# Science Informatics

Dan Fay
Director – Earth, Energy and Environment
dan.fay@microsoft.com

Microsoft®
Research Connections

# MSR eScience Workshop
# Looking Back 8 yrs to the Beginning

Scientific Data Intensive Computing Workshop 2004

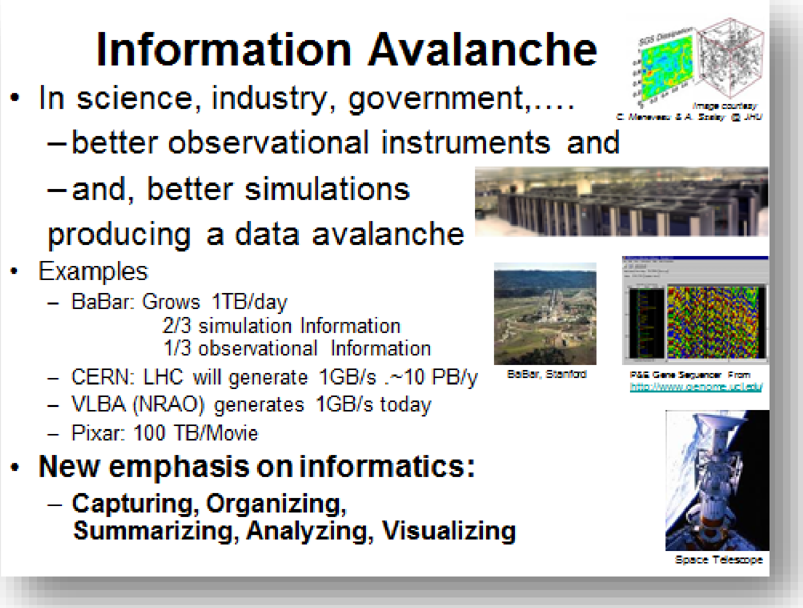
- Keynote: *20 Questions to a Better Application* – Jim Gray

  ***Online Science the New Computational Science***

- Talk: *Data Explosion: Astrophysics with Terabytes of Data*

  - Alex Szalay

# Online Science the New Computational Science



## Information Avalanche

- In science, industry, government,….
  - better observational instruments and
  - and, better simulations

  producing a data avalanche
- Examples
  - BaBar: Grows 1TB/day

    2/3 simulation Information

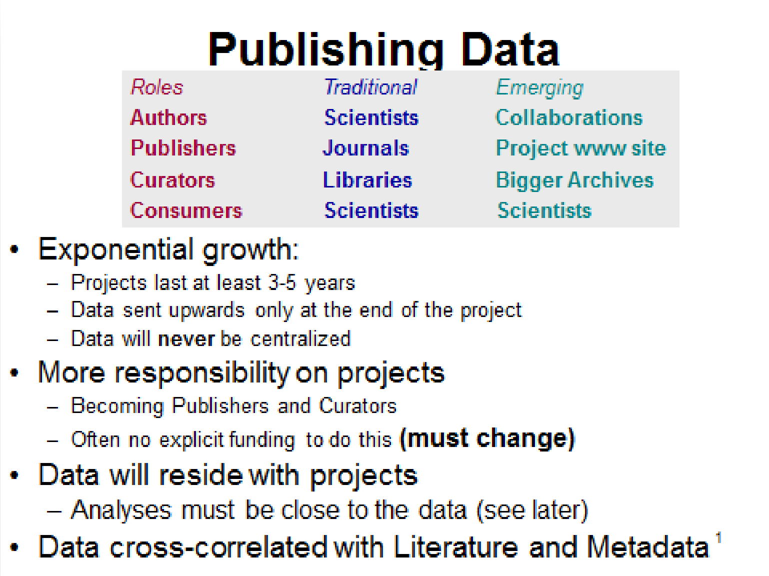
    1/3 observational Information
  - CERN: LHC will generate 1GB/s .~10 PB/y
  - VLBA (NRAO) generates 1GB/s today
  - Pixar: 100 TB/Movie
- **New emphasis on informatics:**
  - **Capturing, Organizing, Summarizing, Analyzing, Visualizing**

Image courtesy C. Meneveau & A. Szalay @ JHU

BaBar, Stanford

P&G Gene Sequencer From http://www.genome.uci.edu/

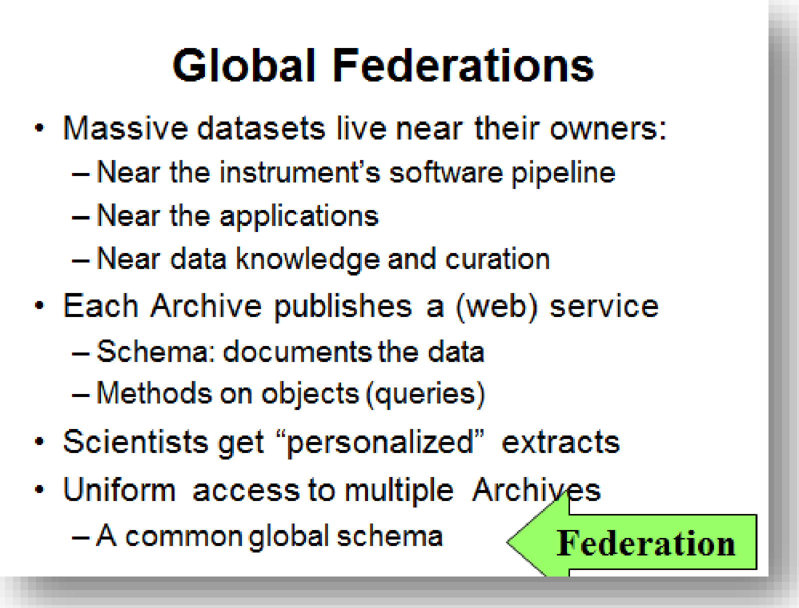
Space Telescope



## Publishing Data

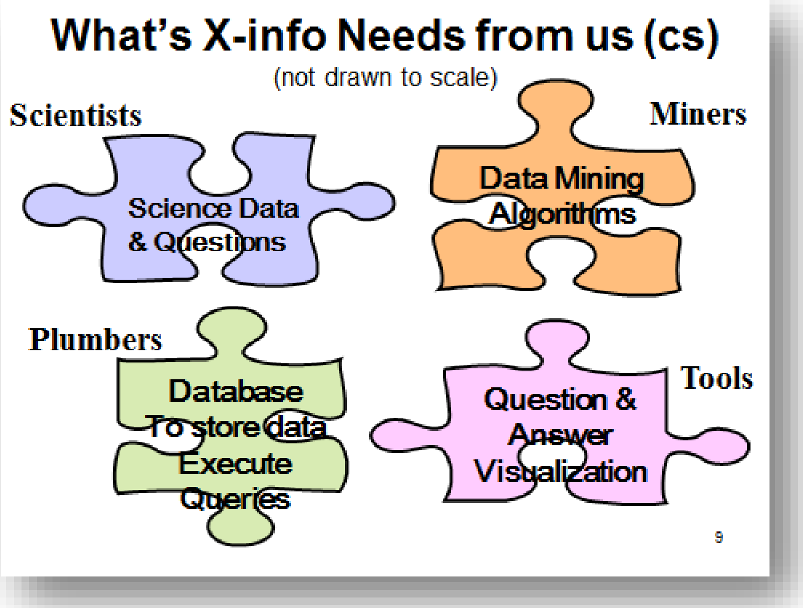
| *Roles* | *Traditional* | *Emerging* |
|---|---|---|
| Authors | Scientists | Collaborations |
| Publishers | Journals | Project www site |
| Curators | Libraries | Bigger Archives |
| Consumers | Scientists | Scientists |

- Exponential growth:
  - Projects last at least 3-5 years
  - Data sent upwards only at the end of the project
  - Data will **never** be centralized
- More responsibility on projects
  - Becoming Publishers and Curators
  - Often no explicit funding to do this **(must change)**
- Data will reside with projects
  - Analyses must be close to the data (see later)
- Data cross-correlated with Literature and Metadata [1]



## Global Federations

- Massive datasets live near their owners:
  - Near the instrument's software pipeline
  - Near the applications
  - Near data knowledge and curation
- Each Archive publishes a (web) service
  - Schema: documents the data
  - Methods on objects (queries)
- Scientists get "personalized" extracts
- Uniform access to multiple Archives
  - A common global schema



## What's X-info Needs from us (cs)

(not drawn to scale)

