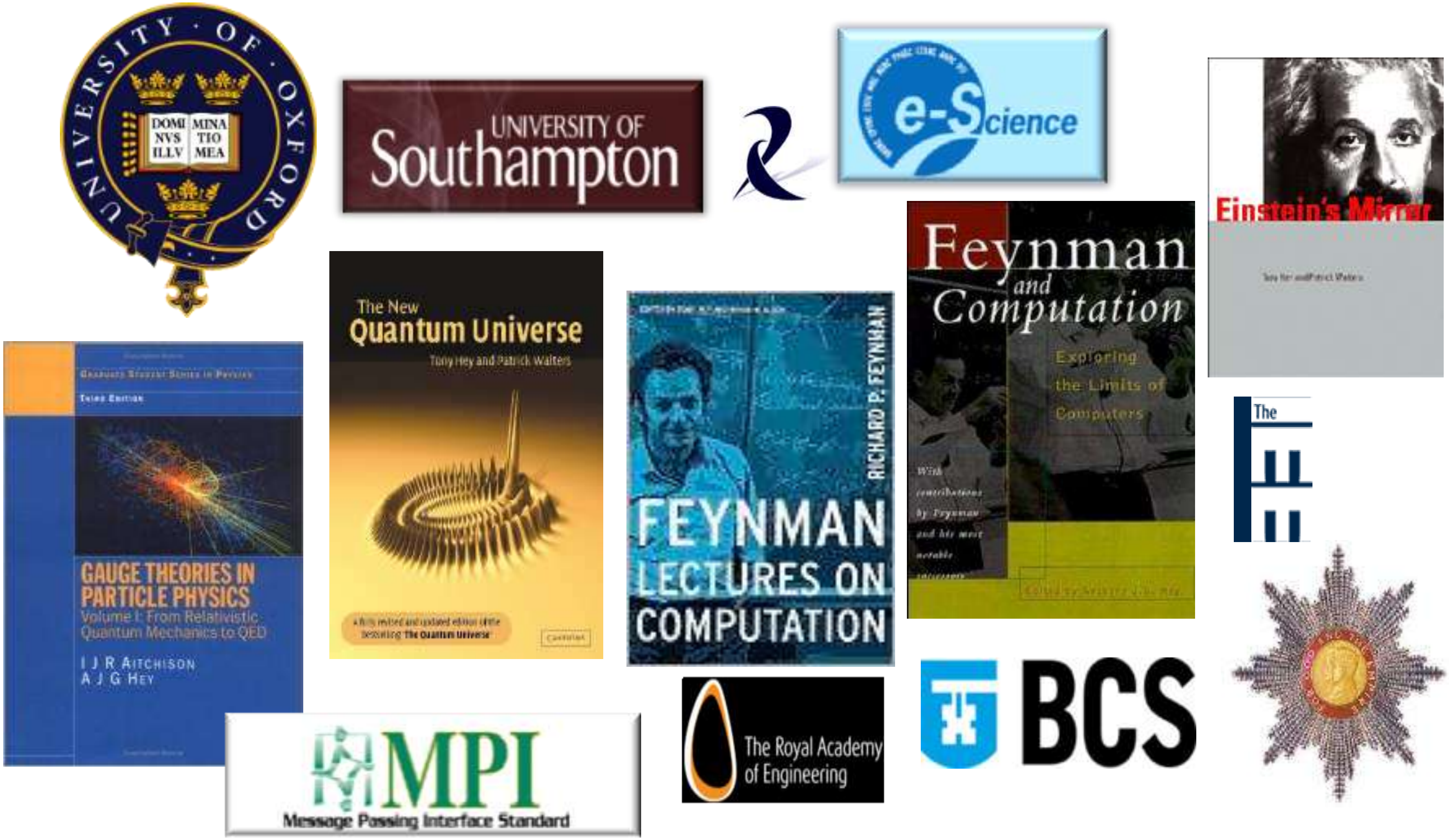
The Fourth Paradigm: Data-Intensive Scientific Discovery

Tony Hey
Vice President
Microsoft Research

Microsoft Research Connections

Tony Hey – An Introduction


UNIVERSITY OF OXFORD
DOMI NVS ILLV MINA TIO MEA
UNIVERSITY OF Southampton
e-Science
Einstein's Mirror
Tony Hey and Patrick Walters
Feynman and Computation
Exploring the Limits of Computers
With contributions by Feynman and his most notable colleagues
The New Quantum Universe
Tony Hey and Patrick Walters
A fully revised and updated edition of the bestselling The Quantum Universe
CAMBRIDGE
RICHARD P. FEYNMAN
FEYNMAN LECTURES ON COMPUTATION
The
GRADUATE STUDENT SERIES IN PHYSICS
THIRD EDITION
GAUGE THEORIES IN PARTICLE PHYSICS
Volume I: From Relativistic Quantum Mechanics to QED
I J R AITCHISON
A J G HEY
MPI
Message Passing Interface Standard
The Royal Academy of Engineering
BCS

Outline

- The Fourth Paradigm and eScience
- Examples of Data-Intensive Science
- Supporting the Data Life Cycle
- Open Data and Open Science
- The Future: Semantic Computing and the Cloud

A Tidal Wave of Scientific Data

COMMUNICATIONS OF THE ACM

CACM.ACM.ORG 12/08 VOL.51 NO.12

Surviving the Data Deluge

Open Information Extraction from the Web

The Economist

The data deluge

ieee spectrum

INSIDE TECHNOLOGY

AEROSPACE BIOMEDICAL COMPUTING CONSUMER ELECTRONICS ENERGY

AT WORK / INNOVATION

OPINION

The Coming Data Deluge

As science becomes more data intensive, so does our language

By PAUL McFEDRIES / FEBRUARY 2011

Email Print Share

Illustration: gluekit

WIRED

SUBSCRIBE » SECTIONS » BLOGS » REVIEWS » VIDEO »

Sign In | RSS

WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

Illustration: Marian Bantjes

nature

THE BITER BIT

TROPICAL CYCLONES

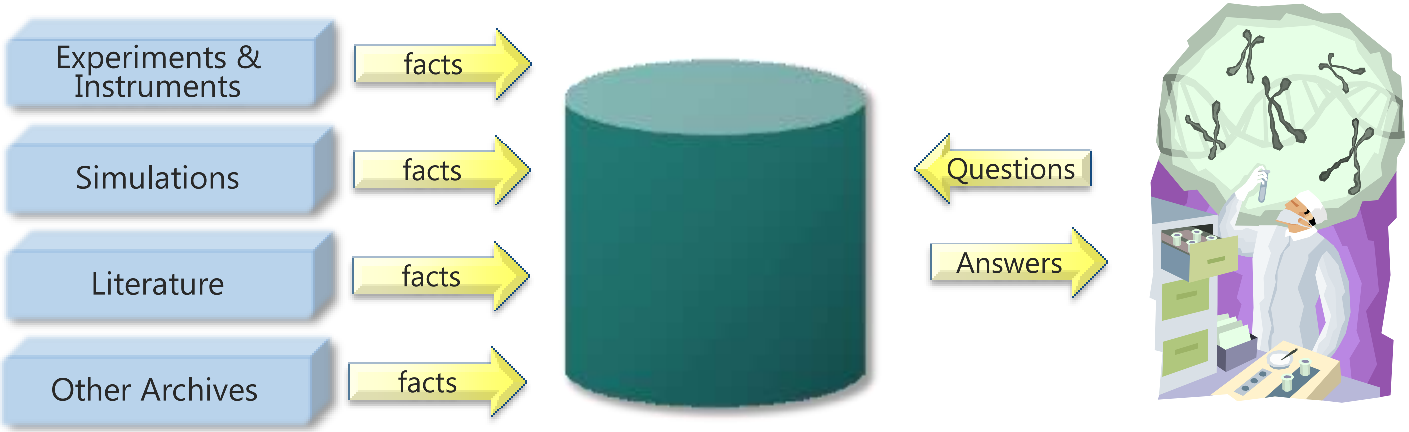
BLACK HOLE PHYSICS

BIG DATA

SCIENCE IN THE PETABYTE ERA

X-Info and Comp-X

- The evolution of X-Info and Comp-X for each discipline X
- How to codify and represent our knowledge



Experiments & Instruments

Simulations

Literature

Other Archives

facts

facts

facts

facts

Questions

Answers

The Generic Problems

- Data ingest
- Managing a petabyte
- Common schema
- How to organize it
- How to *re*organize it
- How to share with others

- Query and Vis tools
- Building and executing models
- Integrating data and Literature
- Documenting experiments
- Curation and long-term preservation

(With thanks to Jim Gray)

Emergence of a Fourth Research Paradigm

Thousand years ago – **Experimental Science**

- Description of natural phenomena

Last few hundred years – **Theoretical Science**

- Newton's Laws, Maxwell's Equations…

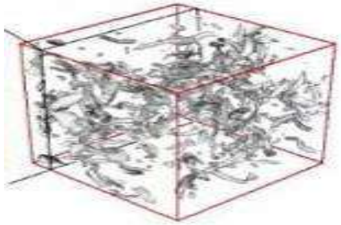
Last few decades – **Computational Science**

- Simulation of complex phenomena

Today – **Data-Intensive Science**

- Scientists overwhelmed with data sets from many different sources
  - Captured by instruments
  - Generated by simulations
  - Generated by sensor networks

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$



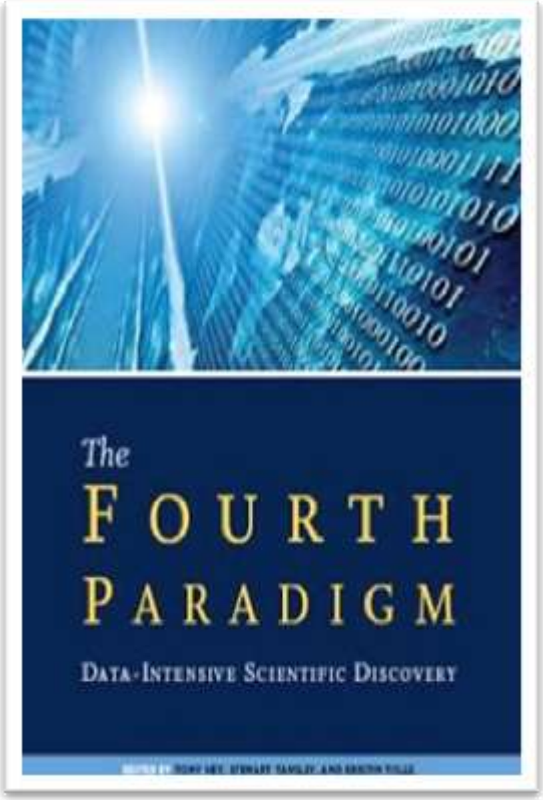
eScience is the set of tools and technologies to support data federation and collaboration

- **For analysis and data mining**
- **For data visualization and exploration**
- **For scholarly communication and dissemination**

(With thanks to Jim Gray)

Changing Nature of Discovery

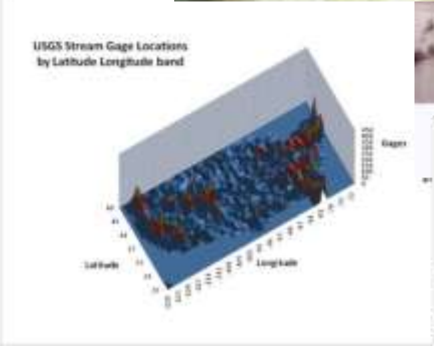
- Complex models
  - Multidisciplinary interactions
  - Wide temporal and spatial scales
- Large multidisciplinary data
  - Real-time steams
  - Structured and unstructured
- Distributed communities
  - Virtual organizations
  - Socialization and management

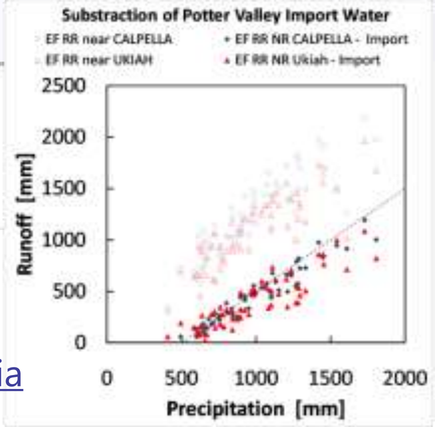


http://fourthparadigm.org

Digital Watersheds

- Russian River watershed challenges: forestry, farming, urbanization, gravel mining, and fish habitat restoration.
  - Can we understand historic and on-going changes using only publically available data sources such as USGS, NOAA, Sonoma Ecology Center, etc?
- Early studies examined overall water balance and changes in suspended sediment
  - Scientific data “mashups” are leading to new and useful results.

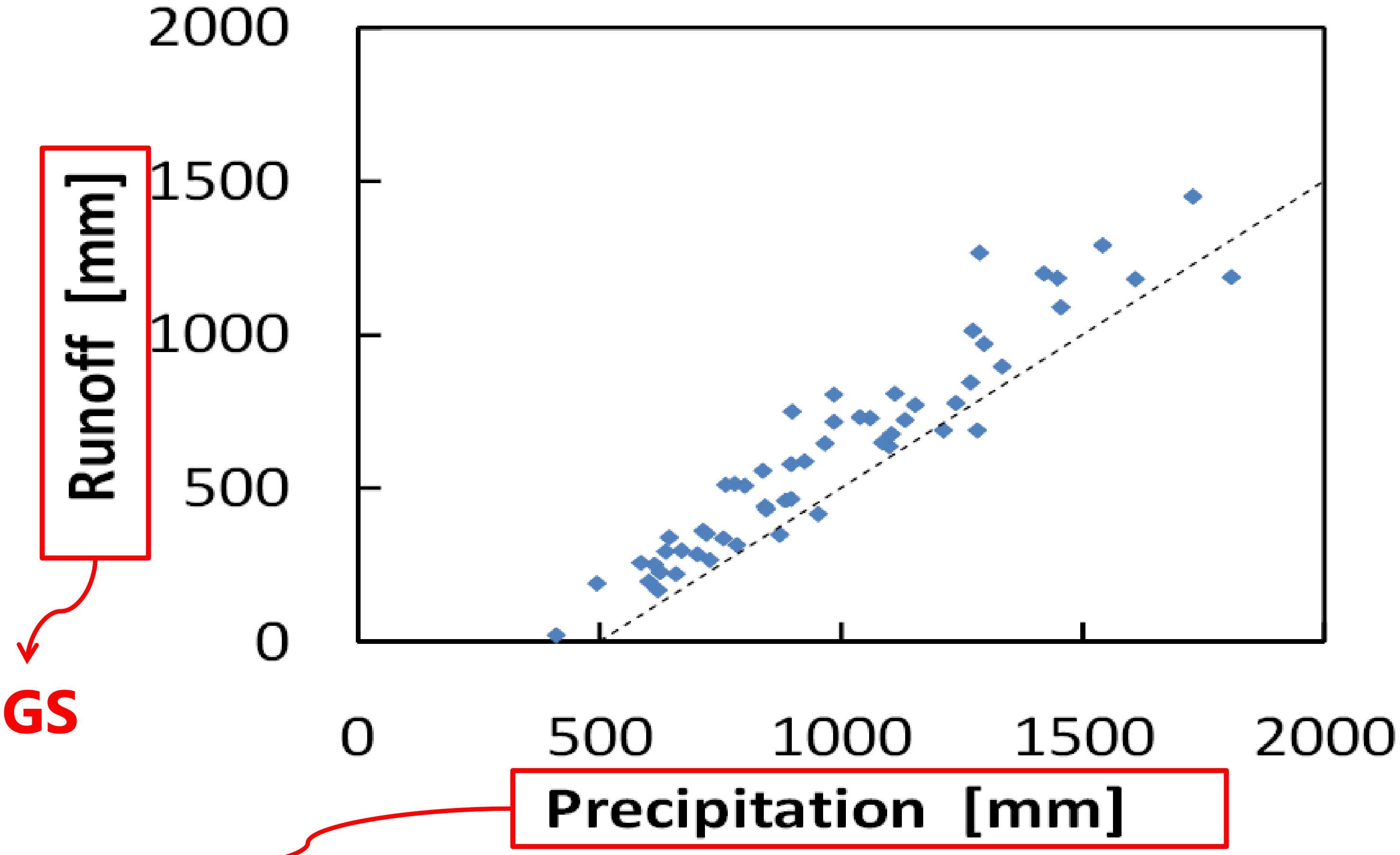




http://bwc.berkeley.edu
http://www-esd.lbl.gov/BWC/California
http://www.cuahsi.org/

