Microsoft Research Connections: An Overview

Tony Hey

Vice President

Microsoft Research Connections

Microsoft Research Asia
Faculty Summit 2012

Microsoft Research

- *Expand the state of the art in each of the areas in which we do research*
- *Rapidly transfer innovative technologies into Microsoft products*
- *Ensure that Microsoft products have a future*

http://research.microsoft.com/
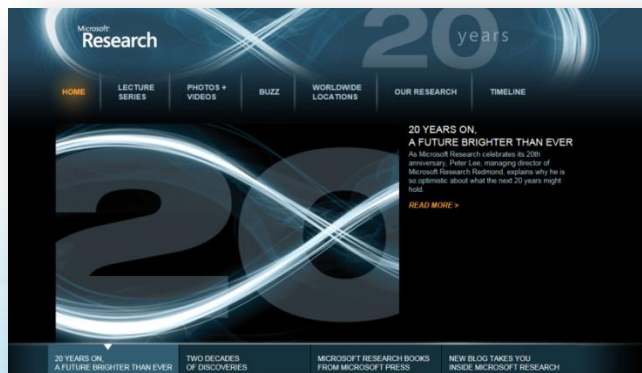
Microsoft Research
Worldwide Presence

MSR Asia

MSR Cambridge, UK

MSR India

MSR New England

Redmond

Silicon Valley, California

NEW:
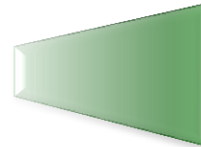MSR New York

XCG

FUSE Labs

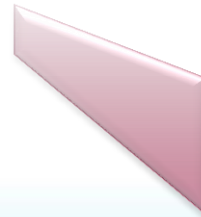# Microsoft Research Connections

Work with the worldwide academic research community to speed research, improve education, and foster innovation
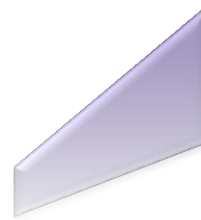
Collaborations to pursue scientific breakthroughs

Inspire computer and information scientists

Accelerate scientific exploration with computing

Microsoft Research Asia
Faculty Summit 2012

Engagement and Collaboration

Core Computer Science	Natural User Interface	Earth Energy Environment	Health and Wellbeing	Education and Scholarly Communication

Research Accelerators

- Worldwide Telescope
- MS Biology Foundation
- Chemistry Add-in for Word
- Zentity
- Trident Workflow Workbench

Global Partnerships

- Centre for Computational and Systems Biology
- Microsoft Research-INRIA Joint Centre
- Microsoft Research Asia (MSRA) Joint Lab Program
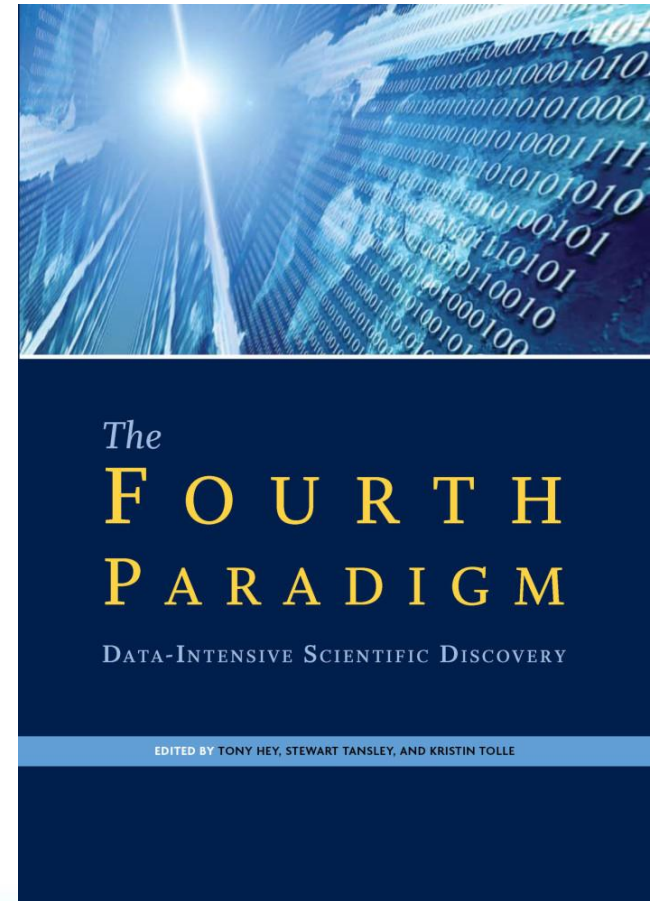- Microsoft & FAPESP

External Researchers

- Faculty Fellows
- Graduate Women Scholars
- Jim Gray eScience Award
- Student Internships
- ACM Student Research Competition