Scientists

Science Data & Questions

Miners

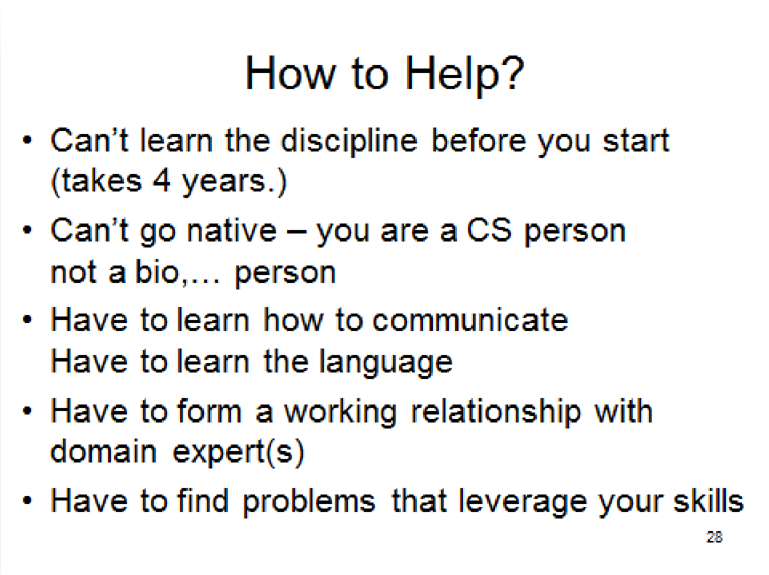
Data Mining Algorithms

Plumbers

Database To store data Execute Queries

Tools

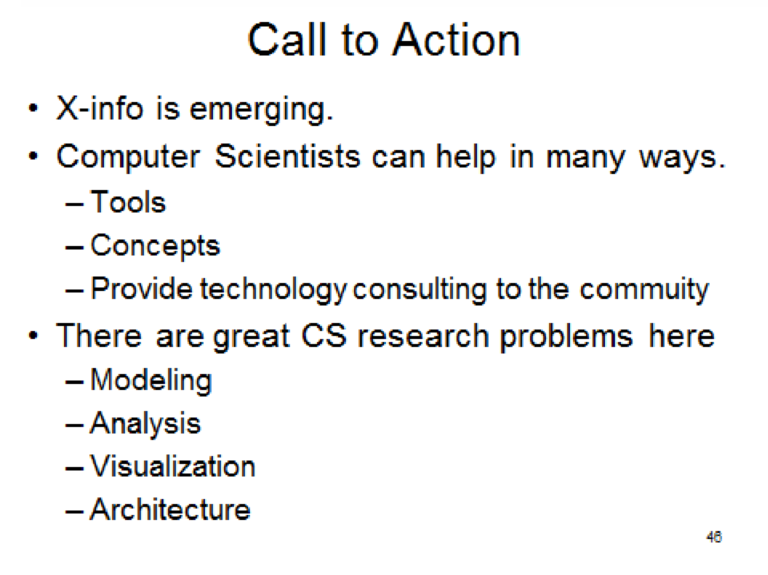
Question & Answer Visualization



## How to Help?

- Can't learn the discipline before you start (takes 4 years.)
- Can't go native – you are a CS person not a bio,… person
- Have to learn how to communicate

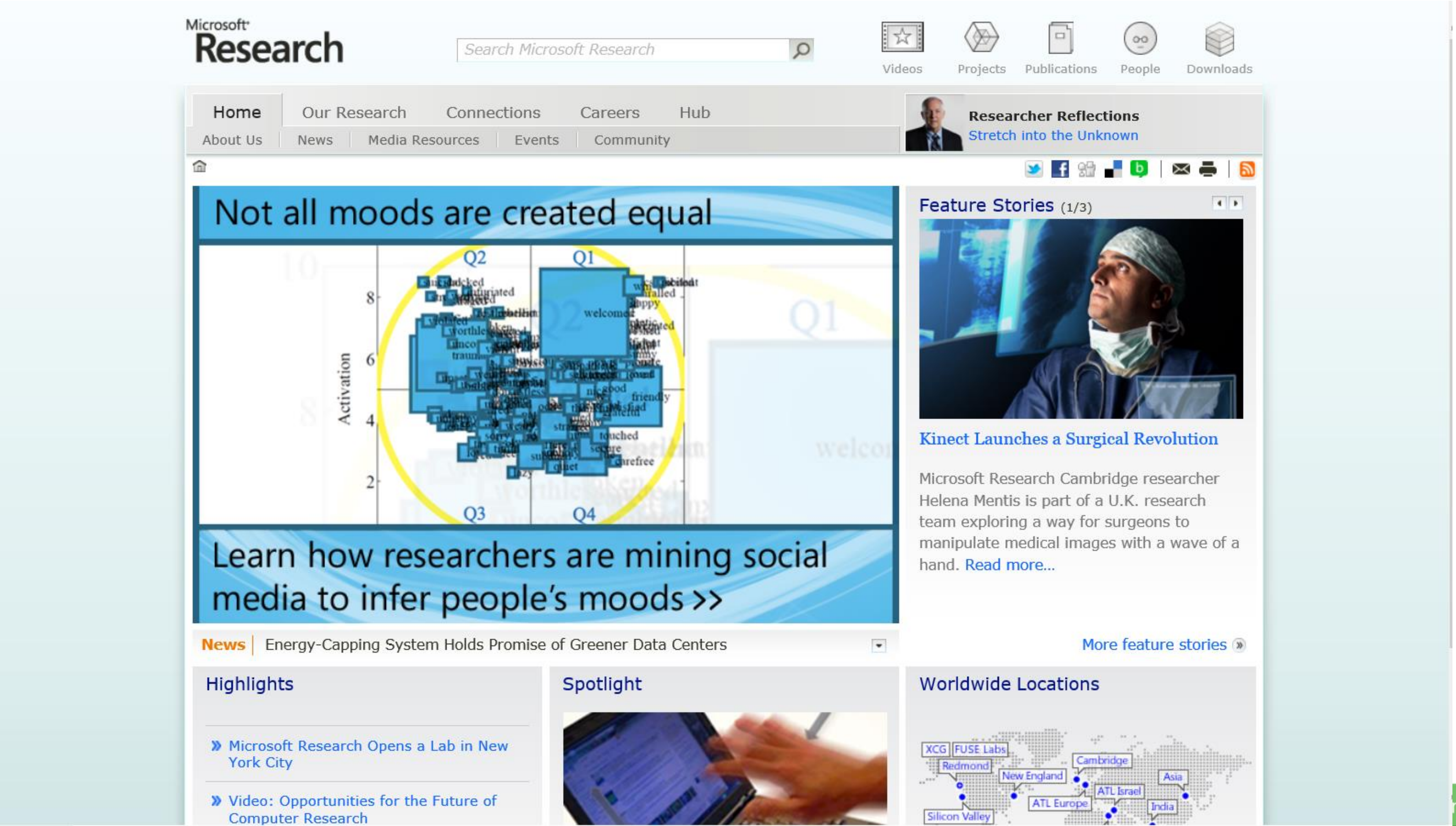
  Have to learn the language
- Have to form a working relationship with domain expert(s)
- Have to find problems that leverage your skills



## Call to Action

- X-info is emerging.
- Computer Scientists can help in many ways.
  - Tools
  - Concepts
  - Provide technology consulting to the commuity
- There are great CS research problems here
  - Modeling
  - Analysis
  - Visualization
  - Architecture

# Microsoft Research



Researcher Reflections

Stretch into the Unknown

Not all moods are created equal

Learn how researchers are mining social media to infer people's moods >>

Feature Stories (1/3)

Kinect Launches a Surgical Revolution

Microsoft Research Cambridge researcher Helena Mentis is part of a U.K. research team exploring a way for surgeons to manipulate medical images with a wave of a hand. Read more...