James Hunt, BWC

Data from Multiple Sources



USGS

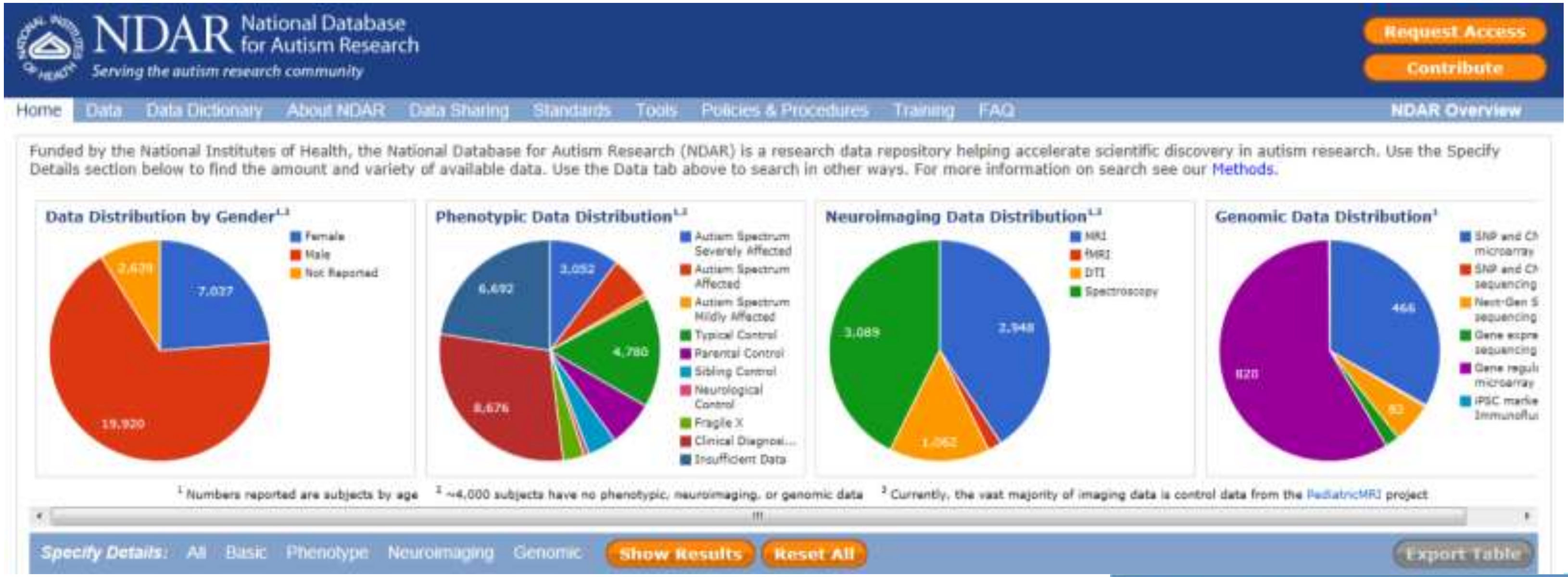
NOAA

National Database for Autism Research

Federated sources of data, tools, & specimens from major US autism research funders and investigators

- partners adopting NDAR standards, e.g.:
  - Global Unique Identifier (GUID)
  - Data dictionary (29,000 elements defined)
  - Data definition & validation tools
  - Authentication scheme
- Data from > 85,000 subjects federated

"Community Science"

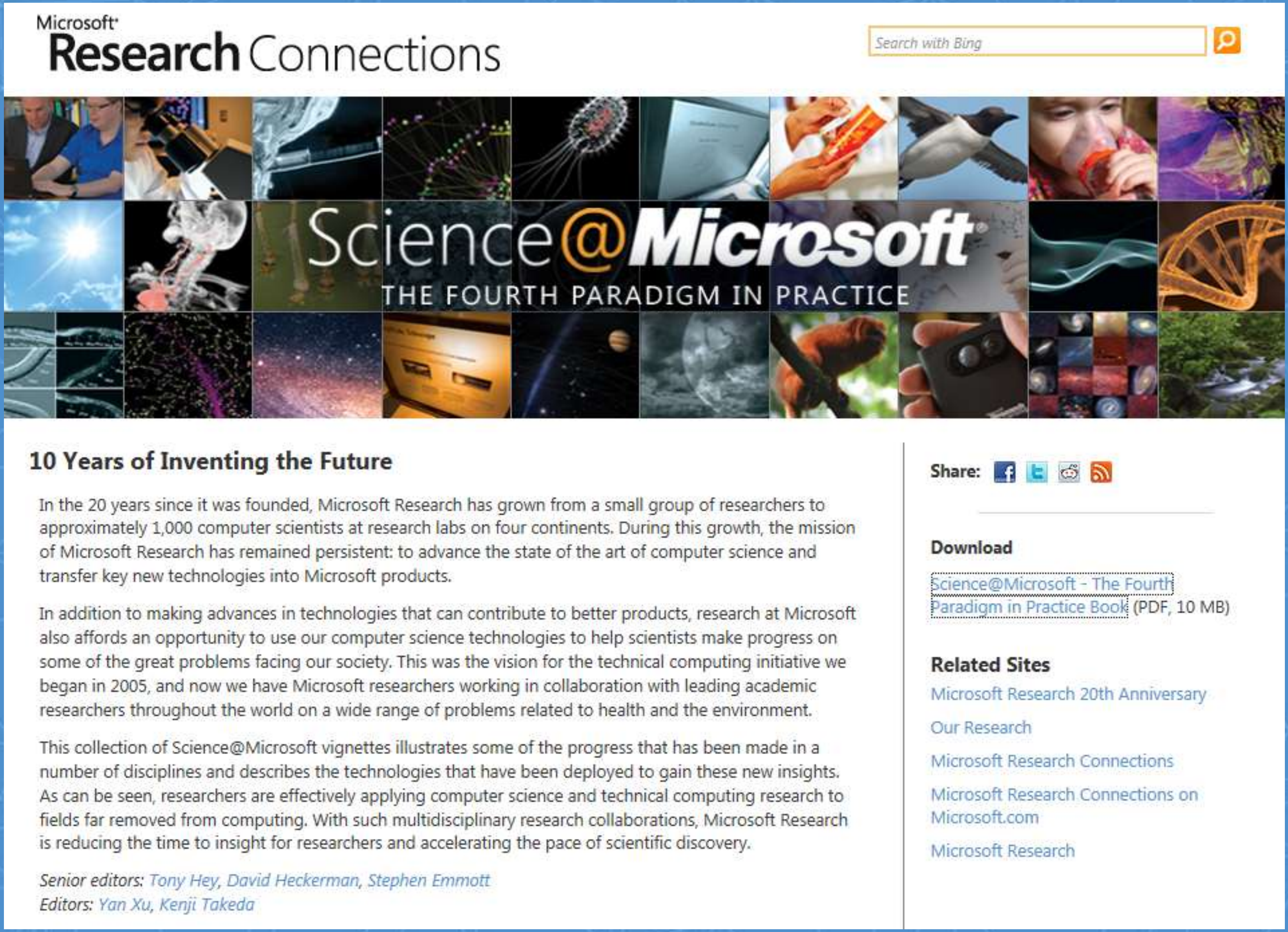


Funded by the National Institutes of Health, the National Database for Autism Research (NDAR) is a research data repository helping accelerate scientific discovery in autism research. Use the Specify Details section below to find the amount and variety of available data. Use the Data tab above to search in other ways. For more information on search see our Methods.

[1] Numbers reported are subjects by age

[2] ~4,000 subjects have no phenotypic, neuroimaging, or genomic data

[3] Currently, the vast majority of imaging data is control data from the PediatricMRI project



Microsoft Research Connections

Science@Microsoft

THE FOURTH PARADIGM IN PRACTICE

10 Years of Inventing the Future

In the 20 years since it was founded, Microsoft Research has grown from a small group of researchers to approximately 1,000 computer scientists at research labs on four continents. During this growth, the mission of Microsoft Research has remained persistent: to advance the state of the art of computer science and transfer key new technologies into Microsoft products.

In addition to making advances in technologies that can contribute to better products, research at Microsoft also affords an opportunity to use our computer science technologies to help scientists make progress on some of the great problems facing our society. This was the vision for the technical computing initiative we began in 2005, and now we have Microsoft researchers working in collaboration with leading academic researchers throughout the world on a wide range of problems related to health and the environment.

This collection of Science@Microsoft vignettes illustrates some of the progress that has been made in a number of disciplines and describes the technologies that have been deployed to gain these new insights. As can be seen, researchers are effectively applying computer science and technical computing research to fields far removed from computing. With such multidisciplinary research collaborations, Microsoft Research is reducing the time to insight for researchers and accelerating the pace of scientific discovery.

Senior editors: Tony Hey, David Heckerman, Stephen Emmott
Editors: Yan Xu, Kenji Takeda

Share:

Download

Science@Microsoft - The Fourth Paradigm in Practice Book (PDF, 10 MB)

Related Sites

Microsoft Research 20th Anniversary

Our Research

Microsoft Research Connections

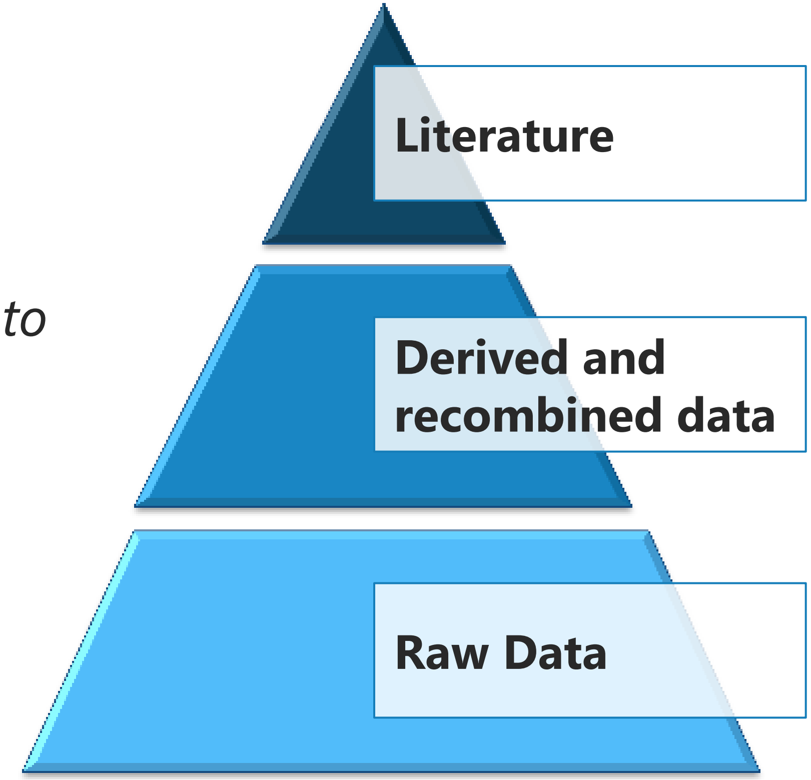
Microsoft Research Connections on Microsoft.com

