Update from Microsoft External Research:

Recent Scholarly Communication Projects

Lee Dirks
Director, Education & Scholarly Communications
Microsoft External Research
Microsoft Corporation
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.
business model (dual benefit)
**The Scholarly Communication Lifecycle**

- **Data Collection, Research & Analysis**
- **Storage, Archiving & Preservation**
- **Authoring**
- **Publication & Dissemination**

**Collaboration**
- SharePoint
- LiveMeeting
- Office Live

**Office 2007:**
- Word
- PowerPoint
- Excel
- OneNote
- Tablet PC/UMPC

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

**Discoverability**
- FAST
- MSR Academic Search
- “Bookweb”
- SharePoint

- Word 2007 + PowerPoint 2007
- WPF & Silverlight
- “Sea Dragon” / “PhotoSynth” / “Deep Zoom”

- Excel 2007
- Windows Server HPC
- “Astoria” / “Pop Fly”

This work is licensed under a [Creative Commons Attribution 3.0 United States License](http://creativecommons.org/licenses/by/3.0/us/).
Goal: Transform Scholarly Communication

- **Interoperability is essential**
  - Actively lobby and drive for consensus around technical standards and standardized protocols proactively adopted by the community; enable broad community engagement
    - Customers have told Microsoft that interoperability is OUR responsibility

- **Leverage Existing Community Protocols, Practices, Guidelines, etc.**
  - Example – metadata conventions / taxonomies / ontologies: a traditional strength for libraries – and a critical component in enabling Web 2.0

- **Optimize for data-driven research**
  - To both data (scientific) and to information (scholarly publications)
  - Reproducible research + computational science
  - Properly document / annotate scholarly output

- **Data preservation (and provenance) should be baseline**
  - Documentation of the data’s provenance
  - Preservation needs to be like “accessibility” features – i.e., assumed as required

- **Semantic knowledge discovery & social networking**
  - Harnessing collective intelligence must be a consideration – since accessing research is a core step in the life-cycle. Enable knowledge discovery
  - Optimize for Web 2.0 scenarios and allow end-users/experts to find things easier

This work is licensed under a Creative Commons Attribution 3.0 United States License.
Scholarly Communications: Projects Overview

Current or Completed Projects:

- **Cornell** – arXiv.org + Word 2007 (and repository interoperability via SWORD)
- **MIT / Broad Institute** – Authoring (Word 2007) + data for research reproducibility
- **MSR** – CMT++ interoperability with data + metadata transfer/exchange (conference management tool enhancements)
- **LiveLabs** – eJournal publishing online service (community publishing tool)
- **UC San Diego / PLoS** – Semantic mark-up of scholarly articles (+ submission)
- **Chem4Word with Office & Cambridge University** – Create add-in to Word 2007 to facilitate drawing of chemical compounds and equations
- **Johns Hopkins University** – Digital Archive for Astronomy/Astrophysics data (storage, preservation and access)
- **Planets Project / EU** (with MSR – Cambridge) OpenXML and file format preservation + interoperability
- **eChemistry Project (Cornell, Penn State, Indiana, Cambridge, Southampton)** – ORE exemplar: access to compound chemical info objects (cross-repository access to open chemistry data)
- **British Library** – Researcher Information Centre (RIC) online workflow tool for scientists and researchers
- **Creative Commons Add-in for Office 2007** – evolving the Word 2003 effort
- **University of Southampton (UK)** – Port ePrints Repository Software for installation on the Windows platform
- **University of Manchester / “MyExperiment” Project** – social networking for scientists
- **ORE Acceleration Project** (OAI – Object Reuse & Exchange) – Alpha spec development
- **UK National Archives** – Virtual PC / Emulation of legacy systems to facilitate preservation
- **National Library of Medicine / NCBI** – “PubMed Int’l” UK version of PubMed + NLM DTD
- **DRIVER 2 (EU)** – Infrastructure integration of across a network of European research repositories

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

Ontology Add-in for Word 2007

- Phil Bourne
- Lynn Fink

Services: Ontology download web service

Relationships: Ontology browser

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

Intent: Term recognition & disambiguation

This work is licensed under a Creative Commons Attribution 3.0 United States License.
Article Authoring Add-in for Word 2007

Structure: Read, convert, and author NLM XML documents

Structure: Client-side XML validation

Services: repository deposit via SWORD

Relationships: ORE Resource Map creation

Relationships: Citation lookup and reference management

Binary (version 2.0):
http://research.microsoft.com/authoring/
Single-dose oral naproxen for acute postoperative pain: a quantitative systematic review

Abstract

Naproxen and naproxen sodium are non-steroidal anti-inflammatory drugs used in a variety of painful conditions, including the treatment of postoperative pain. This review aims to assess the efficacy, safety and duration of action of a single oral dose of naproxen/naproxen sodium for moderate to severe acute postoperative pain in adults, compared with placebo.