Changing Nature of Discovery

- Complex models
  - Multidisciplinary interactions
  - Wide temporal and spatial scales
- Large multidisciplinary data
  - Real-time steams
  - Structured and unstructured
- Distributed communities
  - Virtual organizations
  - Socialization and management

The FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

http://fourthparadigm.org

Microsoft Research Asia
Faculty Summit 2012

Machine Learning and eScience

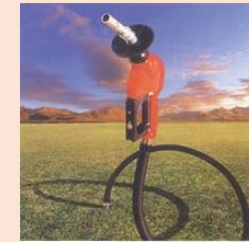*Tackling societal challenges*

Identifying genetic and environmental causes of disease

Fighting HIV/AIDS

Increasing energy yield of sugar cane through genome assembly

Data-intensive Science

Acquisition & modelling

Archiving & preserving

Data

Collaboration & visualization

Dissemination & sharing

Analysis & data mining

Microsoft Research Asia
Faculty Summit 2012
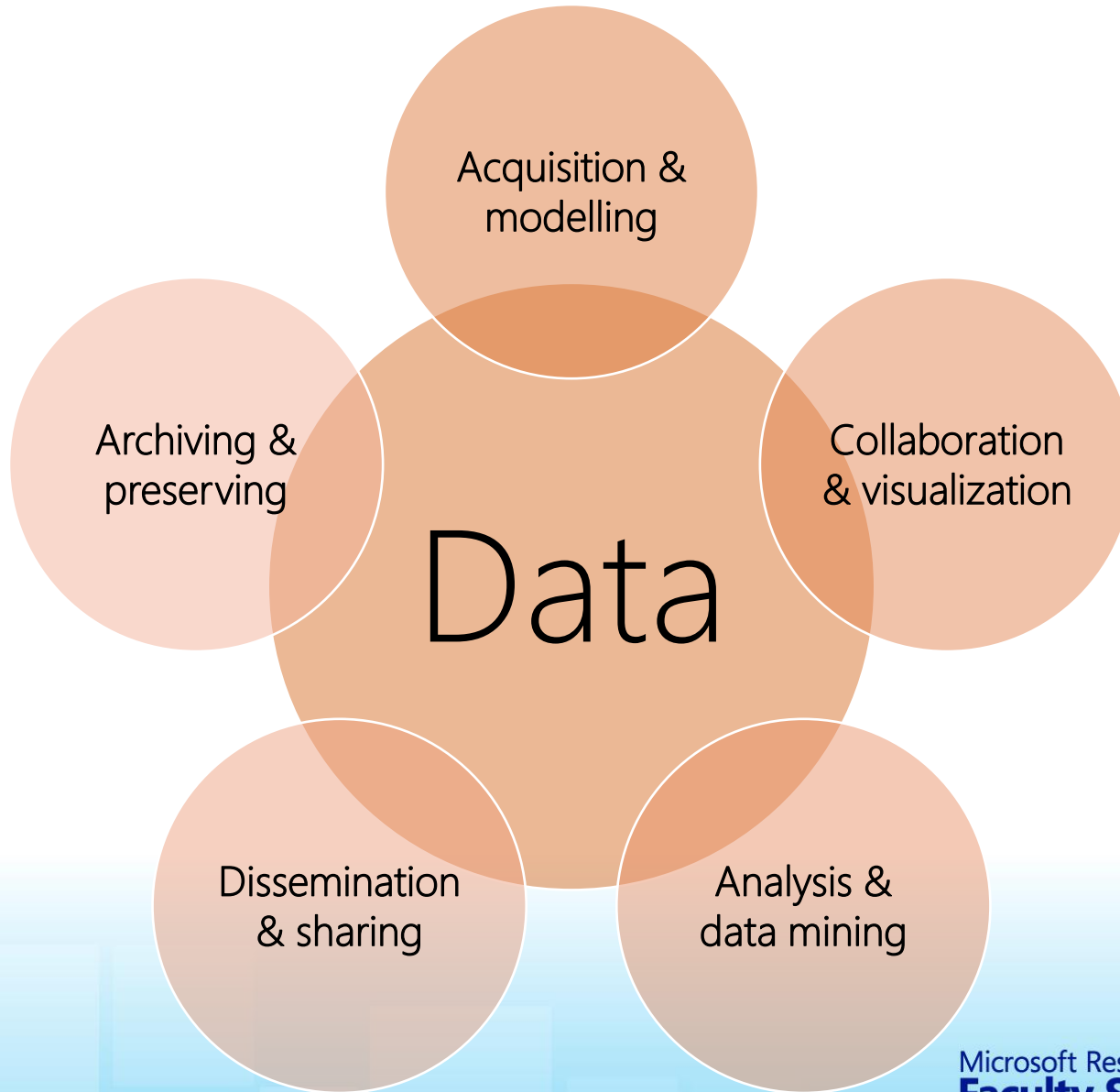
Virtual Research Environment (VRE) Toolkits

