The Fourth Paradigm: Particle Physics, Astronomy, Jim Gray and Open Access

Tony Hey
Vice President
Microsoft Research

Particle Physics and Open Access

Paul Ginsparg and arXiv

- The particle physics research community had a long tradition – since the 1960's - of circulating 'preprints' prior to peer review

- With the advent of word processing and the Web, Paul Ginsparg set up the LANL preprint archive at xxx.lanl.gov in 1991

- The repository expanded to other disciplines and changed its name to arXiv.org in 1999
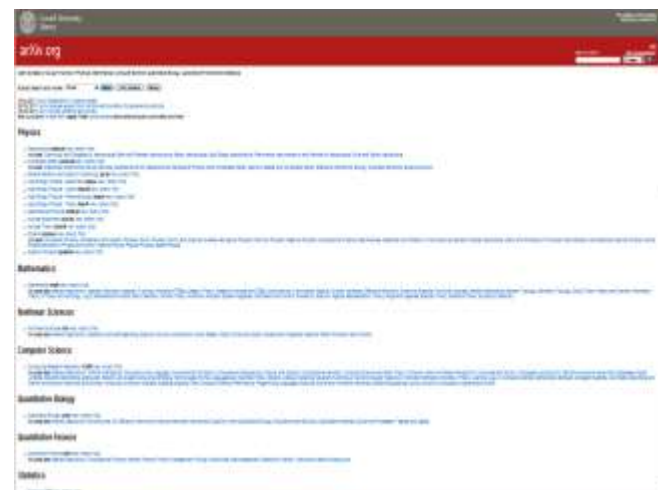
Comments (1)

- The arXiv repository is now over 20 years old and has a submission rate of over 7,000 e-prints per month

- Full text versions of over half a million research papers are available free both to researchers and to the general public.

- More than 200,000 articles are downloaded from arXiv each week by about 400,000 users.

- The arXiv repository 'publishes' e-prints prior to the refereeing process

- Acceptance in a journal is effectively a post-publication quality mark

Comments (2)

- The apparent drawback of multiple, slightly different versions of a paper turns out not to be a serious drawback in practice

- Gentil-Beccot, Mele and Brooks have made a detailed study of the arXiv system in:
  'Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories'.

- The paper shows a significant citation advantage for papers first posted in arXiv, and subsequently published in journals

- The paper is, of course, available as arXiv:0906.5418.

Comments (3)

'arXiv is the primary daily information source for hundreds of thousands of researchers in physics and related fields. Its users include 53 physics Nobel laureates, 31 Fields medalists and 55 MacArthur fellows, as well as people in countries with limited access to scientific materials. The famously reclusive Russian mathematician Grigori Perelman posted the proof for the 100-year-old Poincaré Conjecture solely in arXiv.'

http://phys.org/news142785151.html

Microsoft Research Connections

The American Physical Society and arXiv

"Marty Blume, when he was editor-in-chief in the '90's, was incredibly supportive. They decided that the membership wanted it, so they'd figure a way to work with it, one way or another. It made me proud to be a physicist."

Paul Ginsparg

A Sustainability Model for arXiv?

- The operation of arXiv is currently funded by Cornell University Library.

- In 2010, Cornell broadened funding for arXiv by asking institutions to make an annual contribution based on the amount of downloading by each institution.

- Annual donations vary in size between $2,300 to $4,000, based on usage.

- As of February 2010, 27 institutions have pledged support on this basis.

- The annual budget for arXiv was $400,000 for 2010.

Stevan Harnad and Green Open Access

- His "Subversive Proposal" was an Internet posting by Stevan Harnad on June 27 1994

- His proposal called on all authors of "esoteric" writings—written only for research impact, not for royalty income—to archive them free online

- This proposal led to the 'Green Open Access' or self-archiving movement

Scholarly Journals
at the Crossroads:
A Subversive Proposal
for Electronic Publishing

Oct 1999 – EPrints software announced

- Stevan Harnad and Les Carr from Southampton attended Santa Fe meeting that led to the Open Archiving Initiative (OAI).

- The two-day meeting defined the OAI-PMH protocol for metadata

- Southampton offered to develop turnkey system to help groups start up their own archives

➢ This became Eprints software

EPrints Repository Software

Flexible Repository Software

EPrints supports research papers, theses, teaching materials, arts and more.

UCL Discovery

Welcome to UCL Discovery - home to UCL Research. Use this site to explore the unique scale and diversity of UCL research and our global expertise. UCL authors can access RPS from this site, and use au ...

More ...

Getting Started

Find out more about the EPrints software or download it now:

Latest Version: **v3.3.11**
.tar.gz | .deb | .rpm

Debian/Ubuntu
Install Guide

Redhat/Fedora
Install Guide

Windows XP/Vista/7
Install Guide

Older Versions
3.2.0 | 3.1.3 | 3.0.5

Try Live Demo

Using EPrints

EPrints Services
Support, Hosting, Training and Consultancy

Documentation
All latest documentation can be found on the EPrints wiki

Technical Discussion
Subscribe | View Archive

Announcements
Subscribe

Training Materials
All training materials from our workshops

Dec 2001 – Budapest Open Access Initiative

- **The Budapest Open Access Initiative arose from a small but lively meeting convened in Budapest by the Open Society Institute (OSI) on December 1-2, 2001.**

- **The purpose of the meeting was to accelerate progress in the international effort to make research articles in all academic fields freely available on the internet.**

David Lipman and PubMed Central

- PubMed Central (PMC) is a free archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM)

- Launched in February 2000, PMC was developed and is managed by NLM's National Center for Biotechnology Information (NCBI)

US NIH Open Access Policy

- *Once posted to PubMed Central, results of NIH-funded research become more prominent, integrated and accessible, making it easier for all scientists to pursue NIH's research priority areas competitively.*

- *PubMed Central materials are integrated with large NIH research data bases such as Genbank and PubChem, which helps accelerate scientific discovery.*

- *Clinicians, patients, educators, and students can better reap the benefits of papers arising from NIH funding by accessing them on PubMed Central at no charge.*

- *Finally, the Policy allows NIH to monitor, mine, and develop its portfolio of taxpayer funded research more effectively, and archive its results in perpetuity"*

NIH Open Access Compliance

- PMC Compliance Rate
  - Before legal mandate compliance was 19%
  - After legal mandate compliance up to 75%
  - Signed into law by George W. Bush in 2007

- NIH have taken a further step of announcing that, 'sometime in 2013' they

  '... will hold processing of non-competing continuation awards if publications arising from grant awards are not in compliance with the Public Access Policy.'
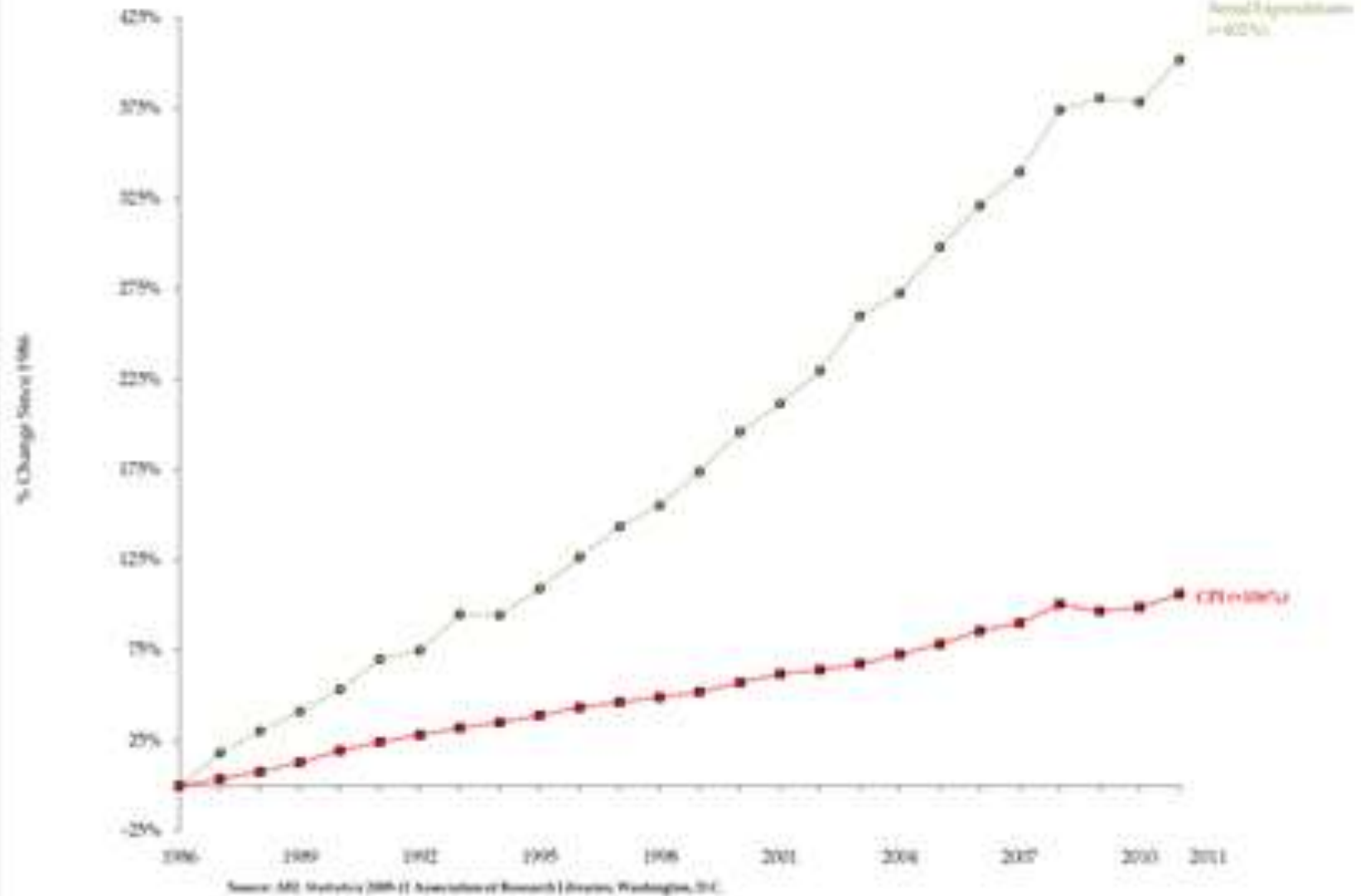
A Personal Journey towards OA

