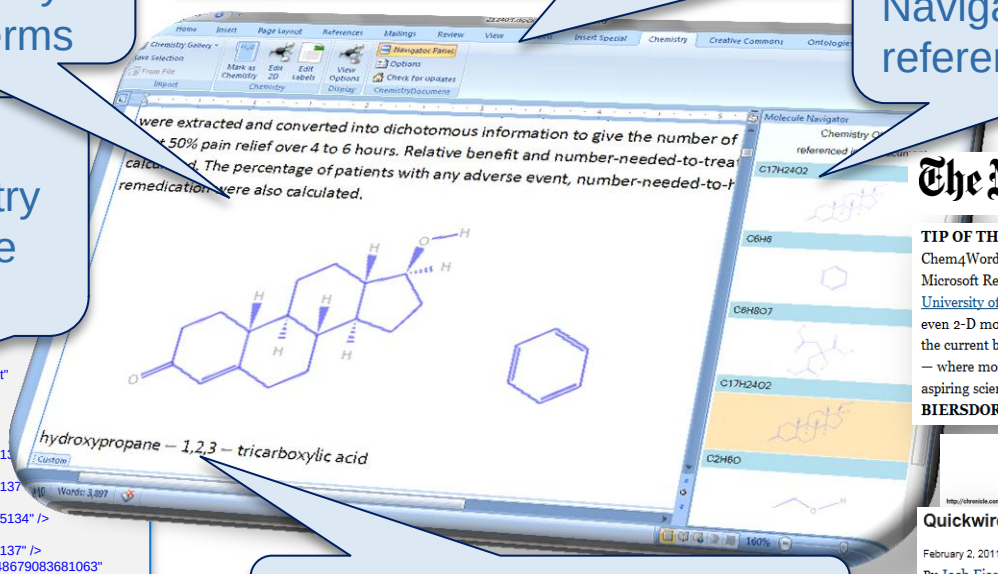
UNIVERSITY OF CAMBRIDGE

Intent: Recognizes chemical dictionary and ontology terms

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

Relationships: Navigate and link referenced chemistry

Data: Semantics stored in Chemistry Markup Language (CML)




```
<?xml version="1.0" ?>
<cml version="3" convent="org-synth-report"
xmlns="http://www.xml-cml.org/schema">
  <molecule id="m1">
    <atomArray>
      <atom id="a1" elementType="C" x2="-
2.9149999618530273" y2="0.7699999809265137" />
      <atom id="a2" elementType="C" x2="-
1.5813208400249916" y2="1.5399999809265137" />
      <atom id="a3" elementType="O" x2="-
0.24764171819695613" y2="0.7699999809265134" />
      <atom id="a4" elementType="O" x2="-
1.5813208400249912" y2="3.0799999809265137" />
      <atom id="a5" elementType="H" x2="-4.248679083681063"
y2="1.5399999809265137" />
      <atom id="a6" elementType="H" x2="-2.914999961853028"
y2="-0.7700000190734864" />
      <atom id="a7" elementType="H" x2="-4.248679083681063"
y2="-1.9073486456910 87E-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

The New York Times — Personal Tech

By J. D. BIERSDORFER
Published: April 7, 2010

TIP OF THE WEEK Chemistry students and teachers might want to check out the new Chem4Word add-on for Microsoft Word. The free software, which was developed by Microsoft Research and the Unilever Centre for Molecular Science Informatics at the University of Cambridge, allows Word users to insert chemical symbols, formulas and even 2-D models of molecules into documents. Chem4Word works with Word 2007 and the current beta version of Word 2010, and is listed as a beta version itself at bit.ly/riK33 — where more information and a demonstration video are also available for scientists, aspiring scientists and those who have chemistry papers due soon. **J. D. BIERSDORFER**

THE CHRONICLE of Higher Education — Wired Campus

Quickwire: Microsoft Word Goes Chemical

February 2, 2011, 2:50 pm
By Josh Fischman

Chem4Word, a free, open-source plug-in that lets authors draw intricate chemical structures—and store information about molecules—within their Word documents, has been released by Microsoft Research (the company's unit that collaborates with universities), the University of Cambridge, and the Outercurve Foundation.