Currently Available Kits

- Cancer Imaging for SharePoint (Oxford University)
- Collaborator Search Kit (British Library)
- Content Archiving Kit (British Library)
- Content Lifecycle Integration Framework (University of Hull)
- Data Management Kit (University of Southampton)
- Document Review Workflow Kit (British Library)
- LabTrove Research Blog (University of Southampton)
- Literature Review Kit (British Library)
- Repository Integration Kit (University of Southampton)
- RSS Feeds Kit (British Library)
- User Administration (FBA) Kit (British Library)
- X-Ray Diffraction Data Converter Kit (University of Delhi)

Templates

- Researcher and Project Templates (British Library)
- LVRE with Active Directory (La Trobe University)

VRE Based on Sharepoint

Microsoft Research Asia
Faculty Summit 2012

World Wide Telescope
www.worldwidetelescope.org

Big data requires new types of data visualization tools

Collaborators:
- Alyssa Goodman; Harvard University
- Alex Szalay; Johns Hopkins University
- Curtis Wong, Jonathan Fay; Microsoft Research

- Integration of data sets and one-click contextual access
- Easy access and use

- As of May 2010, over 4M unique users
- Average number of WWT users is over 8K per day

See TED talks by Roy Gould and Curtis Wong

Microsoft Research Asia
Faculty Summit 2012

Layerscape with Seismic Data

Microsoft Research Asia
Faculty Summit 2012
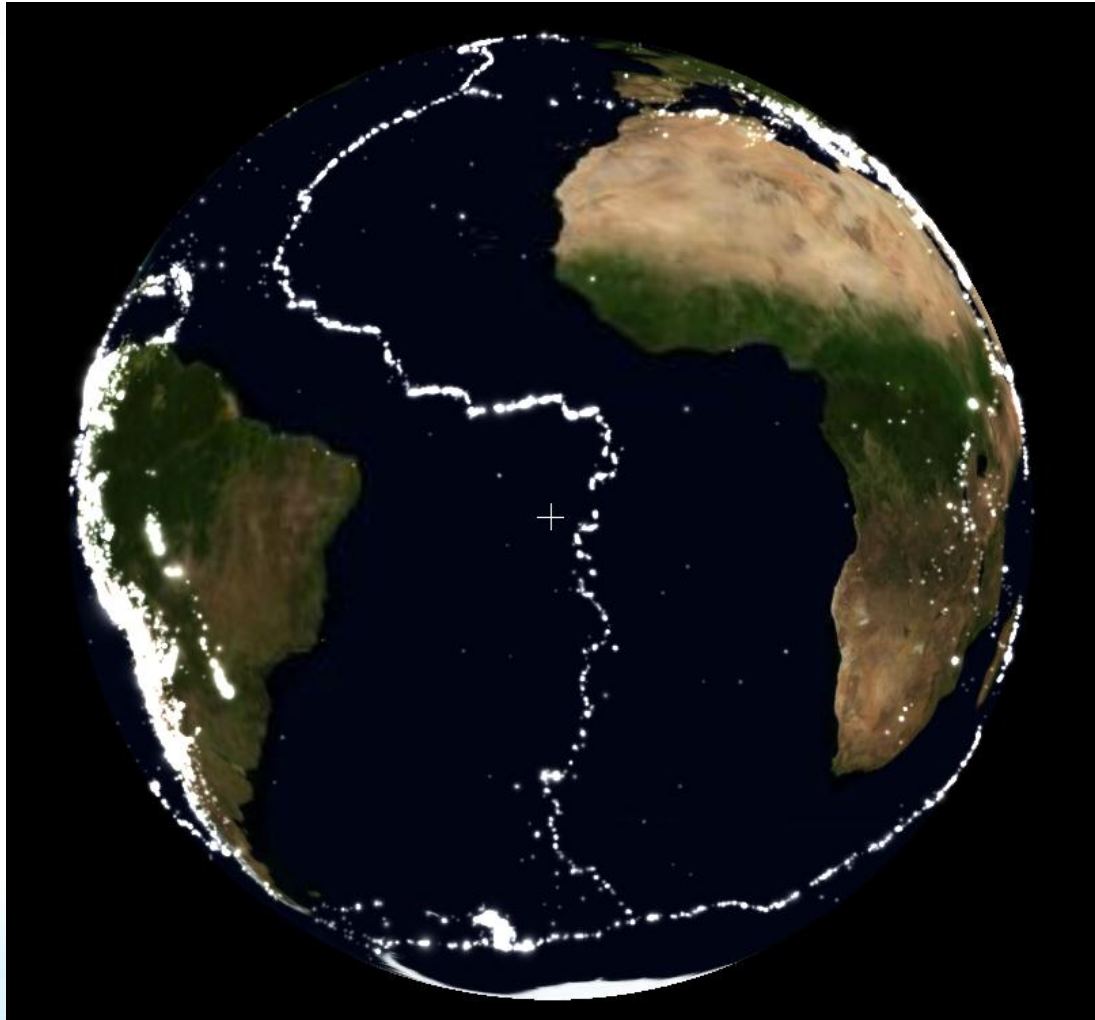
ChronoZoom – History in its broadest possible context …