Microsoft Research

http://www.microsoft.com/en-us/researchconnections/science/stories/

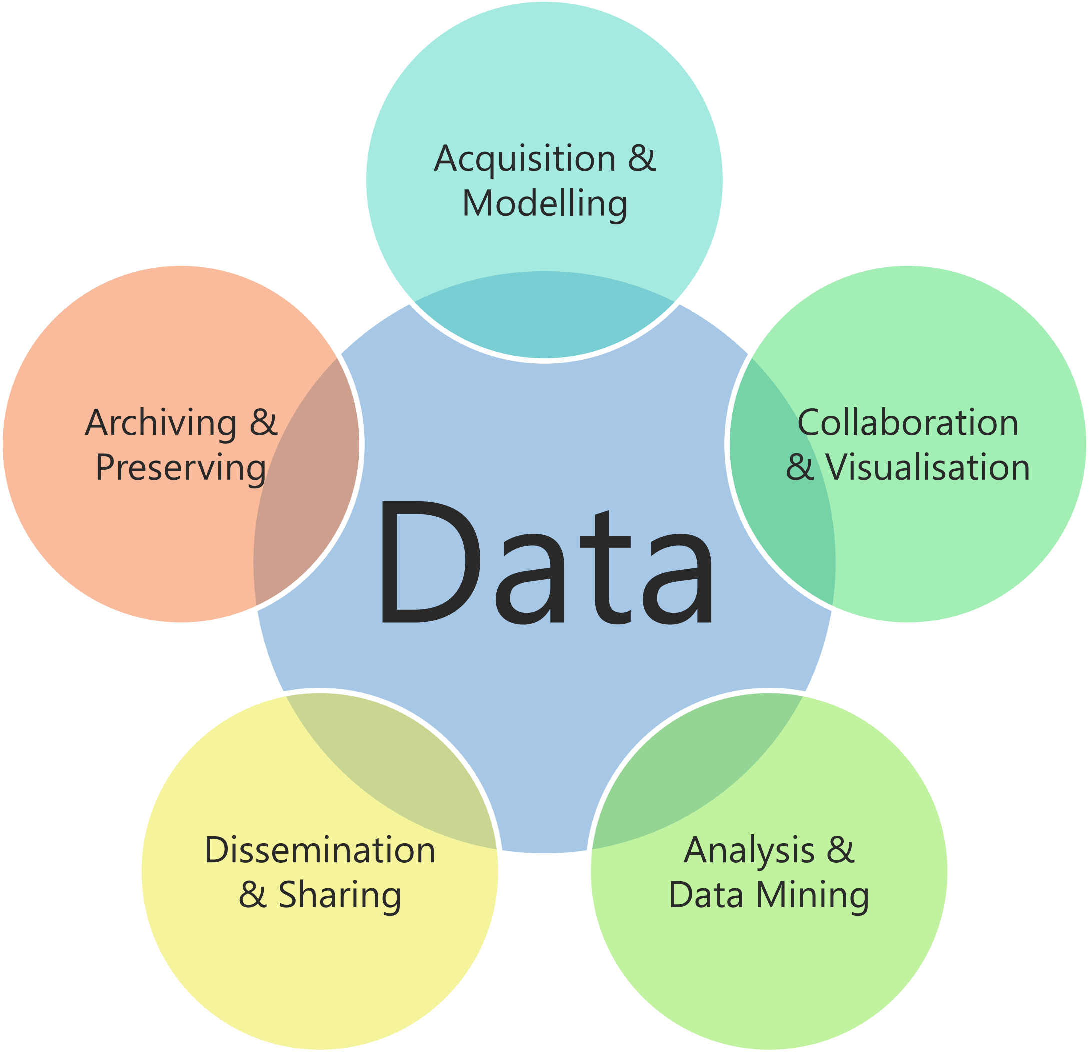
All Scientific Data Online

- 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



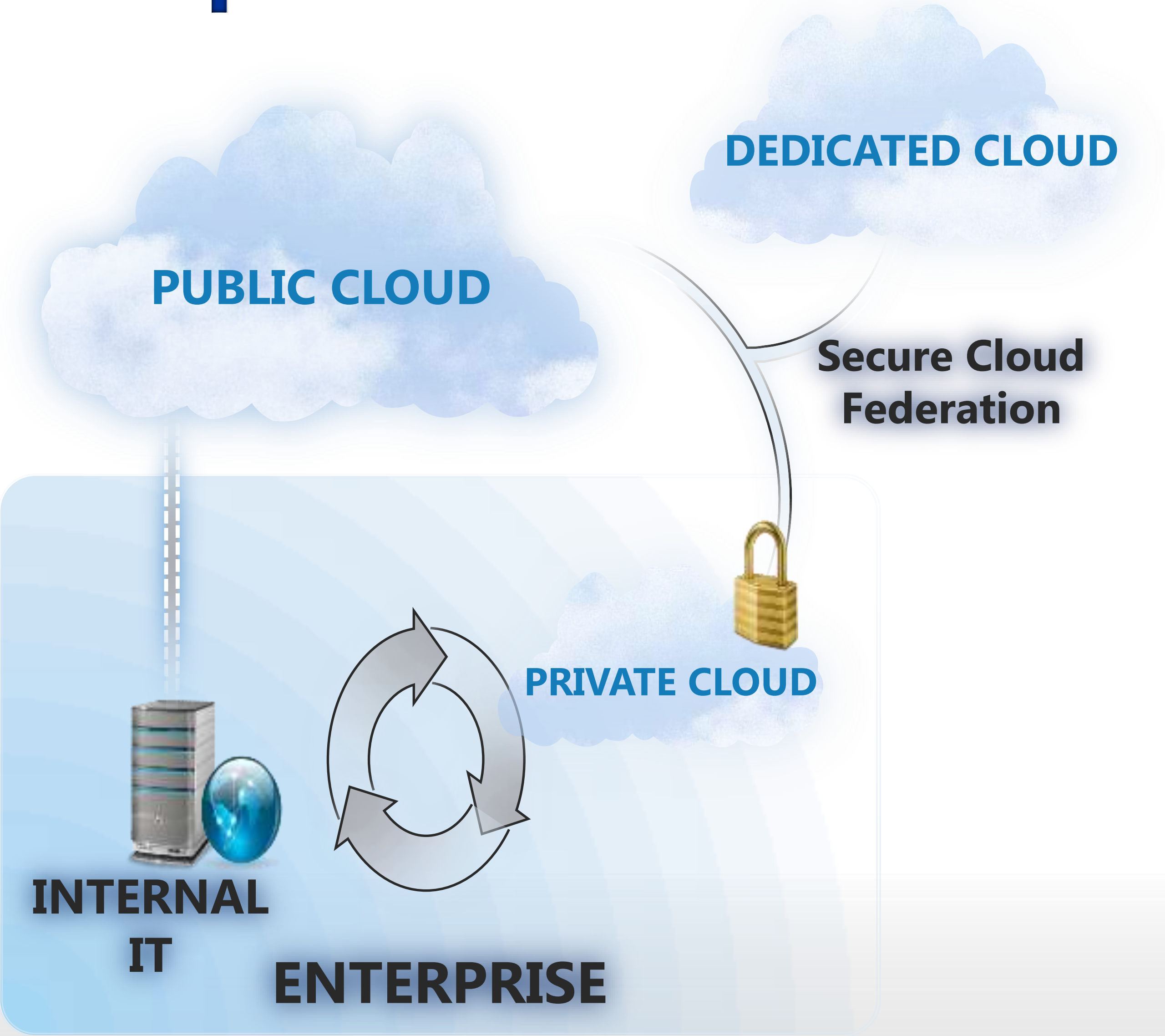
(From Jim Gray's last talk)

The Data-Intensive Research Lifecycle

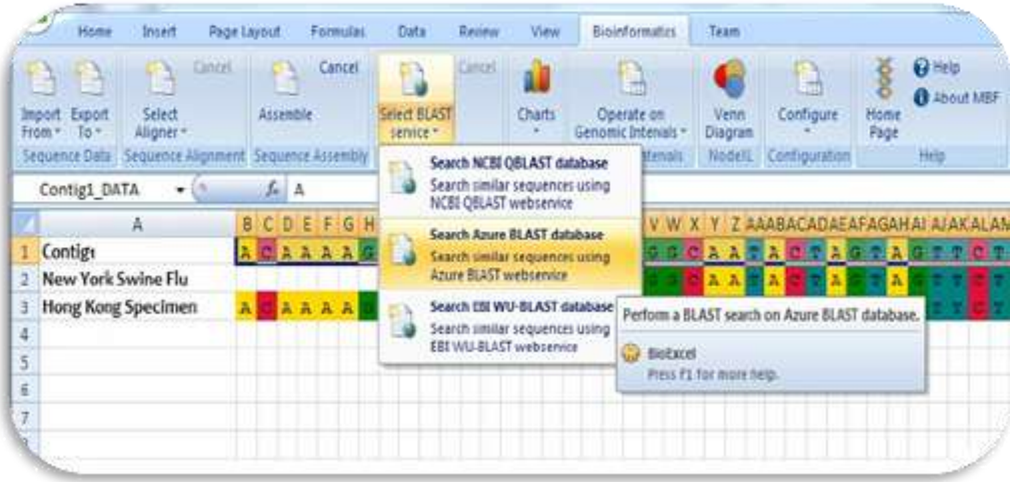


Acquisition & Modeling

The Cloud - Options

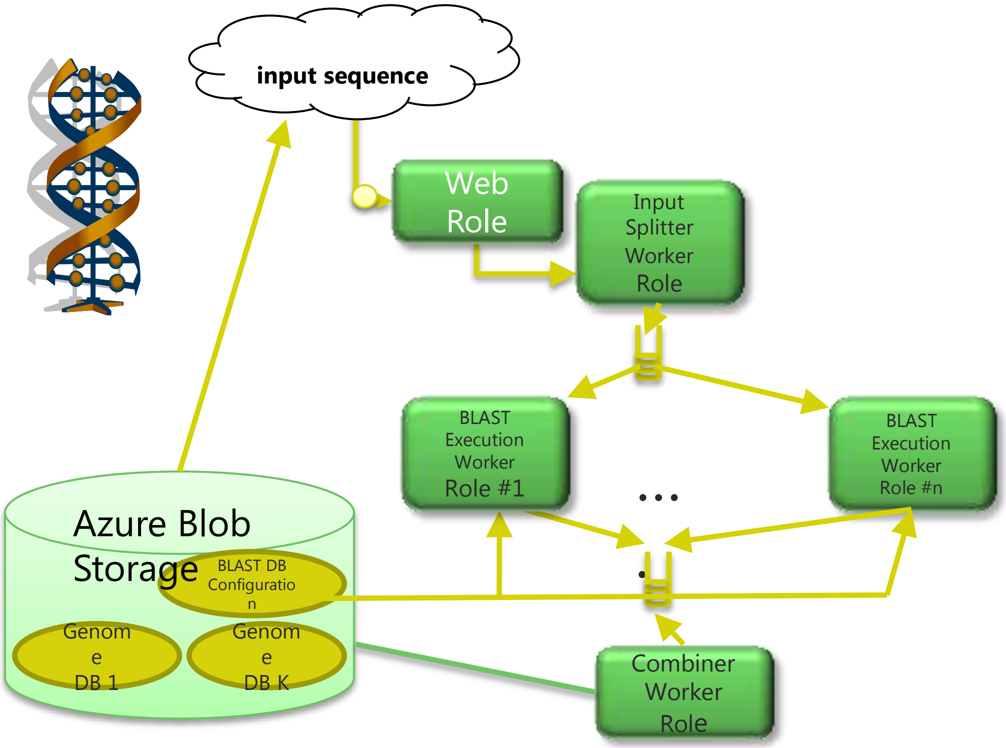


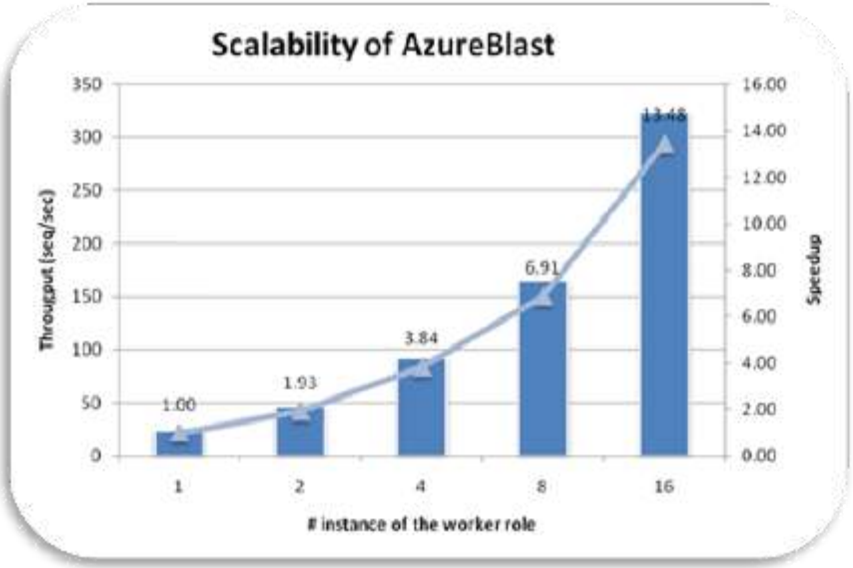
AzureBLAST



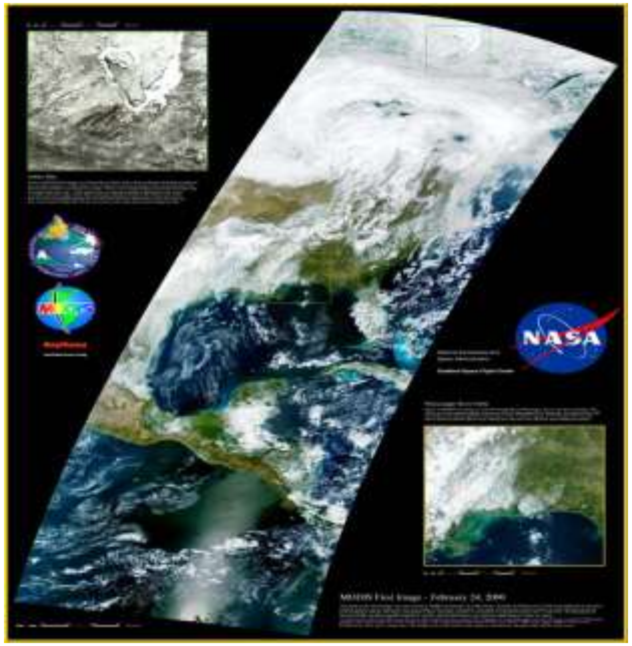
Seamless Experience

- Evaluate data and invoke computational models from Excel.
- Computationally heavy analysis done close to large database of curated data.
- Scalable for large, surge computationally heavy analysis.
- Test local, run on the cloud.