http://chronicle.com/blogs/wiredcampus/quickwire-microsoft-word-goes-chemical/29423

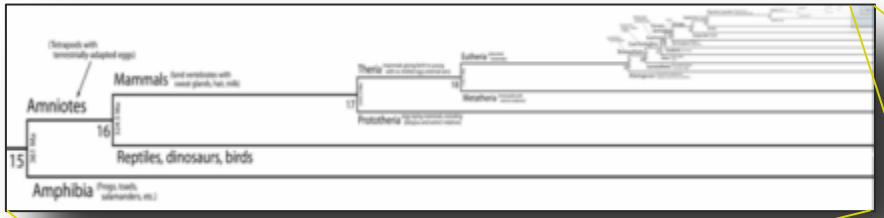
V1.0 now available (binary and open source)

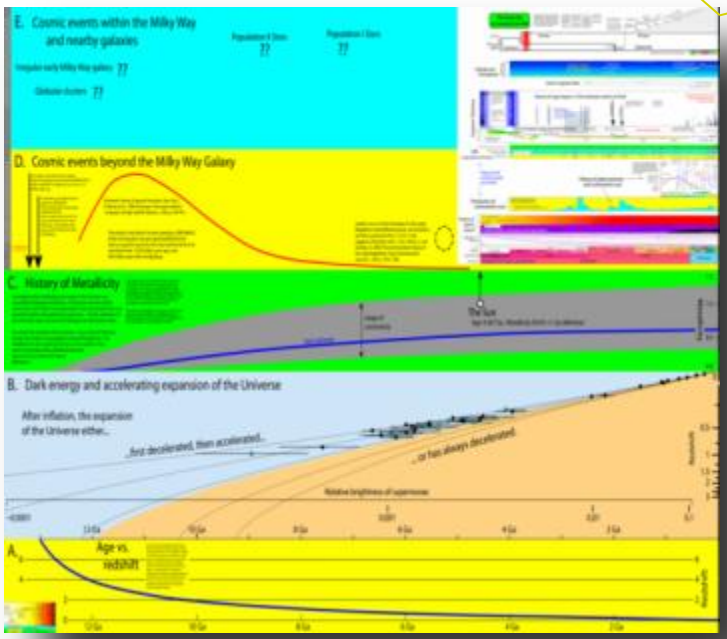
http://research.microsoft.com/chem4word/

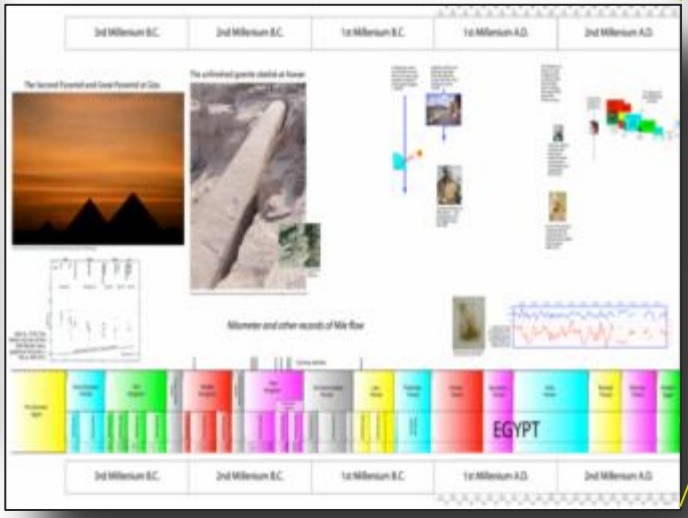
Project 'Big Time' @ Cal-Berkeley

A "zoomable" timescale platform

See the demo live at
www.chronozoomtimescale.org
Watch the video at
http://bit.ly/alvarez97







Microsoft Translator

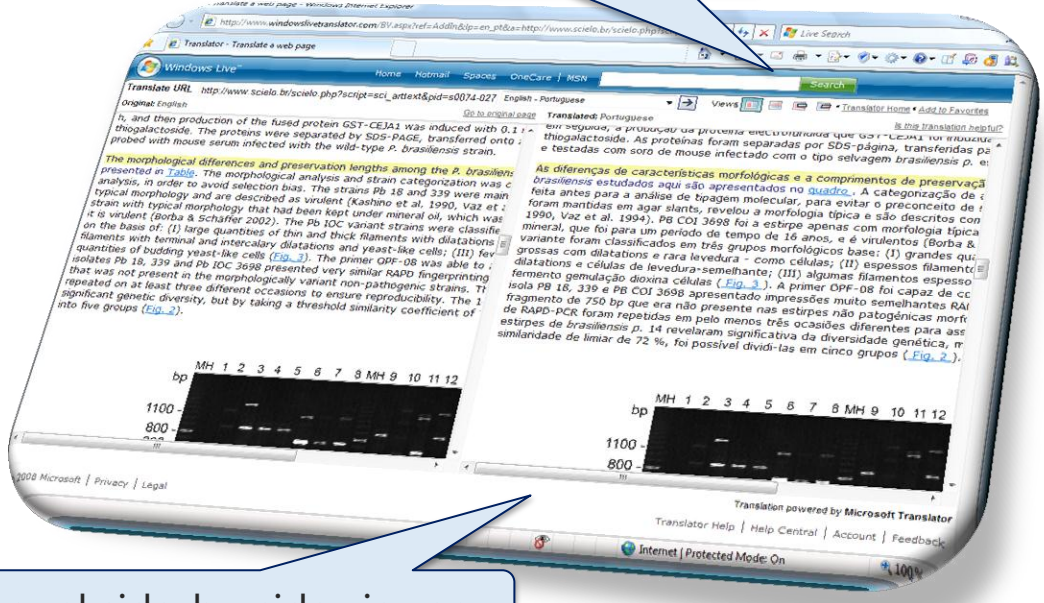
Your website/content in 30 languages

bing

Microsoft Translator

http://worldwidescience.org

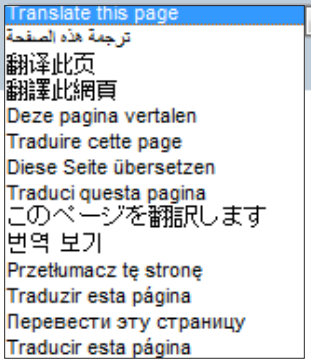
Query-time translation



Bilingual side-by-side viewer

Embeddable widget

Your viewers will see this



SciELO Brazil

JAIRO Japanese Institutional Repositories Online

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

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

Microsoft Research Connections

Research Information Centre (RIC) Project

Virtual Research Environment (VRE) Toolkit for SharePoint

Collaborative environment for research groups

Personal site for each researcher and project site for each project

Document management, federated search, social networking, real-time communication, blogs, wikis

Microsoft Office SharePoint

BRITISH LIBRARY

SoftEdge SOFTEDGE SYSTEMS

Project Overview:
http://research.microsoft.com/ric/
http://research.microsoft.com/vre/

Version 1.1 (Open Source under Ms-PL):
http://ric.codeplex.com/

ScienceCinema

Project MAVIS: powerful audio-visual search

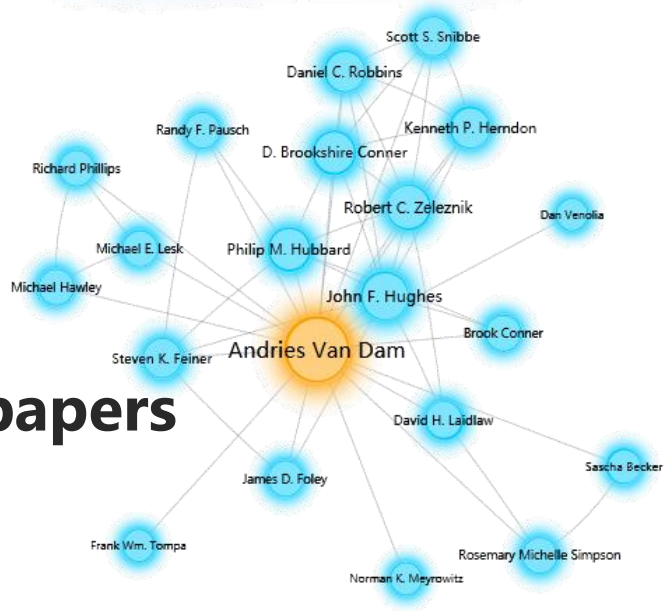
