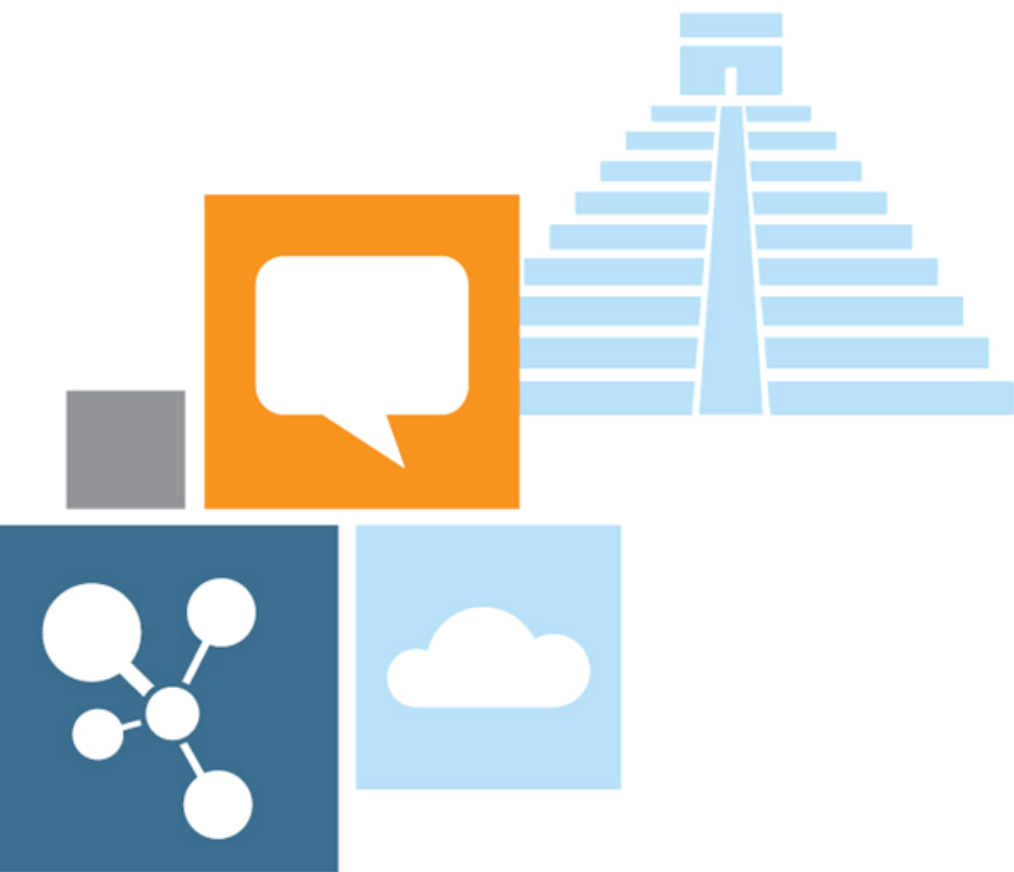


Microsoft



Microsoft® Research

Faculty Summit 2012

Riviera Maya, Mexico | May 23-25 | In partnership with CONACYT



eResearch

Surveying the State of the Art

Lee Dirks
Director, Portfolio Strategy
Microsoft Research Connections

May 24th, 2012 | Riviera Maya, Mexico



Microsoft Research | Connections

Outreach. Collaboration. Innovation.

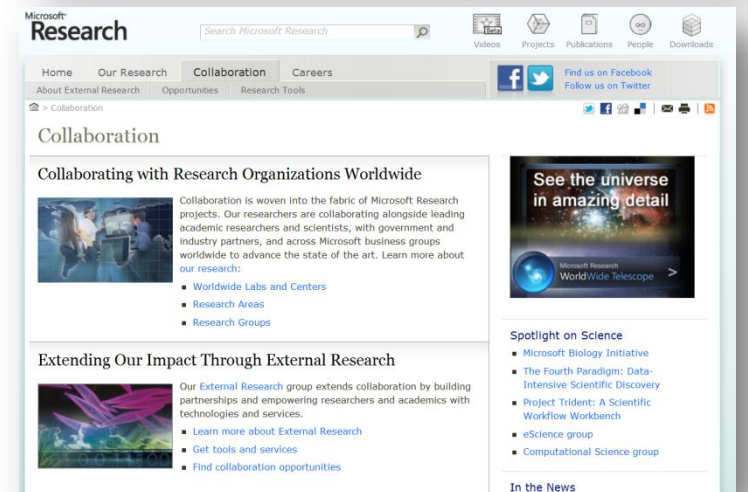
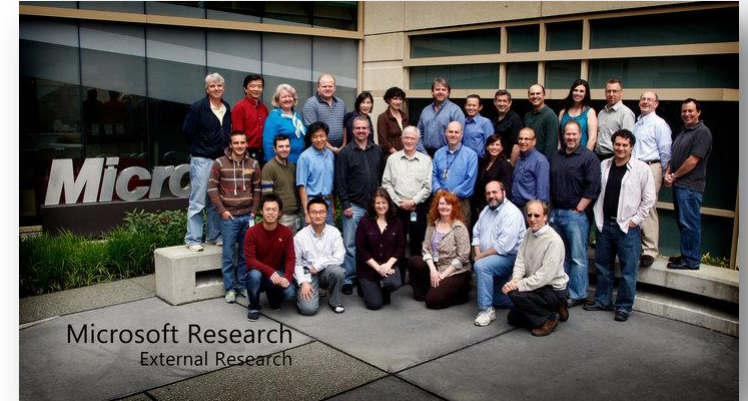
Division within Microsoft Research focused on **partnerships** between academia, industry and government to advance computer science, education, and research in fields that rely heavily upon advanced computing.

Supporting **groundbreaking research** to help advance human potential and the wellbeing of our planet.

Developing **advanced technologies and services** to support every stage of the research process.

Microsoft Research Connections is **committed to interoperability** and to providing open access, open tools, and open technology.

