

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

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

The screenshot shows the Microsoft Research Faculty Summit 2009 website. At the top left is a logo for the 10th anniversary of the Faculty Summit. To its right is a paragraph of introductory text. Below the logo is a section titled 'Addressing World-Scale Challenges' with a paragraph of text. To the right of this section is a 'Faculty Summit 2009 Contents' list with links for Overview, Agenda Day 1, Agenda Day 2, Speaker Biographies, and FAQ. Below the contents is a 'Related Links' section with links for Faculty Summits at Microsoft Research, Faculty Summit 2008 in Redmond, and Collaboration at Microsoft Research. At the bottom of the page is a section titled 'Identifying computational enablers' with a paragraph of text and three bullet points: Energy Sustainability, Addressing climate change, and Transformational improvement in healthcare.

Microsoft Research Faculty Summit 2009

On July 13-14, 2009, the tenth Microsoft Research Faculty Summit brings together more than 400 thought-leaders from academia, government, and Microsoft to reflect on how current computing disciplines open new opportunities for research and development.

### Addressing World-Scale Challenges

Computational approaches provide a powerful means for addressing previously unsolvable problems. Increasingly, computing technologies are what makes the difference in enabling new approaches applied to world-scale challenges in such diverse disciplines as medicine and healthcare, energy and the environment, and educational and social progress. In response to these significant global challenges, the Microsoft Research Faculty Summit 2009 investigates how computing technologies can best help scientists make progress in these important areas. Attendees will have the opportunity to participate in creative, open discourse on research topics.

#### Identifying computational enablers

for solving critical social and scientific problems is a main theme for this year's faculty summit:

- **Energy Sustainability.** Discussions will focus on computing research challenges in data center efficiency and sustainability as well as computational issues related to reducing our carbon footprint.
- **Addressing climate change.** The sessions will examine how to develop greater understanding of climate dynamics through the evaluation of sensor network approaches, advanced data mining and visualization techniques, and improved computationally-driven climate models.
- **Transformational improvement in healthcare.** The availability and delivery of solutions for previously elusive healthcare challenges such as HIV-AIDS, diabetes, asthma, obesity, and aging are showing promising results through the collaboration of medical and computational researchers. Ubiquitous cell phone networks around the world are now opening the way to deliver a new level of health care in rural areas where no doctors are available.

### Faculty Summit 2009 Contents

- Overview
- [Agenda Day 1](#)
- [Agenda Day 2](#)
- [Speaker Biographies](#)
- [FAQ](#)

### Related Links

- [Faculty Summits at Microsoft Research](#)
- [Faculty Summit 2008 in Redmond](#)
- [Collaboration at Microsoft Research](#)

