



Microsoft® Research

# Faculty Summit

10  
YEAR ANNIVERSARY

# Microsoft Research Faculty Summit 2009

Harold Javid  
Conference Chair  
Microsoft External Research



# Microsoft External Research

Tony Hey  
Corporate Vice President  
Microsoft External Research

# Welcome to the Microsoft Research Faculty Summit 2009

Microsoft  
Research

- Addressing World-Scale Challenges
- Computation as a powerful change agent in areas such as Energy, Environment, Healthcare, Education
- Collaboration and Community





# A Deluge of Data = Research Opportunities

- Massive amounts of data collected and aggregated from the internet, satellites, sensors, and other sources
- We need to move from data to knowledge
- Computing technologies are enabling new approaches applied to world-scale challenges in disciplines such as medicine and healthcare, energy and the environment, educational and social progress



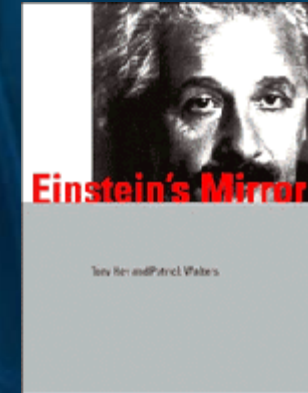
*Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster from the Digitized Sky Survey combined with an image of the moon, synthesized within the WorldWide Telescope service.*

# Tony Hey – An Introduction

Microsoft®  
Research

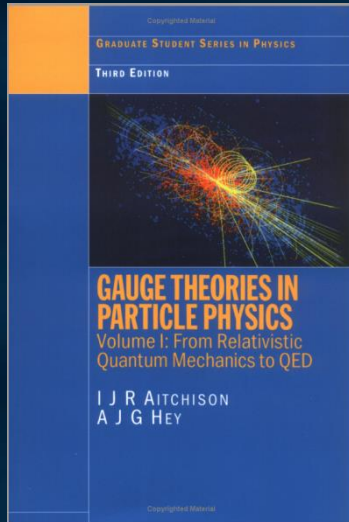


Research Councils UK



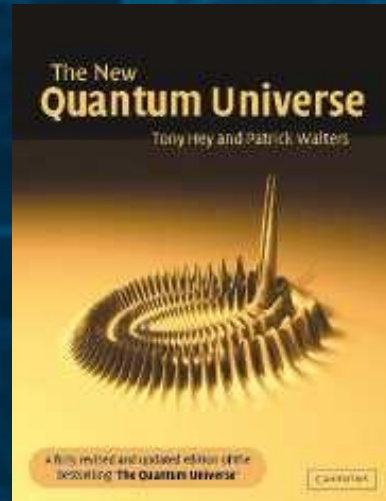
Einstein's Mirror

Tony Hey and Patrick Walters



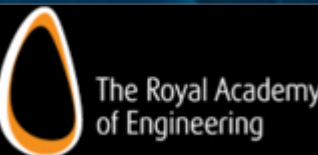
GAUGE THEORIES IN  
PARTICLE PHYSICS  
Volume I: From Relativistic  
Quantum Mechanics to QED

I J R AITCHISON  
A J G HEY

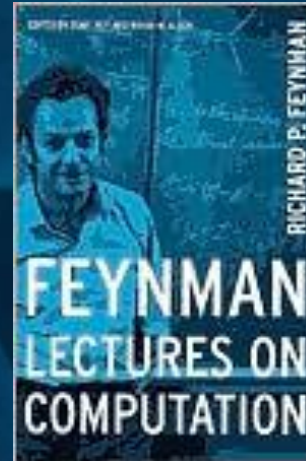


The New  
Quantum Universe  
Tony Hey and Patrick Walters

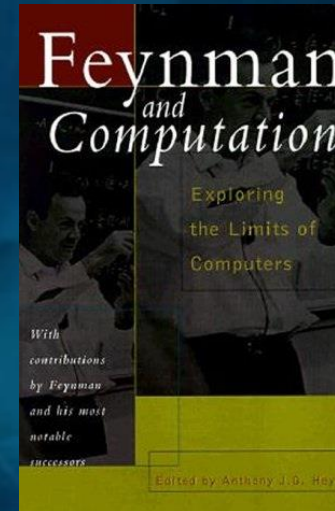
A fully revised and updated edition of the  
bestselling *The Quantum Universe*