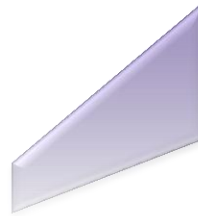
<http://research.microsoft.com/collaboration/>



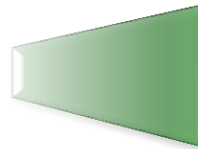


MSR Connections – Our Mission

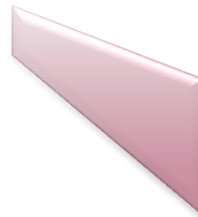
Work with the
worldwide academic
research community to
speed research,
improve education,
and **foster innovation.**



**Advance the
State of the Art**



**Inspire Computer
Scientists**



**Accelerate
Discovery and
Exploration**





What is eResearch?

“The term...refers to the use of information technology to support existing and new forms of research. E-research extends e-Science and cyberinfrastructure to other disciplines, including the humanities and social sciences.”



The main features of eResearch are that it:

- is **collaborative**
- uses **grid or cloud technologies**
- is **data-intensive**

<http://en.wikipedia.org/wiki/Ersearch>



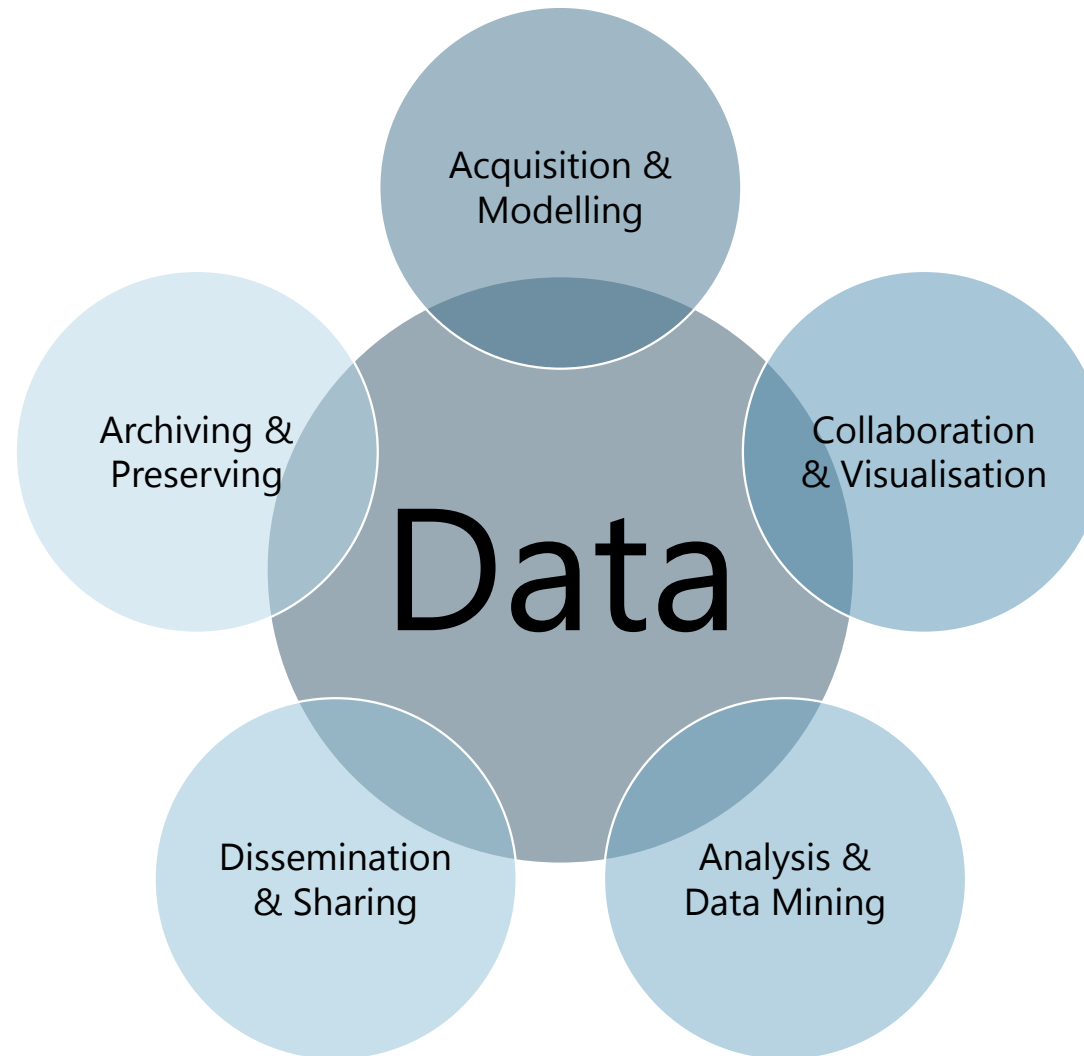
Key Emerging Themes

Scholarship as we know it is changing...

- Promoting the Curation & Sharing of Data
- Encouraging Research Reproducibility (interoperability)
- Changing the Reward System



The (Data-Intensive) Research Lifecycle



Acquisition & Modeling



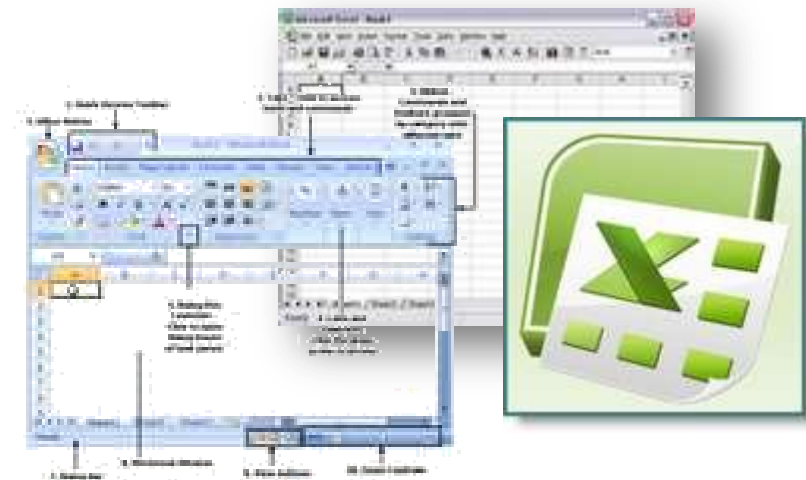
Data Curation for Excel (DCXL)