The challenge: exploration of all known time series data with the ability to smoothly transition from billions of years down to individual nanoseconds…

This is what Walter Alvarez, Professor of Earth and Planetary Science at University of Berkeley set out to do.

Our vision is to create an application that allows researchers to browse, overlay, and explore interdisciplinary data sources.
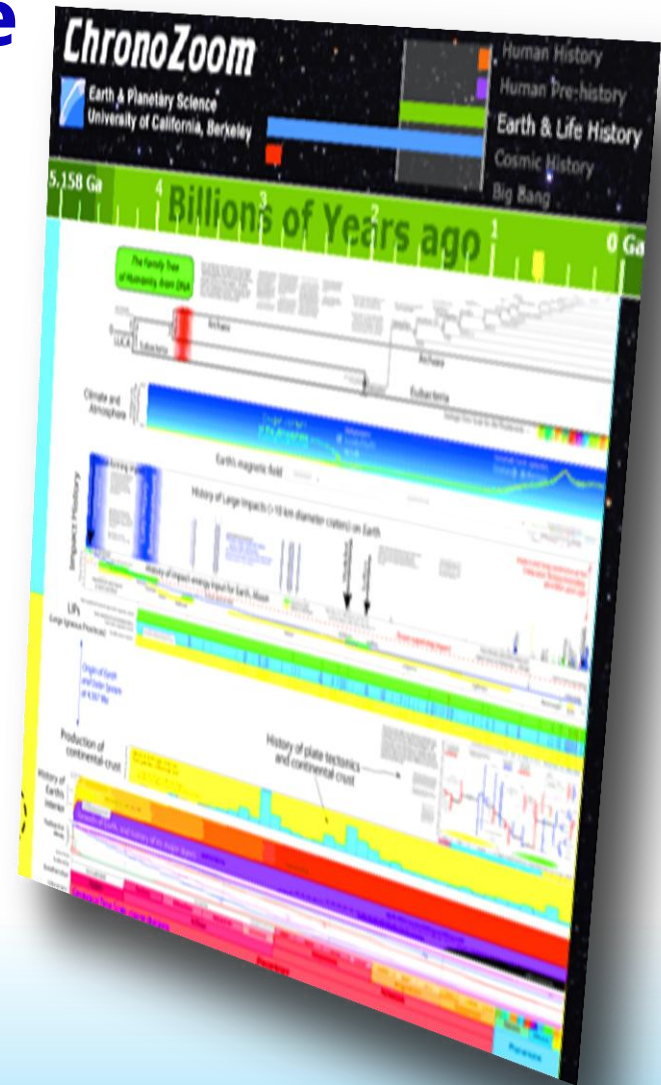
www.chronozoomtimescale.org

Microsoft Research Asia
Faculty Summit 2012

DataUp: Data Curation for Excel

Excel add-in and HTML5 browser app
Microsoft Research with California Digital Library's Curation Center
Linked to the DataONE (an NSF DataNet Project)

- Goal to facilitate data management, sharing, and archiving for scientists.
- Building open-source add-in (extension) for Microsoft Excel to assist in documenting and preparing Excel data for archiving and sharing.
- Targeting environmental scientists but should be useful for wide audiences
- Define archiving as movement or storage of data with metadata, to a repository for long-term retention.