# 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



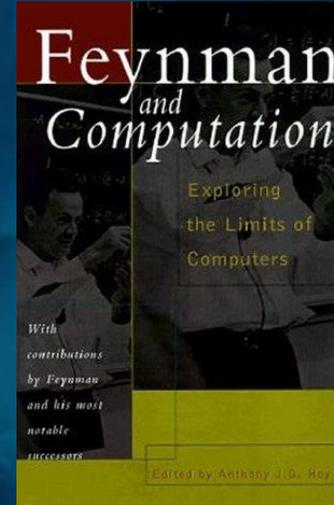
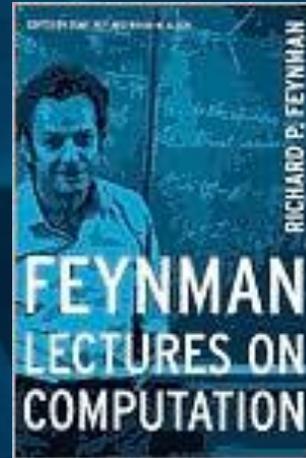
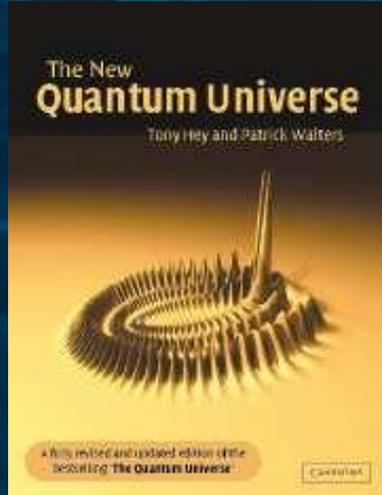
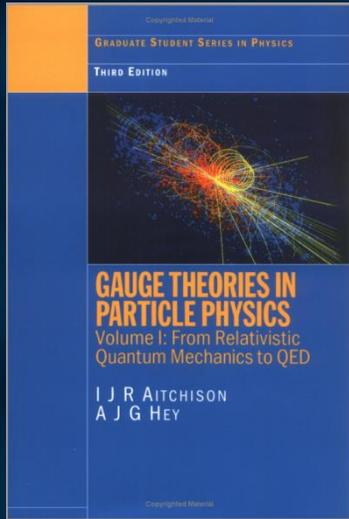
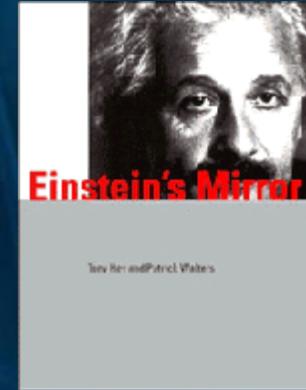
*Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster from the Digitized Sky Survey combined with an image of the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

# Tony Hey – An Introduction

Microsoft  
Research



Research Councils UK



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

Community and Geographic Outreach

Core Computer  
Science



Earth, Energy &  
Environment



Education &  
Scholarly  
Communication

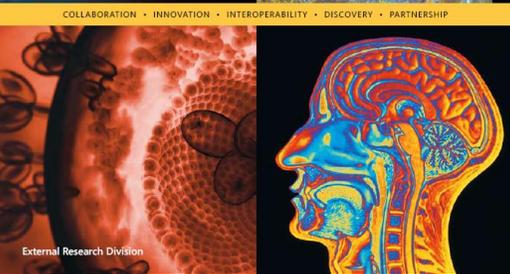
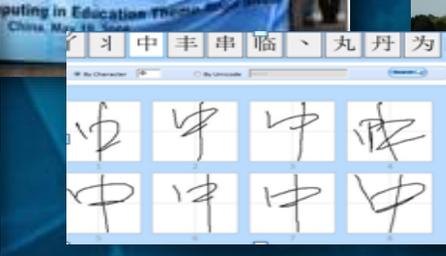
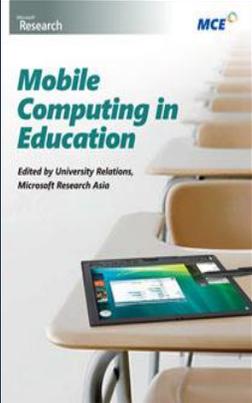


Health &  
Wellbeing



Advanced Research Tools and Services

# Since the Faculty Summit One Year Ago:



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

**Fast Facts**  
Project Principal: David Fletcher, assistant professor, University of California at Berkeley  
Key Words: CellScope, portable, low-cost, disease diagnosis

What started as a basic college class assignment spawned an invention that could help bring the benefits of modern microscopy to the developing world. Student researchers at UC Berkeley and their professor have developed a camera phone microscope that is powerful enough to diagnose diseases such as malaria and tuberculosis.

A few years ago, David Fletcher posed a challenge to the students in his Optics and Microscopy class at the University of California at Berkeley (UC Berkeley). They had three to eight weeks to design a camera phone microscope that was powerful enough to diagnose diseases such as malaria and tuberculosis. Fletcher asked to create the camera phone microscope out of mobile microscope that could be used to diagnose the disease.