News | Energy-Capping System Holds Promise of Greener Data Centers

More feature stories »

Highlights

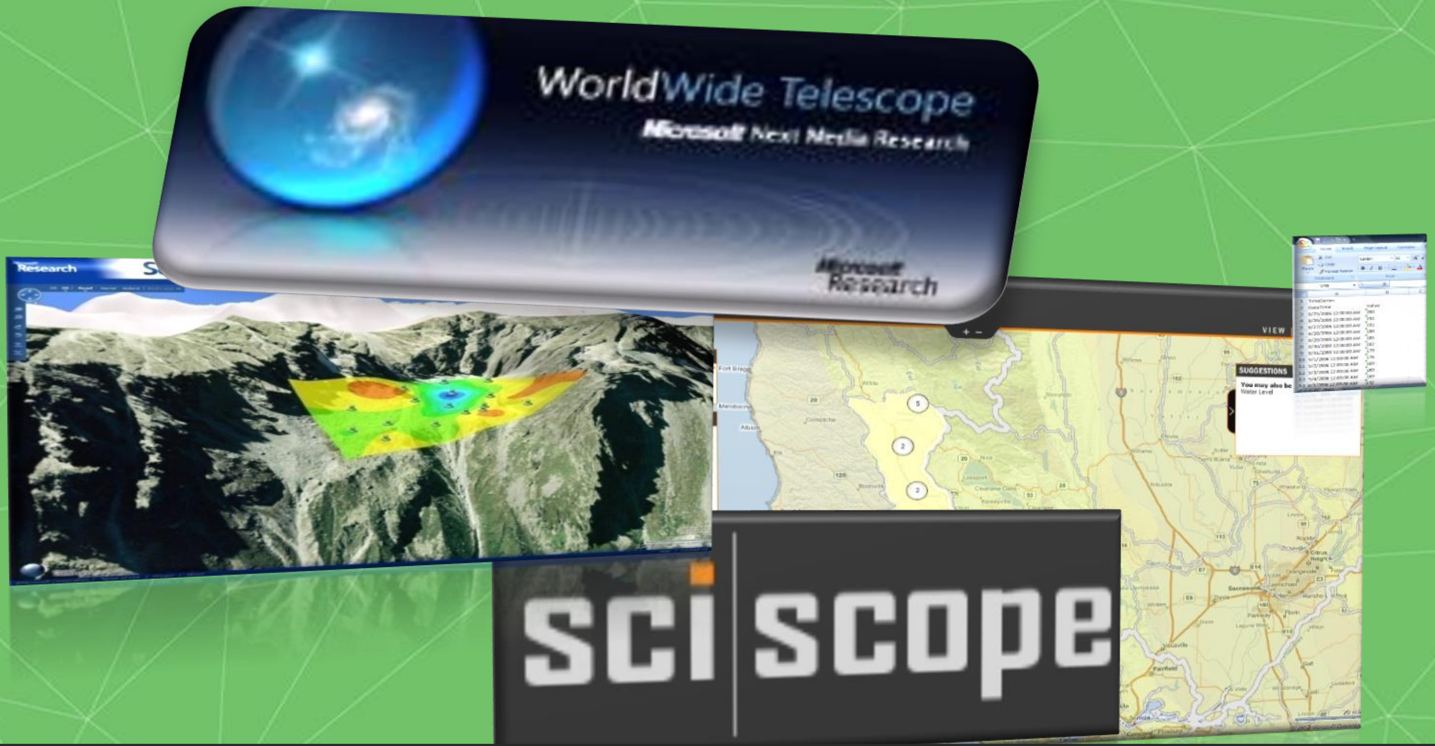
» Microsoft Research Opens a Lab in New York City

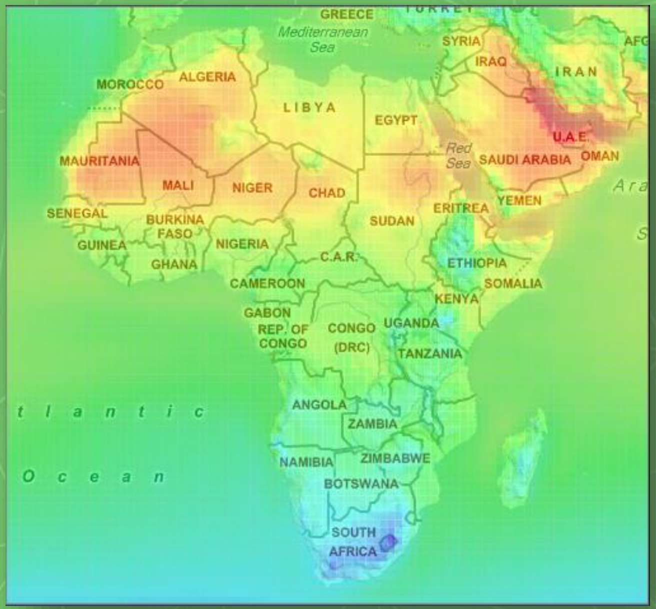
» Video: Opportunities for the Future of Computer Research

Spotlight

Worldwide Locations

# Earth, Energy & Environment





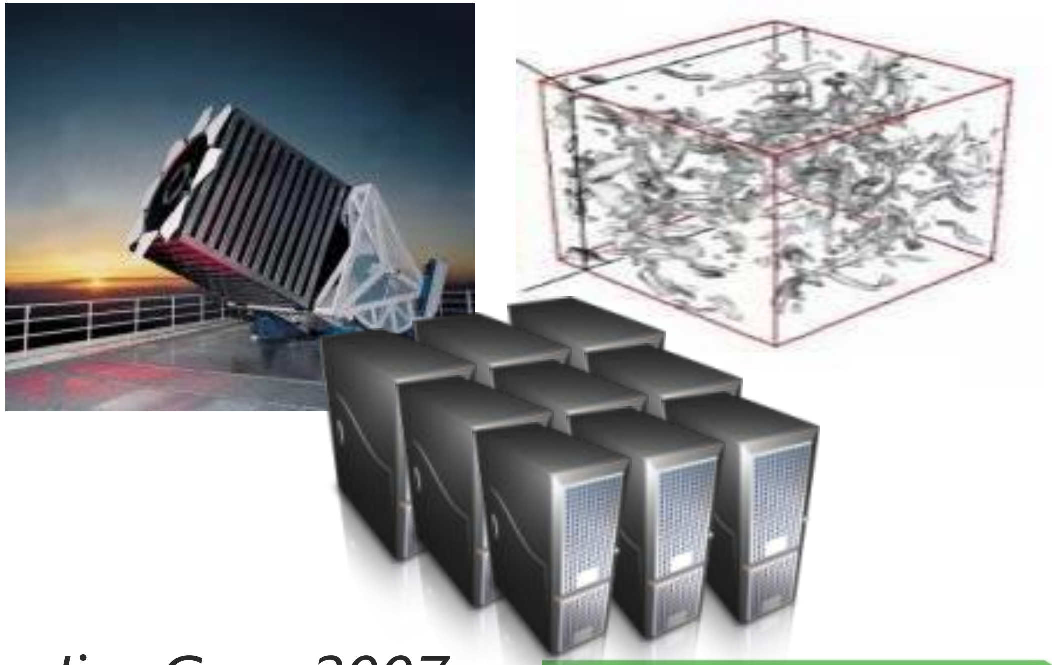
- **Visualizing and Experiencing $E^3$ Data + Information:** Provide a unique experience to reduce time to insight and knowledge through visualizing data and information
- **Accessible Data:** Ensure $E^3$ data (remote and local sensing) is easily discoverable, accessible and consumable in the scientists domain
- **Enabling Scientific Collaboration:** Look at new ways to enable collaboration in scientific virtual organizations

# Emergence of a 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**
- 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

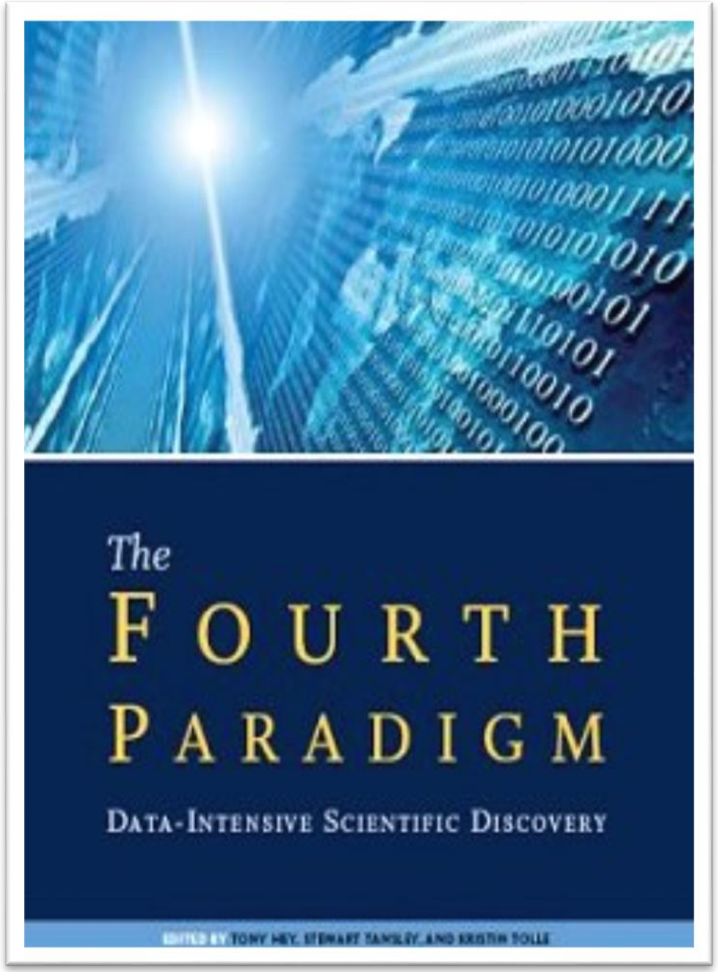
$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