AzureMODIS – *Remote Sensing Geoscience*

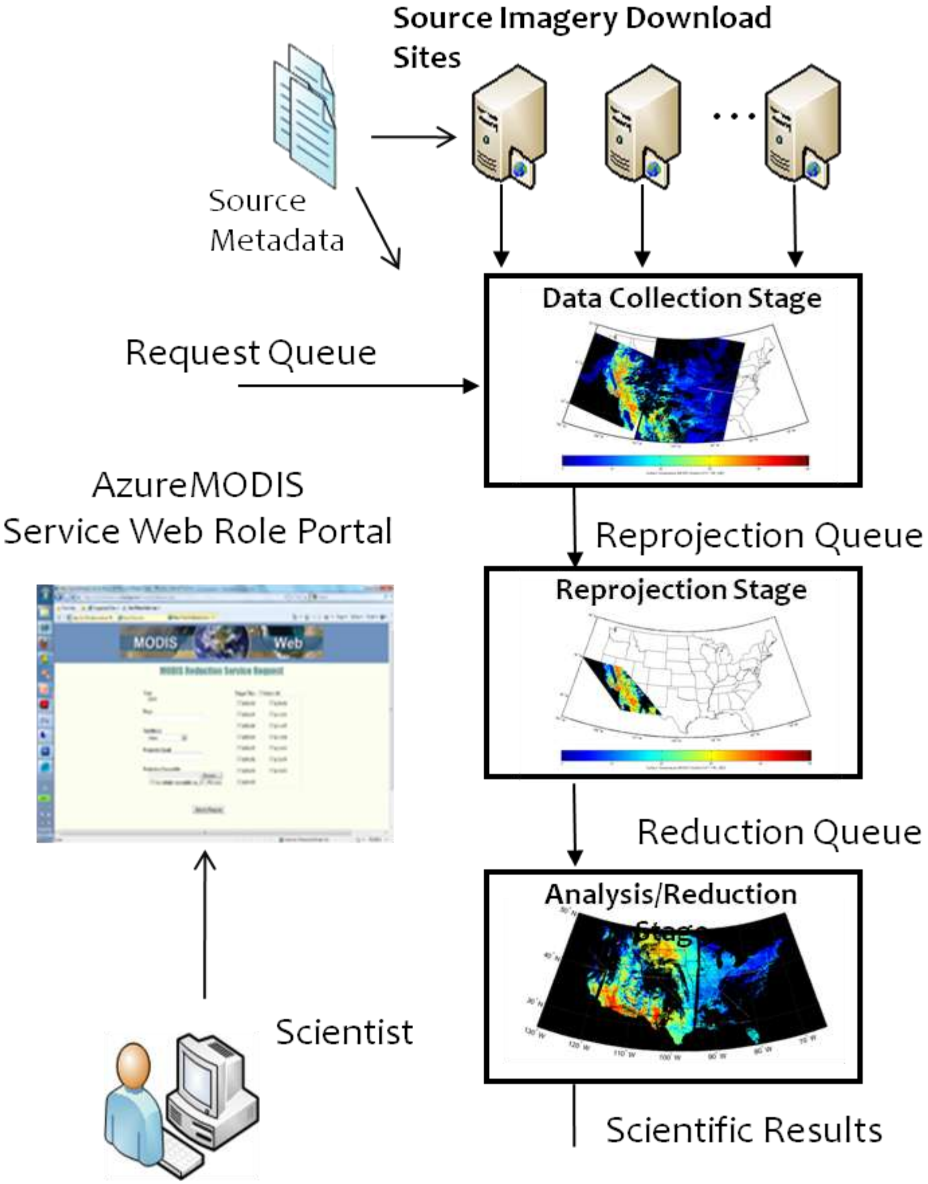


5 TB (~600K files) upload of 9 different imagery products from 15 different locations (~6 days of download)

4 TB reprojected harmonized imagery ~35000 cpu hours

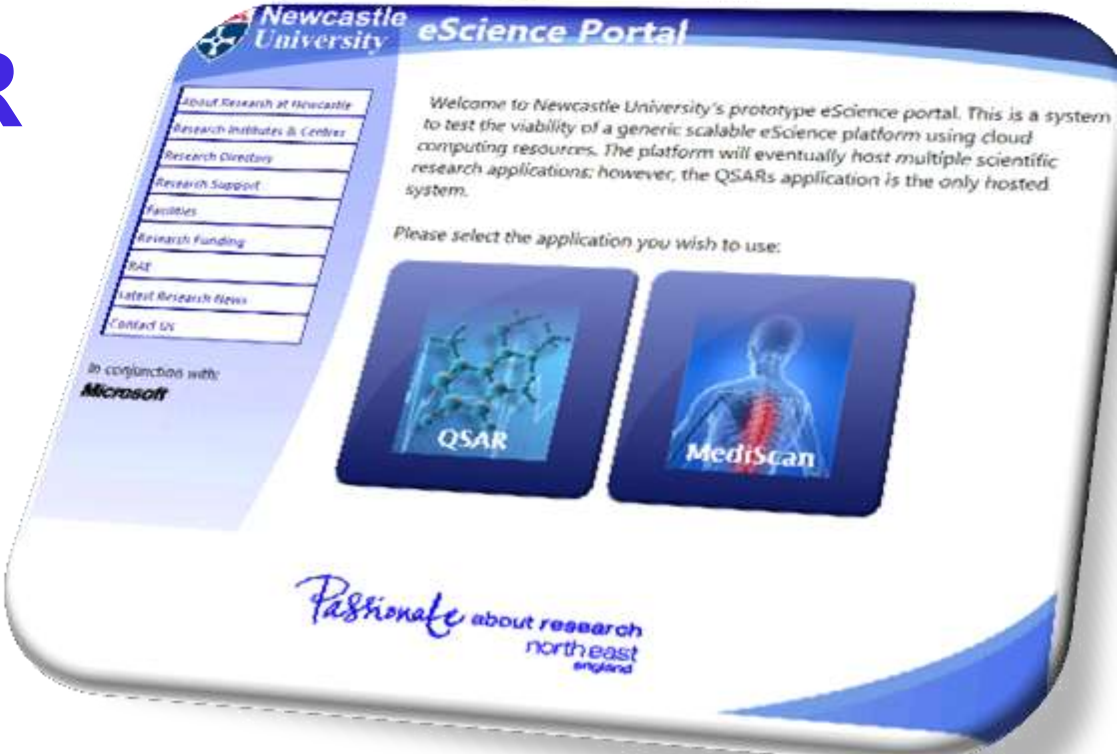
50 GB reduced science variable results ~18000 cpu hours (~14 hour download)

50 GB additional reduced science analysis results ~18000 cpu hours (~14 hour download)



Project JUNIOR

- **Newcastle University, UK**
  Paul Watson
  - Investigating applicability of commercial clouds for scientific research
  - Build a working prototype for use-cases
    in chemo-informatics
  - Uses Microsoft technologies to build science-related services (Windows Azure, Silverlight…)
  - Exploits Azure and Amazon Clouds
- **Built initial proof-of-concept**
  - Silverlight UI for basic Quantitative Structure-Analysis Relationship (QSAR) modeling
  - Demonstrated ability to scale QSAR computations
    in Windows Azure



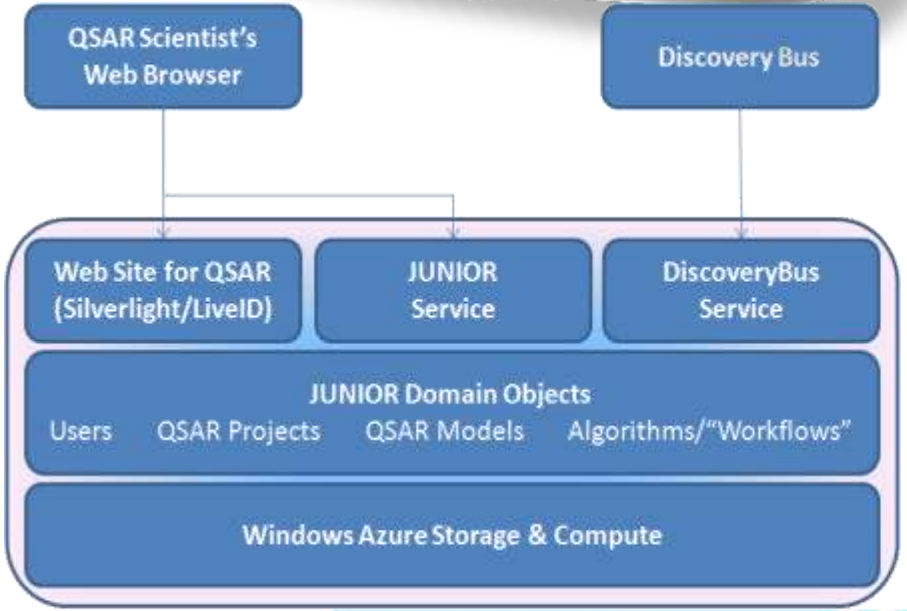
Newcastle University eScience Portal

- About Research at Newcastle
- Research Institutes & Centres
- Research Directory
- Research Support
- Facilities
- Research Funding
- RAE
- Latest Research News
- Contact Us

In conjunction with:
Microsoft

Welcome to Newcastle University's prototype eScience portal. This is a system to test the viability of a generic scalable eScience platform using cloud computing resources. The platform will eventually host multiple scientific research applications; however, the QSARs application is the only hosted system.

Please select the application you wish to use:



Collaboration & Visualization

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

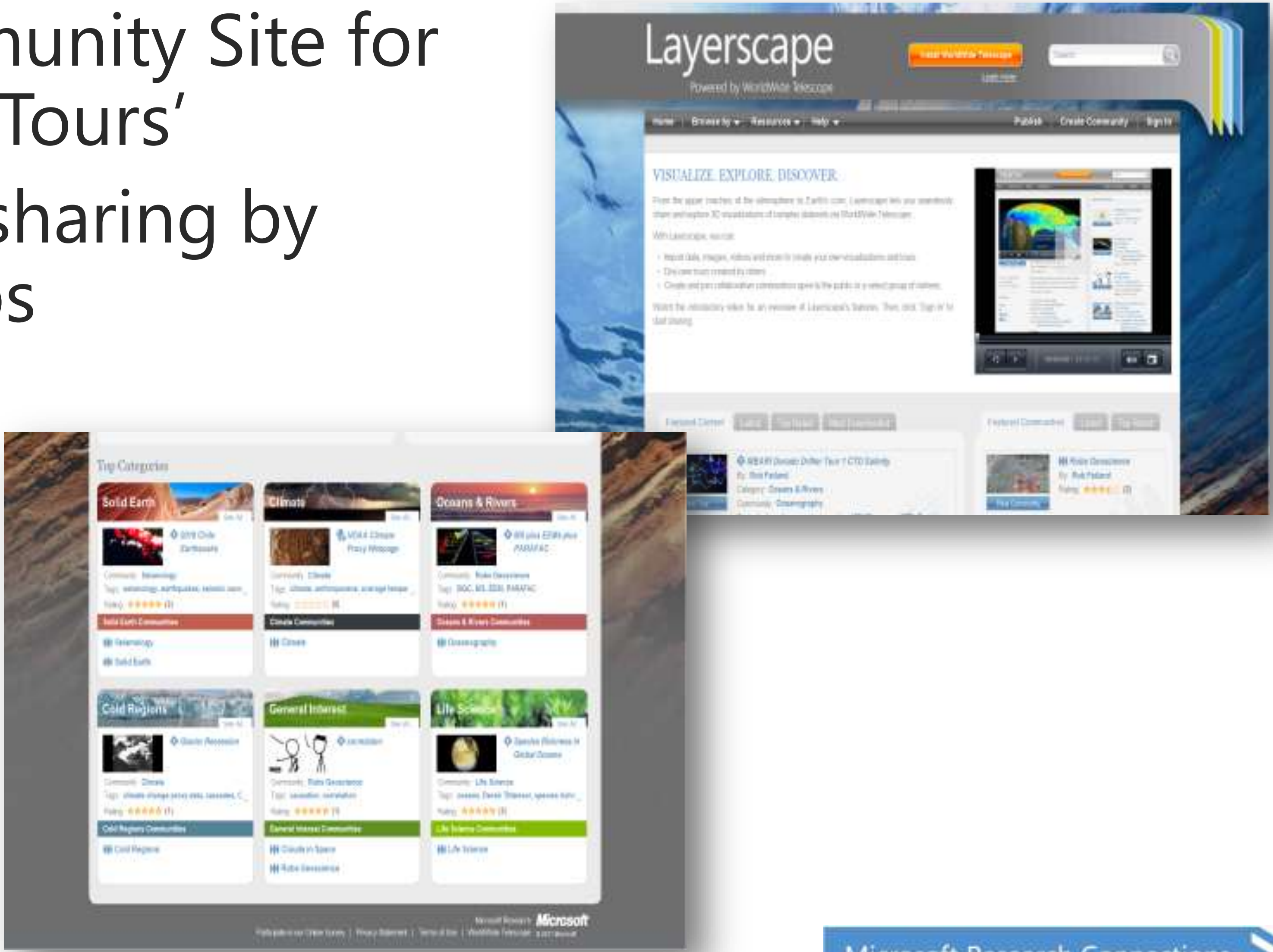
- Over 4M unique users
- Average number of WWT users is over 8K per day

See TED talk by Roy Gould and Curtis Wong

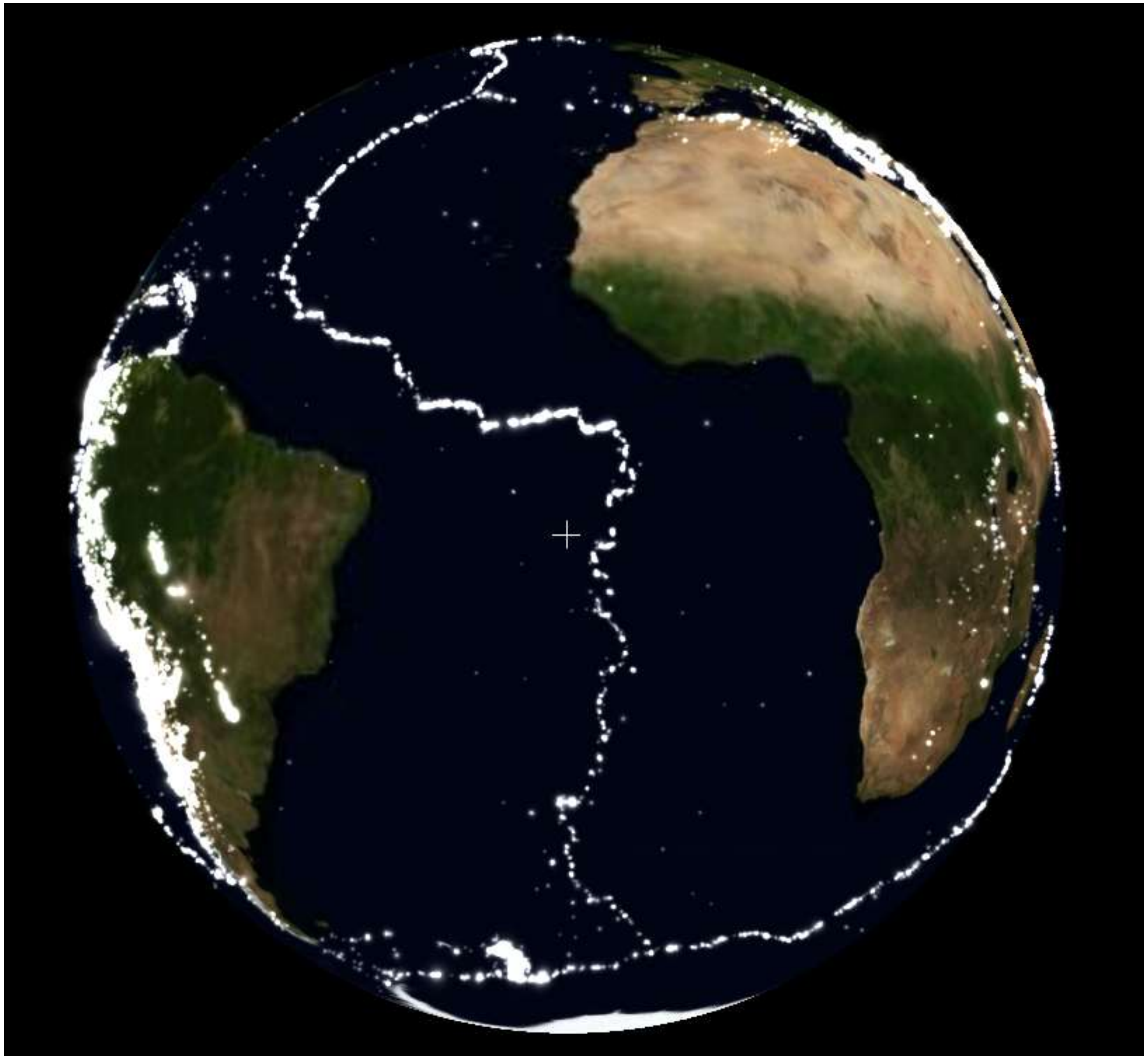
http://www.youtube.com/watch?v=NPu2j3JVmnw&feature=related