- As Dean of Engineering at Southampton in 1999 I was 'responsible' for monitoring the research output of over 200 Faculty and 500 Post Docs and Grad Student but …
  - The University library could not afford to subscribe to all the journals that my staff published in
  - nor afford to purchase conference proceedings and workshop contributions

  ➢ Insisted on keeping a digital copy of all research output in a Repository …

Library Budgets and Journal Prices

MIT Libraries Materials Purchases vs. CPI % Increase 1986-2006
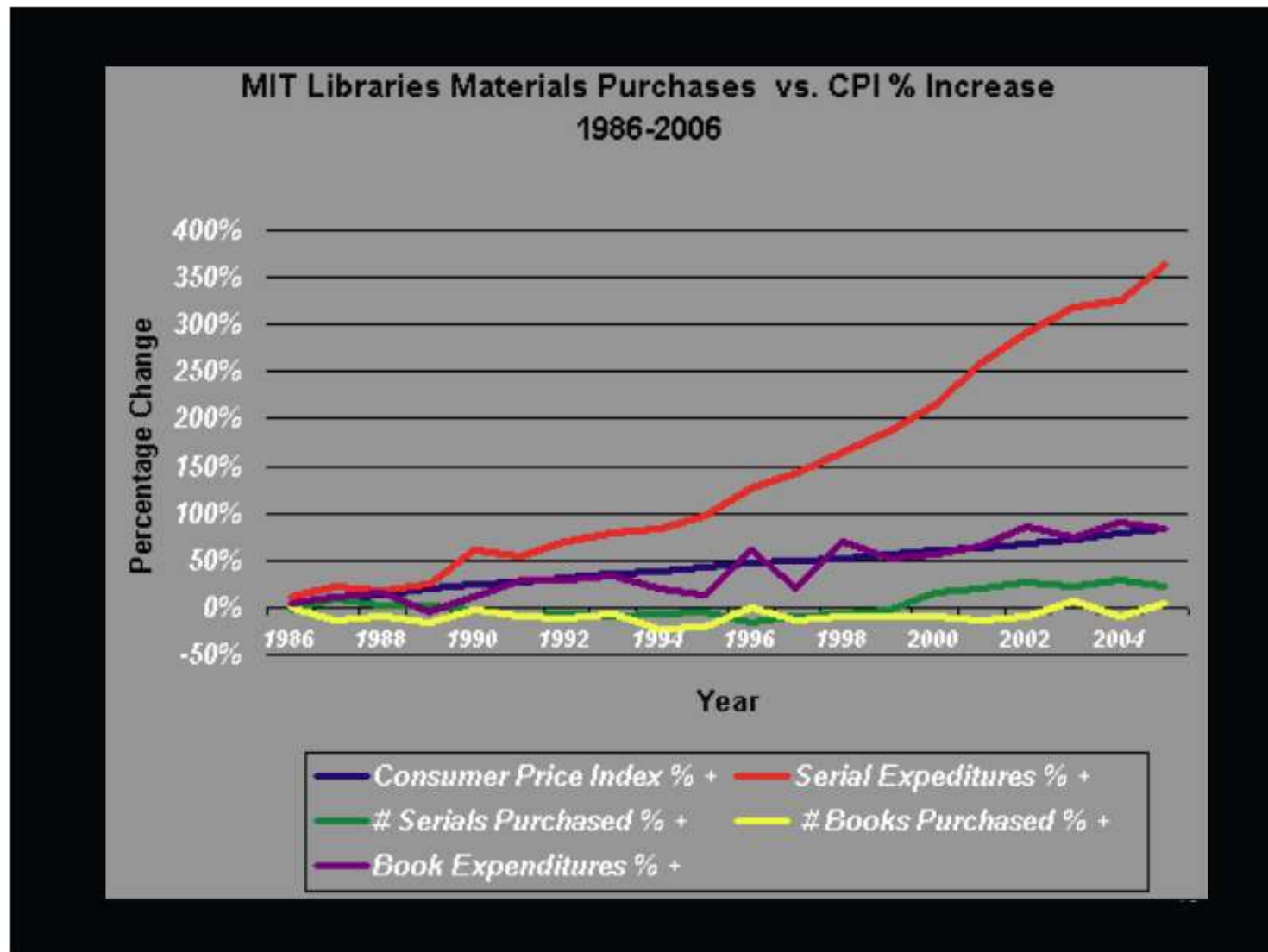
Serial Expenditures and CPI Trends in ARL Libraries, 1986-2011

Microsoft Research Connections

Aug 2002 – The Institutional Repository at the University of Southampton

- **Developed in the TARDis project, the University's Institutional Repository was a novel departure from the mainly discipline-specific repositories of the pre-OAI world.**

- **TARDis Project Team: Pauline Simpson and Jessie Hey**
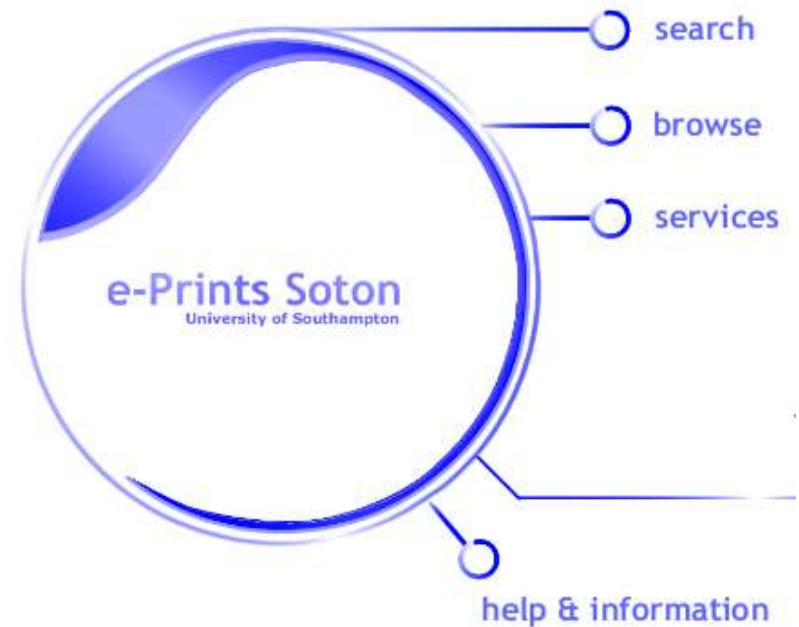
Jan 2003 – The First Open Access Mandate

The Electronics and Computer Science Department at the University of Southampton issued a mandate for OA:

- 'It is our policy to maximise the visibility, usage and impact of our research output by maximising online access to it for all would-be users and researchers worldwide'

- 'It is also our policy to minimise the effort that each of us has to expend in order to provide open online access to our research output'

Jan 2004 – Registry of Open Access Repositories (ROAR)