*Jim Gray 2007*

# Changing Nature of Discovery

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



The FOURTH PARADIGM

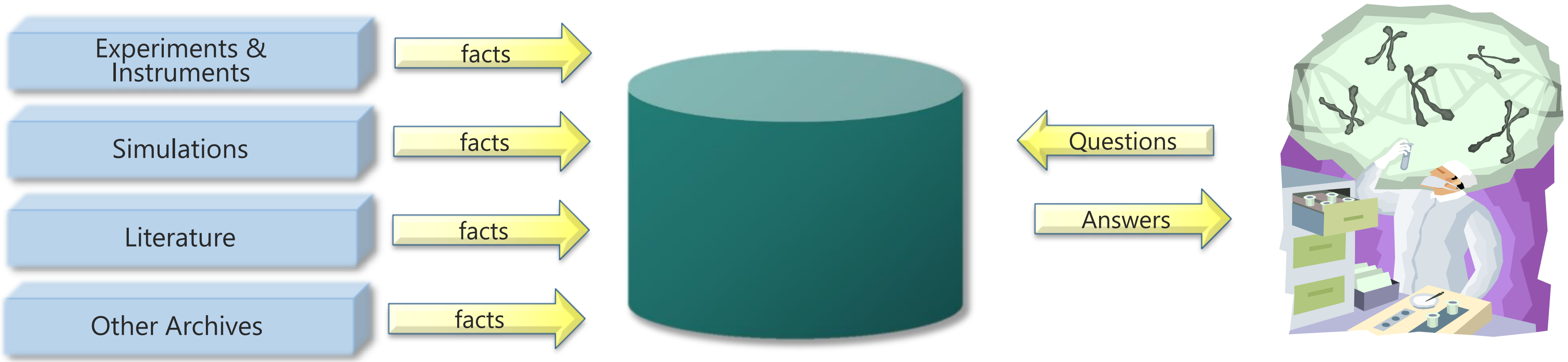
DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

http://fourthparadigm.org

# The Problem for the e-Scientist

How to codify and represent our knowledge



Experiments & Instruments

Simulations

Literature

Other Archives

facts

facts

facts

facts

Questions

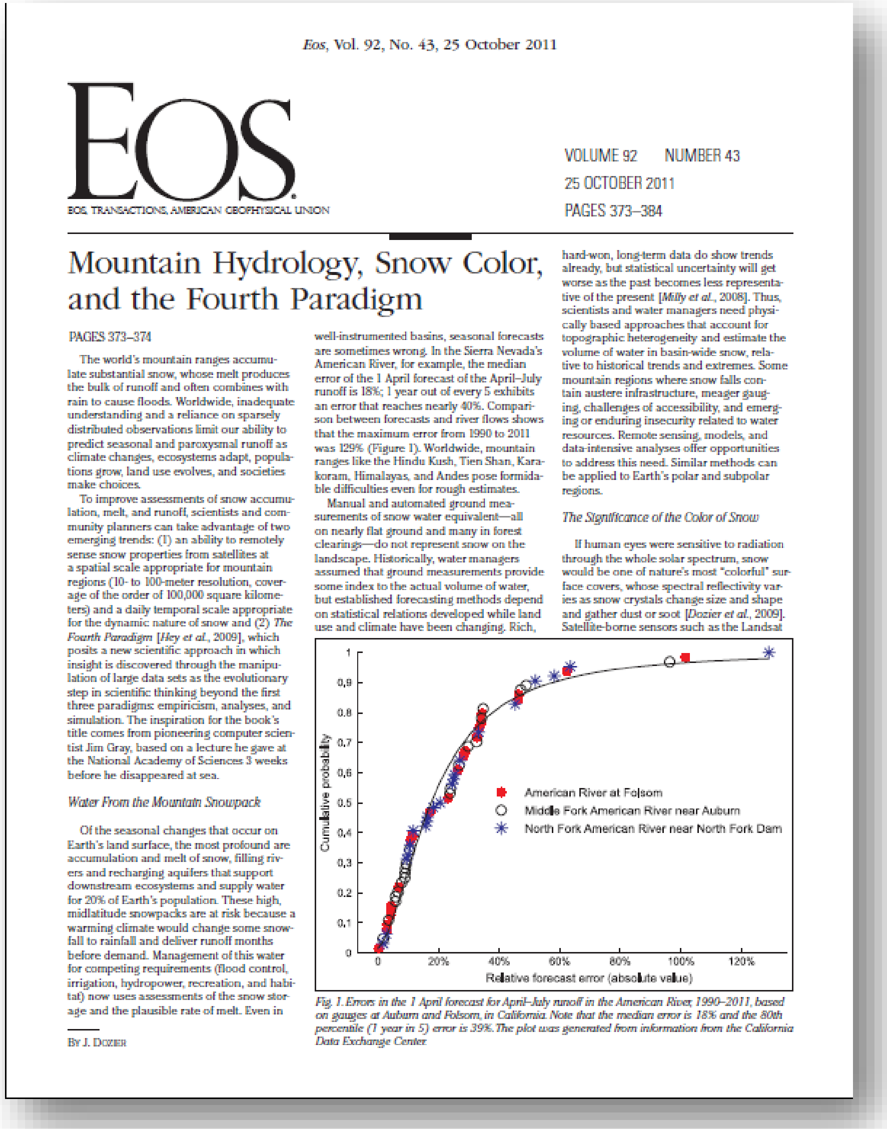
Answers

## The Generic Problems

- Data ingest
- Managing a petabyte
- Common schema
- How to organize it
- How to *re*organize it
- How to share with others
- Query and Vis tools
- Building and executing models
- Integrating data and Literature
- Documenting experiments
- Curation and long-term preservation

*(With thanks to Jim Gray)*

# EOS Article: *Mountain Hydrology, Snow Color, and the Fourth Paradigm* by Jeff Dozier



*Eos*, Vol. 92, No. 43, 25 October 2011

EOS

EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 92 NUMBER 43
25 OCTOBER 2011
PAGES 373–384

## Mountain Hydrology, Snow Color, and the Fourth Paradigm

PAGES 373–374

The world's mountain ranges accumulate substantial snow, whose melt produces the bulk of runoff and often combines with rain to cause floods. Worldwide, inadequate understanding and a reliance on sparsely distributed observations limit our ability to predict seasonal and paroxysmal runoff as climate changes, ecosystems adapt, populations grow, land use evolves, and societies make choices.

To improve assessments of snow accumulation, melt, and runoff, scientists and community planners can take advantage of two emerging trends: (1) an ability to remotely sense snow properties from satellites at a spatial scale appropriate for mountain regions (10- to 100-meter resolution, coverage of the order of 100,000 square kilometers) and a daily temporal scale appropriate for the dynamic nature of snow and (2) *The Fourth Paradigm* [*Hey et al.*, 2009], which posits a new scientific approach in which insight is discovered through the manipulation of large data sets as the evolutionary step in scientific thinking beyond the first three paradigms: empiricism, analyses, and simulation. The inspiration for the book's title comes from pioneering computer scientist Jim Gray, based on a lecture he gave at the National Academy of Sciences 3 weeks before he disappeared at sea.

### *Water From the Mountain Snowpack*

Of the seasonal changes that occur on Earth's land surface, the most profound are accumulation and melt of snow, filling rivers and recharging aquifers that support downstream ecosystems and supply water for 20% of Earth's population. These high, midlatitude snowpacks are at risk because a warming climate would change some snowfall to rainfall and deliver runoff months before demand. Management of this water for competing requirements (flood control, irrigation, hydropower, recreation, and habitat) now uses assessments of the snow storage and the plausible rate of melt. Even in well-instrumented basins, seasonal forecasts are sometimes wrong. In the Sierra Nevada's American River, for example, the median error of the 1 April forecast of the April–July runoff is 18%; 1 year out of every 5 exhibits an error that reaches nearly 40%. Comparison between forecasts and river flows shows that the maximum error from 1990 to 2011 was 129% (Figure 1). Worldwide, mountain ranges like the Hindu Kush, Tien Shan, Karakoram, Himalayas, and Andes pose formidable difficulties even for rough estimates.

Manual and automated ground measurements of snow water equivalent—all on nearly flat ground and many in forest clearings—do not represent snow on the landscape. Historically, water managers assumed that ground measurements provide some index to the actual volume of water, but established forecasting methods depend on statistical relations developed while land use and climate have been changing. Rich, hard-won, long-term data do show trends already, but statistical uncertainty will get worse as the past becomes less representative of the present [*Milly et al.*, 2008]. Thus, scientists and water managers need physically based approaches that account for topographic heterogeneity and estimate the volume of water in basin-wide snow, relative to historical trends and extremes. Some mountain regions where snow falls contain austere infrastructure, meager gauging, challenges of accessibility, and emerging or enduring insecurity related to water resources. Remote sensing, models, and data-intensive analyses offer opportunities to address this need. Similar methods can be applied to Earth's polar and subpolar regions.