The Royal Academy  
of Engineering



RICHARD P. FEYNMAN  
FEYNMAN  
LECTURES ON  
COMPUTATION



Feynman  
and  
Computation

Exploring  
the Limits of  
Computers

With  
contributions  
by Feynman  
and his most  
notable  
successors

Edited by Anthony J.G. Hey



Commander of the British Empire





- 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 External Research is committed to interoperability and to providing open access, open tools, and open technology

# External Research Global Themes

Microsoft<sup>®</sup>  
Research

Community and Geographic Outreach

Core Computer  
Science



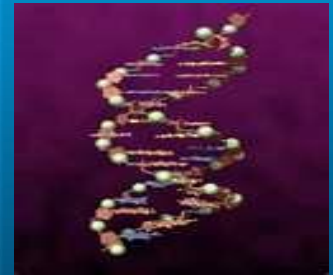
Earth, Energy &  
Environment



Education &  
Scholarly  
Communication



Health &  
Wellbeing



Advanced Research Tools and Services




# Since the Faculty Summit One Year Ago:

Microsoft Research

MCE

## Mobile Computing in Education

Edited by University Relations, Microsoft Research Asia



### Research

#### CellScope Could Offer Low-Cost, Portable Options for Disease Diagnosis

Project Principal: David Fletcher, assistant professor, University of California at Berkeley

Fast Facts: The CellScope is a handheld device that can be used to diagnose diseases such as malaria and tuberculosis. It is a low-cost, portable device that can be used in developing countries. It is a handheld device that can be used to diagnose diseases such as malaria and tuberculosis. It is a low-cost, portable device that can be used in developing countries.



### WorldWide Telescope

Microsoft Next Media Research

### Researchers Use Web Tool to Boost Collaboration in Latin America

Academic researchers throughout Latin America and the Caribbean make important contributions to computer science, yet much of their work involves collaboration with project teams on other continents and does not address locally relevant challenges. Several universities have joined forces to enable greater collaboration within the region and more technology breakthroughs in areas such as healthcare, education, the environment, energy and business development—using Conference@CIC as a Web-based videoconferencing tool.

For computer science researchers at Latin American and Caribbean universities, collaborating with peers halfway around the world can often be easier than trying to initiate a research project with someone in a neighboring country. Lack of regional funding, a smaller pool of colleagues with similar research interests and limited resources to form locally-based research teams are among the hurdles that make Latin American researchers more likely to work on U.S.- or European-led projects than ones based closer to home.

It is a problem of critical mass: Latin America has few researchers and we are very isolated," says Ignacio Cases, an associate professor in the Computer Science Department at Pontificia Universidad Católica de Chile (PUC-CH) in Santiago. Over the past decade, he has led 90 percent of the computer science projects involving Latin American researchers in another continent.

In North America and Europe, the situation is different. Researchers at the Latin American and Caribbean Universities are helping Conference@CIC address research challenges as a forum for live discussion and information sharing.

Other academic researchers are the region using Conference@CIC.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

### COMPUTER SCIENCE

#### Beyond the Data Deluge

Gordon Bell, Tony Hey, Alex Szalay

Since at least Newton's laws of motion in the 17th century, scientists have organized experimental and theoretical science in the basic research paradigm for understanding nature. In recent decades, computer simulations have become an essential tool for extending the paradigm to explore domains that are inaccessible to theory and experiment, such as the evolution of the universe, car passenger crash testing, and predicting climate change. As simulations and experiments yield ever more data, a fourth paradigm is emerging, consisting of the techniques and technologies needed to perform data-intensive science (1). For example, new types of computer clusters are emerging that are optimized for data movement and analysis rather than computing, while in astronomy and other sciences, integrated data systems allow data analysis and storage on one instead of requiring download of large amounts of data.

Today, some areas of science are facing a deluge of data. In astronomy, for example, the volume of data generated by a single telescope is increasing at a rate of about 100 terabytes per year. In genomics, the volume of data generated by a single experiment is increasing at a rate of about 100 terabytes per year. In genomics, the volume of data generated by a single experiment is increasing at a rate of about 100 terabytes per year.