Excel add-in + accompanying web service
Microsoft Research with California Digital
Library's Curation Center, and linked to the
DataONE (an NSF DataNet Project)

- The goal is to facilitate data management, sharing, and archiving for scientists. To do this, we are building an open-source add-in (extension) for Microsoft Excel that will assist individuals in documenting and preparing their Excel data for archiving and sharing. Although we are targeting earth, environmental, and ecological scientists, we envision that this add-in will be useful for a wide range of people who wish to create collections based on Excel spreadsheet data. We define archiving as movement or storage of data that is not actively being used, and its accompanying metadata, to a repository for long-term retention.



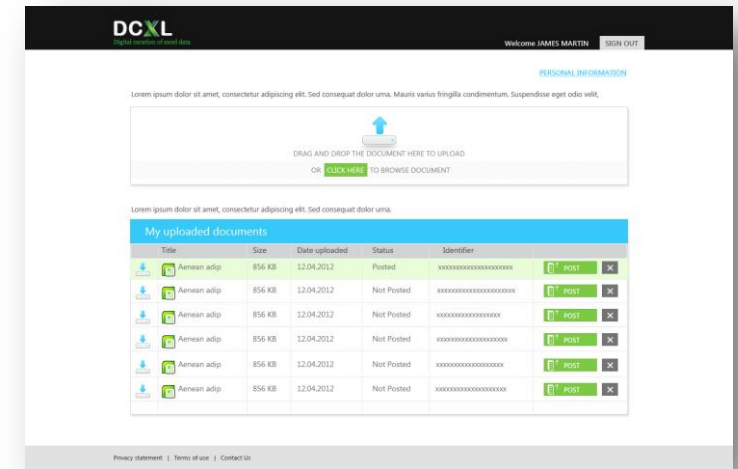
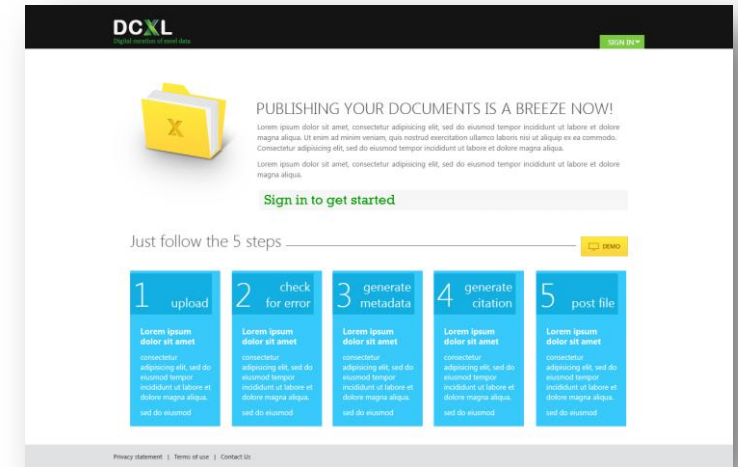
<http://dcxl.cdlib.org/>



DCXL Requirements http://dcxl.cdlib.org/?page_id=374



1. Check data file for .csv compatibility & create .csv version data file
2. Generate metadata that is linked to the data file
3. Generate a citation for the data file
4. Authenticate user to set up designated repository
5. Link an identifier to the data file
6. Ensure data file is ready for deposition into a repository
7. Submit the data file for deposition to the designated repository
8. Ensure compatibility for Excel users without the add-in



Collaboration & Visualization



Project VIVO

"Enabling National Networking of Scientists"

- Originally developed at Cornell University in 2003
- Expanded in 2009 through a \$12.2M stimulus grant from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH)

An open source semantic web application that enables the discovery of research and scholarship across disciplines at a particular institution and beyond.

- Populated with detailed profiles of faculty and researchers including information such as publications, teaching, service, and professional affiliations.
- Provides search functionality for locating people and information within the institution



<http://vivoweb.org/>

Partner Institutions



Cornell University



INDIANA UNIVERSITY



Weill Cornell Medical College

Virtual Research Environment (VRE) Toolkit for SharePoint

A collaborative, multi-institutional effort to create a set of researcher-focused extensions to Microsoft Office SharePoint Server 2010

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)
- La Trobe Templates for VRE – RIC with Active Directory (La Trobe University)

Collaborative environment
for research groups



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

Personal site for each
researcher and project
site for each project





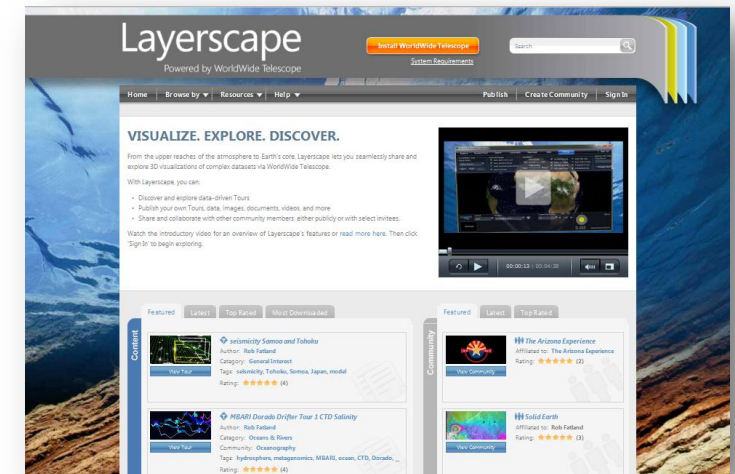
Layerscape (from Microsoft Research)

Layerscape has been set up primarily to enable Earth scientists to **collaborate with their peers in visualizing their data, their studies, and their scientific ideas**. Data, such as images, videos or collections of location based data, can be published easily. Studies and ideas can be articulated through the **creation of tours (animated slide shows)**. In addition, the site enables educators to collaborate with their faculty and students both in the **sharing of their own data and studies**, and also in the viewing of the scientific data where it has been made available for general viewing.