Methods

The Cochrane Library (issue 4 2002), EMBASE, PubMed, MEDLINE and an in-house database were searched for randomised, double blind, placebo controlled trials of a single dose of orally administered naproxen or naproxen sodium in adults with acute postoperative pain. Pain relief or pain intensity data were extracted and converted into dichotomous information to give the number of patients with at least 50% pain relief over 4 to 6 hours. Relative benefit and number-needed-to-treat were then calculated. The percentage of patients with any adverse event, number-needed-to-harm, and time to remedication were also calculated.
Chem4Word - Chemistry Drawing in Word

Intent: Recognizes chemical dictionary and ontology terms

Data: Semantics stored in Chemistry Markup Language

Intelligence: Verifies validity of authored chemistry

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

Relationships: Navigate and link referenced chemistry

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

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

• Peter Murray-Rust
• Joe Townsend
• Jim Downing
GenePattern Reproducible Research Add-in

Services: Connects to GenePattern database

Relationships: Inline graphics are synchronized to dataset

Data: Control and execute query pipelines into GenePattern

Data: Resulting data (and provenance) stored within Word document

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

This work is licensed under a Creative Commons Attribution 3.0 United States License.
Research Information Centre (RIC)

Personal site for each researcher and project site for each project

Collaborative environment for researchers

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

Project site navigation and tool based on project lifecycle

Social networking, real-time communication, blogs, wikis

Available soon: http://research.microsoft.com/ric/
eJournal Editorial Workbench
Module for the Research Information Centre

Targeted at scholarly societies and small to medium-sized publishers

Editorial and peer review management tool

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

Support and tracks online collaboration between authors

http://research.microsoft.com/ejournal/
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/
Discovery & Visualization Tools

- **NodeXL**
  - Network visualization & exploration for Excel
  - [http://nodexl.codeplex.com](http://nodexl.codeplex.com)

- **Gazer**
  - Silverlight Control for graphical network browsing

- **FacetLens**
  - WPF control for faceted browsing
  - [http://research.microsoft.com/cue/facetlens/](http://research.microsoft.com/cue/facetlens/)
PLANETS
Tools and methods for sustainable long-term preservation of digital objects

Organization
• EU Commission Project, €14M for 4 years
• Consortium of 5 national libraries, 4 national archives, 4 universities and 4 industry partners

Goals
• Preservation of Office Documents based on OpenXML

Execution
• Delivered prototype converters for MS Office binary formats and ODF to OpenXML
• Demonstrating the benefits of Office OpenXML and endorsement of ISO standardization by key stakeholders: British Library, Swiss National Archives, UK National Archives
Document Conversion Service

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

View documents in various formats

Compare original and converted documents

http://odf-converter.sourceforge.net/
The National Archives and Microsoft join forces to Preserve the UK’s Digital Heritage
Unique relationship ensures sustainability of digital information for future generations of Britons

LONDON, UK. — July 04, 2007 — Microsoft and The National Archives today announced a Memorandum of Understanding (M.O.U) ensuring preservation of the nation’s digital records from the past, present and into the future. Microsoft will make available to The National Archives a system, which combines previous versions of Windows and Office, to help solve problems of managing historical records based on legacy Microsoft Office formats. The ongoing collaboration will also see The National Archives further its innovative approach to sustaining the nation’s digital heritage, by contributing their expertise and insight in digital preservation towards the development of future releases of Microsoft products. Having addressed some of the key format sustainability issues through the implementation of the ecma Open XML format in its latest products, the announcement represents a further step in Microsoft’s commitment to digital preservation.
Microsoft and The UK National Archives joined forces to ensure preservation of the United Kingdom’s digital records.

To help solve problems of managing historical records based on legacy Office formats, Microsoft is making available to The UK National Archives a new system combining all previous versions of Windows and Office on one PC.

The ongoing collaboration will also see The UK National Archives contribute expertise and insight in digital preservation towards the development of future releases of Microsoft products.

Having addressed some of the key format sustainability issues through the implementation of the ECMA Open XML format in its latest products, the announcement represents a further step in Microsoft’s commitment to digital preservation.
"DuraCloud"

durable digital assets in the cloud

DuraCloud Mediating Service

Durable Store Service Layer (Foundation Service)  Durable Store Service Layer (Commercially Operated)  Durable Store Service Layer (Locally Operated)

Foundation provided Software and Configuration (Open Source)

Amazon EC2/S3  Google  HP  Local IT

Yahoo  Microsoft  Offline Backup

Compute and Storage [Cloud/Grid] Providers
Questions?

Lee Dirks
Director—Education & Scholarly Communication
Microsoft External Research
lee.dirks@microsoft.com

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

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