Other research fields also face major data management challenges. In almost every laboratory, "born digital" data proliferate in files, databases, or databases stored on hard drives, digital notebooks, Web sites, blogs, and so on. The management, curation, and archiving of these digital data are becoming increasingly important for research scientists.

Over the past 40 years or more, Moore's Law has enabled researchers to obtain data at a smaller and smaller scale. At the same time, technology improvements for data storage cannot keep up with the ever increasing flood of scientific data generated by the faster computers. In university research labs, research clusters—groups of mostly identical, inexpensive PCs connected that can be used for parallel computations—have



More and More from the VO Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a virtual site. The image shows the Trifid, the Helix, and the Ring galaxies from the Digitized Sky Survey combined with an image of the Moon, synthesized with the World Wide Telescope.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.

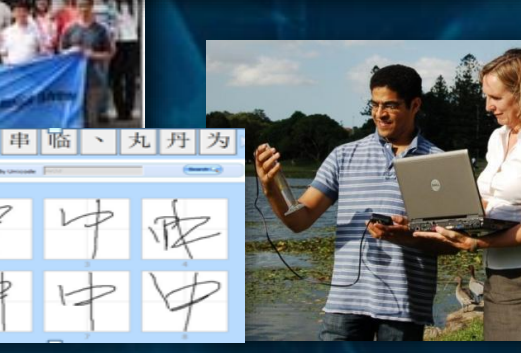
Microsoft External Research Division within Microsoft Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide. Microsoft External Research is committed to providing research expertise to enable ground-breaking projects worldwide.



COLLABORATION • INNOVATION • INTEROPERABILITY • DISCOVERY • PARTNERSHIP



External Research Division





# Introducing: Prof. Judith Bishop – Core Computer Science theme lead

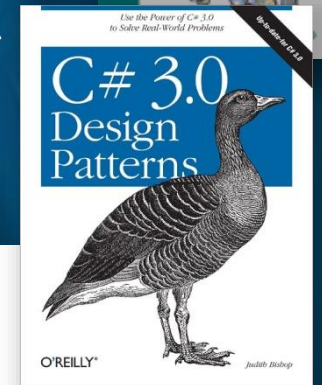
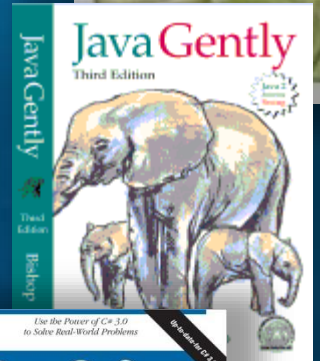
Microsoft®  
Research



*Judith Bishop was in the first group to study computer science in South Africa in 1970 and has stayed at the front of her field of programming languages for distributed systems ever since.*

*She wrote the first BASIC compiler for ICL computers in 1972 and was involved in the first Pascal compiler for the in 1976. Her doctorate investigated the relationship between the new languages of the 1970s (such as Ada and occam) and the stack and descriptor based mainframes of the time. She wrote the first Java textbook to become widely used in 1997 and **one of the first C# textbooks in 2004**. After having contributed to the field of configuration description languages in the 1990s, **she now works on the principles of adaptive software in a multi-lingual and mobile environment, in collaboration with Microsoft Research, local companies and collaborators in Germany and Italy**. Professor Bishop is the top NRF rated woman computer scientist in South Africa and has published over 70 journal and conference papers. Her 14 books are available in six languages and read worldwide.*

[www.cs.up.ac.za/cs/jbishop/](http://www.cs.up.ac.za/cs/jbishop/)





# Accelerating time to insight with advanced research tools and services



Our goal is to accelerate research by collaborating with academic communities to create open tools and services based on Microsoft platforms and productivity software.