Leverages **Worldwide Telescope (WWT)** for visualizations.



<http://layerscape.org/>



Worldwide Telescope (WWT)



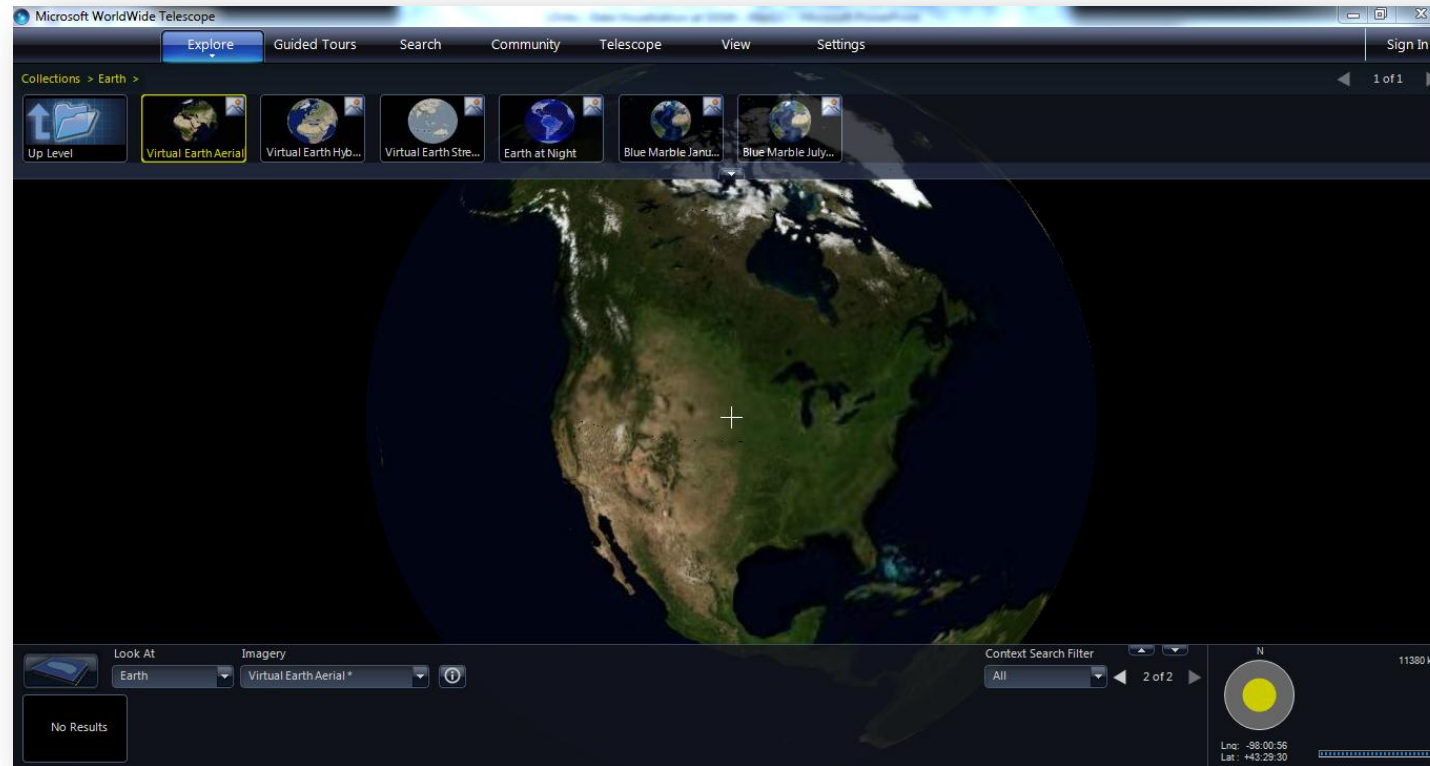
Launched
February 2008

TED
IDEAS WORTH SPREADING

<http://www.worldwidetelescope.org/>



WWT for Earth-based visualization



Earth Mode allows users to view a 3D model of the Earth, with a default data set with near global coverage and resolution down to sub-meter in high-population centers. Unlike most Earth viewers, WorldWide Telescope supports many different map projections including Mercator, Equirectangular and TOAST. There are also map layers for seasonal, night, streets, hybrid and science oriented MODIS imagery. *The new layer manger can be used to add data visualization on the Earth or other planets.*

Layerscape

demo

Analysis & Data Mining



PivotViewer (Silverlight)

PivotViewer makes it easier to interact with massive amounts of data on the web in ways that are powerful, informative, and fun. By visualizing thousands of related items at once, users can see trends and patterns that would be hidden when looking at one item at a time.

- Because PivotViewer leverages Deep Zoom, it displays full, high-resolution content without long load times, while the animations and natural transitions provide context and prevent users from feeling overwhelmed by large quantities of information.
- **The Silverlight PivotViewer control** is available now and can be accessed by developers and designers to begin creating collections and deploying solutions.

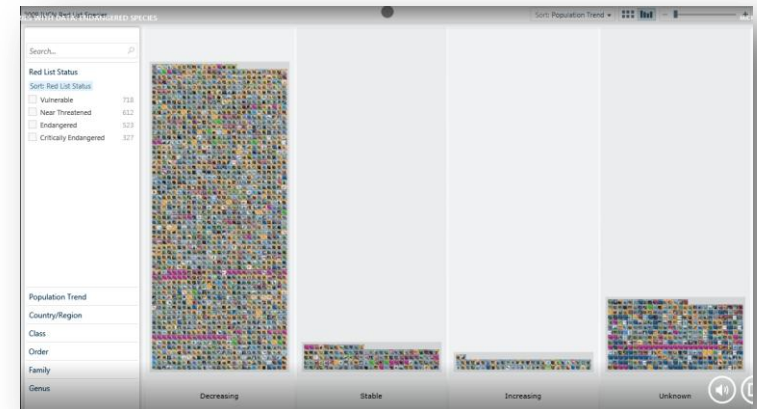


PivotViewer

<http://www.microsoft.com/silverlight/pivotviewer/>



Microsoft
Silverlight



Dissemination & Sharing



Sharing Data

"Data should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property."