Layerscape

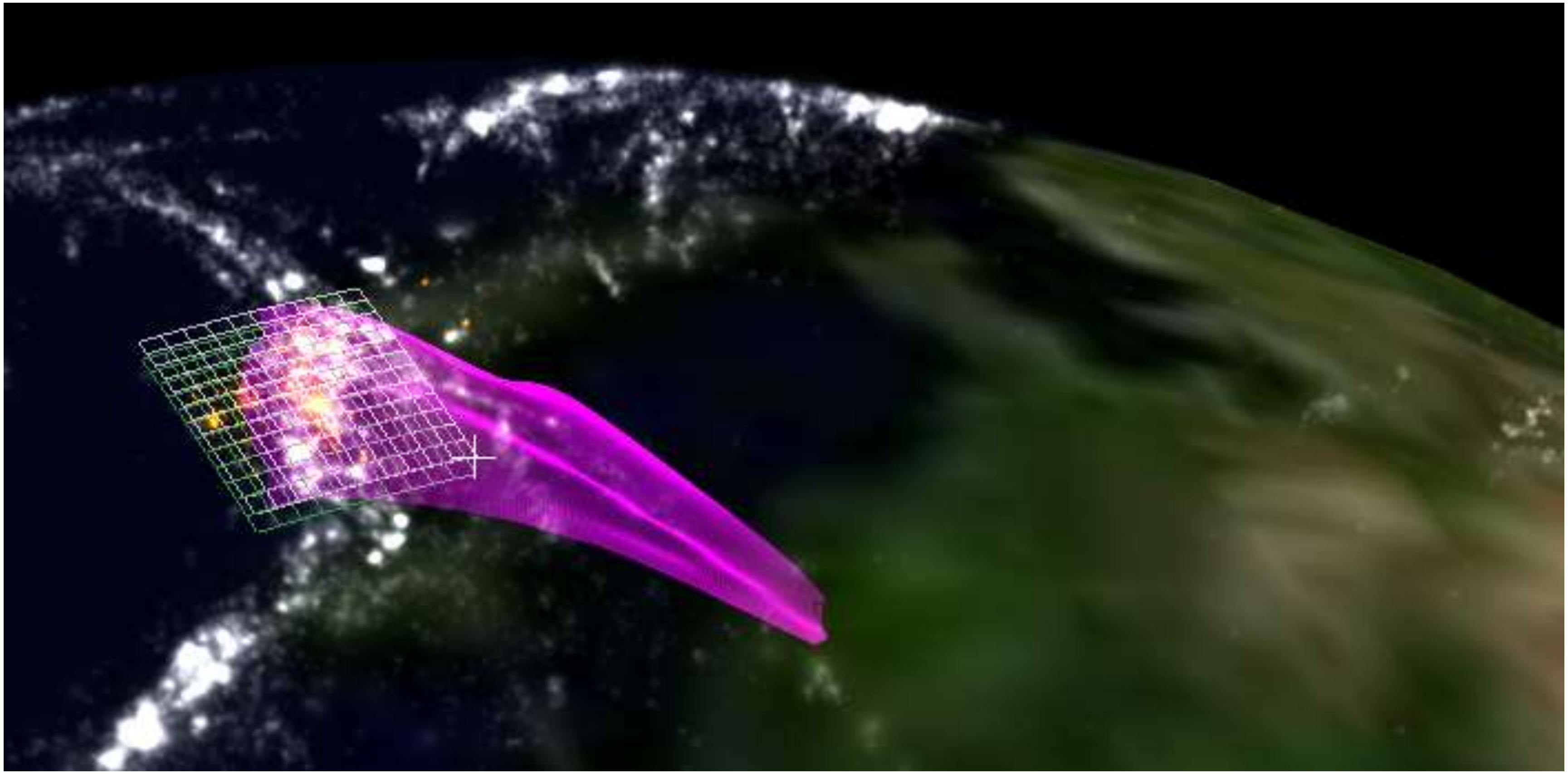
- Community Site for 'Data Tours'
- Data sharing by groups



Layerscape with Seismic Data



Visualization of Models



Tohoku events (shallow) and subduction slab

Analysis & Data Mining

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

Fighting HIV with ML and HPC

- PhyloD.Net is a Bayes-net-based tool that deciphers evolution of HIV within a patient
- Developed by eScience research group and published in *Science*, March 2007
- Used by dozens of HIV research groups
- Analysis requires HPC to do tens of thousands of independent computations
- Integrated into .NET Bio open source library

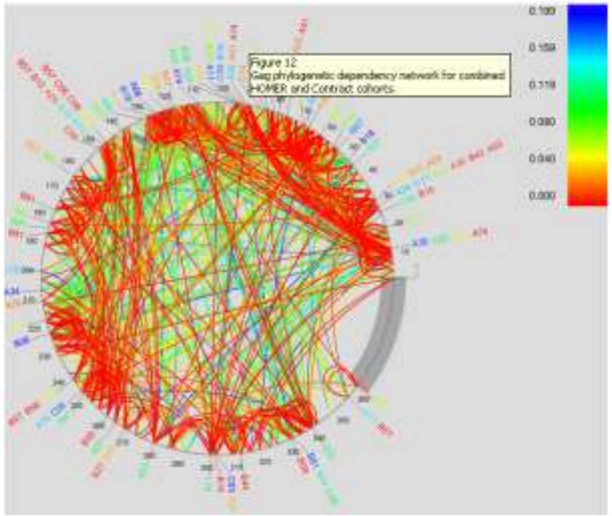


Figure 12
Gag phylogenetic dependency network for combined HOMER and Contract cohorts.
0.199
0.159
0.119
0.080
0.040
0.000

PhyloD.Net on cover of *PLoS Comp Bio*, Nov 2008
Carlson, Kadie, & Heckerman et al.

Open Data Protocol

"Web protocol for querying and updating data"

- Based on HTTP/ATOM
  - + Metadata
  - + Query options
  - + URI conventions
  - + JSON representation
- Uses REST semantics
- Open specification
- Submitted to OASIS

UK Met Office Weather Open Data

Data

Unlimited

£0.00

https://api.datamarket.azure.com/DataGovUK/MetOfficeWeatherOpenData/

www.odata.org

Dissemination & Sharing

The Berlin Declaration 2003

- ‘To promote the Internet as a functional instrument for a global scientific knowledge base and for human reflection’
- Defines open access contributions as including:
    - ‘original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material’

NSF Data Sharing Policy 2010

“Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.”

All future grant proposals now require a two-page Data Management Plan that addresses the above requirement and the Plan will be subject to peer review.

Datacite and ORCID

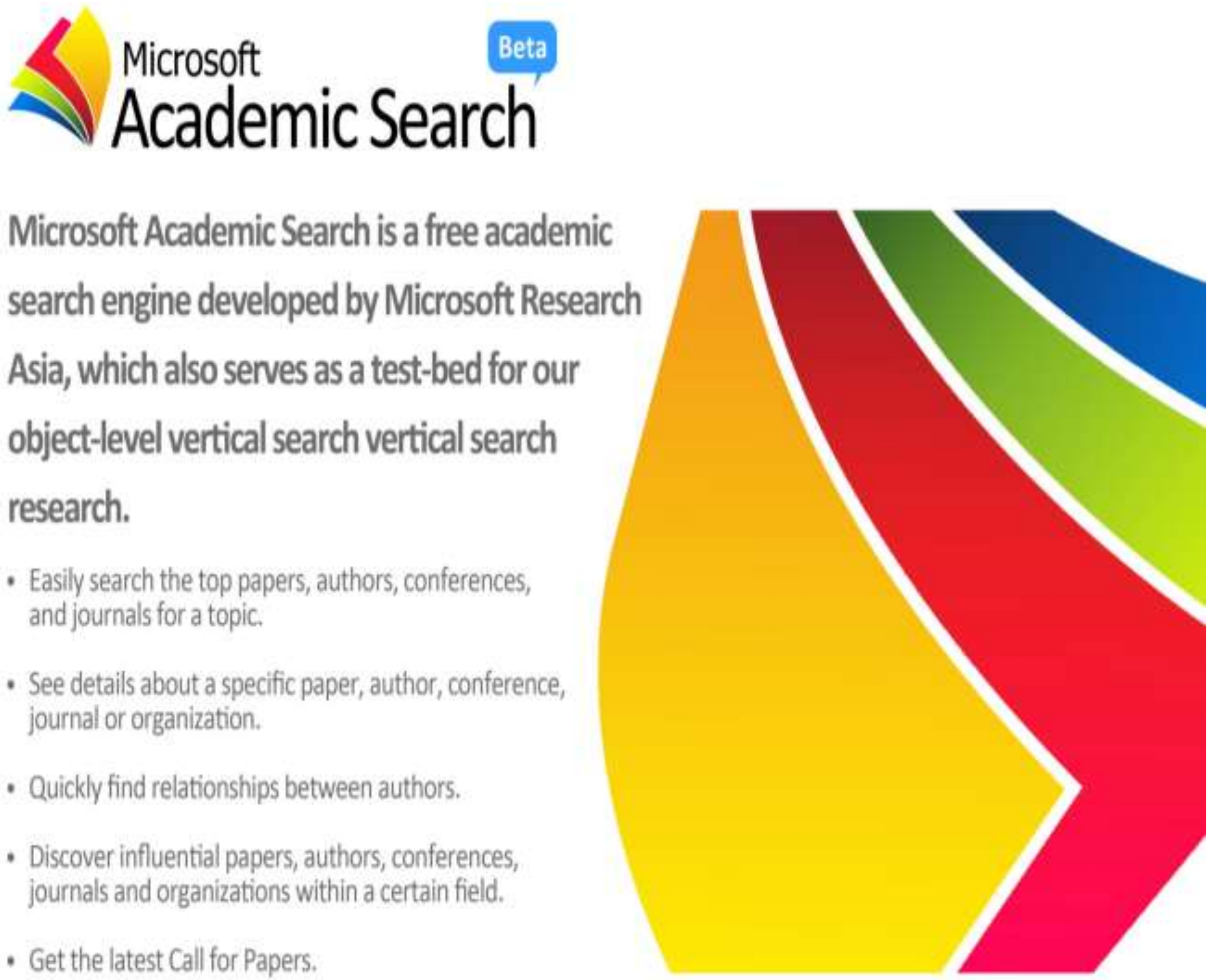
DataCite

- International consortium to establish easier access to scientific research data
- Increase acceptance of research data as legitimate, citable contributions to the scientific record
- Support data archiving that will permit results to be verified and re-purposed for future study.

ORCID - Open Research & Contributor ID

- Aims to solve the author/contributor name ambiguity problem in scholarly communications
- Central registry of unique identifiers for individual researchers
- Open and transparent linking mechanism between ORCID and other current author ID schemes.
- Identifiers can be linked to the researcher’s output to enhance the scientific discovery process

Microsoft Academic Search



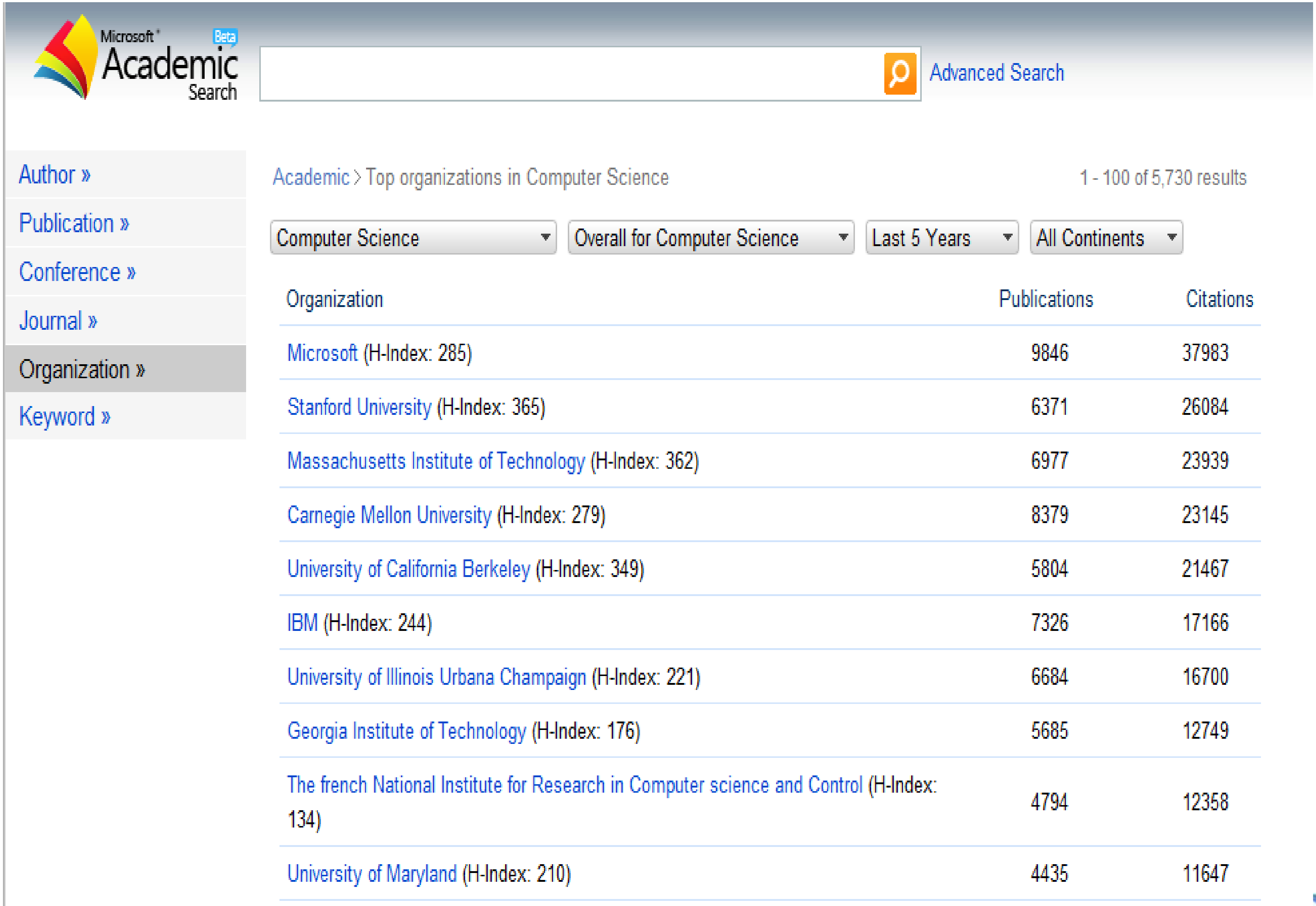
Microsoft Academic Search Beta

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/

Top 10 Computer Science Organizations



Academic > Top organizations in Computer Science

1 - 100 of 5,730 results

Computer Science ▾ | Overall for Computer Science ▾ | Last 5 Years ▾ | All Continents ▾

Organization	Publications	Citations
Microsoft (H-Index: 285)	9846	37983
Stanford University (H-Index: 365)	6371	26084
Massachusetts Institute of Technology (H-Index: 362)	6977	23939
Carnegie Mellon University (H-Index: 279)	8379	23145
University of California Berkeley (H-Index: 349)	5804	21467
IBM (H-Index: 244)	7326	17166
University of Illinois Urbana Champaign (H-Index: 221)	6684	16700
Georgia Institute of Technology (H-Index: 176)	5685	12749
The french National Institute for Research in Computer science and Control (H-Index: 134)	4794	12358
University of Maryland (H-Index: 210)	4435	11647

Public API for Academic Search

- **Application Programming Interface**
  - Supports queries against all academic entities and their basic information
- **With the API, you can**
  - Work with others to share information
  - Help users to build useful clients
- **Openly available to everyone**
  - Targeting the academic community
  - API is available for non-commercial use only