By building open software solutions in collaboration with the research community, we help scientists spend more time on their research and less time on IT issues

*announcing*

# Project Trident: A Scientific Workflow Workbench

## Dryad and DryadLINQ



# Project Trident: A Scientific Workflow Workbench

## Accelerating the pace of discovery

Microsoft  
Research

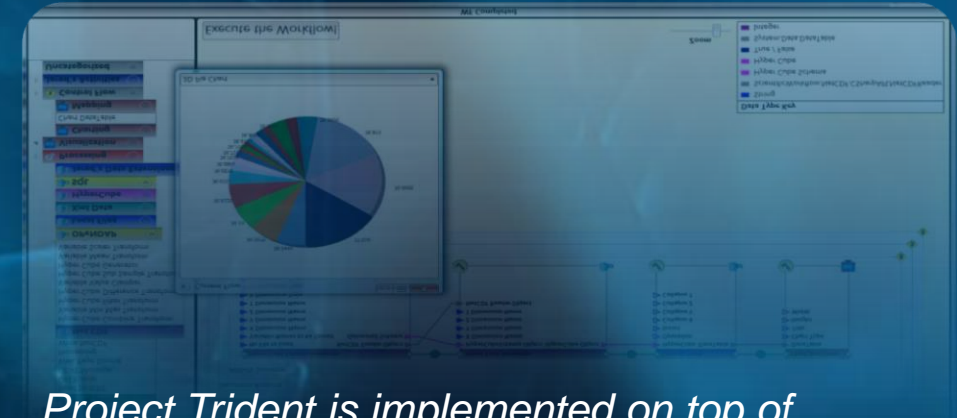
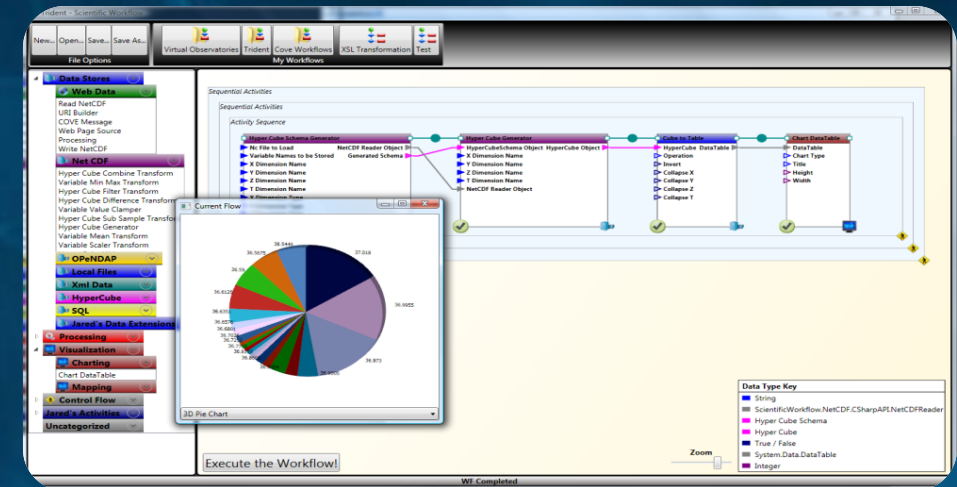
- Makes it easier for scientists to ingest and make sense of data
- Get answers to questions at a rate not previously possible
- Capture provenance
- Scientists in data-intensive fields such as oceanography, astronomy, environmental science and medical research can use these tools to manage, integrate and visualize volumes of information.
- The tools are available as no-cost downloads to academic researchers and scientists



*What once required weeks  
or months of custom coding,  
now takes just hours*

# Project Trident for Researchers

- Visually program workflows
- Libraries of versioned activities and workflows
- Social annotations and search, export entire workflow libraries to share their methodology.
- Automatically schedules workflows over HPCS
- Support for administering and monitoring workflows
- Automatic provenance capture, for both workflows and results
- Cost model, including elapsed time, CPU, memory, data transfer
- Integrated data storage and access, from SQL to S3 and SDS
- Integrated visualization tools
- Fault tolerance, also used to facilitate smart reruns and what-if analysis
- Supports reproducible research



*Project Trident is implemented on top of Microsoft's Windows Workflow Foundation, using the existing functionality of a commercial workflow engine based on SQL Server and Windows HPC cluster technologies.*