- Created initially to track the growth of EPrints repositories
- Expanded to deal with repositories based on any software
- Currently over 300 open access mandates internationally
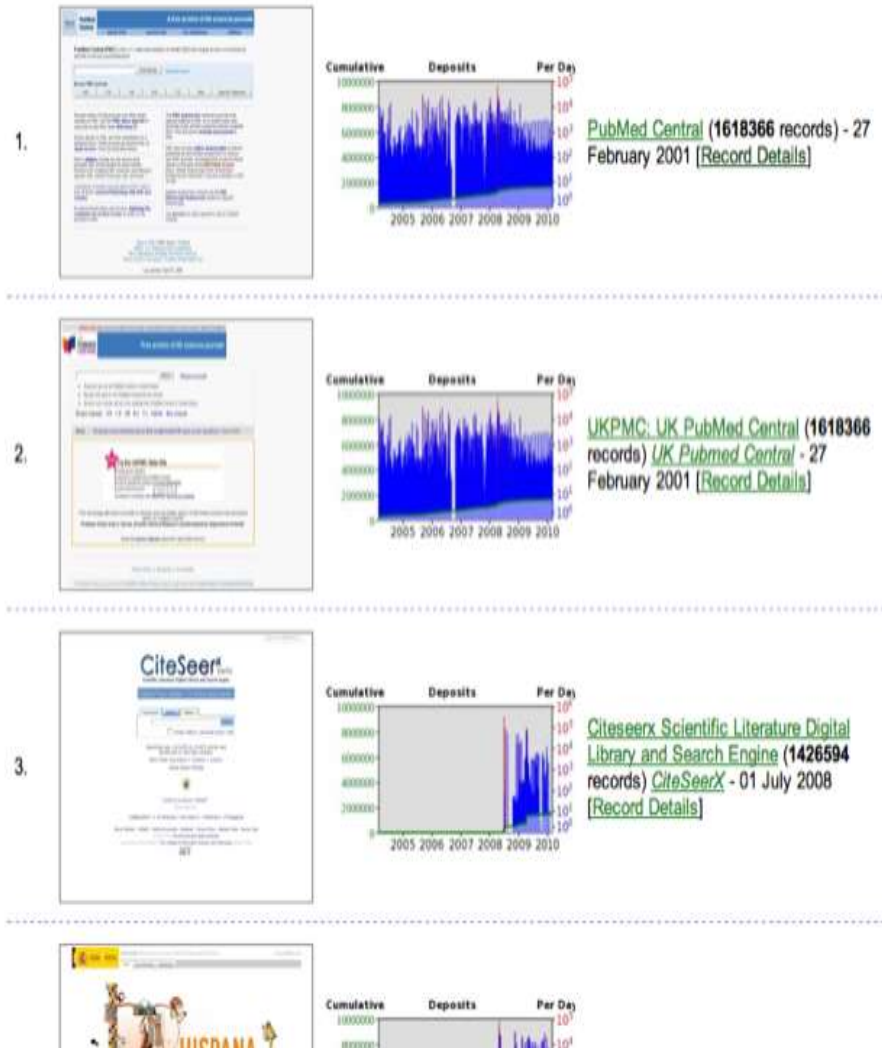
Institutional Repositories Worldwide

http://roar.eprints.org/

Microsoft Research Connections

Green Open Access

'Green Open Access' or 'Self-Archiving' requires authors to make peer-reviewed final drafts of their articles accessible by depositing them in their Institution's OA Repository

- either on submission

- or on acceptance for publication

Notes:

- Individual papers need not be set to be immediately visible outside the institution - can be set to 'delayed open access' as in NLM's PubMedCentral.

- Web copies of non-journal versions are allowed by over 80% of publishers …
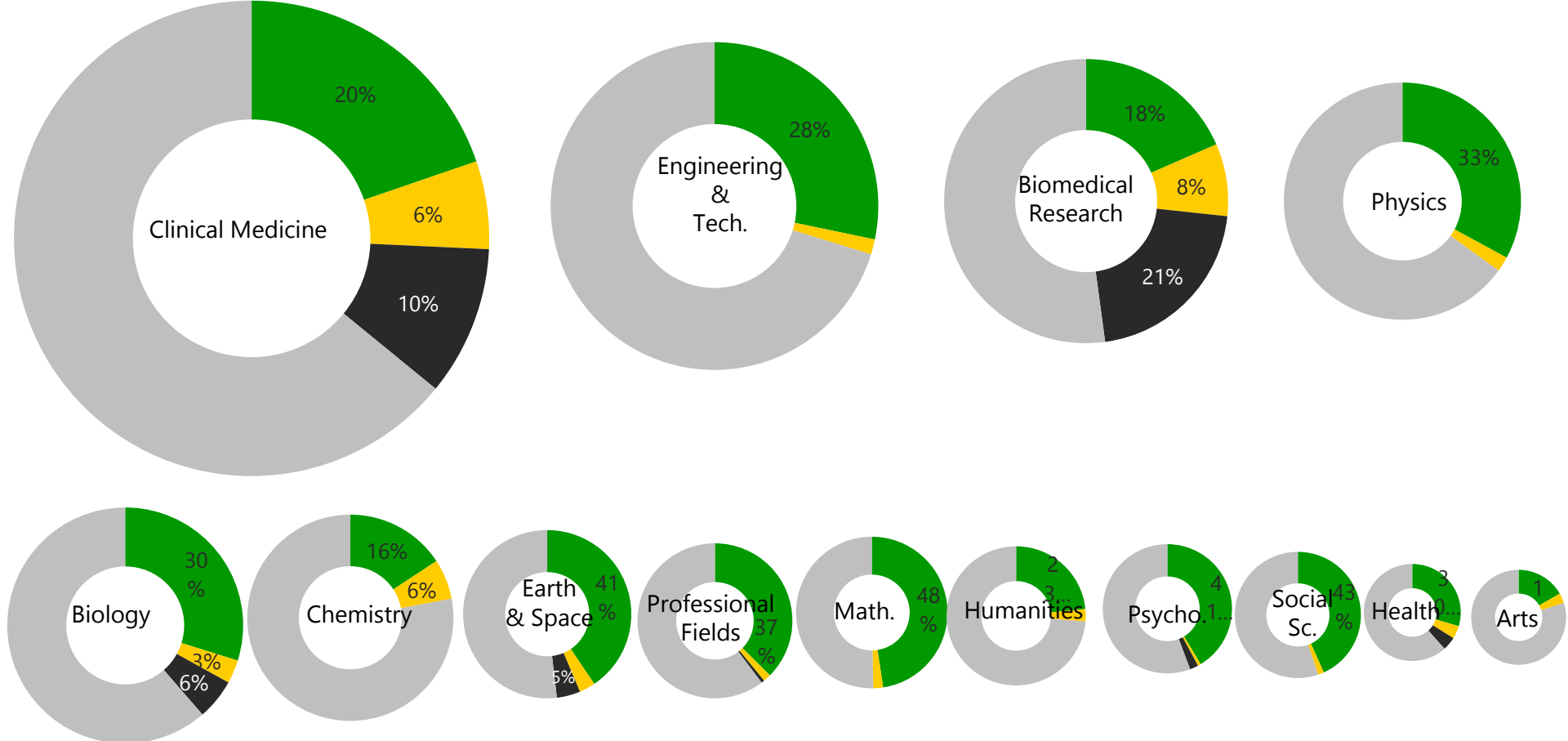
Worldwide Percentage Open Access (OA) by Field (2005-10)

Green OA :   26.3%
Gold OA :   4.3%
Delayed   Access : 6 %
Total: 36.6%

Microsoft Research Connections

Green OA Mandates can triple Green OA
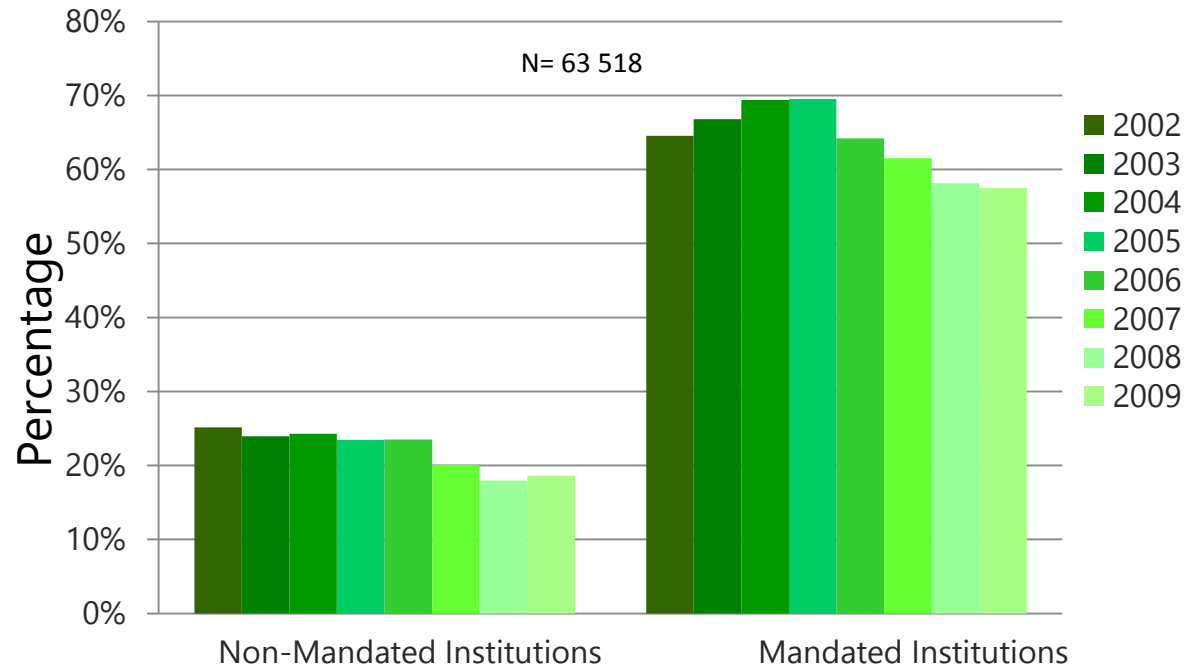
University Reputation Management

- Organizations like Webometrics will be using the university's visible web presence to produce global and national rankings of universities
  - ➢ Many different ranking methodologies, need to ensure university displays its best data.