While the students were working on their projects, Fletcher noticed the "microscope" team was having trouble. They had a camera phone that always came up with a black screen. He thought they had the camera, but there was no light source. They needed a light source. He thought they had the camera, but there was no light source. They needed a light source.

It became clear to Fletcher that something was about the phone and not the microscope. He started thinking about how to use the phone's camera to take pictures of the microscope. He started thinking about how to use the phone's camera to take pictures of the microscope.



### 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@P, a set of Web-based videoconferencing tools.**

For computer science researchers in Latin America 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 fund 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-CHILE) in Santiago. Over the past decade, he led 90 percent of the computer science projects involving Latin American researchers in another continent.

In North America and European parts of the Latin American and Caribbean, however, such projects are abundant and opportunities to Latin American researchers are plentiful.

Other academic researchers are the region using Conference@P.



### Beyond the Data Deluge

**COMPUTER SCIENCE**

Since at least Newton's time of motion in the 17th century, scientists have recognized the need for powerful tools for understanding nature. In recent decades, computer simulations have become an essential tool for scientists to explore domains that are inaccessible to theory and experiments, 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 humongous data deluges. Increases in data volume from satellites, telescopes, high-throughput instruments, sensor networks, accelerators, and supercomputers, compared to the volume generated only a decade ago (2), in astronomy and particle physics, these are experiments generate petabytes (1 petabyte = 10<sup>15</sup> bytes) of data per year. In bioinformatics, the human genome is 3 Gb.

Other research fields also face major data management challenges. In almost every laboratory, "born digital" data proliferate in files, spreadsheets, 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 transitions on silicon chips to get smaller and processes to get faster. At the same time, technology improvements for disks for storage cannot keep up with the ever increasing flood of scientific data generated by the faster computers. In astronomy research, however, clusters—groups of mutually identical, increasing PC computers that can be used for parallel computation—have



### 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@P, a set of Web-based videoconferencing tools.**

For computer science researchers in Latin America 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 fund 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-CHILE) in Santiago. Over the past decade, he led 90 percent of the computer science projects involving Latin American researchers in another continent.

In North America and European parts of the Latin American and Caribbean, however, such projects are abundant and opportunities to Latin American researchers are plentiful.

Other academic researchers are the region using Conference@P.

### Microsoft External Research

The Microsoft External Research Division within Microsoft Research works with academic, government and industry to advance computer science and technology research aimed at helping address some of the world's most urgent and significant social and technological challenges. Along with investing cash, software, hardware and research expertise to enable ground-breaking projects worldwide, Microsoft External Research is committed to providing the resources, training and support needed to support every stage of the research process. Microsoft External Research is a research organization that explores new tools and methods that empower solutions and students and develops software that supports the entire scholarly research and communication process.

Microsoft External Research  
<http://research.microsoft.com/en-us/div/external/>

### Microsoft External Research

**Fast Facts**  
Project Principal: Sergio Chelico, assistant professor, Computer Science Department, Universidad de Chile  
Ignacio Cases, assistant professor, Computer Science Department, PUC-CHILE  
Claudia Linares, project manager, Computer Science Department

**Web Sites:**  
[www.testing.microsoft.com/latam/en/projects/conference@p/](http://www.testing.microsoft.com/latam/en/projects/conference@p/)

**Profile:**  
To help encourage greater collaboration among computer science researchers throughout the region, Microsoft American and Caribbean universities are helping Conference@P adoption. Meeting software as a forum for file discussion and information sharing.