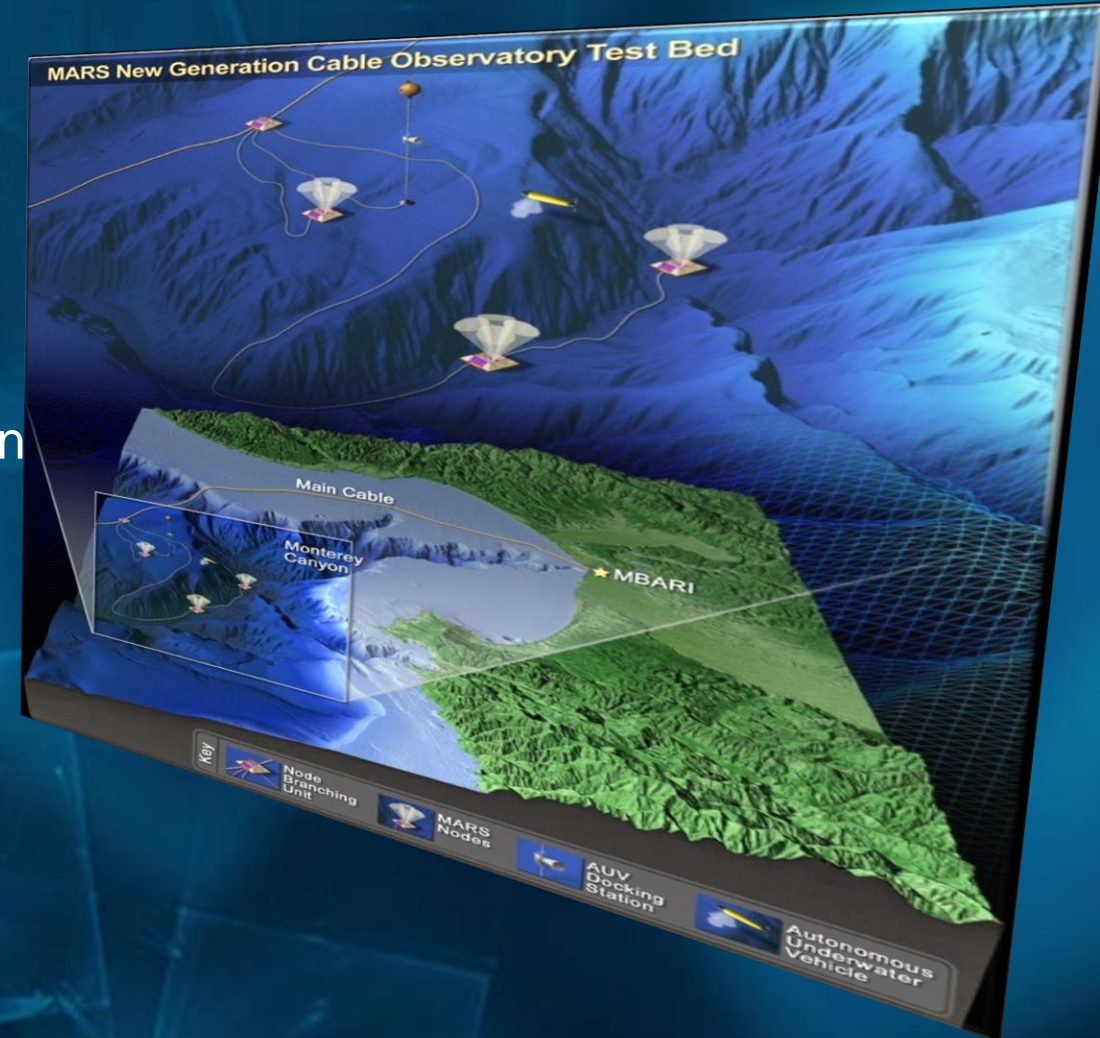
# Project Trident: Scientific Workflow Workbench

University of Washington and Monterey Bay Aquarium Research Institute

Microsoft®  
Research

*Scientific workflow workbench to automate the data processing pipelines of the world's first plate-scale undersea observatory*

- From raw data to useable data products (visualizations)
- Focusing on cleaning, analysis, re-gridding, interpolation
- Support real time, on-demand visualizations
- Custom activities and workflow libraries for authoring
- Visual programming accessible via a browser