- With the advent of open access to both full text of papers and data, university research repositories will be an important part of the university's reputation management strategy
  - ➢ COAR – Coalition of Open Access Repositories

The Big Data Bandwagon

Microsoft®
Research Connections

A Tidal Wave of Scientific Data
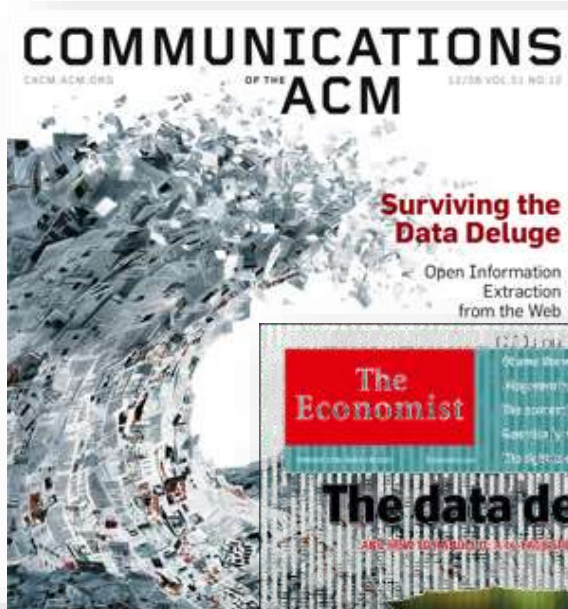
"data scientist" 🔍

254,000 RESULTS

The **Data Scientist** role is a role of the future!
www.**datascientists**.net ▾
The **Data Scientist** role is a role of the future! Future proof your career and start transitioning today.

Data Scientist: The Hottest Job You Haven't Heard Of - Careers ...
jobs.aol.com/articles/2011/08/10/**data**-**scientist**-the-hottest-job... ▾
Aug 10, 2011 · Data **scientists** are an integral part of competitive intelligence, a newly emerging field that encompasses a number of activities

LinkedIn's Monica Rogati On "What Is A **Data Scientist**?" - Forbes
www.forbes.com/.../linkedins-monica-rogati-on-what-is-a-**data**-**scientist** ▾
Nov 27, 2011 · To continue our series on the emerging role of the **data scientist** in today's data-driven organizations, we spoke with Monica Rogati, Senior Data ...

Related searches for **"data scientist"**

Data Scientist **Seattle**	Data Scientist **Fortune**
Data Scientist **Salary**	Data Scientist **Jobs**
Data Scientist **Interview Ques**...	**Introduction to** Data **Science**

Data scientist: The hot new gig in tech - Fortune Tech
tech.fortune.cnn.com/2011/09/06/**data**-**scientist**-the-hot-new-gig-in-tech ▾
Sep 06, 2011 · Companies that want to make sense of all their bits and bytes are hiring so-called data **scientists** - if they can find any. FORTUNE -- The unemployment rate ...

The **Data Scientist** | Mine, Visualize, and Learn
www.the**datascientist**.com ▾
As I jumped from room to room on Turntable.fm last night my eyes caught a glimpse of a rare room titled "AOKIxSOLREPUBLIC" . I clicked it with a fury.

What is a Data Scientist?

Data Engineer	**People who are expert at** • Operating at low levels close to the data, write code that manipulates • They may have some machine learning background. • Large companies may have teams of them in-house or they may look to third party specialists to do the work.
Data Analyst	**People who explore data through statistical and analytical methods** • They may know programming; May be an spreadsheet wizard. • Either way, they can build models based on low-level data. • They eat and drink numbers; They know which questions to ask of the data. Every company will have lots of these.
Data Steward	**People who think to managing, curating, and preserving data.** • They are information specialists, archivists, librarians and compliance officers. • This is an important role: if data has value, you want someone to manage it, make it discoverable, look after it and make sure it remains usable.

What is a data scientist? Microsoft UK Enterprise Insights Blog, Kenji Takeda
http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/01/31/what-is-a-data-scientist.aspx

Microsoft Research Connections

School of Information Studies
SYRACUSE UNIVERSITY

SYR.EDU NEWS EVENTS ABOUT CONTACT **LOGIN**

HOME **FUTURE STUDENTS** **CURRENT STUDENTS** **ALUMNI & FRIENDS** **FACULTY & STAFF** **RESEARCH** **ISCHOOL LIFE**

Newsroom

CURRENT NEWS   **ARCHIVED NEWS**

iSchool Introduces New Graduate Certificate in Data Science

3/21/2012

By: Diane Stirling
(315) 443-8975

The School of Information Studies (iSchool), an international leader in the information field, is boosting its cutting-edge academic offerings once again with the introduction of a new graduate Certificate of Advanced Studies in Data Science.

The program is the first state-approved certificate of advanced study in this topic area in New York. Content focuses on equipping students and professionals for the next looming issue in the information field – having the skills sets and the knowledge to analyze, problem-solve and take advantage of the proliferation of increasing magnitudes of information – dealing with "big data" for all types of organizations.

Microsoft Research Connections

Data Size and Speed are Growing

Entire sequence of DNA for the human body, consists of around 3 billion of these base pairs.

The human genome requires ~750 megabytes of storage

Large Hadron Collider

150 million sensors delivering data 40 million times per second.

Data flow: ~700 MB/sec
            ~15 PB/year

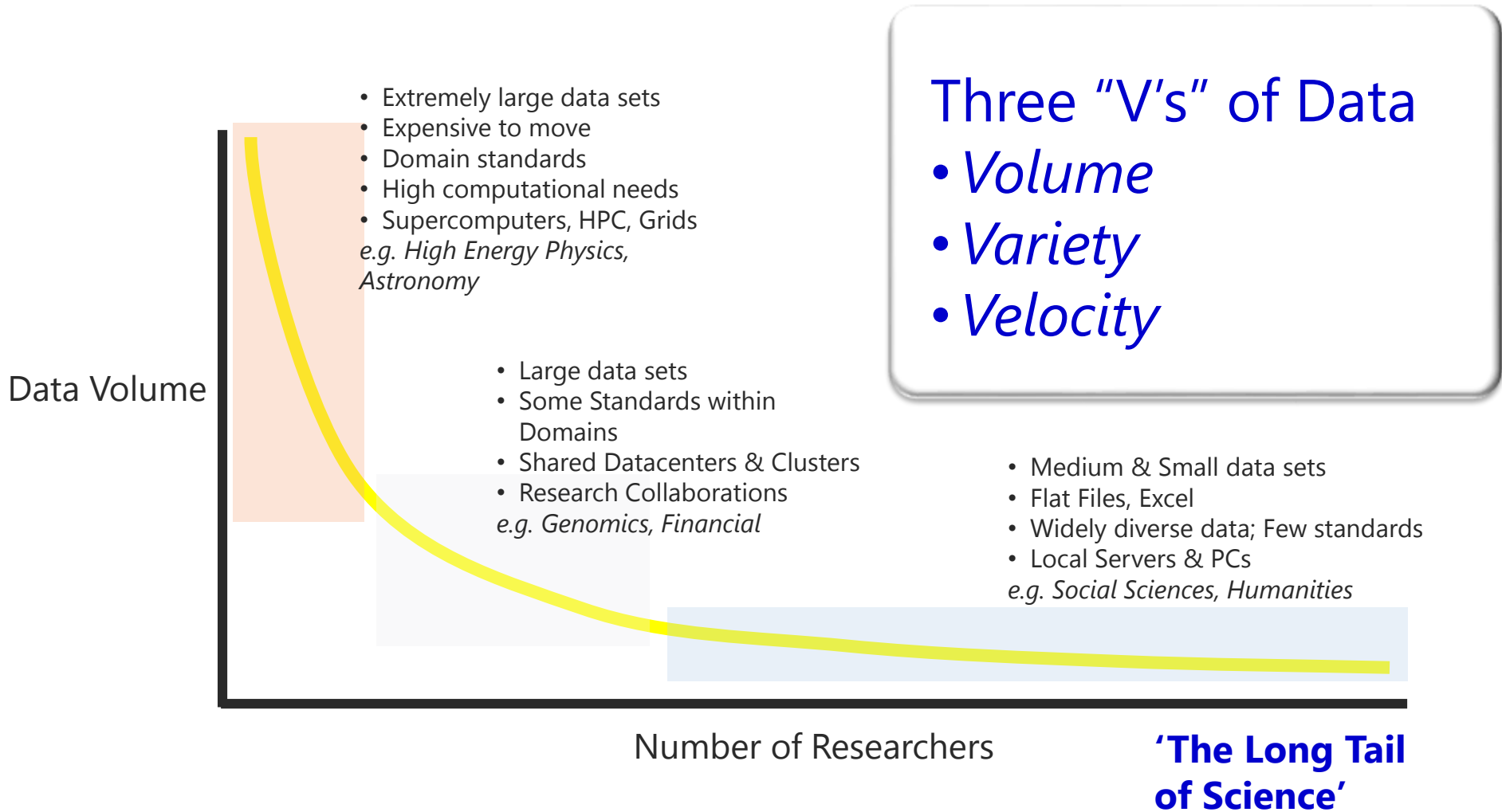1000's of scientists around the world; Institutions in 34 different countries:

Thousands of small antennas spread over a distance of more than 3000km.