API details at:

http://academic.research.microsoft.com/About/Help.htm#5

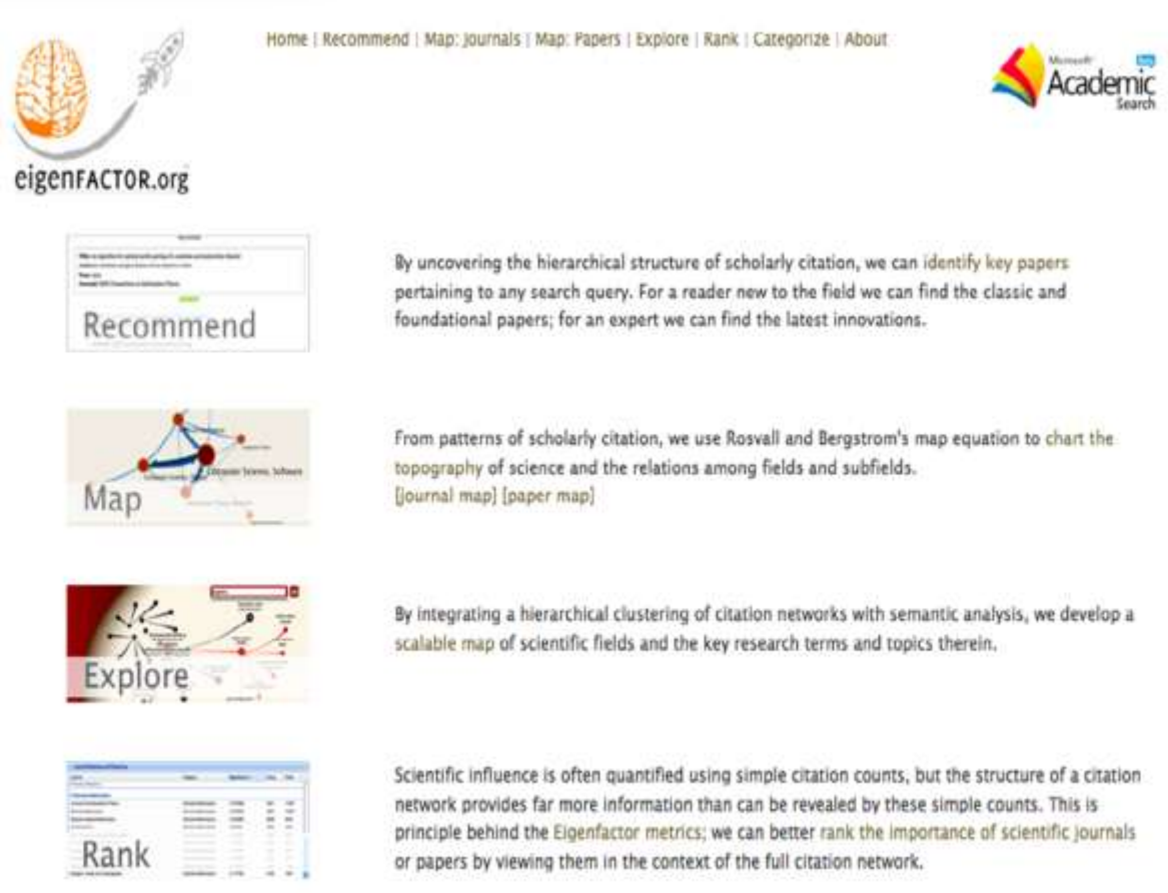
The Eigenfactor Project

eigenFACTOR.org™

RANKING AND MAPPING SCIENTIFIC KNOWLEDGE

http://mas.eigenfactor.org/

- The *Eigenfactor Project*™ is a non-commercial academic research project in the Department of Biology at the University of Washington.
- We aim to use recent advances in network analysis and information theory to develop novel methods for evaluating the influence of scholarly periodicals and for mapping the structure of academic research.
- We are committed to broadly disseminating our research findings and technological developments, while respecting the confidentiality of the data sources we use.

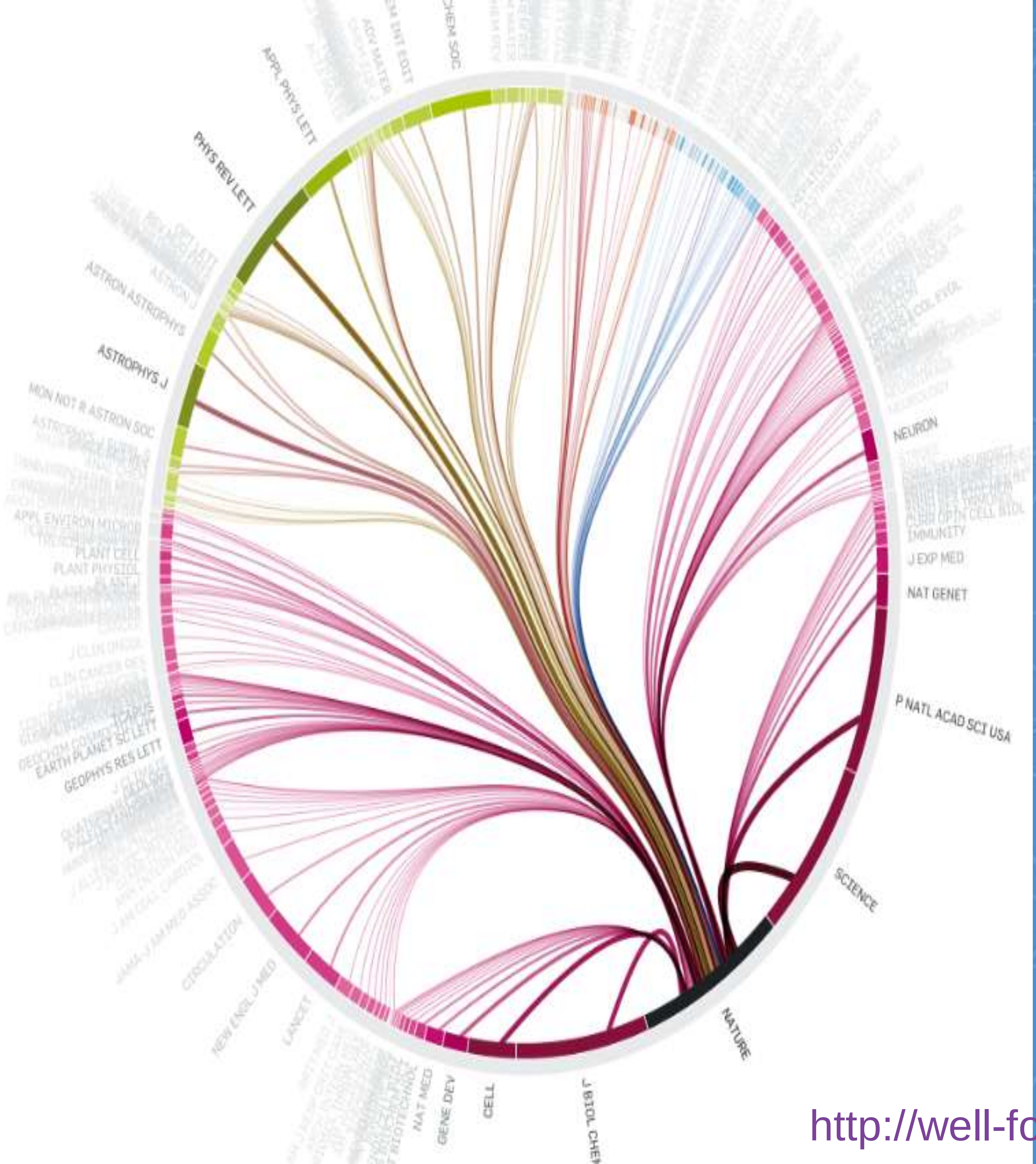

Home | Recommend | Map: Journals | Map: Papers | Explore | Rank | Categorize | About
Microsoft Academic Search
eigenFACTOR.org
Recommend
Map
Explore
Rank

By uncovering the hierarchical structure of scholarly citation, we can identify key papers pertaining to any search query. For a reader new to the field we can find the classic and foundational papers; for an expert we can find the latest innovations.

From patterns of scholarly citation, we use Rosvall and Bergstrom's map equation to chart the topography of science and the relations among fields and subfields.
[journal map] [paper map]

By integrating a hierarchical clustering of citation networks with semantic analysis, we develop a scalable map of scientific fields and the key research terms and topics therein.

Scientific influence is often quantified using simple citation counts, but the structure of a citation network provides far more information than can be revealed by these simple counts. This is principle behind the Eigenfactor metrics; we can better rank the importance of scientific journals or papers by viewing them in the context of the full citation network.



http://well-formed.eigenfac

Archiving & Preservation

Key driver from a UK Research Council

EPSRC Policy Framework on research data (May 2011)

- “all institutions in receipt of their funding should develop a clear roadmap for research data management, which should be implemented by May 1st 2015”
- “organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10 years”

Data Curation for Excel (DataUp)

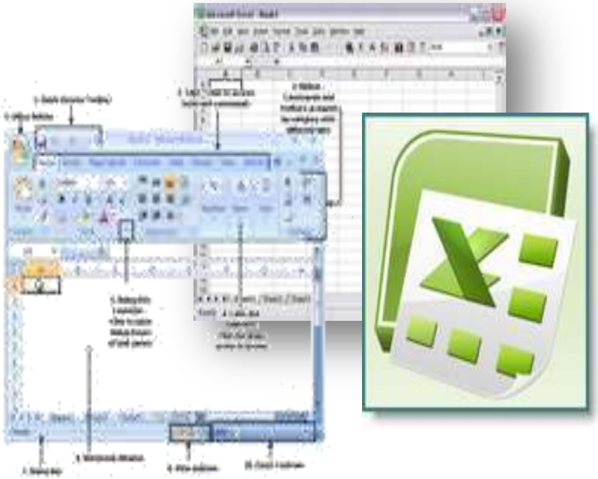
- Excel add-in and HTML5 browser app
- Microsoft Research with California Digital Library's Curation Center
- Part of the DataONE (an NSF DataNet Project)
  - Goal to facilitate data management, sharing, and archiving for scientists.
  - Open-source add-in and service for Microsoft Excel to assist in 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.

University of California CDL California Digital Library

GORDON AND BETTY MOORE FOUNDATION

DataONE

http://dcxl.cdlib.org/



- Ensure long-term access to Europe’s cultural and scientific heritage
  - Improve decision-making about long term preservation
  - Ensure long-term access to valued digital content
  - Control the costs through automation, scalable infrastructure
  - Ensure wide adoption across the user community
  - Establish market place for preservation services and tools
- Build practical solutions
  - Integrate existing expertise, designs and tools
  - Share and build

planets

Preservation and

Long-term

Access through

NETworked

Services

PLANETS Partners

The British Library
National Library, Netherlands
Austrian National Library
State and University Library, Denmark
Royal Library, Denmark

National Archives, UK
Swiss Federal Archives
National Archives, Netherlands

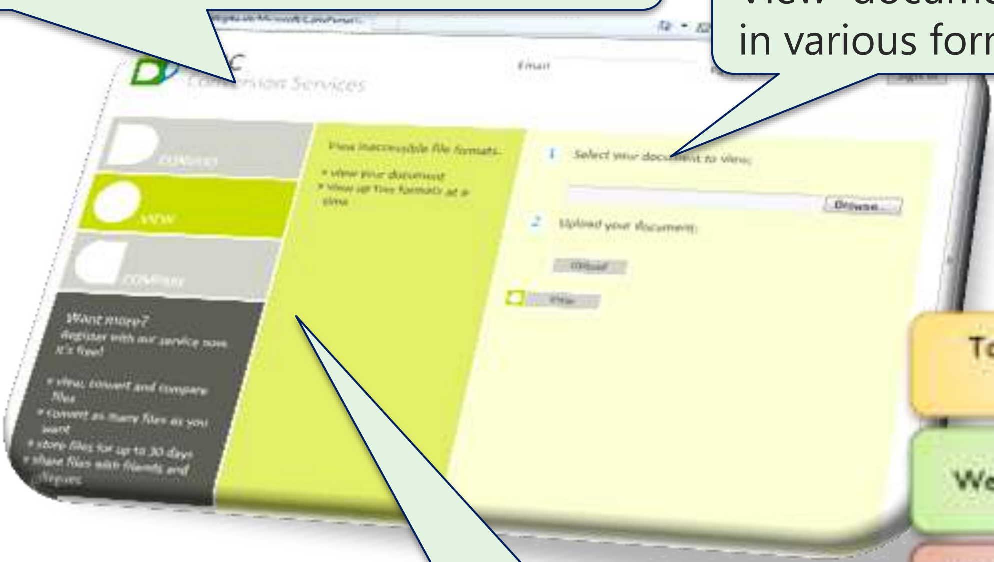
Hatii at University of Glasgow
University of Freiburg
Technical University of Vienna
University at Cologne

Tessella Plc
IBM Netherlands
Microsoft Research, Cambridge
ARC Seibersdorf research

Archiving and Preservation:
A Document Conversion Service

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

View documents in various formats



Conversion Services

Email

convert

view

compare

View inaccessible file formats.

- view your document
- view up five formats at a time

1 Select your document to view:

Browse...

2 Upload your document:

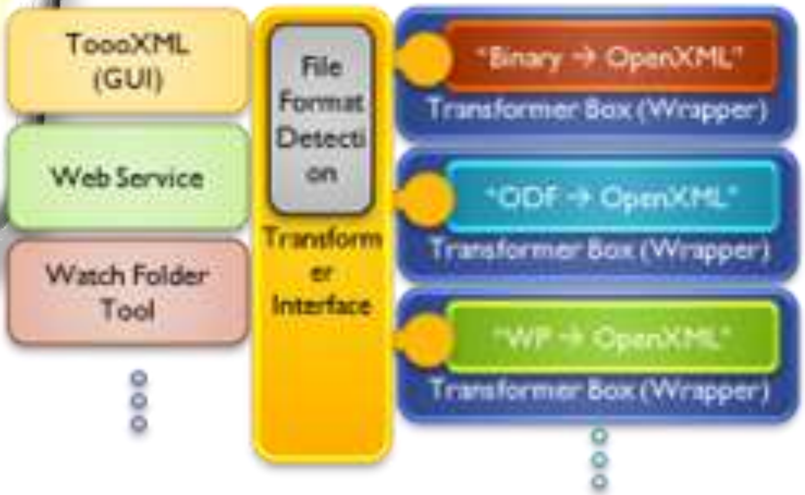
Upload

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

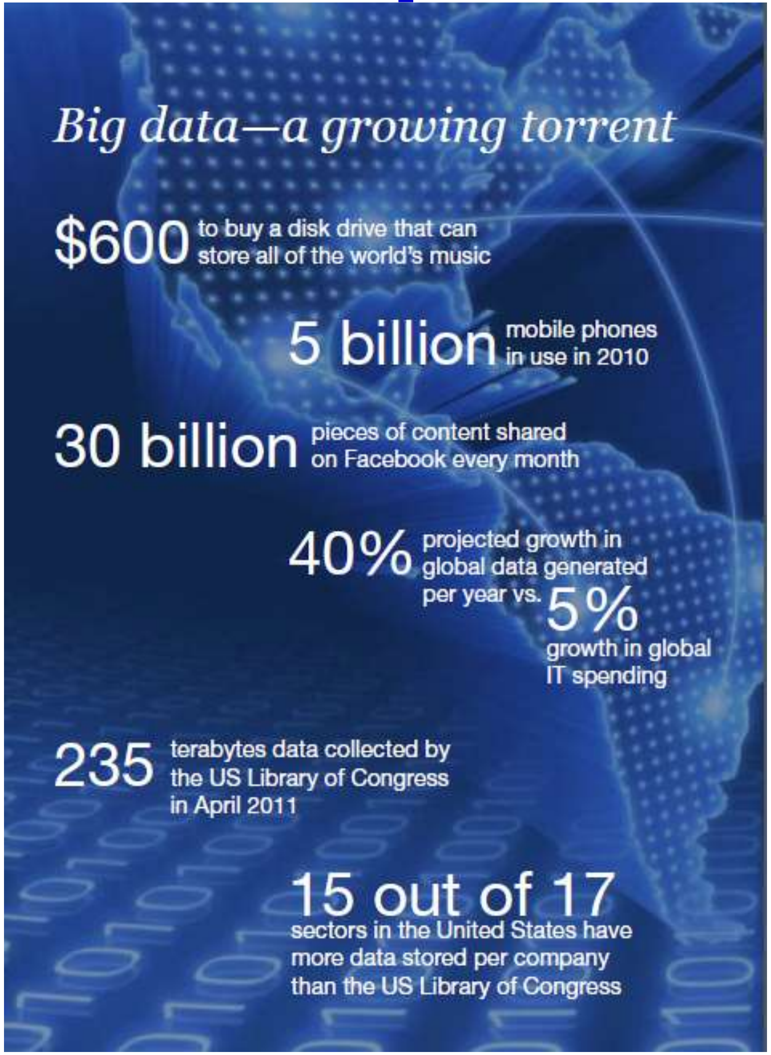
Compare original and converted documents