# Word Add-In for Reproducible Research

Microsoft  
Research

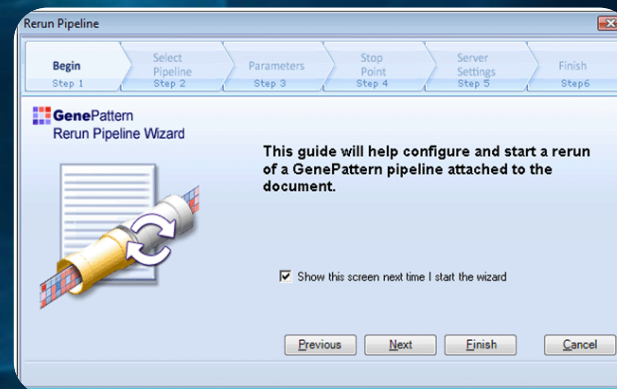
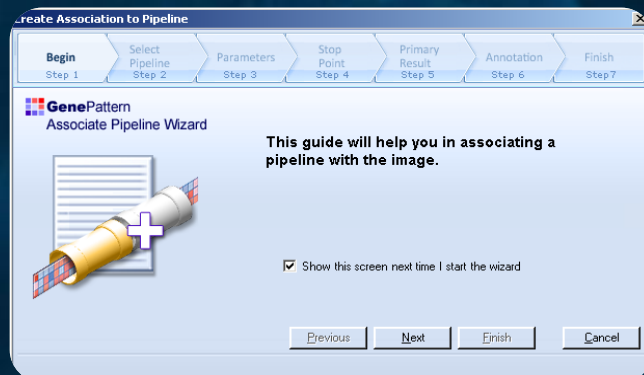
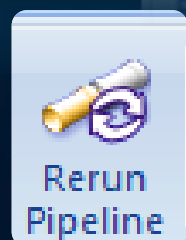
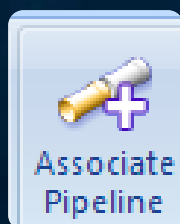


Allows users to connect to a Trident server;

To insert the output of a workflow (pipeline) into a document;

Each pipeline in a Word document is associated with an image or text;

A person reading the document can click on such an image or icon to view the associated pipelines and its input files, and rerun the pipelines on a Trident server while remaining in the Word application.





Document1 - Microsoft Word

Home Insert Page Layout References Mailings Review View **GenePattern** Trident

Dashboard View Protect Associate Insert Rerun  
Associations Document Pipeline GenePattern Link Pipeline  
Document Tasks

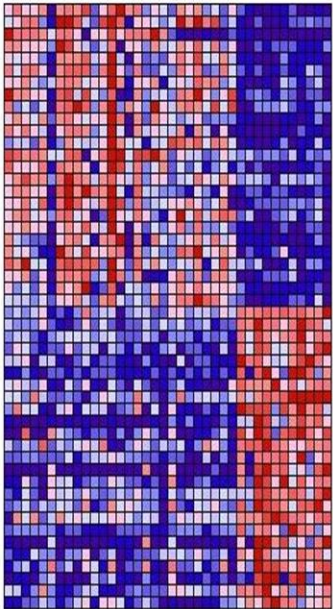
Remove All Pipeline Associations  
Use Wizard UI

Connections Current Connection: Create Connection  
GenePattern Server Connections

**Rerun of GenePattern pipeline association Demo**  
Added at 1/7/2009 11:30:22 AM

Pipeline parameters:  
data filename: C:\Downloads\GenePattern\Test Files\all\_aml\_train  
all\_aml\_train.gct  
class filename: C:\Downloads\GenePattern\Test Files\all\_aml\_train  
all\_aml\_train.cls

Inserted results:



U22378\_cdi2\_1\_at  
X59417\_at  
U05155\_inal\_1\_at  
M59287\_at  
X74261\_at  
L13278\_at  
M31211\_1\_at  
D26156\_1\_at  
M31522\_at  
U12944\_at  
L47739\_at  
M28170\_at  
Z11115\_at  
X52142\_at  
X15949\_at  
AF009426\_at  
S50223\_at  
M91432\_at  
U19175\_at  
D18073\_at  
X61469\_at  
U29998\_at  
Z69881\_at  
J05242\_at  
M55150\_at  
U50126\_inal\_1\_at  
X95735\_at  
M16038\_at  
M23197\_at  
M84526\_at  
Y11670\_at  
U82759\_at  
D49950\_at  
M27891\_at  
X17642\_at  
U12477\_cdi1\_1\_at  
U46751\_at  
Y06787\_1\_at  
L08240\_at  
M80256\_at  
M62782\_at  
M81923\_at  
M96316\_inal\_1\_at  
M28110\_inal\_1\_at  
M63138\_at  
M11147\_at  
M57710\_at  
M81995\_1\_at  
X85116\_inal\_1\_at