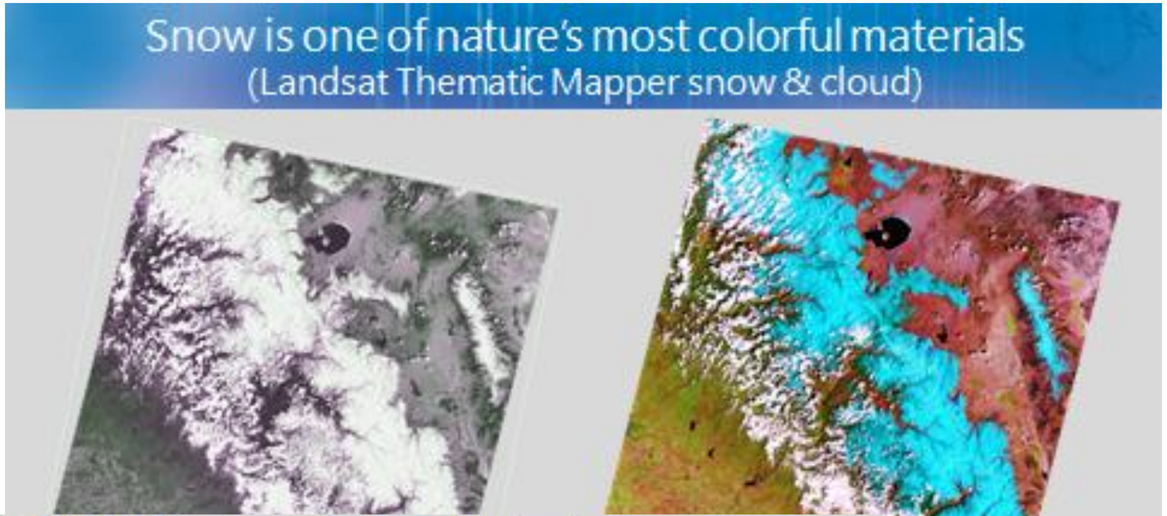
### *The Significance of the Color of Snow*

If human eyes were sensitive to radiation through the whole solar spectrum, snow would be one of nature's most "colorful" surface covers, whose spectral reflectivity varies as snow crystals change size and shape and gather dust or soot [*Dozier et al.*, 2009]. Satellite-borne sensors such as the Landsat

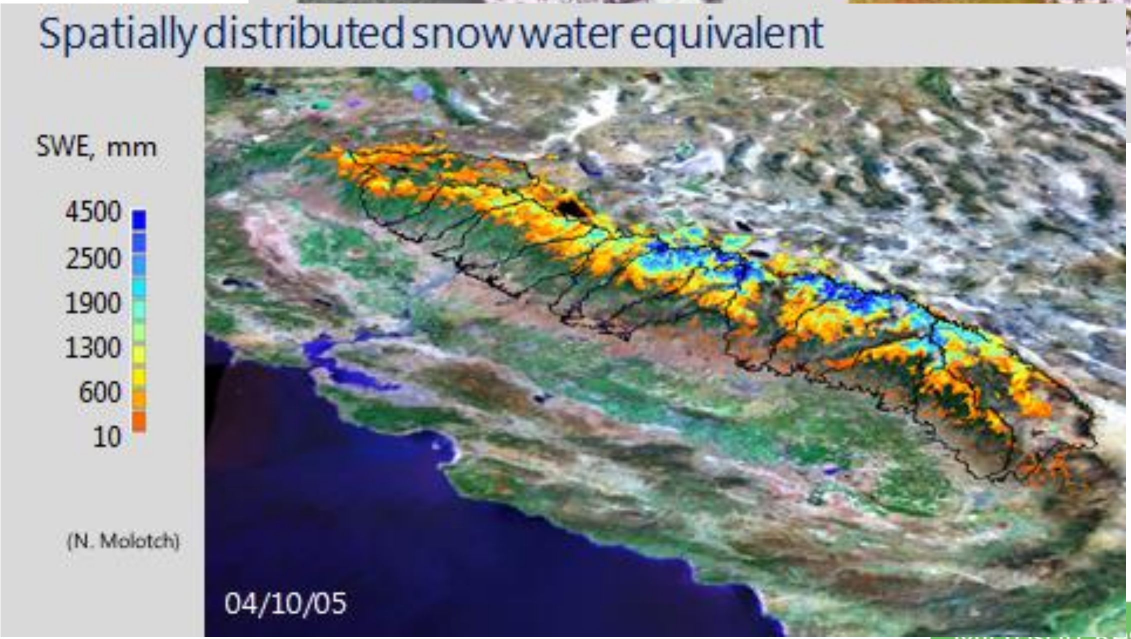
By J. Dozier

1
0,9
0.8
0,7
0,6
0.5
0,4
0,3
0.2
0,1
0
Cumulative probability
0
20%
40%
60%
80%
100%
120%
Relative forecast error (absolute value)
■ American River at Folsom
○ Middle Fork American River near Auburn
✱ North Fork American River near North Fork Dam

*Fig. 1. Errors in the 1 April forecast for April–July runoff in the American River, 1990–2011, based on gauges at Auburn and Folsom, in California. Note that the median error is 18% and the 80th percentile (1 year in 5) error is 39%. The plot was generated from information from the California Data Exchange Center.*



Snow is one of nature's most colorful materials
(Landsat Thematic Mapper snow & cloud)



Spatially distributed snow water equivalent

(N. Molotch)

04/10/05

# Information about water is more useful as we climb the value ladder

Data >>> Information >>> Insight

>>> Increasing value >>>

Forecasting

Reporting

Analysis

Integration

**Done poorly, but a few notable counter-examples**

Distribution

Aggregation

**Done poorly to moderately, not easy to find**

Quality assurance

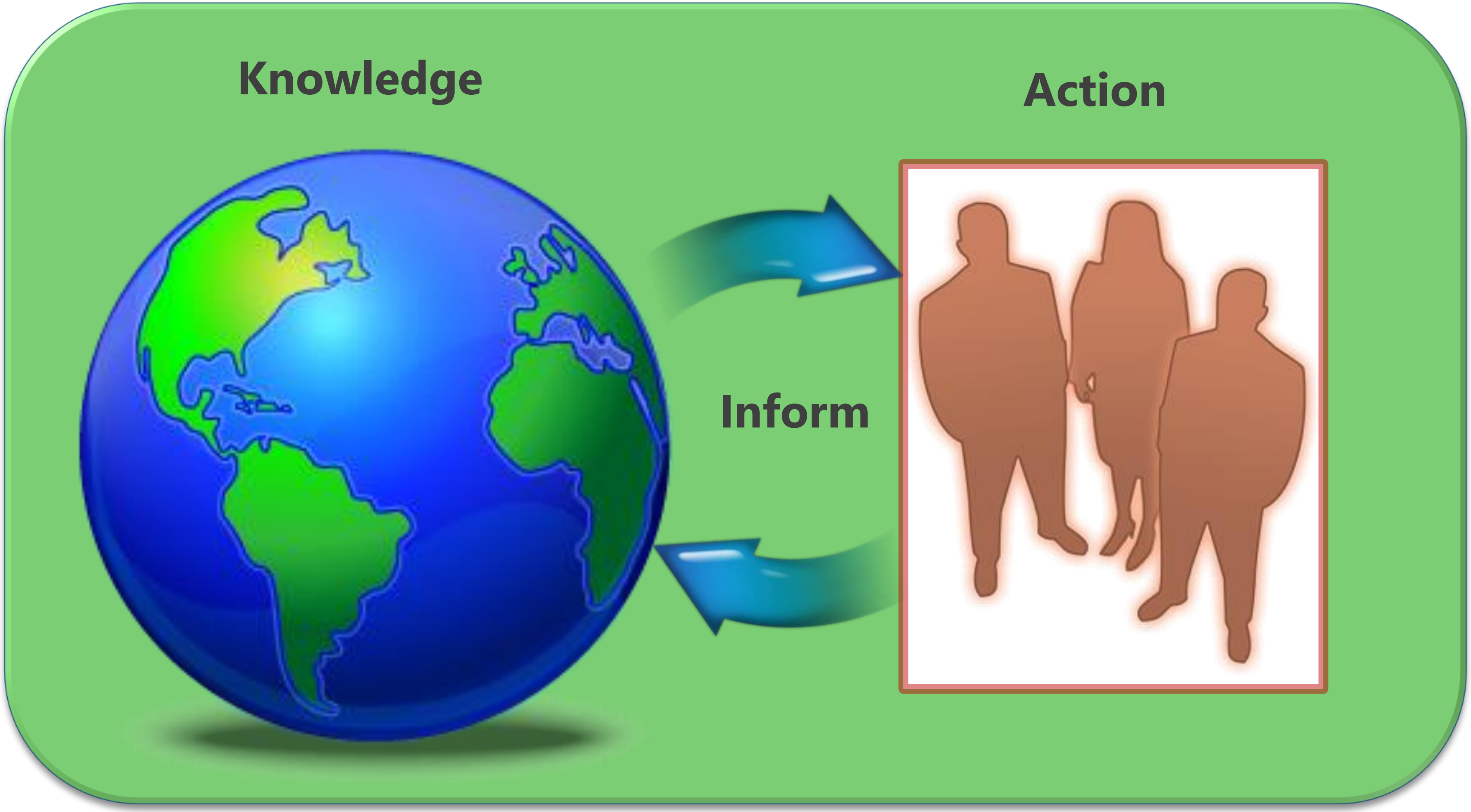
Collation

Monitoring

**Sometimes done well, generally discoverable and available, but could be improved**

(I. Zaslavsky & CSIRO, BOM, WMO)