Science & Technology
Facilities Council

"Researchers are to maximise the availability of data with as few restrictions as possible."

wellcometrust

"Investigators are expected to share the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants."



"It is NIH policy that the results and accomplishments of the activities that it funds should be made available to the public."



"Research data should be made available for use by other researchers. Researchers must retain research data, including electronic data, in a durable, indexed and retrievable form."



Australian Government
National Health and Medical Research Council

"Data should be considered for sharing and made as widely and as freely accessible as possible whilst safeguarding intellectual property, the privacy of the patients and confidential data."



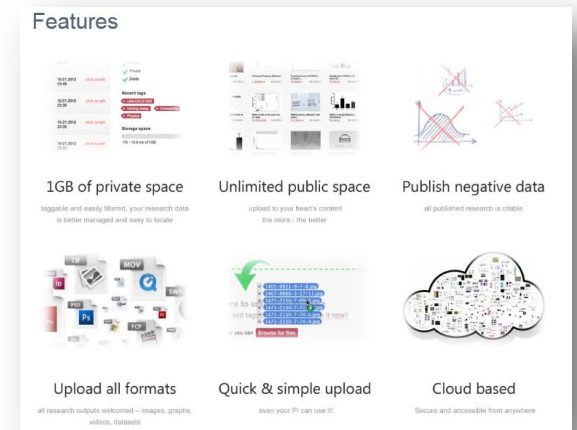


FigShare (a DigitalScience project)

- **Figshare** is the first online repository for storing and sharing all of your *preliminary* findings in the form of individual figures, datasets, media or filesets. Post preprint figures on Figshare to claim priority and receive feedback on your findings prior to formal publication.
- Figshare allows researchers to publish all of their research outputs in seconds in an easily citable, sharable and discoverable manner. All file formats can be published, including videos and datasets that are often demoted to the supplemental materials section in current publishing models.
- Figshare uses Creative Commons licensing to allow frictionless sharing of research data while allowing users to maintain their ownership. Figshare gives users unlimited public space and 1GB of private storage space for free.



<http://figshare.com/>



<http://www.digital-science.com/>



DataCite ("DOIs for data")

- An emerging protocol for citing data, based on DOIs.
- Founded jointly by TIB (Germany) and the British Library
- Goals:
 - *Establish easier access to research data on the Internet*
 - *Increase acceptance of research data as legitimate, citable contributions to the scholarly record*
 - *Support data archiving that will permit results to be verified and re-purposed for future study*
- Currently:
 - 15 members
 - Over 1,000,000 DOIs registered [Digital Object Identifier]
 - Metadata specifications released
 - Shared technical infrastructure established



<http://datacite.org/>

Databib

<http://databib.org/>

Registry of research data
Repositories (hosted by
Purdue University)

Microsoft Academic Search



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

Explore over 38.8 million publications...and growing

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.

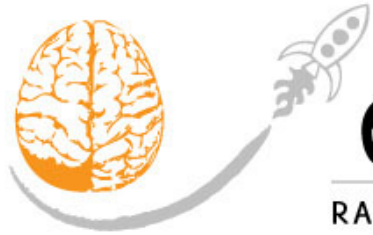
- Easily search the top papers, authors, conferences, and journals for a topic.
- See details about a specific paper, author, conference, journal or organization.
- Quickly find relationships between authors.
- Discover influential papers, authors, conferences, journals and organizations within a certain field.
- Get the latest Call for Papers.



Academic Search

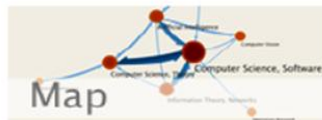
demo

The Eigenfactor Project



eigenFACTOR.org
RANKING AND MAPPING SCIENTIFIC KNOWLEDGE

The *Eigenfactor Project*TM 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.



Recommend | Map: Journals | Map: Papers | Explore | Rank | Categorize | About



<http://mas.eigenfactor.org/>

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.



The “AltMetrics” Movement

- Total Impact = A website that allows one to **view the impact of a wide range of research output**. It goes beyond traditional measurements of research output -- citations to papers -- to embrace a much broader evidence of use across a wide range of scholarly output types. **The system aggregates impact data from many sources and displays it in a single report**, which is given a permaURL for dissemination and can be updated any time.
- Total Impact can track a wide range of research artifacts—including **papers, datasets, software, preprints, and slides** – as well as the related **social media**.
- See also:
 - altmetric.com API demo
 - CitedIn
 - [PLOS](http://PLOS.org) “Article-Level Metrics” application
 - ReaderMeter
 - Science Card
 - Press Forward initiative

total·Impact

<http://total-impact.org/>



J. Priem, D. Taraborelli, P. Groth, C. Neylon (2010), **Altmetrics: A manifesto**, (v.1.0), 26 October 2010. <http://altmetrics.org/manifesto/>



Total Impact—a deeper dive

uncovering the invisible impacts of your research.

Enter the identifiers for a collection of artifacts you want to track. We'll provide you a permanent URL to track statistics about this collection. You can peruse [a sample](#) and [recently-shared reports](#).

what kind of research artifacts can be tracked?

[back to contents](#)

Total-Impact currently tracks a wide range of research artifacts, including papers, datasets, software, preprints, and slides.