- Indexing automatic transcripts as text
  - Automatic transcription accuracy is only 50-80%
- MAVIS techniques
  - Word-level **lattice** indexing
    - Index word alternatives – robust to recognizer errors
    - 50-140% accuracy improvement
    - Index timing – navigate to exact point in video
  - Vocabulary Adaptation
    - Use NLP and Bing Search to expand word dictionary
  - Automatic keywords to expose to search engines
    - Enables discovery of speech content through search engines
    - Bi-product of vocabulary adaptation
- Department of Energy's Office of Scientific & Technical Information
  - 1,000 hours of AV content currently available
- See http://research.microsoft.com/mavis

http://www.osti.gov/sciencecinema

Microsoft Confidential

Microsoft Research – Academic Search Beta

http://academic.research.microsoft.com



- **Powerful search tool for academic papers**
- **From MSR Asia Lab (Beijing)**
  - For researchers, by researchers
- **Historically focused on Computer Science**
  - Now targeting all academic fields/domains
- **Key functionality includes**
  - Find top papers in a domain
  - Easily search the top papers, authors, conferences, and journals for a topic
  - See details about a specific paper, author, conference or journal
  - Quickly find relationships between authors (with Visual Explorer)
  - Get a related Bing Answer
  - Public API currently in testing

Project Garibaldi

Large Art Display on a Surface (LADS)

A digital humanities tool for accessing large or fragile objects – for research and/or teaching purposes