http://dcxl.cdlib.org/

Microsoft Research Asia
Faculty Summit 2012

Envisioning a New Era of Research Reporting

Reproducible Research

Collaboration

Reputation & Influence
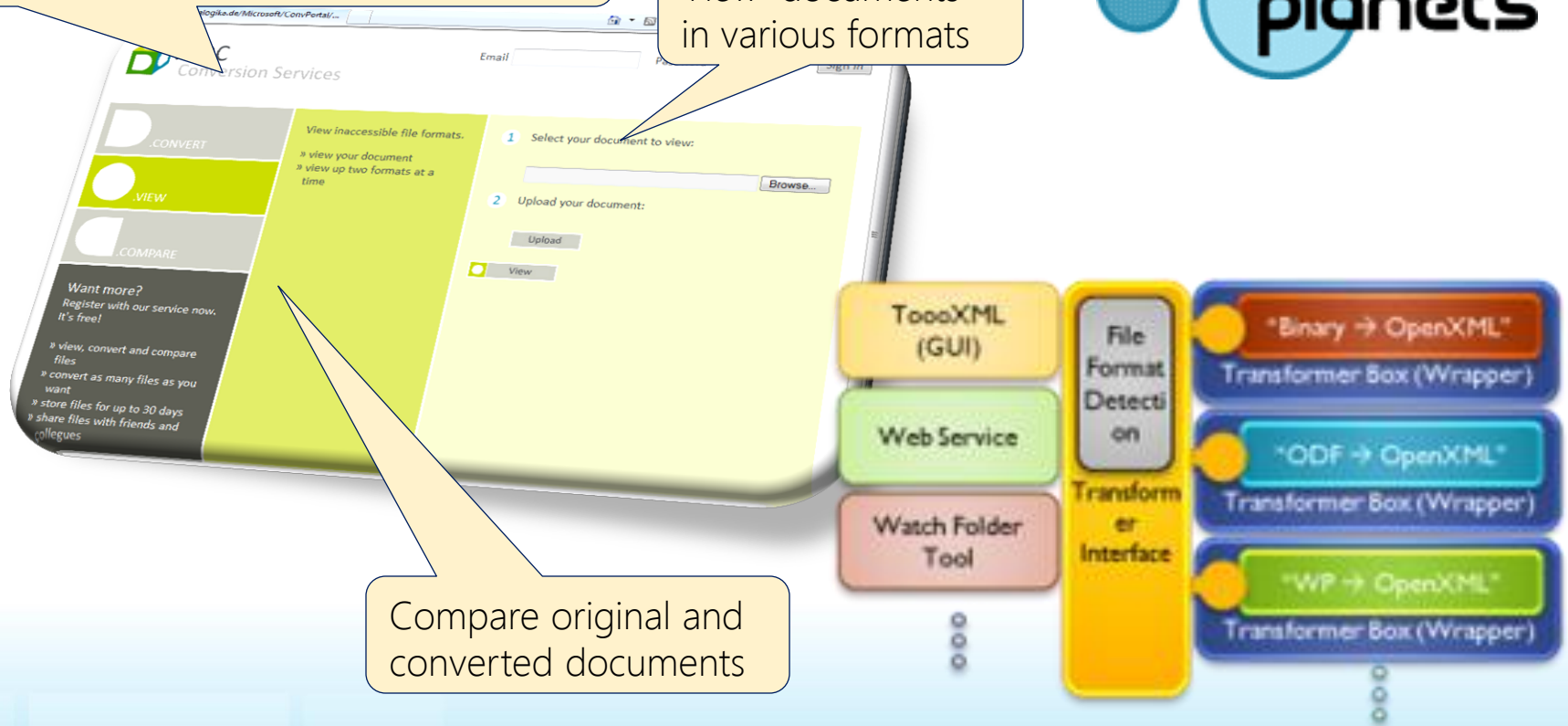
Dynamic Documents