Data flow: ~60 GB/sec
            1 Million PB/day

The SKA supercomputer will perform 1018 operations per second ~ 100M PCs

1990   2000   2010   2020

Science is now Data-Intensive

Data Volume

- Extremely large data sets
- Expensive to move
- Domain standards
- High computational needs
- Supercomputers, HPC, Grids

e.g. High Energy Physics, Astronomy

- Large data sets
- Some Standards within Domains
- Shared Datacenters & Clusters
- Research Collaborations

e.g. Genomics, Financial

- Medium & Small data sets
- Flat Files, Excel
- Widely diverse data; Few standards
- Local Servers & PCs

e.g. Social Sciences, Humanities

Three "V's" of Data
- *Volume*
- *Variety*
- *Velocity*

Number of Researchers
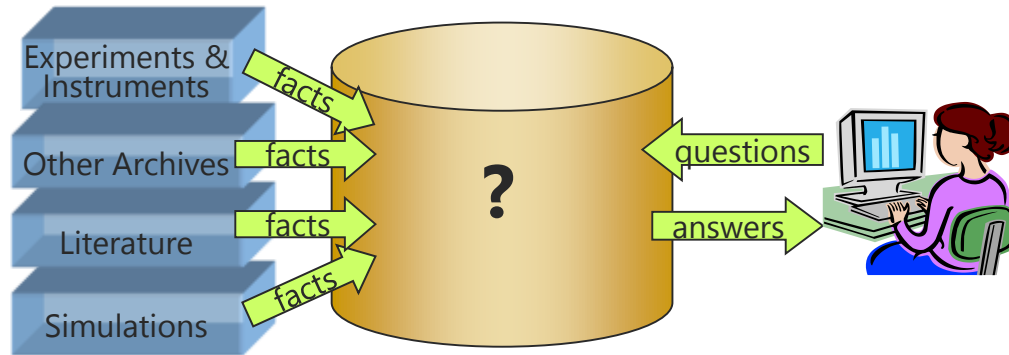
'The Long Tail of Science'

Jim Gray, Turing Award Winner

X-Info

- The evolution of X-Info and Comp-X for each discipline X

- How to codify and represent our knowledge

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)

eScience and the Fourth Paradigm

Thousand years ago – **Experimental Science**

- Description of natural phenomena

Last few hundred years – **Theoretical Science**

- Newton's Laws, Maxwell's Equations...
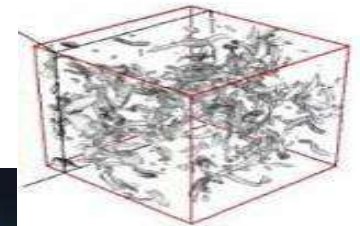
Last few decades – **Computational Science**

- Simulation of complex phenomena

Today – **Data-Intensive Science**

- Scientists overwhelmed with data sets from many different sources
  - Data captured by instruments
  - Data generated by simulations
  - Data generated by sensor networks

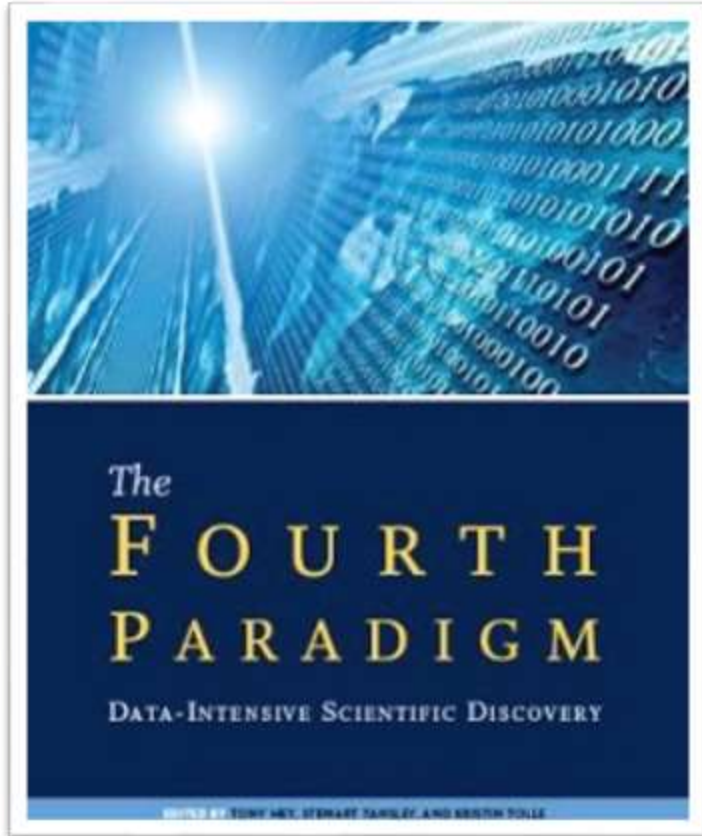$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$

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)

eScience and Data-Intensive Scientific Discovery

**Available online from The Fourth Paradigm and
Science@Microsoft at http://research.microsoft.com
and on Amazon.com**

Astronomy leads the way

The 'Cosmic Genome' Project

- The Sloan Digital Sky Survey was the first major astronomical survey project:

  - 5 color images and spectra of ¼ of the sky
  - Pictures of over 300 million celestial objects
  - Distances to the closest 1 million galaxies

- Jim Gray from Microsoft Research worked with astronomer Alex Szalay to build the public 'SkyServer' archive for the survey

- New model of scientific publishing - publish the data <u>before</u> astronomers publish their analysis

World Wide Telescope
www.worldwidetelescope.org

> Seamless Rich Social Media Virtual Sky Web application for science and education

Participants
- 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 (someone that has downloaded, installed, and successfully used WWT)
- The average number of WWT users over 8K per day

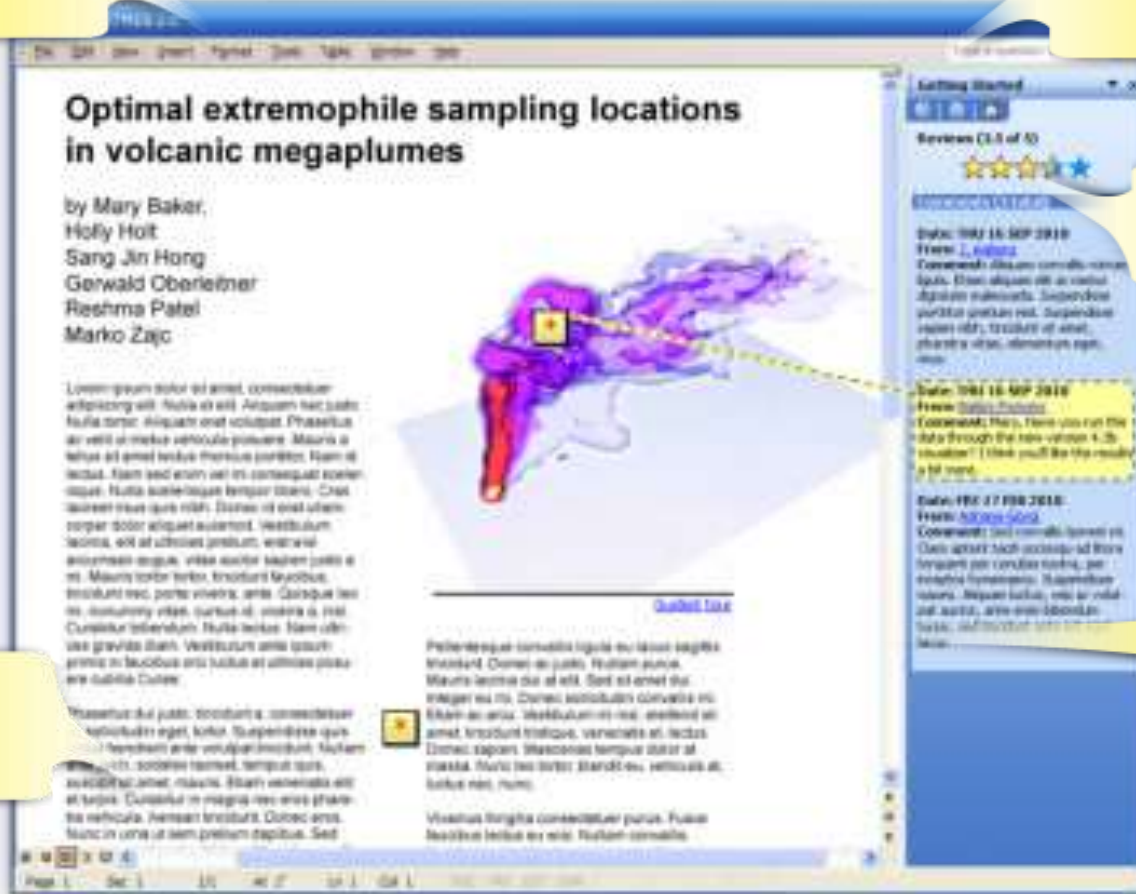# Vision for a New Era of Research Reporting