The Garibaldi Panorama: a 140 year old scroll / painting that is almost five feet tall, nearly the length of a football field, and printed on both sides.

Surface application in use at the British Library and the Harvard-Smithsonian Astrophysics Library

Currently on display at the 'Growing Knowledge' exhibit at The British Library

Unique functionality:

- Two screens for work and display running off Surface. The Surface screen contains navigation and additional information. The vertical screen removes the clutter, and is synchronized with the scrolling of the Surface screen used for sharing.
- Hovering a hand over an information hot spot in the image opens a menu that lets you access additional related documents, videos, etc.
- Flicking a document to the top of the Surface display will make it visible on the vertical display. Multiple documents, videos, etc. can be interacted with in this fashion. Web browser for external content with gesture support.
- Magnification lens via object interaction. Lens movement and sizing is duplicated onto large vertical display.
- IR-Light pen for notation and pen based gestures for moving and erasing notes.

Project overview:
http://dl.lib.brown.edu/garibaldi/about.html

Zentity – *Semantically-enabled repository software*

Built on top of SQL Server & Entity Framework

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

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

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

Binary & Source Code (Version 2.0):
http://research.microsoft.com/zentity/

demo

Proyecto RITA | Repositorio Tadeísta

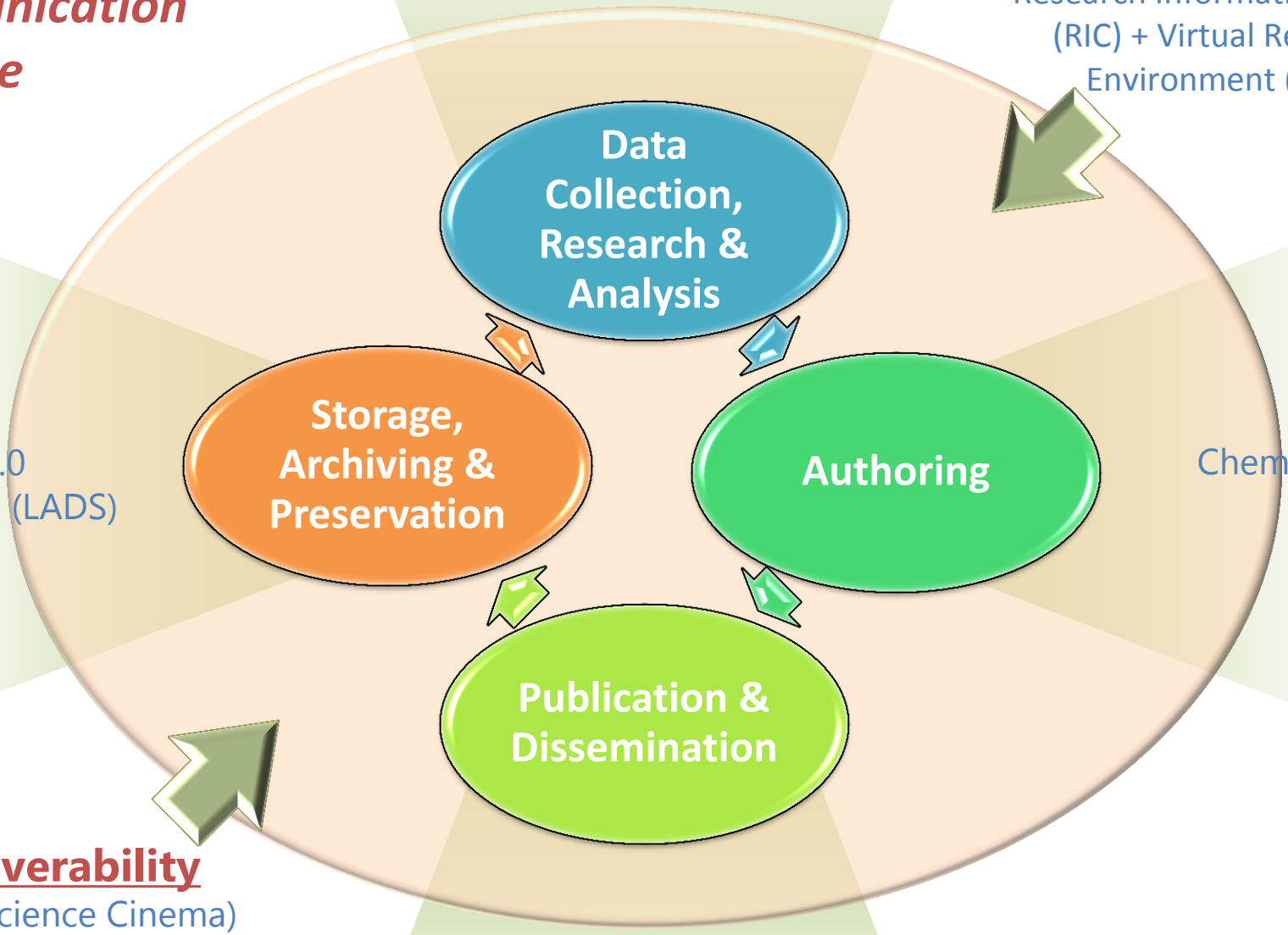
Zentity 2.0 implementation