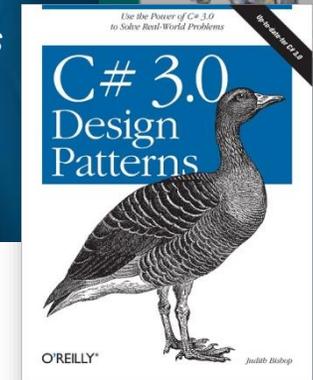
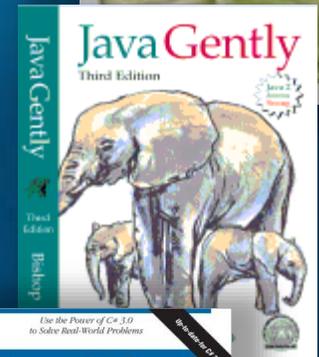
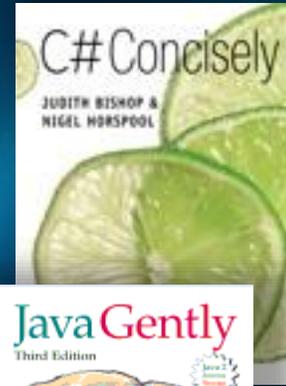
# Introducing: Prof. Judith Bishop – Core Computer Science theme lead

*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/)



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

# 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*

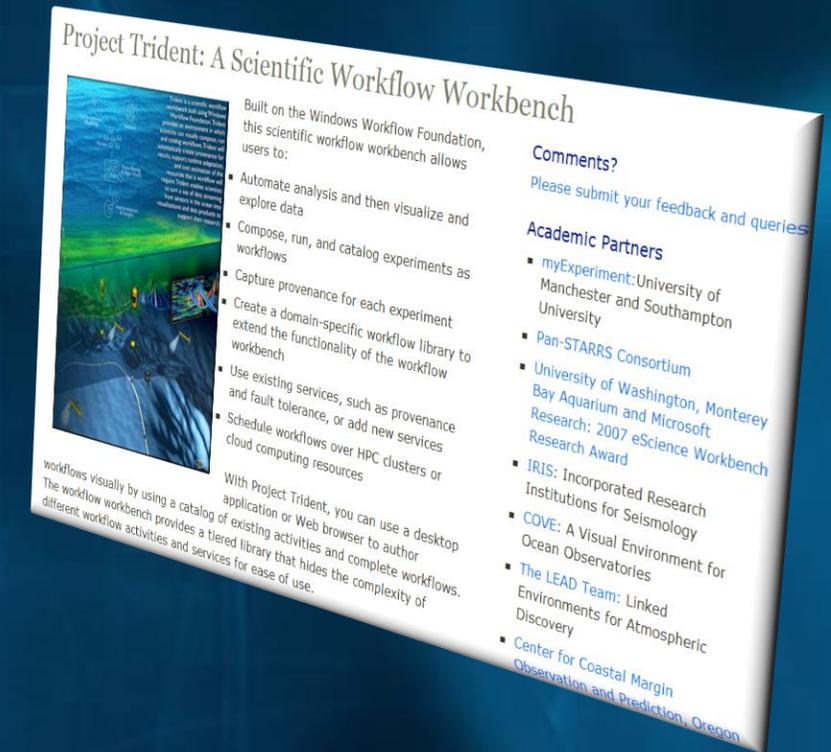
Project Trident: A Scientific  
Workflow Workbench