(Thanks to Bill Gates SC05)

Astrophysics Data System ADS

· **Find Similar Abstracts** (with **default settings below**)   Toggle Highlighting

· **Custom Format**
· **Electronic Refereed Journal Article (HTML)**
· **Full Refereed Journal Article (PDF/Postscript)** ←——————— Links to e-resources
· FIND IT ⑤ HARVARD

· **arXiv e-print** (arXiv:astro-ph/0412451)
· **On-line Data** ←——————— Links to data
· **References in the article**
· **Citations to the Article (84)** (Citation History)
· **Refereed Citations to the Article**
· **SIMBAD Objects (3)** ←——————— Links to objects
· **NED Objects (1)**
· **Also-Read Articles** (Reads History)
·
· **Translate This Page**

Title:	Bow Shock and Radio Halo in the Merging Cluster A520
Authors:	Markevitch, M.; Govoni, F.; Brunetti, G.; Jerius, D.
Affiliation:	AA(Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138; Space Research Institute, Russian Academy of Sciences, 84/32 Profsoyuznaya Street, Moscow 117997, Russia. maxim@head.cfa.harvard.edu), AB(Istituto di Radioastronomia del CNR, via Gobetti 101, 40129 Bologna, Italy.), AC(Istituto di Radioastronomia del CNR, via Gobetti 101, 40129 Bologna, Italy.), AD(Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138 maxim@head.cfa.harvard.edu)
Publication:	The Astrophysical Journal, Volume 627, Issue 2, pp. 733-738. (ApJ Homepage)
Publication Date:	07/2005
Origin:	UCP
Astronomy Keywords:	Galaxies: Clusters: Individual: Alphanumeric: A520, Galaxies: Intergalactic Medium, Radio Continuum: General, X-Rays: Galaxies: Clusters
DOI:	10.1086/430695
Bibliographic Code:	2005ApJ...627..733M

CDS

"*Seamless Astronomy*"
(Tools)

Literature

Data

arXiv.org

nature
A BIG YEAR FOR ASTRONOMY

ads labs
NASA

VizieR

ALADin

SIMBAD

World Wide Telescope

Astro Grid
VO
IVOA
EURO VO
AIDA Astronomical Infrastructure for Data Access

"Registries"

Astrometry.net
flickr

TOPCAT

WIKIPEDIA
The Free Encyclopedia

ds9

AstroBetter

DataScope

Blogs, Wikis, etc.

Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.

Dataverse: Re-inforcing the Link between Research Publications and Research Data

The Dataverse project at Harvard has been awarded an Alfred Sloan Foundation grant for the next 2 years to enhance the link between journals and data

Open Journal System

Seamless integration between the two systems:

Deposit Data from Journal to Dataverse through a standard API (based on SWORD)
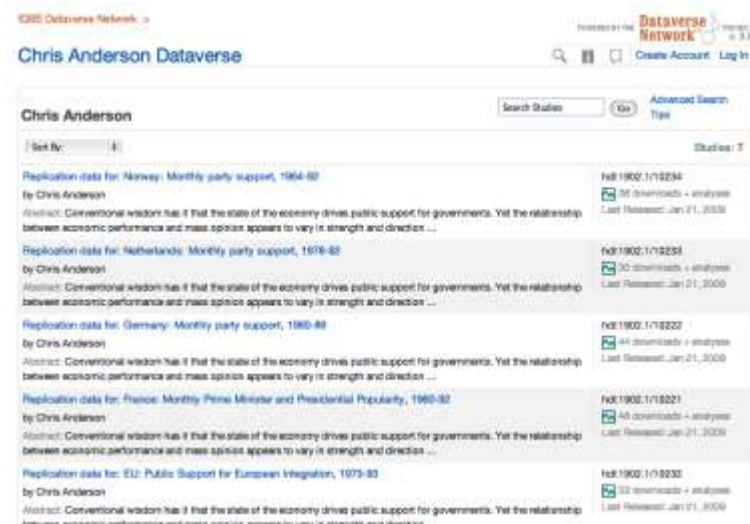
The Dataverse Network
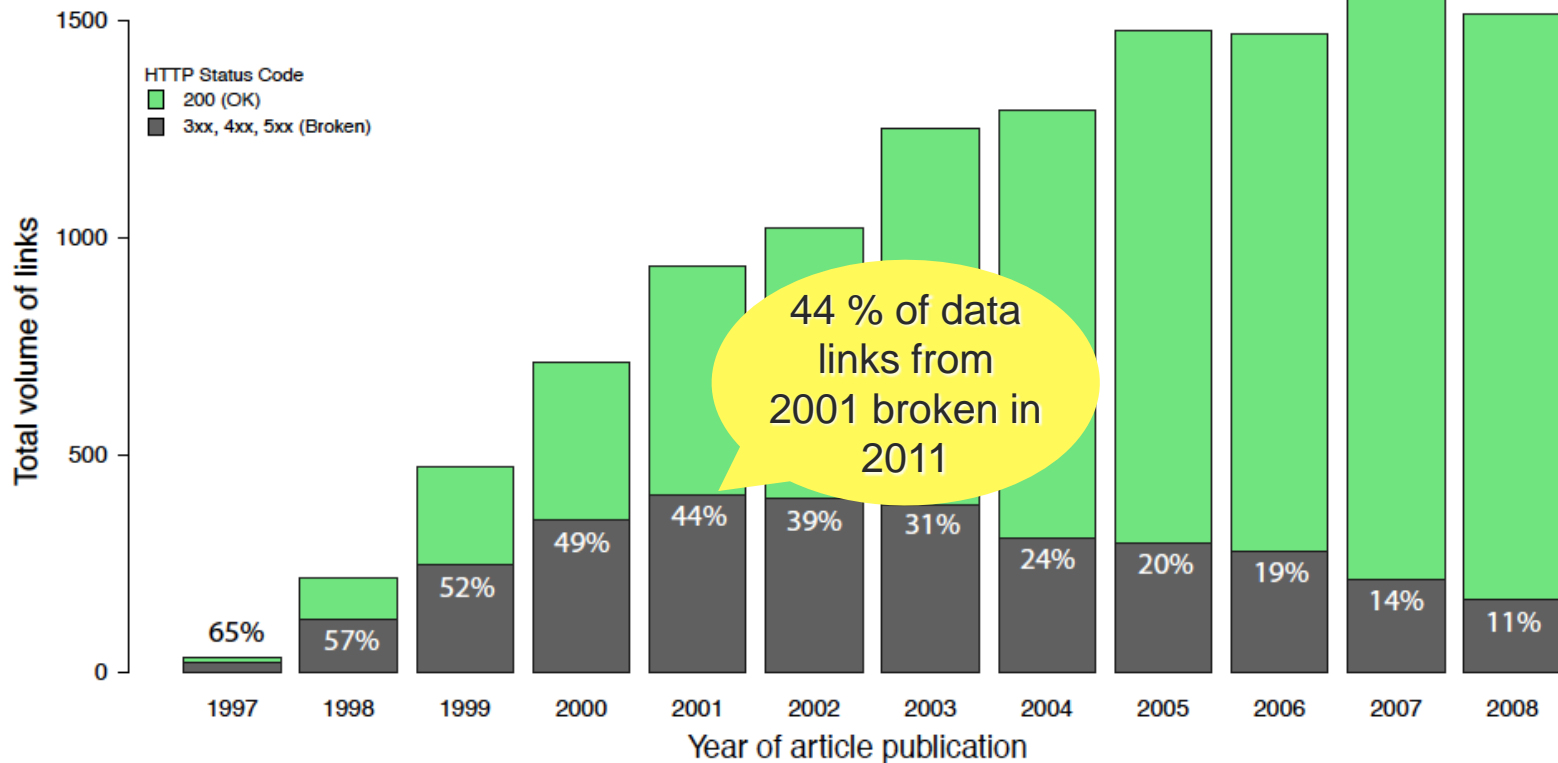
Figure 1. Volume of potential data links in astronomy publications. Total volume of external links in all articles published between 1997 and 2008 in the four main astronomy journals, color coded by HTTP status code. Green bars represent accessible links (200), grey bars represent broken links. .