# Environmental Ecosystem

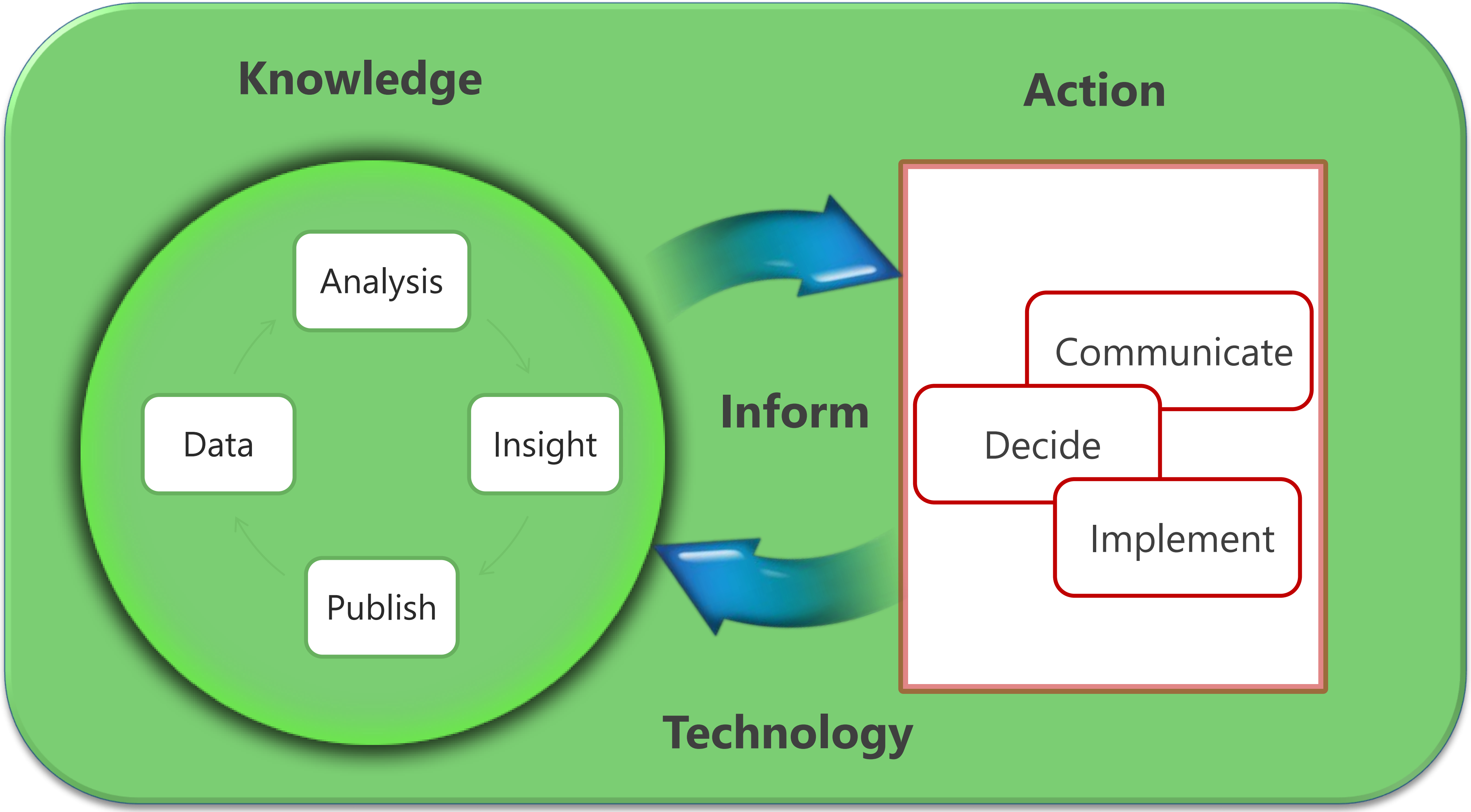


Knowledge

Action

Inform

# Environmental Ecosystem



Knowledge

Action

Analysis

Data

Insight

Publish

Inform

Communicate

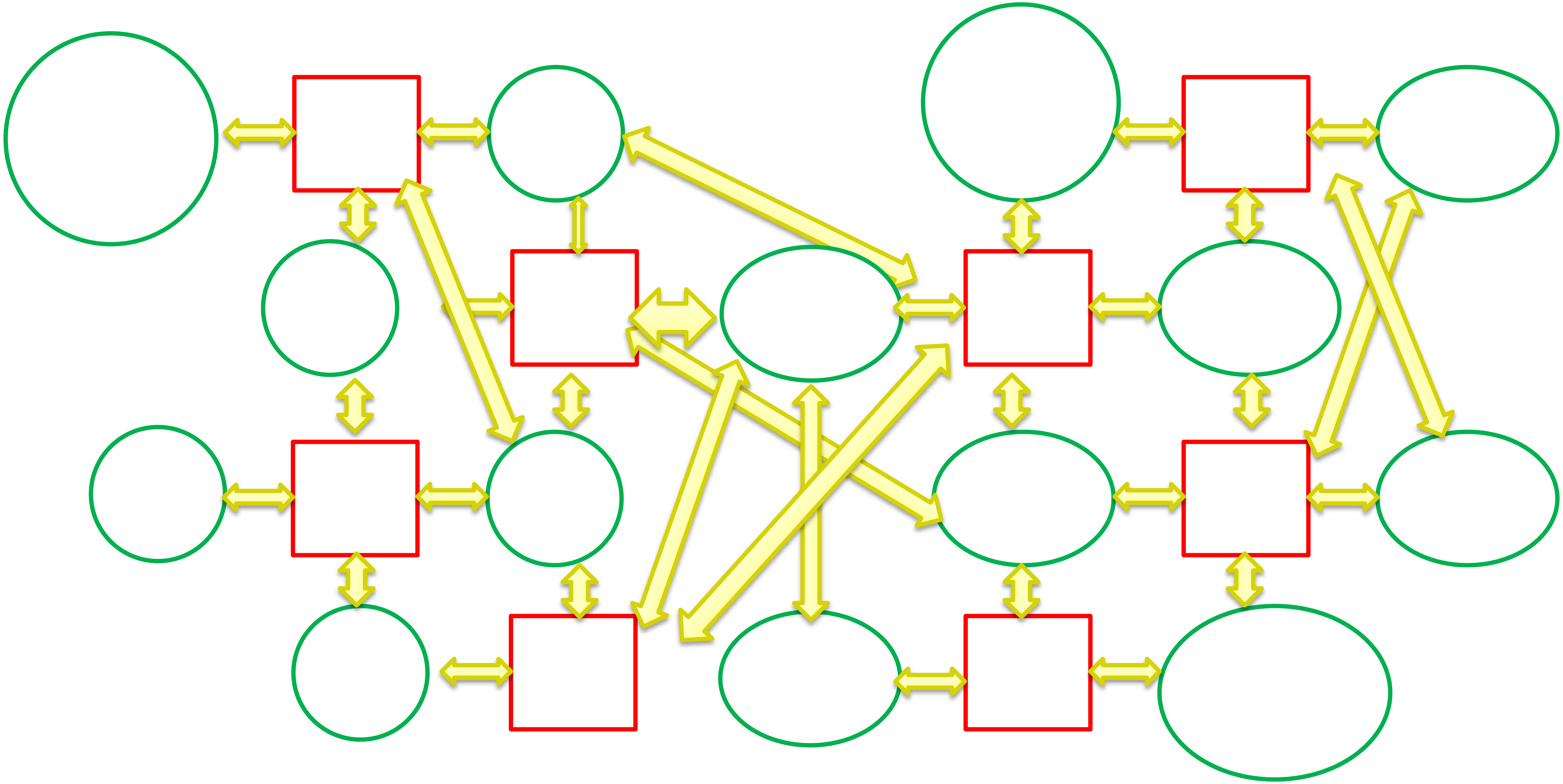
Decide

Implement

Technology

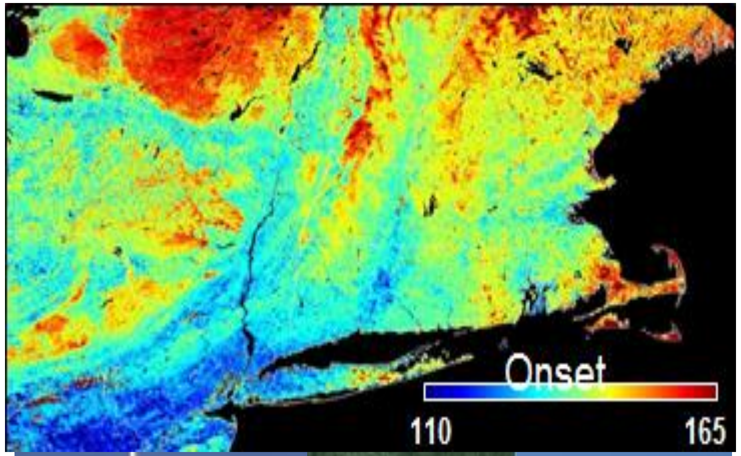
# Information ecosystem:

It is chaotic, unstructured and ad hoc



# The Ecological Data Flood

- We’re living in a perfect storm of remote sensing, cheap ground-based sensors, internet data access, and commodity computing
- Yet deriving and extracting the variables needed for science remains problematic
  - Specialized knowledge for algorithms, internal file formats, data cleaning, etc, etc
  - Finding the right needle across the distributed heterogeneous and very rap growing haystacks



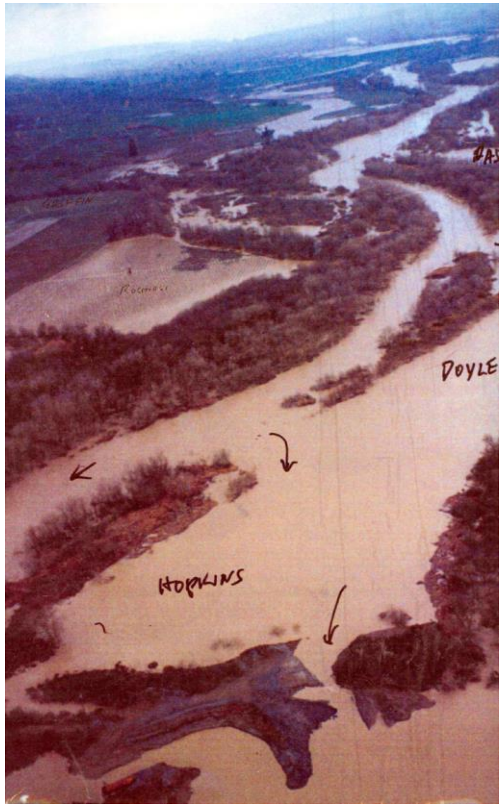
Science Center
Data server
AC power
The US
Long Term Ecological Research
Network
NOAA Satellite and Information Service
National Climatic Data Center
World's Largest Archive of Climate Data
National Climatic Data Center
Weather/Climate, Events, Information & Assessments
Paleoclimatology
Disclaimer
National Weather Service
Hydrologic Information Center
USGS
FIRSTGOV