- Deployed at MIT BROAD, Gene Pattern Analysis Server, Jill Meserov PI.
- Trident Scientific Workflow Workbench

Document1 - Microsoft Word

Home Insert Page Layout References Mailings Review View **GenePattern** Trident **Format**

Dashboard View Protect Associate Insert Rerun  
Associations Document Pipeline Workflow Workflow Trident Link Workflow  
Document Tasks

Connections Current Connection: Trident Create Connection  
Trident Server Connections

Workflow Associations

Association	Job	Job Id
Association 1	Job 1	052bbe...

Edit Delete Export

Activities Description

- My Ocean Current
- Hyper Cube Schema Generator
- Hyper Cube Generator
- Cube to Table
- Create DataTable Chart

Run

Outputs Data Products Parameters

ChartImage

Insert into Document Save Result

**Original run of Trident Workflow Association 'Association 1'**  
Added at: 6/24/2009 7:03:14 PM  
Comment:  
Scheduled by: Roger Barga

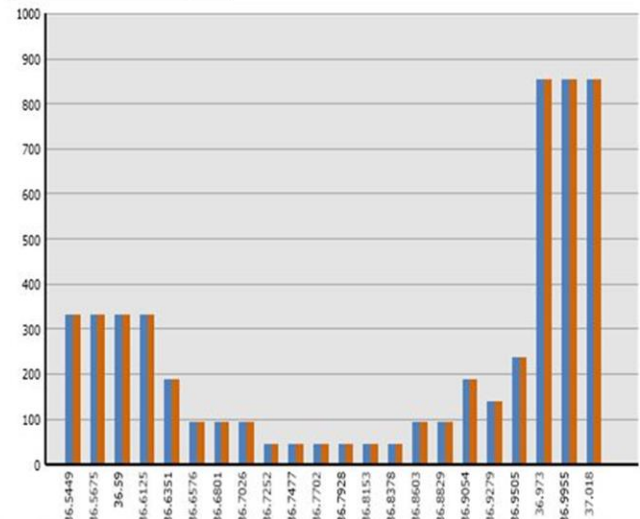
Workflow Required Inputs:

Input Parameter	Value
HyperCubeSchemaGeneratorActivity.NfileName	..\Workflows\Qcc\codar_mnty_4.nc

Workflow Required Outputs:

Output Parameter	Value
ChartDataTableEx.ChartImage	image/gif

Inserted Workflow Outputs:



Page: 1 of 1 Words: 37 English (United States)

# DyradLINQ on HPCS for academic research

- Turning a cluster into an easy-to-use tool:
  - Dryad was designed to simplify the task of implementing distributed applications on clusters of Windows computers
  - DyradLINQ is an abstraction layer, which simplifies the process of implementing Dryad-based applications
- The Academic Release includes:
  - Installation guide, programming samples, tutorials.
  - Client SDK Installer – *installs DyradLINQ, docs and code samples.*
  - Dryad & Dryad Management Tools installer (cluster-side installation)
- The Pre-release was installed at Indiana University and the University of Washington
  - Successfully developed bioinformatics application (pair-wise alignment of genetic sequences) with virtually no support
  - Successfully developed queries for LSST data
- Small community of internal DyradLINQ developers tested on a shared infrastructure (k18 cluster, 70 nodes)



*Dryad Wood Nymph*



# Where to download the tools

[research.microsoft.com/en-us/collaboration/tools](https://research.microsoft.com/en-us/collaboration/tools)

## Other resources:

- **Tools to Access Petabytes of Data**
  - **Beyond Search with Data Driven Intelligence (11 AM, Cascade)**  
Harry Shum, Corporate Vice President, Microsoft
  - “The future of search focusing on data-driven research to help advance the state-of-the-art in the online world”
- **DemoFest Booth 4**
  - **Tools and Services for Data Intensive Research**

# Have a Great Faculty Summit 2009

<http://research.microsoft.com/en-us/events/fs2009>

Microsoft®  
Research

- Agenda
  - On-line
  - Printed
- Please let us know if you have any questions or need any help





# ***Microsoft®***

*Your potential. Our passion.™*