Dr. Camilo A. Acosta-Márquez
Director Centro de Robótica e Informática
Universidad Jorge Tadeo Lozano (Bogotá, Colombia)

Microsoft Research Connections

Servers are the
new shelves.

The Scholarly Communication Lifecycle



DCXL Add-in for Excel

Collaboration
Research Information Centre (RIC) + Virtual Research Environment (VRE)

Zentity 2.0
Garibaldi (LADS)

Chem4Word

Discoverability
MAVIS (Science Cinema)
Academic Search

BigTime (Chronozoom)
Bing Translator

How to ***engage*** with Microsoft Research Connections

- **Tell us more about your projects, your workflows, your issues. We're *always* in "requirements gathering mode"**
  - Email us at scholar@microsoft.com with questions and ideas
  - Especially if you are already utilizing Microsoft technologies
- **Download our add-ins and try them, and then *give us feedback!***
  - Let us know if we can facilitate a connection with the appropriate product group(s)
- **Follow announcements via our RSS feed or via our Facebook group**

Thank you!

Lee Dirks

Education & Scholarly Communication

Microsoft Research | Connections

ldirks@microsoft.com or scholar@microsoft.com

URL – http://www.microsoft.com/scholarlycomm/

Facebook: Scholarly Communication at Microsoft

Thank you!

UNIVERSIDAD DE BOGOTÁ
JORGE TADEO LOZANO

Dr. Camilo A. Acosta-Márquez PhD

Research Fellow Rehabilitation Robotics

Director Centro de Robótica e Informática | Universidad Jorge Tadeo Lozano

Coordinador Académico RUMBO (Red Universitaria Metropolitana de Bogotá)

Proyecto RITA, Repositorio Tadeísta

camilo.acosta@utadeo.edu.co

Microsoft®
Research Connections