Because the software is in early development it has limited robustness for input variations: please pay close attention to the expected format and follow it exactly. For example, inadvertently including a "doi:" prefix, or omitting "http" from a url may render the IDs unrecognizable by the system. Add each ID on a separate line in the input box.

artifact type	host	supported ID format	example
a published paper	any journal that issues DOIs	DOI (simply the DOI alone)	10.1371/journal.pcbi.1000361
a published paper	PubMed	PubMed ID (no prefix)	17808382
a published paper	Mendeley	Mendeley UUID	ef35f440-957f-11df-96dc-0024e8453de8
dataset	Genbank	accession number	AF313620
dataset	PDB	accession number	2BAK
dataset	Gene Expression Omnibus	accession number	GSE2109
dataset	ArrayExpress	accession number	E-MEXP-88
dataset	Dryad	DOI	10.5061/dryad.1295
software	GitHub	URL (starting with http)	https://github.com/mhahnel/total-impact
software	SourceForge	URL	http://sourceforge.net/projects/aresgalaxy
slides	SlideShare	URL	ttp://www.slideshare.net/phylogenomics/eisenall-hands
generic url	A conference paper, website resource, etc.	URL	http://opensciencesummit.com/program/

report for Heather Piwowar

[download](#) [refine](#) [run update](#) updated 31 Oct, 2011 created 31 Oct, 2011 47 artifacts

Permalink: <http://total-impact.org/report.php?id=Sllysw> [copy](#) [Tweet](#) 3

article

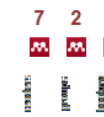
10.1371/journal.pmed.0050183

(2008) Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers *Plos Med.*



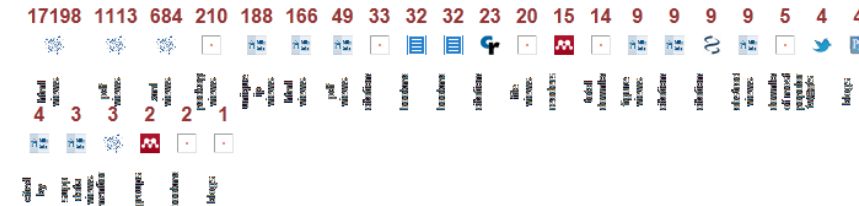
10.1038/npre.2008.1701.1

Piwowar, Chapman (2008) Prevalence and Patterns of Microarray Data Sharing *Nature Precedings.*



10.1371/journal.pone.0000308

Piwowar, Day, Fridsma (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate *PLoS ONE.*
Piwowar HA, Day RS, Fridsma DB Data from: Sharing detailed research data is associated with increased citation rate. *Dryad Data Repository.*



Something missing on import?
See a list of [current limitations](#).

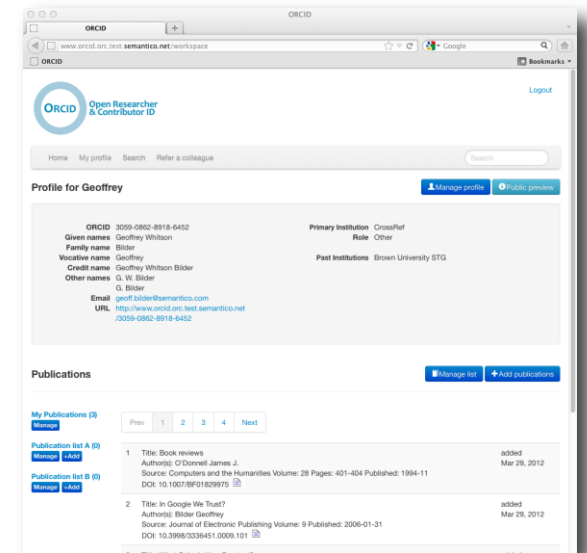
Open Researcher & Contributor ID (ORCID) Initiative

ORCID is a non-profit organization dedicated to solving the name ambiguity problem in scholarly research.

- Problem: Reliable attribution of authors and contributors is impossible without unique person-level identifiers
- Goal: Establish a registry that is adopted and embraced as the de facto standard by the whole of the community.
- Initiated by Thomson-Reuters & Nature Publishing in 2009, ORCID now has 50 Founding Sponsors and 330+ participating organizations.
- May 17th 2012: Provided API and sandbox server for third parties—publishers, institutions, funders—to test system integration.
- Fall 2012: Incorporate requirements determined during integration testing, onboard third party systems, pre-populate profiles, launch ORCID v1.1



<http://www.orcid.org/>



Open Researcher & Contributor ID - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Open Researcher & Contrib... Open Researcher & Contrib...

http://localhost:8080/orcid-web/Workspace.action

Google



Home My Researcher Profile Refer a Colleague Logout Search

Carberry, Joisah

Manage Profile

Preview Public Version

ORCID ID: A-1001-2011

Other Names: J. Carberry; J. S. Carberry; jcarberry

E-mail:

URL: <http://www.orcid.org/orcid-web/A-1001-2011>

Role:

Subject: Anthropology; Archaeology; Area Studies; Arts & Humanities - Other ; Astronomy & Astrophysics; Behavioral Sciences; Biophysics

Keywords: psychoceramics; bilocation; high-energy metaphysics

Description: [Enter a Description](#)

My URLs:

My Institutions [\(more details\)](#)

Primary Institution: Brown University

Sub-org./Dept: Psychoceramics

Role: Professor

Joint Affiliation:

Sub-org./Dept:

Role:

Past Institutions:

My External Identifiers:

Publications

: View

Add Publications

0

[View Publications](#)

Manage | [Add](#)

0

[View Publications](#)

Manage | [Add](#)

Welcome to your ""

The "" is an area to which you can add publications that you have authored.