Pepe et al. 2012

Open Access to Data

Collaboration and Sharing of Data is Expected and Growing

… expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work.

NIH reaffirms its support for the concept of data sharing. We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health … The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers.

A primary goal of Data.gov is to improve access to Federal data and expand creative use of those data beyond the walls of government by encouraging innovative ideas (e.g., web applications). Data.gov strives to make government more transparent and is committed to creating an unprecedented level of openness in Government.

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.

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"

CERN's ZENODO

Zenodo: A new grey literature and data publication solution from CERN

Written by Chris Erdmann
September 12th, 2013

Galactic Gazette
Subscribe RSS

Close

I've been with the Harvard-Smithsonian Center for Astrophysics as the Head Librarian for 3+ years, but a patron request that the John G. Wolbach Library received during my first few weeks still clings to the back of mind. It involved a graduate student simply wishing to submit her dissertation to the Library in electronic format. At the time, we had no solution to manage and disseminate her dissertation as part of an online collection, so we ultimately took the PDF and placed it on our shared network drive. It drove me crazy that we didn't have a solution for both preservation and dissemination of the dissertation. I've continuously revisited the problem, always scanning for potential solutions, only to find that they fell short in some way. That all changed though when I recently started working with the talented development team at CERN behind Zenodo.org, led by Tim Smith and Lars Holm Nielsen.

Image Credit:
Image Above, *The Large Hadron Collider/ATLAS at CERN*, CC Image Courtesy of Image Editor on Flickr

Posted in Wolbach Library

No Comments

SCIENTIFIC DATA

Helping you publish, discover, and reuse research data

Credit
Credit, through a citable publication, for depositing & sharing your data

Reuse
Complete, curated & standardized descriptions enable the reuse of your data

Quality
Rigorous community based peer review

Discovery
Find datasets relevant to your research

Open
Promotes & endorses open science principles & available to all through a Creative Commons license

Service
In-house curation, rapid peer review & publication of your data descriptions

Calling for submissions in Fall 2013, launching in Spring 2014
nature.com/scientificdata

nature publishing group npg

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

Literature

Derived and recombined data

Raw Data
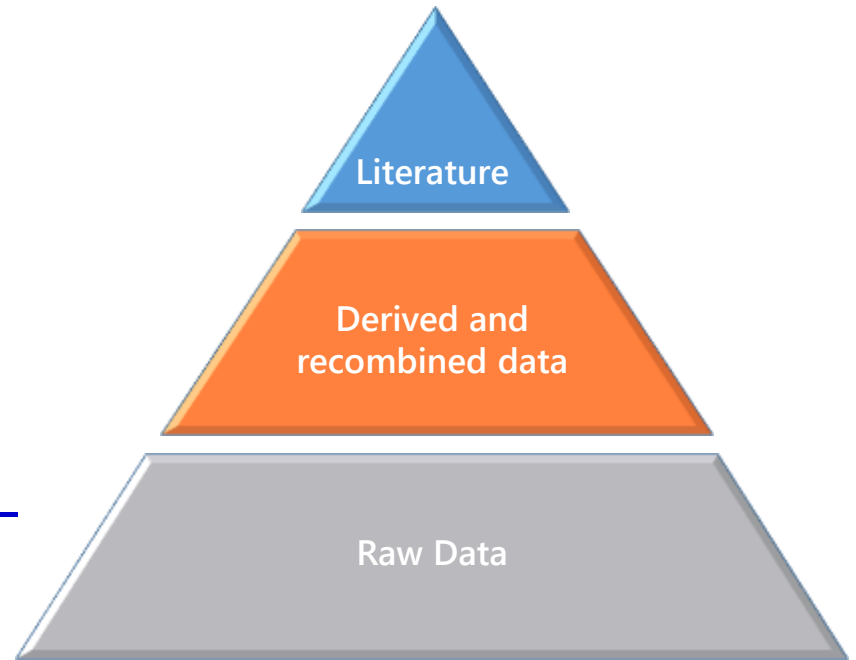
(From Jim Gray's last talk)

The US National Library of Medicine

- The [NIH Public Access Policy](#) ensures that the public has access to the published results of NIH funded research.

- Requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive [PubMed Central](#) *upon acceptance for publication.*

- Policy requires that these papers are accessible to the public on PubMed Central no later than 12 months after publication.
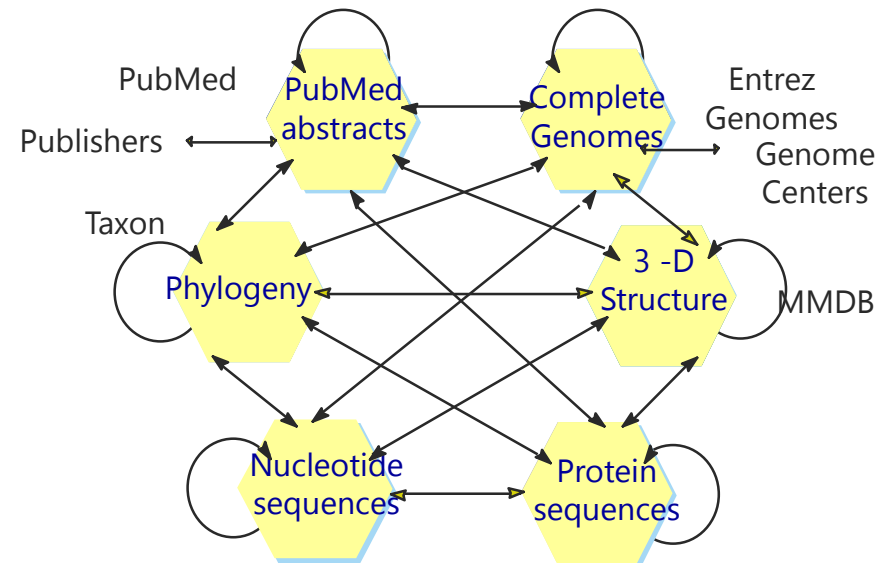
Entrez cross-database search

Open Access: The Tipping Point

Microsoft®
Research Connections

US White House Memorandum

- Directive requiring the major Federal Funding agencies *"to develop a plan to support increased public access to the results of research funded by the Federal Government."*

- The memorandum defines digital data *"as the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens."*

22 February 2013

Global Research Council
http://www.globalresearchcouncil.org/

- Newly founded network of national research funders from all over the world has endorsed an Action Plan towards Open Access.

- Action Plan consists of the 14 points including

  - Collect and document best practices for rewarding the provision of open access

  - Work with scholarly societies to transition society journals into open access

  - Work with repository organisations to develop efficient mechanisms for harvesting and ac-cessing information

30 May 2013

G8 Science Ministers

- "In a joint statement proposing "new areas" of scientific collaboration for the countries, the ministers say they "recognise the potential benefits of immediate global access to and unrestricted use of published peer-reviewed, publicly funded research results."

- "We share the intention, therefore, to continue our cooperative efforts and will consider how best to address the global promotion of increasing public access to the results of publicly funded published research including to peer-reviewed published research and research data."

12 June 2013

University of California approves Open Access

- UC is the largest public research university in the world and its faculty members receive roughly 8% of all research funding in the U.S.

- UC produces 40,000 publications per annum corresponding to about 2 – 3 % of all peer-reviewed articles in world each year

- The faculty remains committed to working with publishers to transform the publishing landscape in ways that are sustainable and beneficial to both the University and the public.

2 August 2013

Open Access in Australia

- The Australian Research Council (ARC) announced their open access policy http://www.arc.gov.au/applicants/open_access.htm

- The Australian Open Access Support Group (AOASG) began operations http://aoasg.org.au

- Chief Investigators for funded projects should have a link to an open access version within 12 months of publication.

1 January 2013

Commentary

This is one of a series of commentaries on the future of scientific publishing. For a listing of the other commentaries, see http://www.jneurosci.org/cgi/content/full/26/36/9077.

As We May Read

Paul Ginsparg

Departments of Physics and Information Science, Cornell University, Ithaca, New York 14853

The e-print arXiv (http://arXiv.org/), initiated in August 1991, has effectively transformed the research communication infrastructure of multiple fields of physics and could play a prominent role in a unified set of global resources for physics, mathematics, and computer science. It has grown to contain >375,000 articles (as of July 2006), with >50,000 new submissions expected in calendar year 2006 and >40,000,000 full-text downloads per year. It is an international project, with dedicated mirror sites in 17 countries and

ders of magnitude. Even with the majority of science research journals now on-line, researchers continue to enjoy both the benefits of the rapid availability of the materials, even if not yet reviewed, and open archival access to the same materials, even if held in parallel by conventional publishers. The methodology works within copyright law, as long as the depositor has the authority to deposit the materials and assign a nonexclusive license to distribute at the time of deposition, because such a license takes precedence over any subse-

helps ensure that the arXiv remains a forum for communication among research professionals, not a mechanism for outsiders to communicate to that community. Additionally, a small group of volunteer "moderators," consisting of interested experts from around the world, cursorily prescans new submissions, typically only at the level of title and abstract, for appropriateness to the proposed primary subject area.