Dryad and DryadLINQ

# Project Trident: A Scientific Workflow Workbench

## Accelerating the pace of discovery

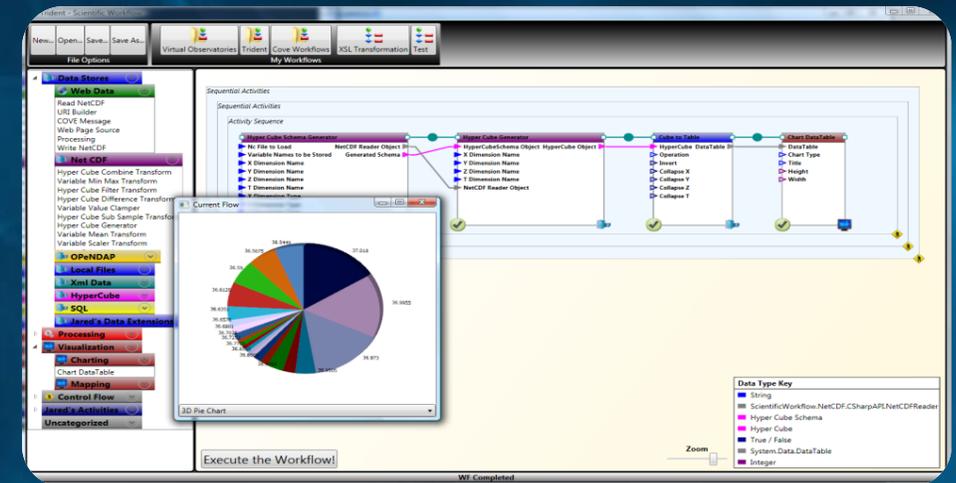
- Makes it easier for scientists to ingest and make sense of data
- Get answers to questions at a rate not previously possible
- Capture provenance
- Scientists in data-intensive fields such as oceanography, astronomy, environmental science and medical research can use these tools to manage, integrate 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

*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



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;

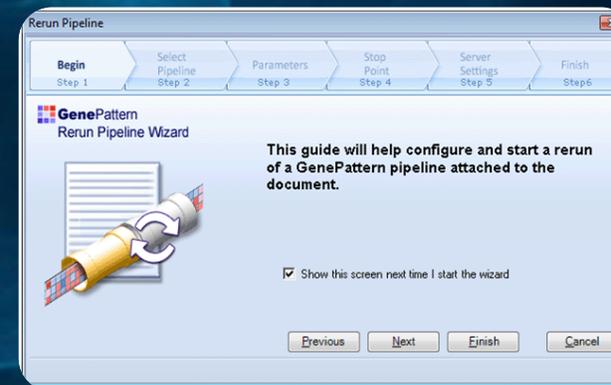
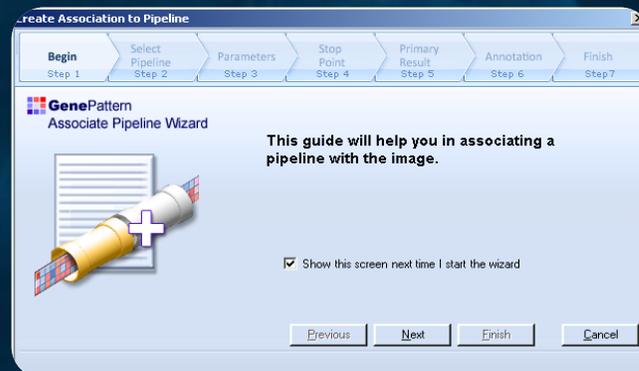
A person reading the document can click on such an image or icon to view the associated pipelines and its input files, and rerun the pipelines on a Trident server while remaining in the Word application.



Associate Pipeline



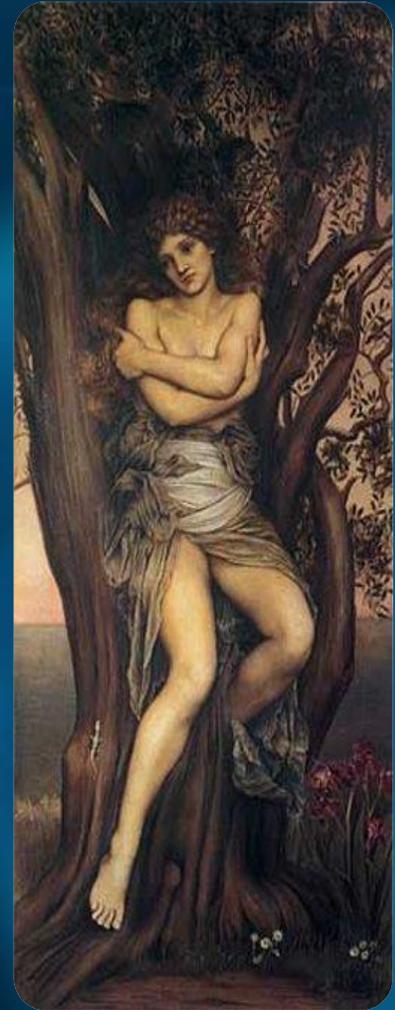
Rerun Pipeline





# DryadLINQ 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
  - DryadLINQ 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 DryadLINQ, 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 DryadLINQ 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