Interactive Data

(Thanks to Bill Gates SC05)
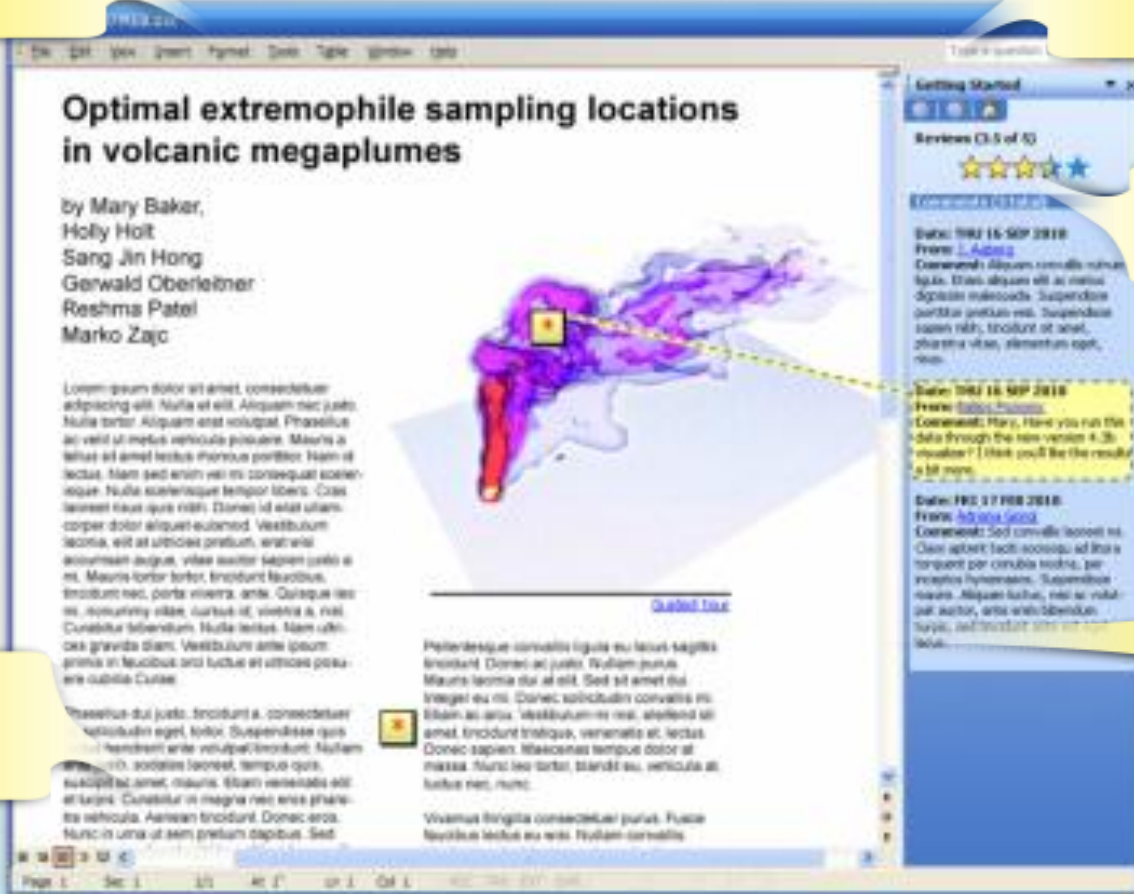
Archiving and Preservation:
A Document Conversion Service

Convert Word Perfect and Word documents to OOXML, ODF and UOF

View documents in various formats

Compare original and converted documents

planets

Conversion Services

.CONVERT

.VIEW

.COMPARE

View inaccessible file formats.
» view your document
» view up two formats at a time

1   Select your document to view:

Browse...

2   Upload your document:

Upload

View

Want more?
Register with our service now.
It's free!

» view, convert and compare files
» convert as many files as you want
» store files for up to 30 days
» share files with friends and colleagues

ToooXML (GUI)

Web Service

Watch Folder Tool

File Format Detection

Transformer Interface

"Binary → OpenXML" Transformer Box (Wrapper)

"ODF → OpenXML" Transformer Box (Wrapper)

"WP → OpenXML" Transformer Box (Wrapper)

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

Microsoft Research Asia
Faculty Summit 2012

EU VENUS-C Project:
## *Virtual multidisciplinary EnviroNments USing Cloud infrastructures*
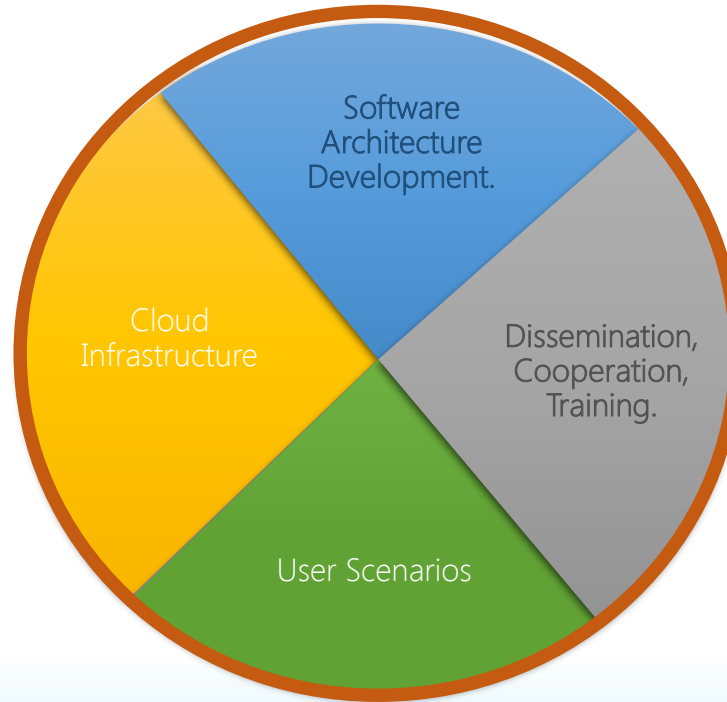
Part of the MSR Cloud Initiative for Research led by Dennis Gannon

Microsoft Research Asia
Faculty Summit 2012

All Scientific Data Online

- Many disciplines overlap and use data from other sciences.

- Internet can unify all literature and data

- Go from literature *to* computation *to* data *back to* literature.