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

Future
NEXT EXIT ↗

Big data: The next frontier for innovation, competition, and productivity



Big data—a growing torrent

$600 to buy a disk drive that can store all of the world's music

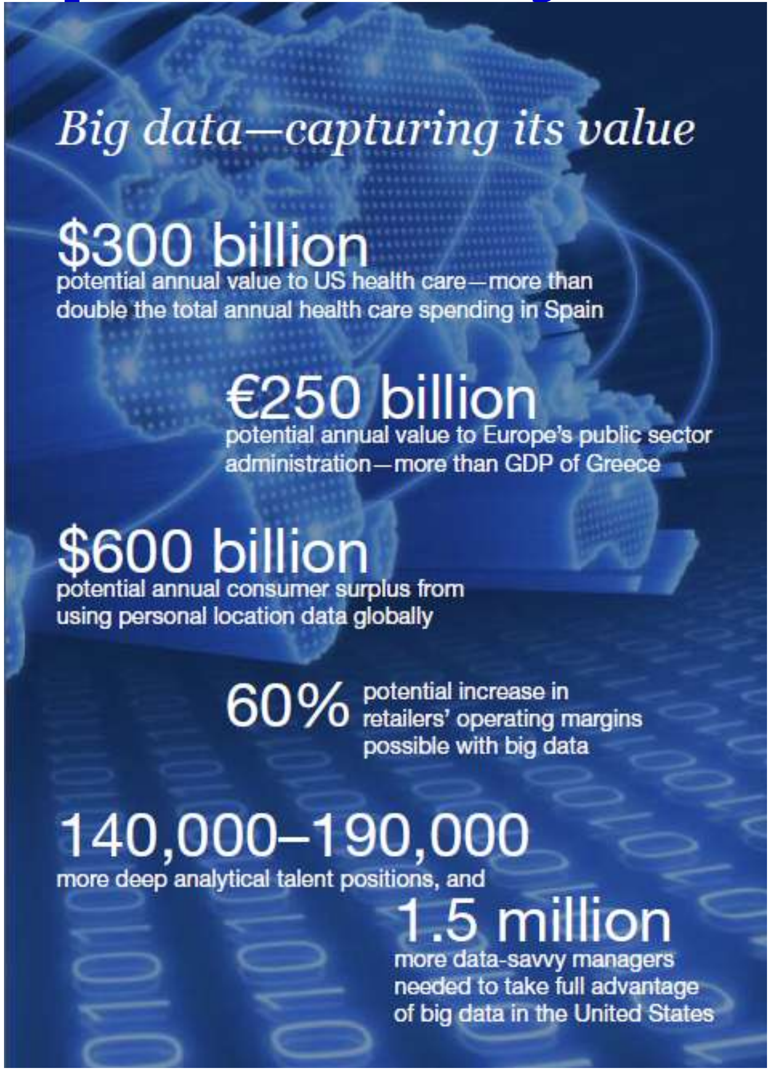
5 billion mobile phones in use in 2010

30 billion pieces of content shared on Facebook every month

40% projected growth in global data generated per year vs. **5%** growth in global IT spending

235 terabytes data collected by the US Library of Congress in April 2011

15 out of 17 sectors in the United States have more data stored per company than the US Library of Congress



Big data—capturing its value

$300 billion potential annual value to US health care—more than double the total annual health care spending in Spain

€250 billion potential annual value to Europe's public sector administration—more than GDP of Greece

$600 billion potential annual consumer surplus from using personal location data globally

60% potential increase in retailers' operating margins possible with big data

140,000–190,000 more deep analytical talent positions, and **1.5 million** more data-savvy managers needed to take full advantage of big data in the United States

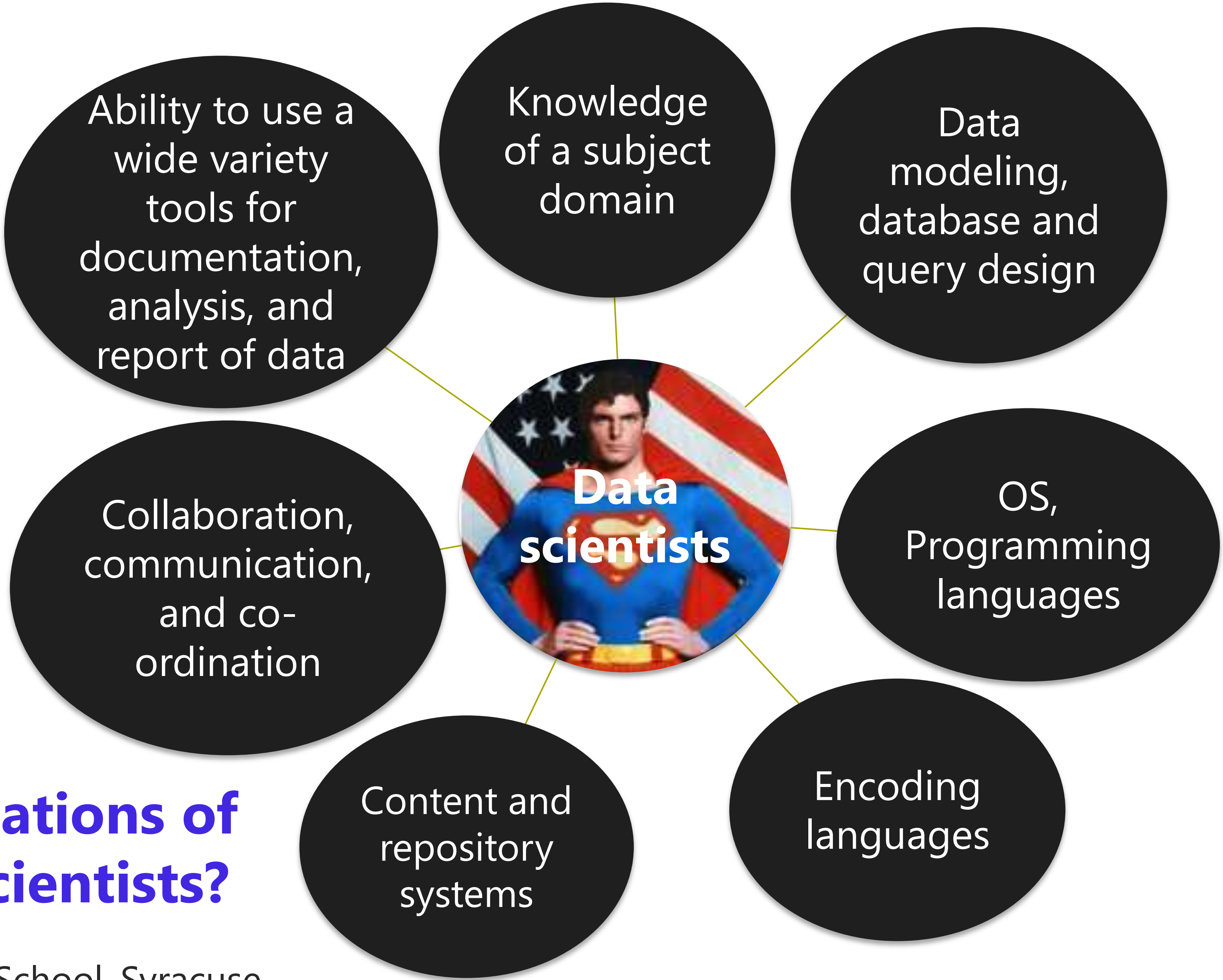
McKinsey Global Institute Report
May 2011

Educating a New Breed of Data Scientists for Scientific Data Management

Jian Qin

School of Information Studies
Syracuse University

Microsoft eScience Workshop, Chicago,
October 9, 2012



Data scientists

Ability to use a wide variety tools for documentation, analysis, and report of data

Knowledge of a subject domain

Data modeling, database and query design

Collaboration, communication, and co-ordination

OS, Programming languages

Content and repository systems

Encoding languages

Expectations of data scientists?

Jian Qin, i-School, Syracuse

From talk at 2012 Microsoft eScience Workshop

Microsoft Research Connections

Semantic Computing?

computers are great **tools** for

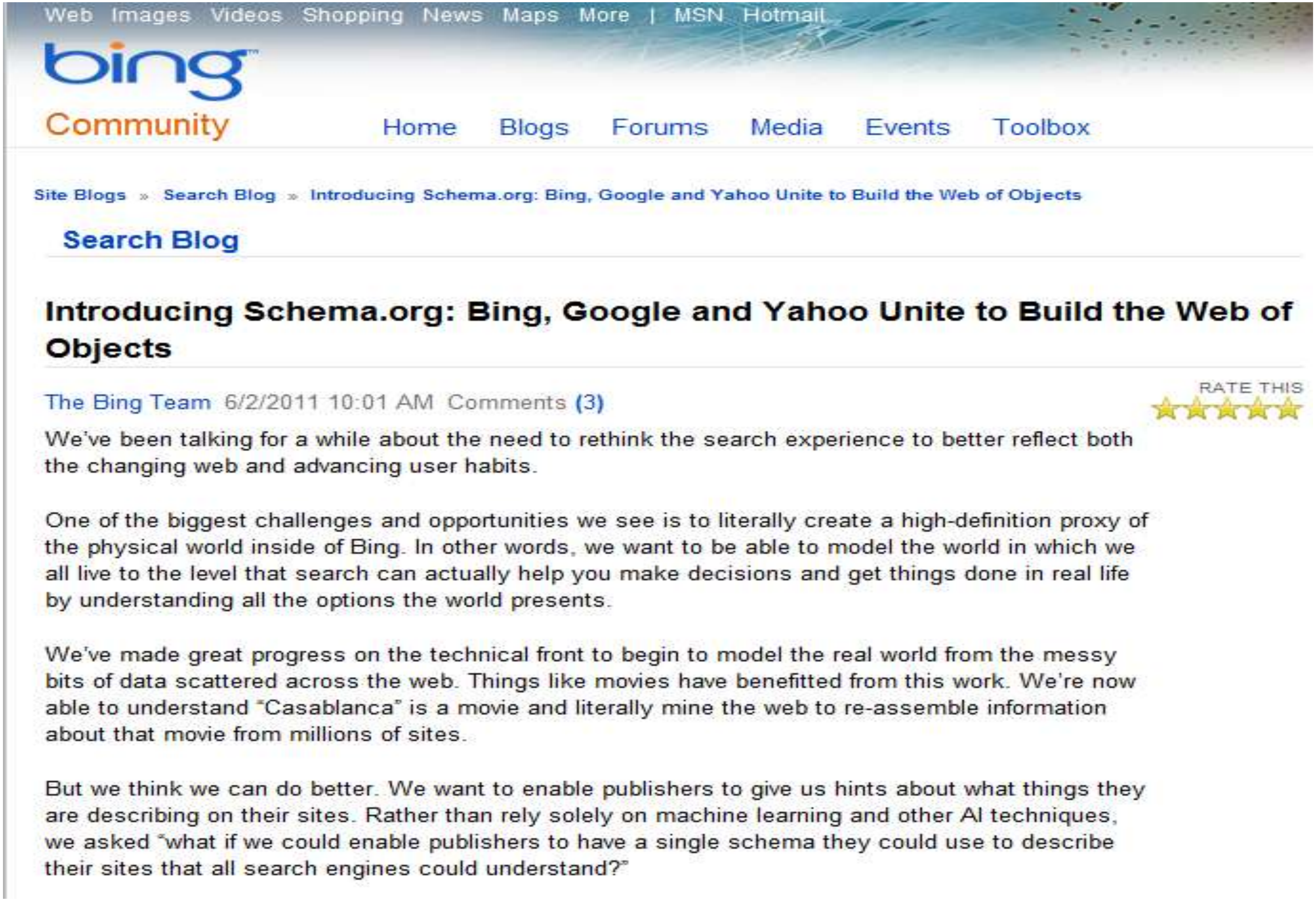
storing	computing
managing	indexing

huge amounts of **data**

we would like computers to also help with the **automatic**

acquisition	discovery
aggregation	organization
correlation	analysis
interpretation	inference

of the world's **information** and **knowledge**


Web Images Videos Shopping News Maps More | MSN Hotmail
bing
Community
Home Blogs Forums Media Events Toolbox
Site Blogs » Search Blog » Introducing Schema.org: Bing, Google and Yahoo Unite to Build the Web of Objects
Search Blog
RATE THIS

Introducing Schema.org: Bing, Google and Yahoo Unite to Build the Web of Objects

The Bing Team 6/2/2011 10:01 AM Comments (3)

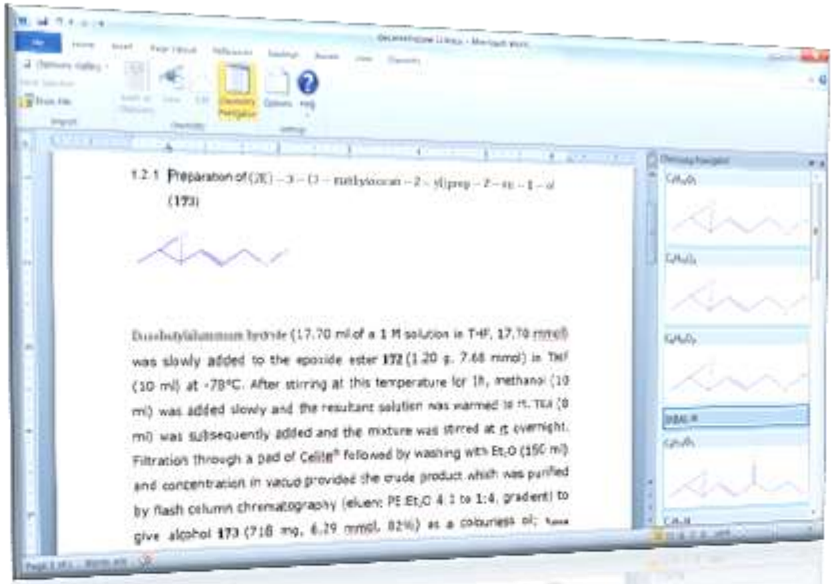
We've been talking for a while about the need to rethink the search experience to better reflect both the changing web and advancing user habits.

One of the biggest challenges and opportunities we see is to literally create a high-definition proxy of the physical world inside of Bing. In other words, we want to be able to model the world in which we all live to the level that search can actually help you make decisions and get things done in real life by understanding all the options the world presents.

We've made great progress on the technical front to begin to model the real world from the messy bits of data scattered across the web. Things like movies have benefitted from this work. We're now able to understand "Casablanca" is a movie and literally mine the web to re-assemble information about that movie from millions of sites.