# Data Variety – The Spice of Life

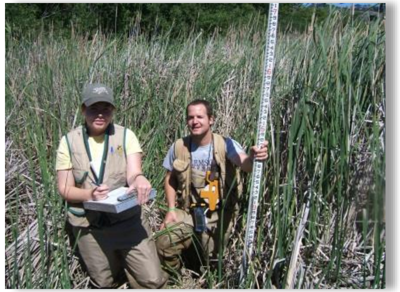
Manual Measurement

Automated Measurement

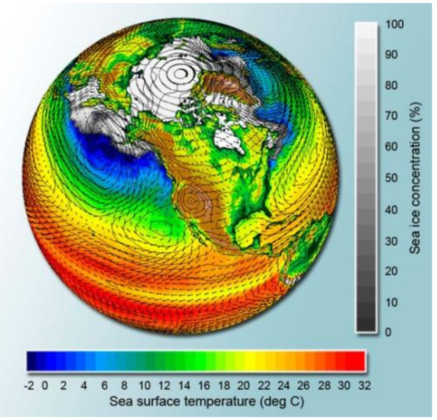
Sample Collection



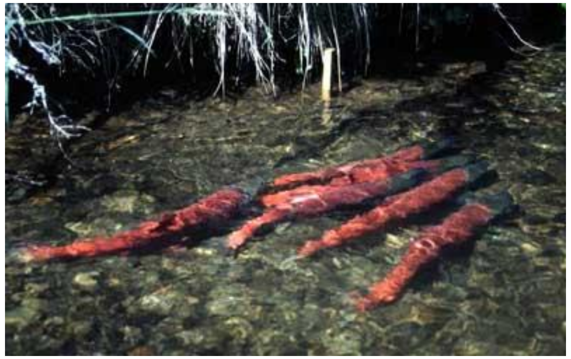
Historical Photographs



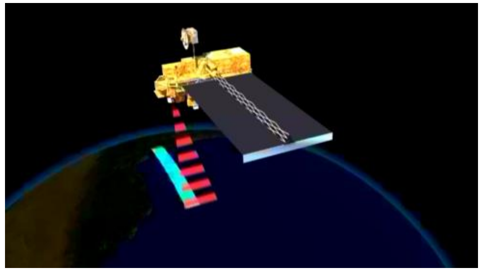
Typing



Model Output



Counting



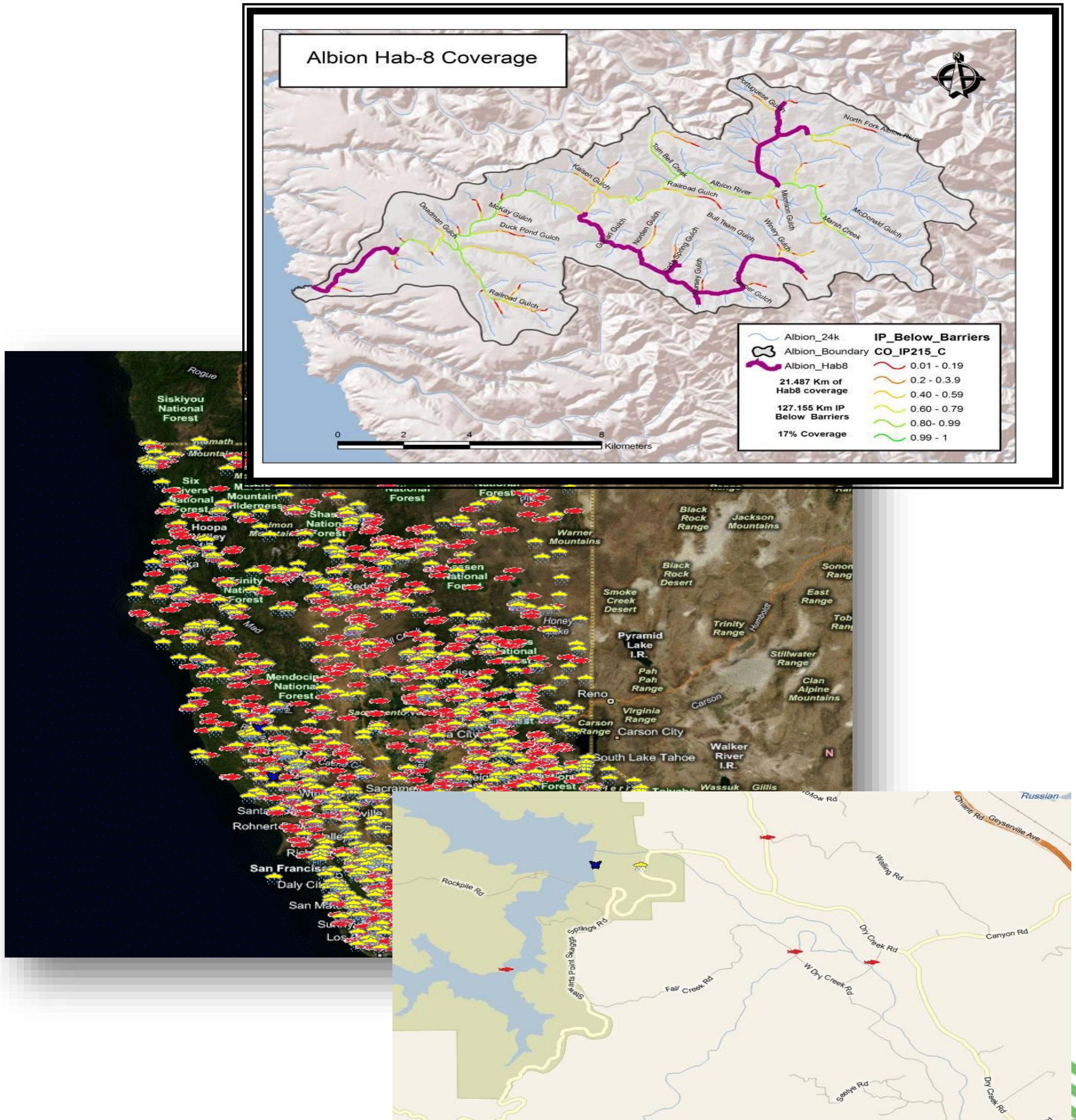
Satellite

Aircraft Surveys

Relatively Ubiquitous Motes

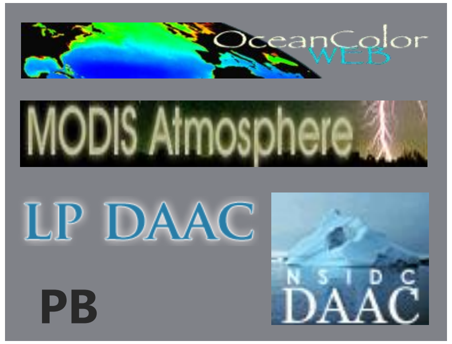
# Data Integration Challenges

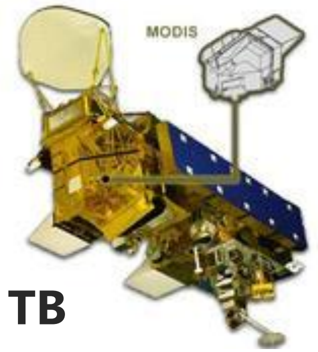
- Regular rasters, points, and spatial features
- Time series and intermittent
- Vocabulary meanings (ontology)
- Sparse in time, duration, or location
- Science variable derivation
- Gaps
- Spatial/temporal harmonization



# Why Make this Distinction?

- Provenance and trust widely varies
  - Data acquisition, early processing, and reporting ranges from a large government agency to individual scientists.
  - Smaller data often passed around in email; big data downloads can take days (if at all)
- Data sharing concerns and patterns vary
  - Open access followed by (non-repeatable and tedious) pre-processing
  - True science ready data set but concerns about misuse, misunderstanding particularly for hard won data.
- Computational tools differ.
  - Not everyone can get an account at a supercomputer center
  - Very large computations require engineering (error handling)
  - Space and time aren't always simple dimensions





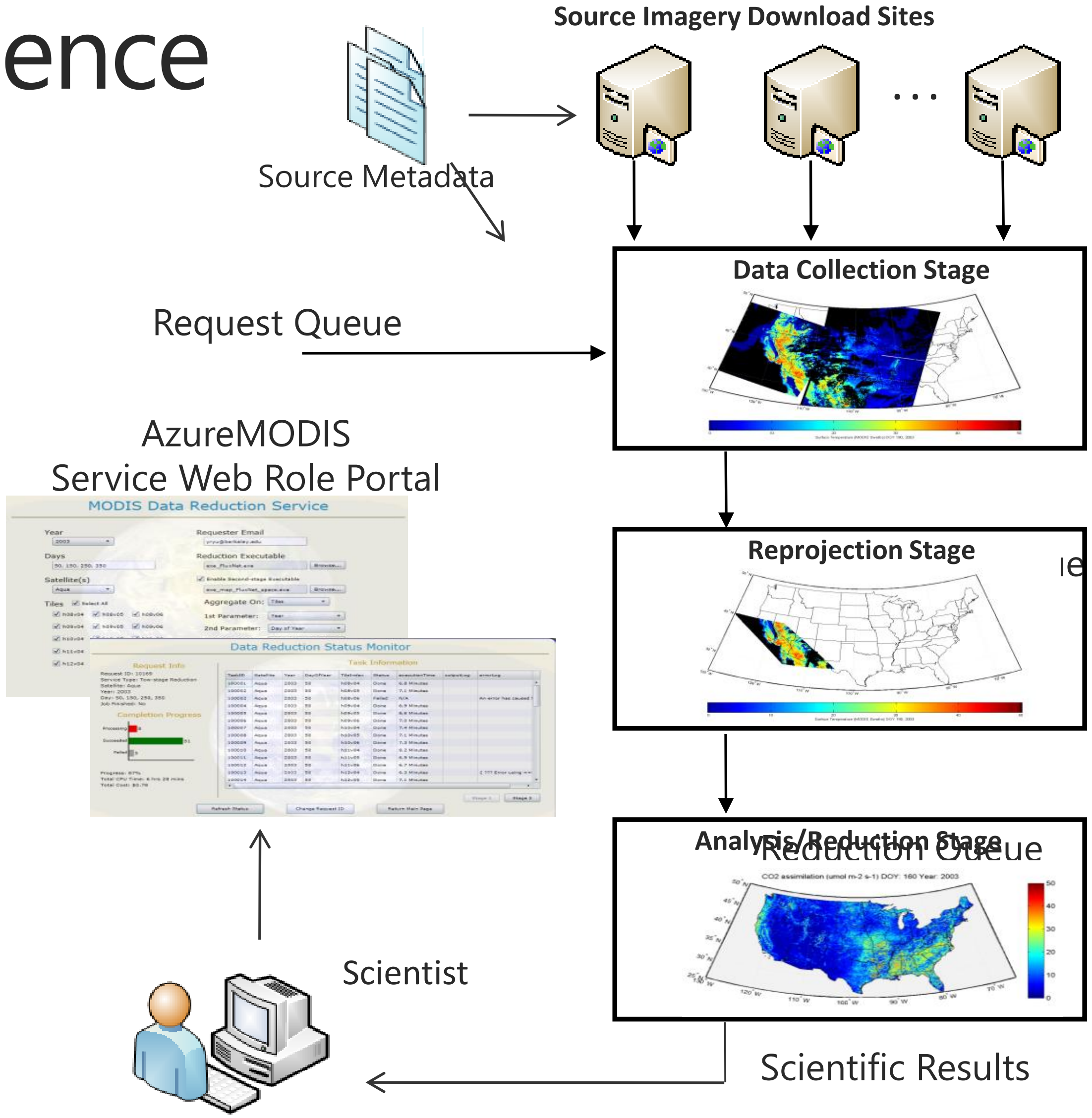
Complex shared detector | Simple instrument (if any)