- Information at your fingertips – For everyone, everywhere

- Increase Scientific Information Velocity

- Huge increase in Science Productivity

(From Jim Gray's last talk)

Literature

Derived and recombined data

Raw Data

Microsoft Research Asia
Faculty Summit 2012
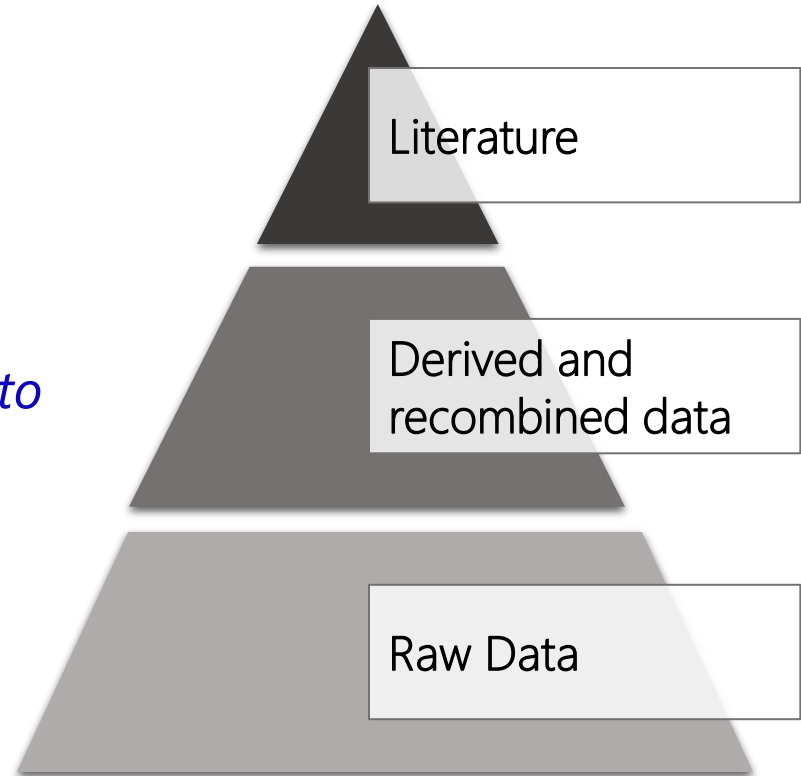
Microsoft Academic Search

Microsoft Academic Search is a free academic search engine developed by Microsoft Research Asia, which also serves as a test-bed for our object-level vertical search vertical search research.

- Easily search the top papers, authors, conferences, and journals for a topic.

- See details about a specific paper, author, conference, journal or organization.

- Quickly find relationships between authors.

- Discover influential papers, authors, conferences, journals and organizations within a certain field.

- Get the latest Call for Papers.

http://academic.research.microsoft.com/

Microsoft Research Asia
Faculty Summit 2012

Research Accelerators

Publishing

DataUp Add-in for Excel
Article Authoring Add-in for Word
Chemistry Add-in for Word

Creative Commons Add-in for Word
Ontology Add-in for Word
WikiBhasha

Scientific Computing

.NET Bio
Azure BLAST
NodeXL
Trident Workflow

Research Management

Research Information Center
VRE Toolkit
Zentity

Learning and Exploration

ChronoZoom
Layerscape
Try F#
WW Telescope for Scientific Data

•.NET Gadgeteer
Project Hawaii for Windows Phone
Kinect for Windows
Touch Develop

http://research.microsoft.com/Accelerators

Microsoft Research Asia
Faculty Summit 2012

Microsoft Research Connections Contributions

Project Trident: Toolset based on Windows Workflow Foundation that provides scientists' need for a flexible, powerful way to analyze large, diverse datasets.

Chemistry Add-in for Word: Chem4Word is an add-in for Microsoft Word that enables semantic authoring of chemical structures.

ConferenceXP: Platform for real-time collaboration that seamlessly connects people or groups over a network, providing high-quality, low-latency videoconferencing and a rich set of collaboration capabilities.