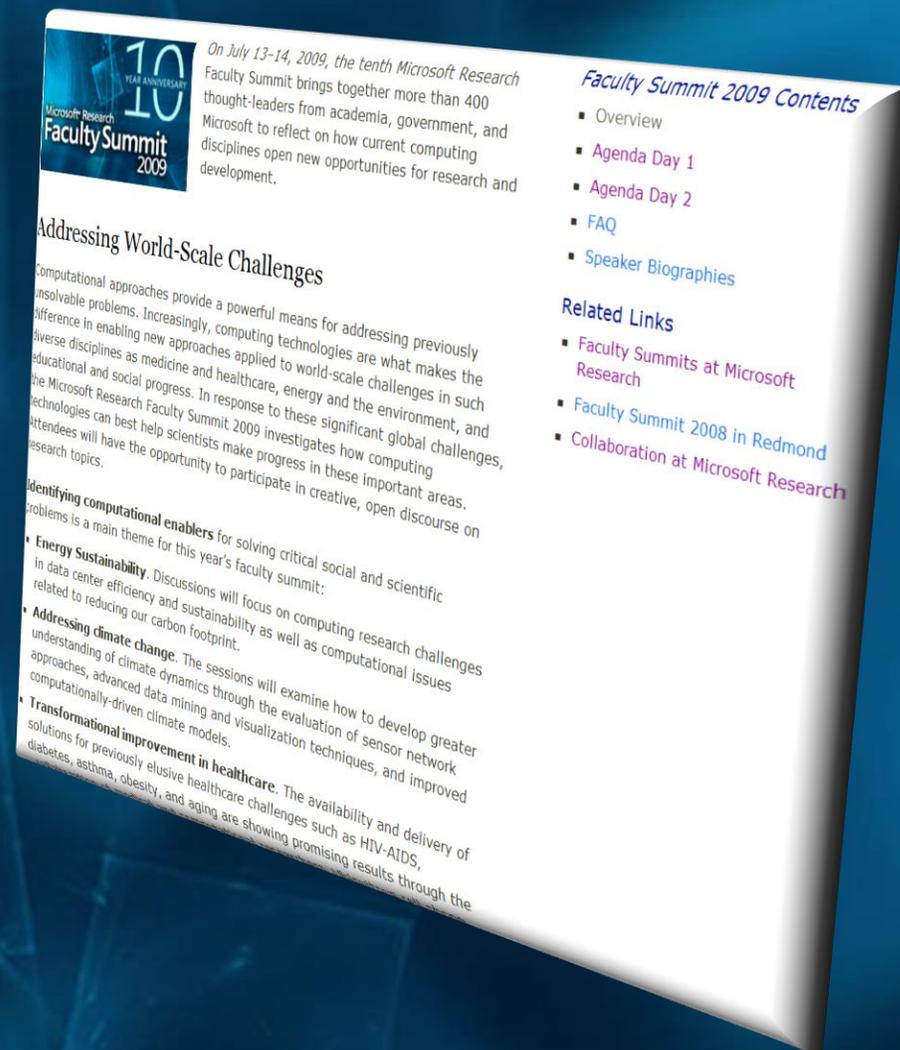
<http://research.microsoft.com/en-us/events/fs2009>

- Agenda

- On-line

- Printed

- Please let us know if you have any questions or need any help



**Microsoft<sup>®</sup>**

*Your potential. Our passion.<sup>™</sup>*