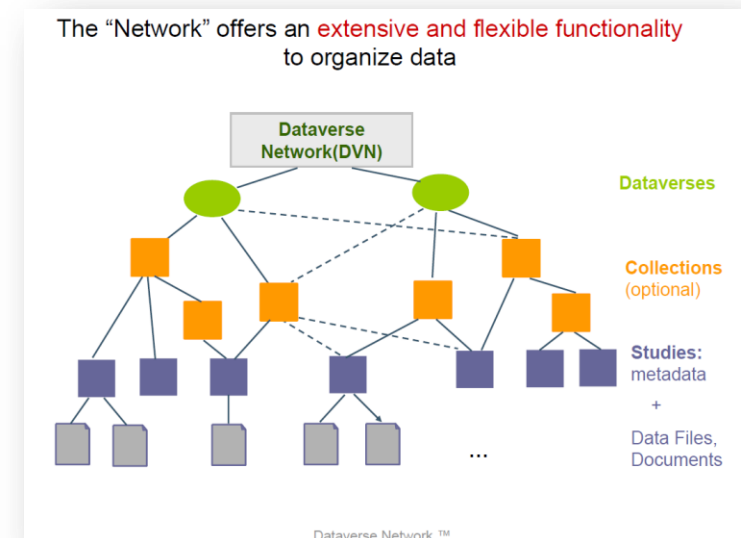
To add publications, click the **Add** link located in the menu on the left-hand side of your screen, or by clicking on the **Add Publications** above.

You have the ability to make your "" public to visitors of ORCID for promotion of your scholarly output. Click on the **Manage Profile** button at the top-right corner of the page to change the public and private settings of your data. Note that

Archiving & Preservation

Harvard's "DataVerse" Project

- Developed by the Institute of Quantitative Social Science (IQSS) at Harvard University
- Leveraging web application software, data citation standards, and statistical methods, the **Dataverse Network** project increases scholarly recognition and distributed control for authors, journals, archives, teachers, and others who produce or organize data; facilitates data access and analysis for researchers and students; and ensures long-term preservation whether or not the data are in the public domain.



<http://thedata.org/>

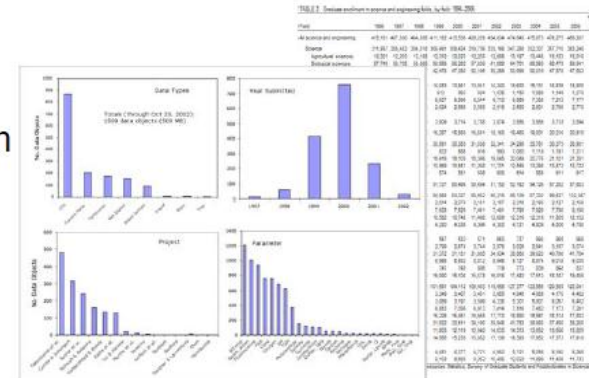


Dataverse (cont'd)

A **data citation** is created for each research study in the Dataverse



Allows to cite
research data from
published work



Printed Citation for published work:

Verba, Sidney; Nie, Norman H., 1984, "Political Participation in America, 1967",
The Social Science Journal, v1, p.1

On-line citation for research data:

Verba, Sidney; Nie, Norman H., 1984, "Political Participation in America, 1967",
[hdl:1902.2/7015](https://doi.org/10.1002/2/7015) UNF:3:+DNr7jVq/5XmsPAmls4KQg==
Inter-university Consortium for Political and Social Research [Distributor]

[hdl:1902.2/7015](https://doi.org/10.1002/2/7015) → Persistent Identifier
[UNF:3:+DNr7jVq/5XmsPAmls4KQg==](https://nfdi.ub.uni-due.de/urn:nbn:de:hbz:5:1-63862-p0011-9) → Universal Numerical Fingerprint

Dataverse Network TM

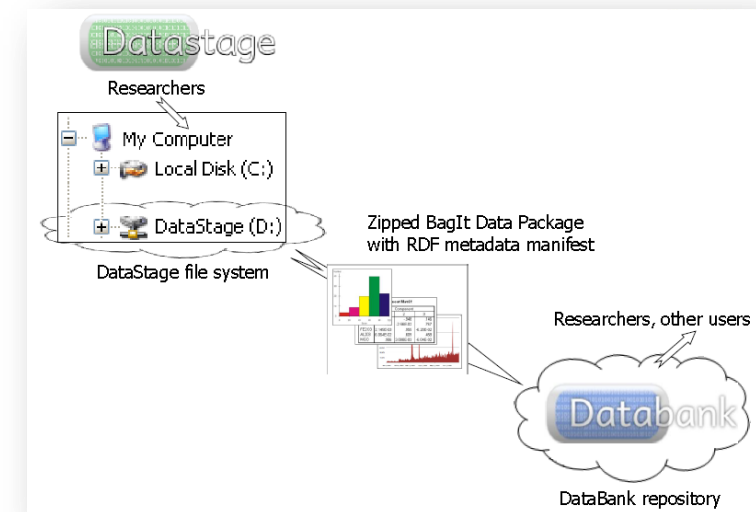


DataFlow (University of Oxford)

- **DataFlow** is creating a two-stage data management infrastructure that makes it easy for you and your research group to work with, annotate, publish, and permanently store your research data. You manage this locally using your own instance of **DataStage**, while allowing your institution to deploy **DataBank** easily to preserve and publish your most valuable datasets.
- **DataBank** is a scalable data repository designed for institutional deployment.
- **DataStage** is a secure personalized 'local' file management environment for use at the research group level, appearing as a mapped drive on the end-user's computer.