.NET Bio: Library of commonly used bioinformatics functions plus applications built upon that framework.
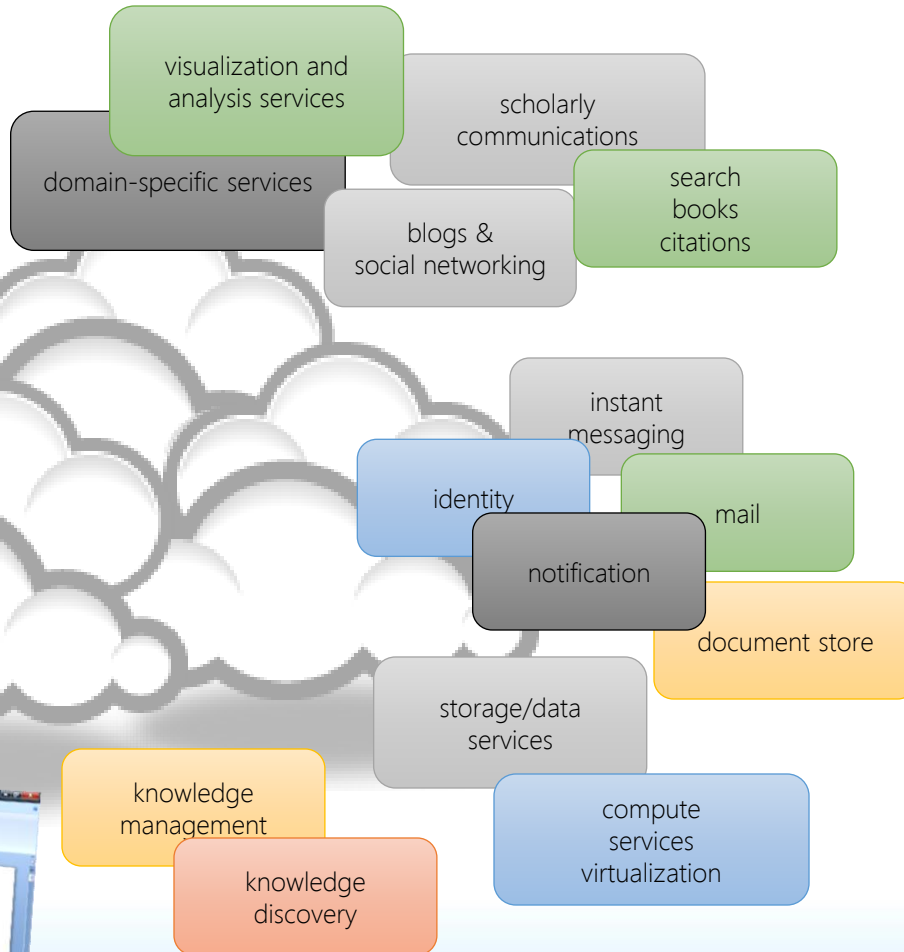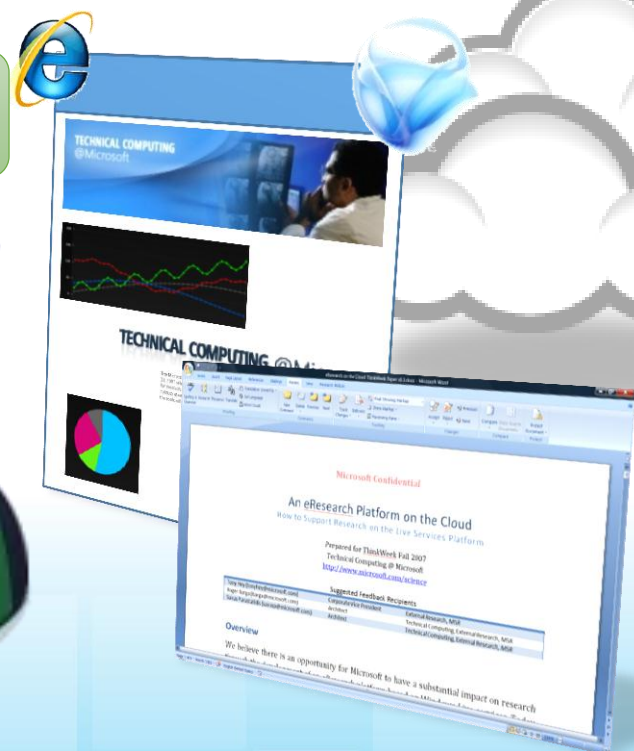
ChronoZoom: Visualization tool with zoom capability for browsing and exploring multidisciplinary data.

http://www.outercurve.org/

Microsoft Research Asia
Faculty Summit 2012

Towards a Semantic Cyberinfrastructure ...

Future Research Infrastructure will use semantic knowledge services on Client + Cloud

visualization and analysis services

scholarly communications

domain-specific services

search books citations

blogs & social networking

Reference management

Project management

instant messaging

identity

mail

notification

document store

storage/data services

knowledge management

compute services virtualization

knowledge discovery

Microsoft Research Asia
Faculty Summit 2012

Some Resources

- Microsoft Research
  - http://research.microsoft.com
  - Microsoft Research downloads:
    http://research.microsoft.com/research/downloads
- Microsoft External Research
  - http://research.microsoft.com/en-us/collaboration/
- Science at Microsoft
  - http://www.microsoft.com/science
- Scholarly Communications
  - http://www.microsoft.com/scholarlycomm
- CodePlex
  - http://www.codeplex.com

Microsoft Research Asia
Faculty Summit 2012