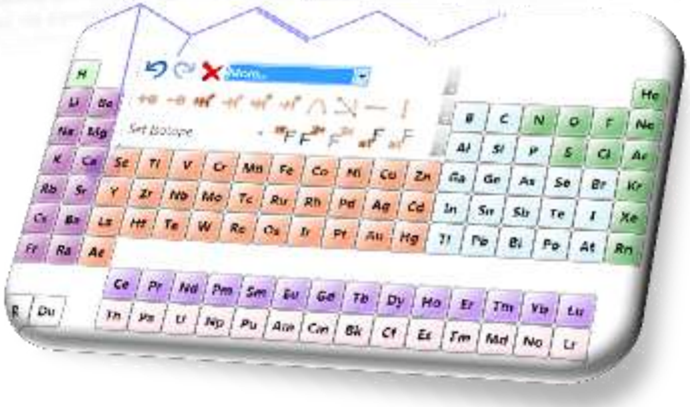
But we think we can do better. We want to enable publishers to give us hints about what things they are describing on their sites. Rather than rely solely on machine learning and other AI techniques, we asked "what if we could enable publishers to have a single schema they could use to describe their sites that all search engines could understand?"

Semantic Chemistry Add-in for Word



1.2.1 Preparation of (2E) – 3 – (3 – methyloxiran – 2 – yl)prop – 2 – en – 1 – ol (173)

Diisobutylaluminium hydride (17.70 ml of a 1 M solution in THF, 17.70 mmol) was slowly added to the epoxide ester 172 (1.20 g, 7.68 mmol) in THF (10 ml) at -78°C. After stirring at this temperature for 1h, methanol (10 ml) was added slowly and the resultant solution was warmed to rt. TEA (8 ml) was subsequently added and the mixture was stirred at rt overnight. Filtration through a pad of Celite® followed by washing with Et_2O (150 ml) and concentration in vacuo provided the crude product which was purified by flash column chromatography (eluent: PE:Et_2O 4:1 to 1:4, gradient) to give alcohol 173 (718 mg, 6.29 mmol, 82%) as a colourless oil;



- Authoring and rendering of semantic-rich chemical information (CML)
- In partnership with the University of Cambridge
- Support for Office 2007 and Office 2010
- Available under Apache 2.0
- Over **360K** downloads since March 22nd, 2010

Ontology Add-in for Word

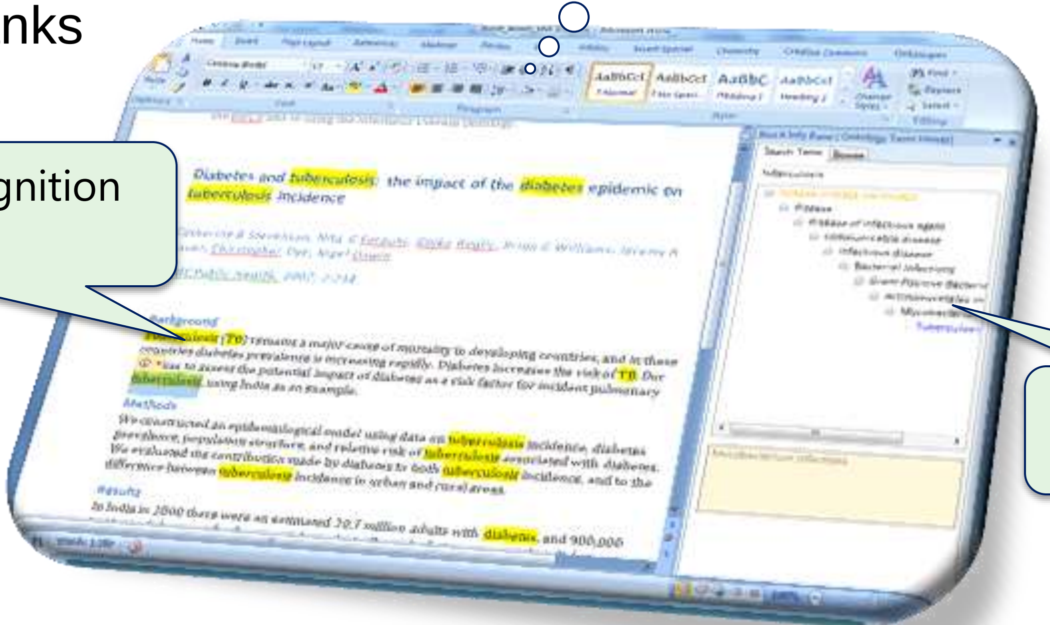
THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

science commons

- John Wilbanks

University of California San Diego

- Phil Bourne
- Lynn Fink



Services: Ontology download web service

Intent: Term recognition & disambiguation

Relationships: Ontology browser

Diabetes and tuberculosis: the impact of the diabetes epidemic on tuberculosis incidence

[illegible] Stevenson, Nisha [illegible] Christopher Dye, Nigel Unwin

BMC Public Health, 2007, 7:234

Background

Tuberculosis (TB) remains a major cause of mortality in developing countries, and in these countries diabetes prevalence is increasing rapidly. Diabetes increases the risk of TB. Our aim was to assess the potential impact of diabetes as a risk factor for incident pulmonary tuberculosis, using India as an example.

Methods

We constructed an epidemiological model using data on tuberculosis incidence, diabetes prevalence, population structure, and relative risk of tuberculosis associated with diabetes. We evaluated the contribution made by diabetes to both tuberculosis incidence, and to the difference between tuberculosis incidence in urban and rural areas.

Results

In India in 2000 there were an estimated 20.7 million adults with diabetes, and 900,000

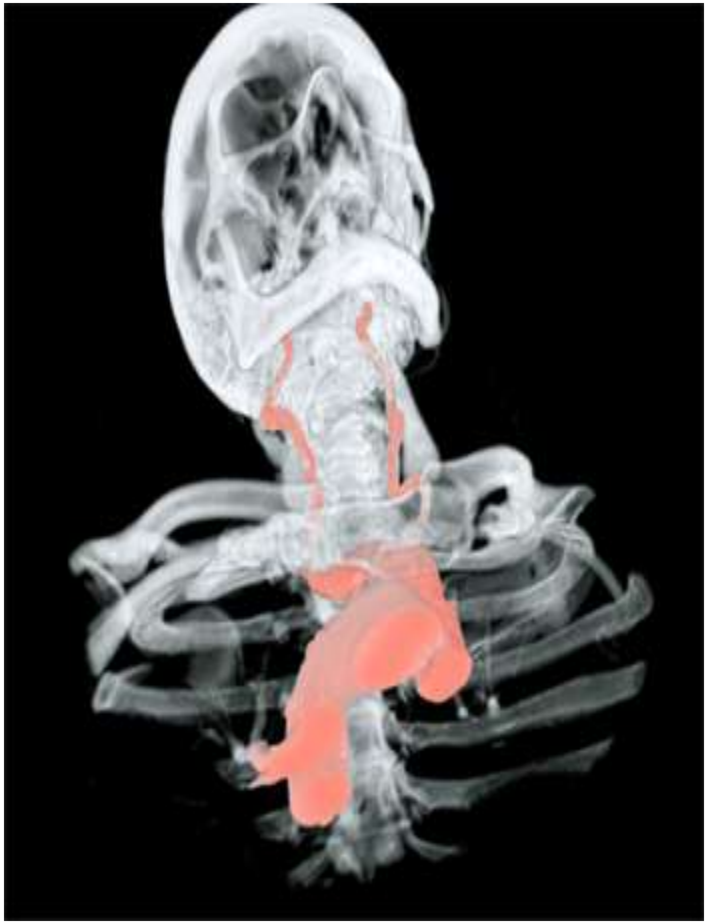
Downloads = 4,000+

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

InnerEye: Semantic Understanding of Medical Images

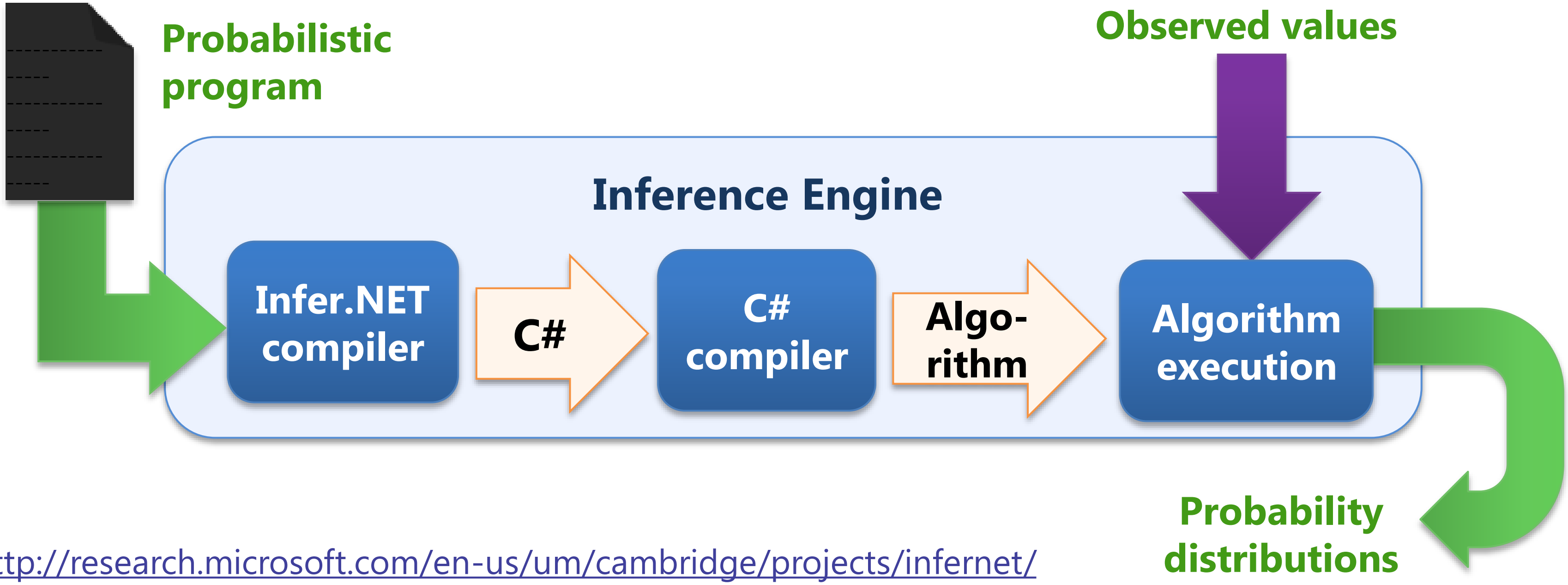
- InnerEye focuses on the analysis of patient scans using machine learning techniques for automatic detection and segmentation of healthy anatomy as well as anomalies.

- In this image, we see that InnerEye can separate a carotid artery visually from adjacent parts of a human body



Infer.NET
Machine Learning as a Service?

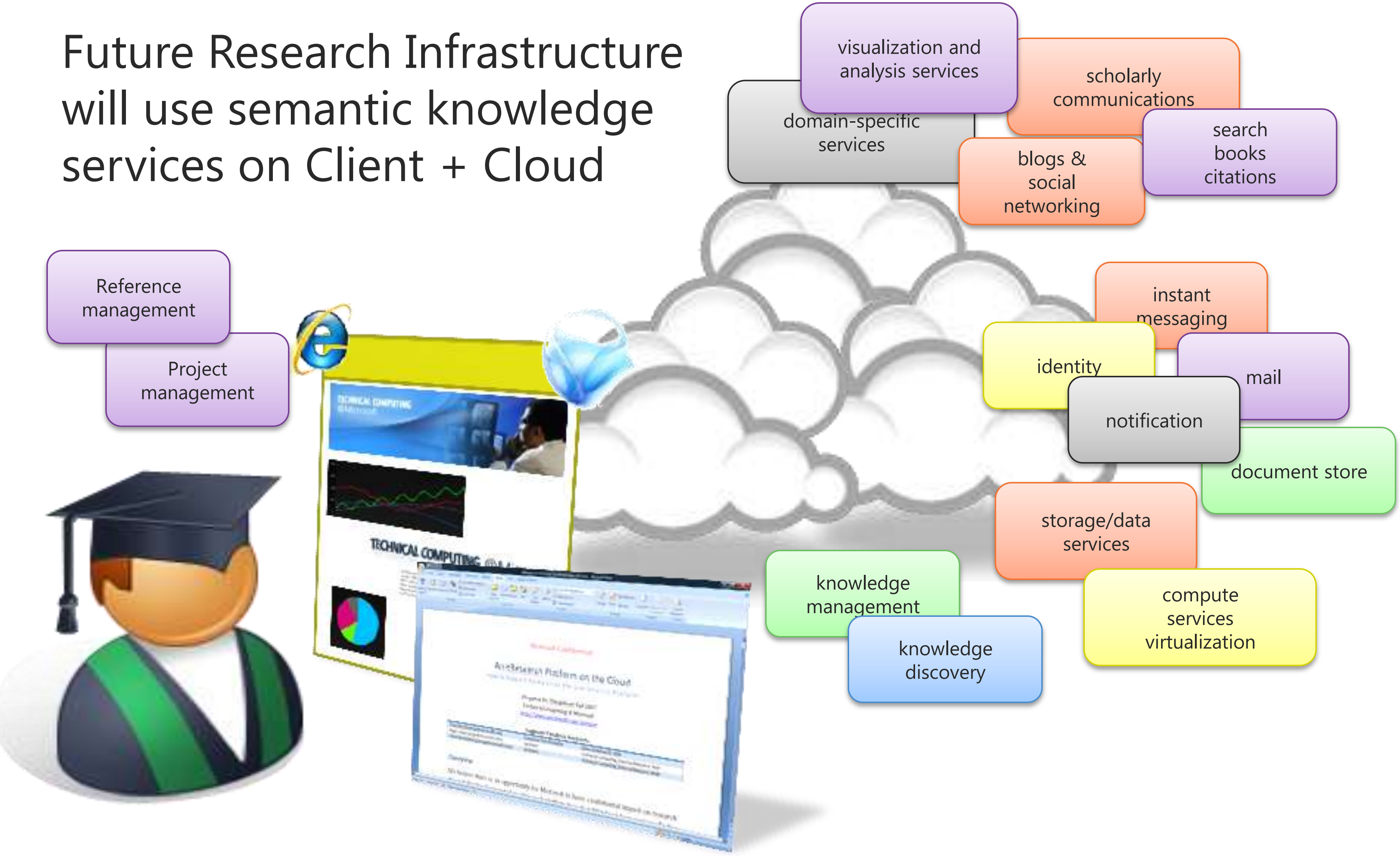
- Provides a probabilistic programming language allowing rapid development of new models
- Goal is to provide a platform for modern applications of Bayesian inference



http://research.microsoft.com/en-us/um/cambridge/projects/infernet/

Towards a Semantic Cyberinfrastructure?

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



Some Resources

- Microsoft Research
  - http://research.microsoft.com
  - Microsoft Research downloads: http://research.microsoft.com/research/downloads
- Microsoft Research Connections
  - 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
- Outercurve Foundation
  - http://www.outercurve.org/
- Tony Hey on eScience
  - http://tonyhey.net/

Microsoft®

Your potential. Our passion.™