<http://www.dataflow.ox.ac.uk/>





DuraCloud (a DuraSpace service)

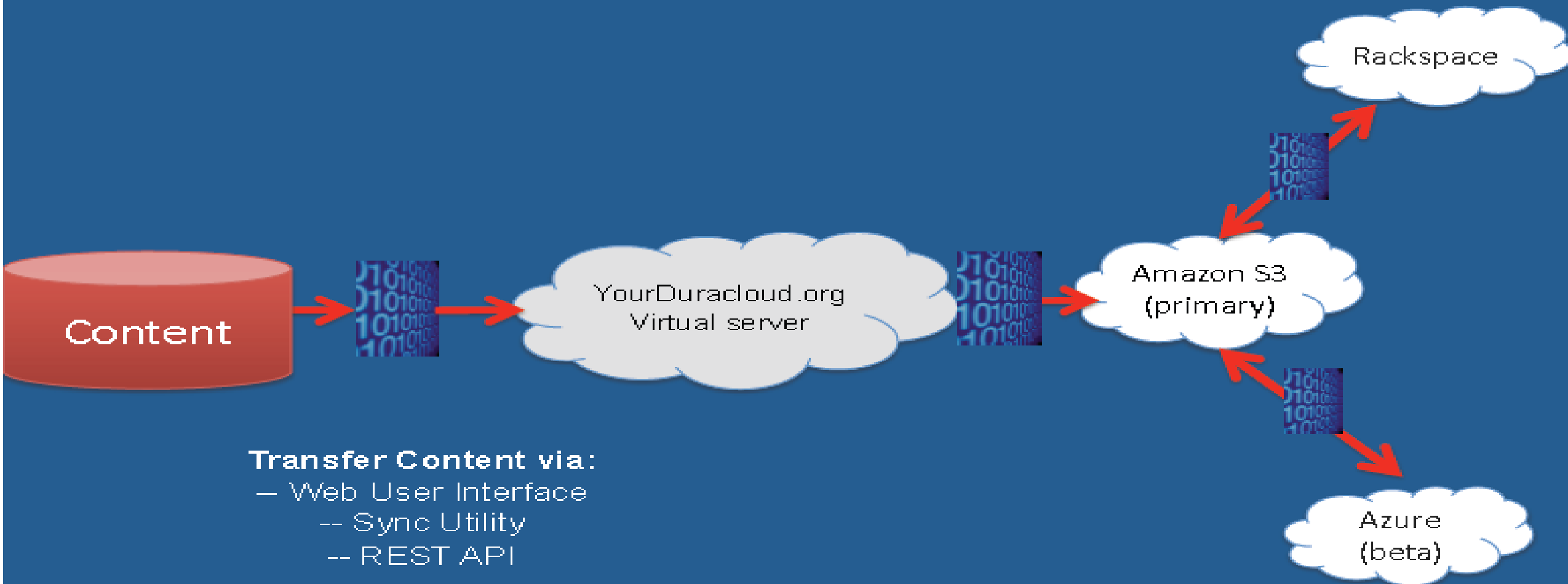
- DuraCloud is a hosted service and open technology developed by DuraSpace that makes it easy for organizations and end users to use cloud services. It is a cloud-based service that leverages existing cloud infrastructure to enable durability and access to digital content.
- DuraCloud builds on pure storage from expert storage providers by overlaying the access functionality and preservation support tools that are essential to ensuring long-term access and ease of use. DuraCloud offers cloud storage and replication of content across multiple providers, via one web-accessible interface.
- V1.0 Launched in Nov11 with 11 pilot partners, including—Columbia University, Northwestern University, Rice University, and many others.



<http://duracloud.org/>

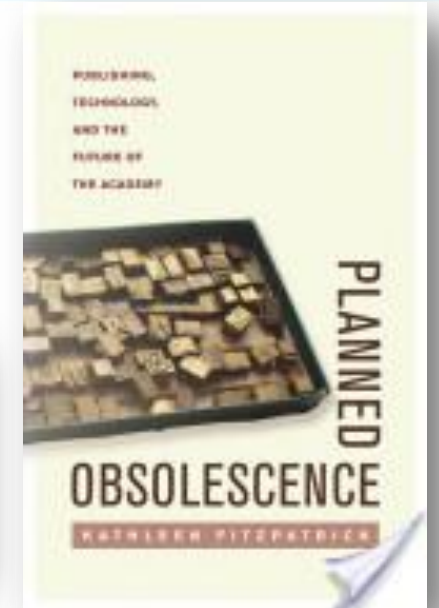
 **DURASPACE.ORG**

How does DuraCloud work?



Recommended Reading

A dramatically changed peer-review system such as the one that I propose, however, would require us to think about new structures of authorship. In chapter 2, I argue that a turn from pre-publication review to post-publication review will almost certainly necessitate a parallel turn from thinking about academic publishing as a system focused on the production and dissemination of individual *products* to imagining it as a system focused more broadly on facilitating the *processes* of scholarly work, as the time and effort required to maintain a community-oriented, gift-economy-driven system of peer-to-peer review will oblige scholars, much like the developers of large-scale open-source software projects, to place some portion of their emphasis not on their own individual achievements, but rather on finding their self-interest served by the advancement of the community as a whole. This is a utopian ideal, of course, and it largely goes against our training as scholars, particularly within the humanities; what we accomplish, we accomplish alone.



Planned Obsolescence: *Publishing, Technology, & the Future of the Academy*

By Kathleen Fitzpatrick
(New York University Press, 2011)



Change Hurts

"People change when the pain of the status quo becomes greater than the fear of making the change."

– Helen Harkness

<http://www.amazon.com/Career-Chase-Creative-Control-Chaotic/dp/0891060987>

What will this future look like?



Evolved Infrastructure

- Easily shareable and findable data sets.
- Data + publications that are preserved for posterity.

New Tools & Resources

- Quickly consumable and re-useable data sets.
- Powerful tools for visualization and analysis across data sets (perhaps even across disciplines).

Social and Behavioral Change

- Automatic attribution for specific creators who get credit for sharing.



¡Gracias!



Lee Dirks



Director, Portfolio Strategy

Microsoft Research | Connections



ldirks@microsoft.com



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

The logo features the Microsoft logo in a small, white, sans-serif font. Below it, the word "Research" is written in a large, bold, white, sans-serif font. To the right of "Research", the word "Connections" is written in a white, sans-serif font, slightly smaller than "Research". The entire text is centered horizontally against a blue background with a white geometric pattern of interconnected lines forming a mesh of triangles.

Microsoft®

Research Connections

The Microsoft logo is displayed in a large, bold, white, italicized sans-serif font. The letters are closely spaced, and the 'i' in 'Microsoft' has a distinct dot. A registered trademark symbol (®) is located at the top right of the 't'.

© 2012 Microsoft Corporation. All rights reserved. Microsoft, Windows, and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.