The arXiv repository functions are flexible enough either to coexist with the

Published in the Journal of Neuroscience, September 20, 2006

Ginsparg's Conclusions?

"On the one-decade time scale, it is likely that more research communities will join some form of global unified archive system without the current partitioning and access restrictions familiar from the paper medium, for the simple reason that it is the best way to communicate knowledge and hence to create new knowledge."

"Ironically, it is also possible that the technology of the 21st century will allow the traditional players from a century ago, namely the professional societies and institutional libraries, to return to their dominant role in support of the research Enterprise."
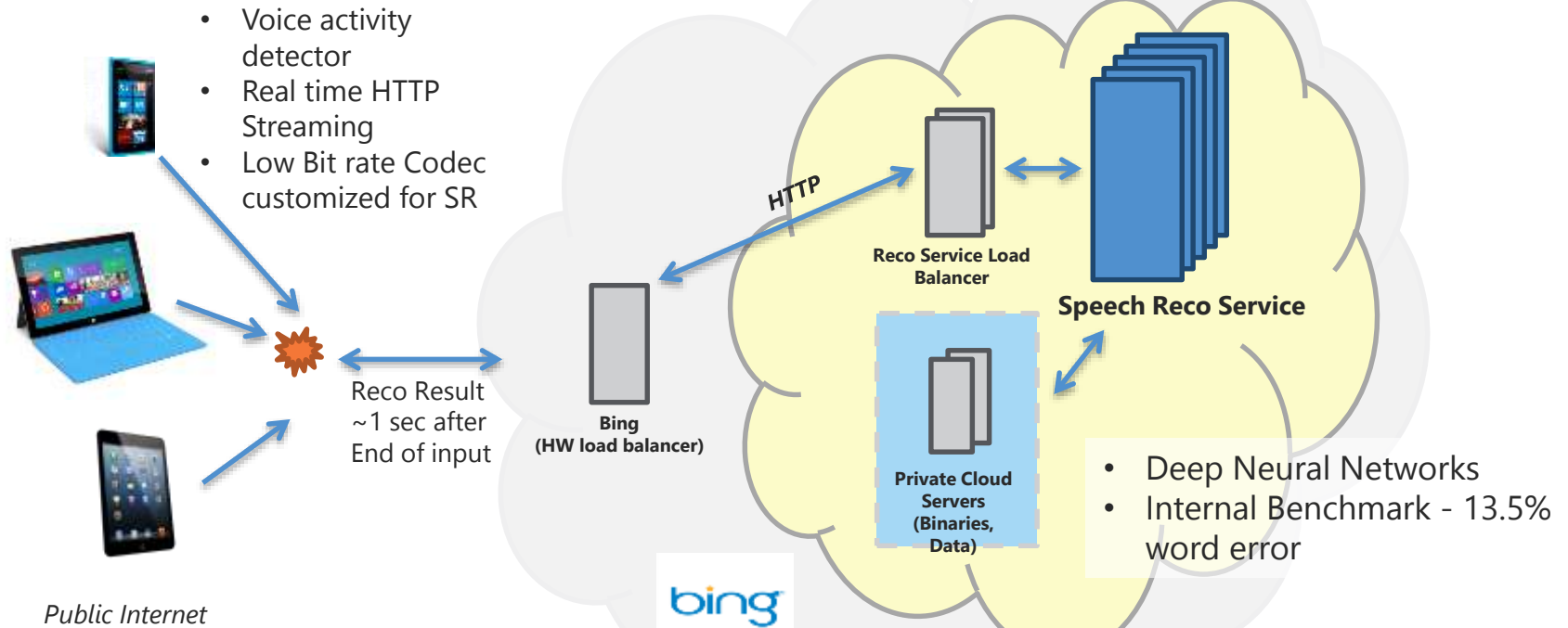
☐ **Big Data**

☐ **Machine Learning**

☐ **Visualization**

☐ **Cloud Services**

Microsoft Research Connections

Bing Speech Recognition Service Architecture

- Voice activity detector
- Real time HTTP Streaming
- Low Bit rate Codec customized for SR

Reco Result ~1 sec after End of input

Public Internet

HTTP

Reco Service Load Balancer

Bing (HW load balancer)

Speech Reco Service

Private Cloud Servers (Binaries, Data)

bing

- Deep Neural Networks
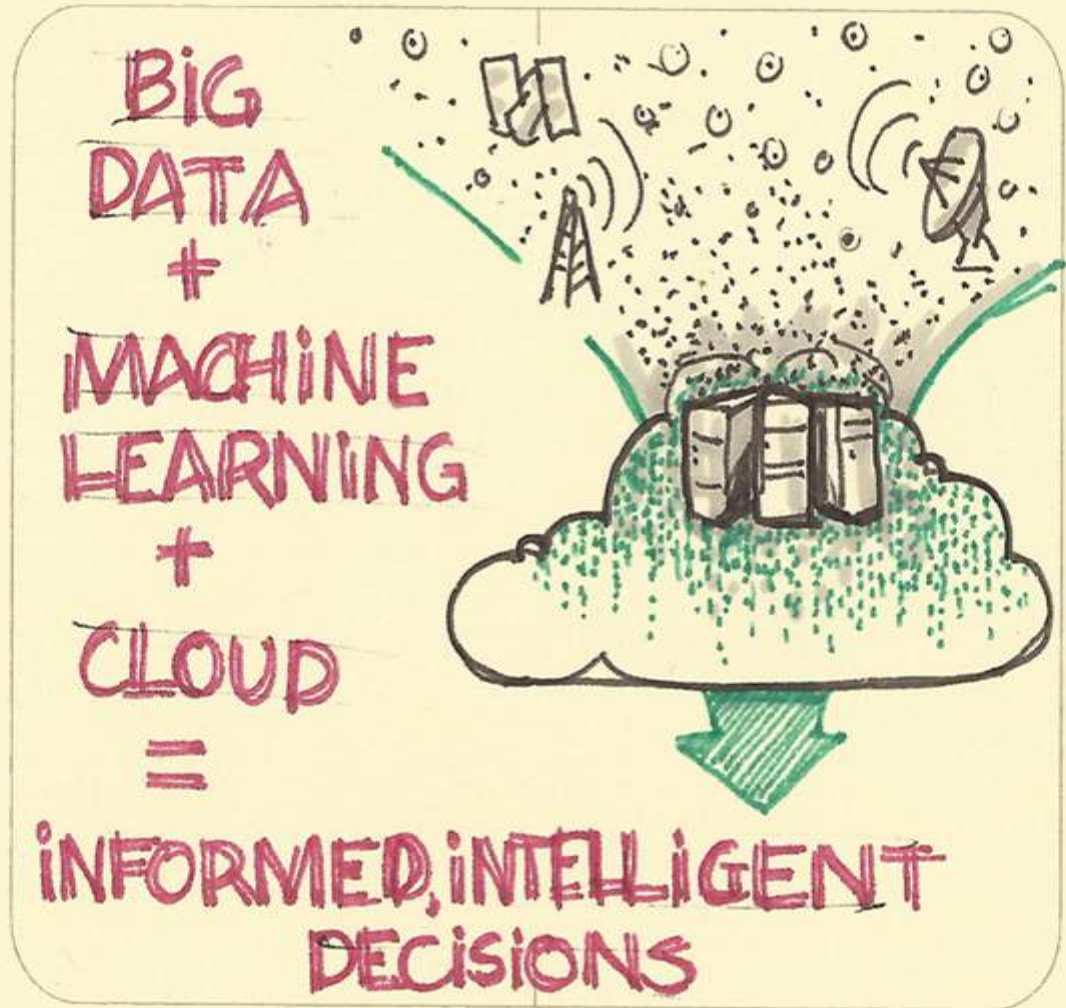- Internal Benchmark - 13.5% word error

Greek Virtual Fire

Virtual Fire (VF) is an early warning and decision support system for integrated forest fire management, based on geoinformatics and modeling fire risk

University of the Aegean, University of Athens, MS Hellas/MIC in Greece, and Microsoft Research

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
- Azure Cloud for Research
  - [http://research.microsoft.com/en-us/projects/azure/default.aspxOutercurve Foundation](http://research.microsoft.com/en-us/projects/azure/default.aspx)
- Tony Hey on eScience
  - http://tonyhey.net/

Windows Azure for Research
Accelerate the Speed of Scientific Discovery

Windows Azure provides researchers with the power and scalability of cloud computing for collaboration, computation, and data-intensive processing. This open and flexible global cloud platform supports any language, tool, or framework.

Windows Azure

The Windows Azure for Research program:

- Free access to Windows Azure cloud computing and storage (submit proposals for Windows Azure Research Awards)
- Windows Azure for Research training classes
- Support and technical resources

Apply the power of cloud computing to your computational and data challenges. Experiment at **azure4research.com.**

Microsoft Research Connections