*Science happens when PBs, TBs, GBs, and KBs can be mashed up simply*

Complex and Heavy process by experts | Ad hoc observations and models

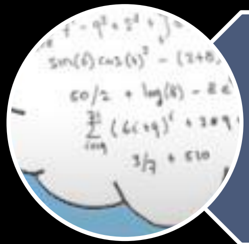
# AzureMODIS – Azure Service for Remote Sensing Geoscience

- Science pipeline for download, initial processing, and reduction of satellite imagery. Developed by MSR, UVa, UCB.
- Dramatically lowers resource and complexity barriers to use satellite imagery for terrestrial hydrology and geoscience.
  - Common imagery location determination and upload from diverse sources
  - Optional scientist-provided reduction algorithm (.NET, Java, or MatLab)
  - On-demand scalability beyond local desktop or cluster
- In use now to compute 10 year continental scale water balance for North America. Per year:
  - 500 GB (~60K files) upload of 9 different source imagery products from 15 different locations
  - 400 GB reprojected harmonized imagery consuming ~3500 cpu hours
  - 5 GB reduced science result leveraging reported field data aggregates consuming ~60 cpu hour



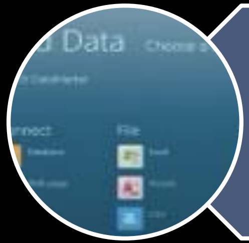
Catharine van Ingen (Microsoft Research), Jie Li, Marty Humphreys (UVA), Youngryel Ryu (UCB), Deb Agarwal (BWC/LBL)

# Microsoft® SQL Azure™ Labs



## Microsoft Codename "Cloud Numerics‟

numerical and data analytics library for data scientists, quantitative analysts, and others



## Microsoft Codename "Data Explorer‟

organize, manage, mash up and gain new insights from your data.

## Microsoft Codename "Data Hub‟

Organizationsto curate and publish its data on a private data marketplace

## Microsoft Codename "Trust Services‟

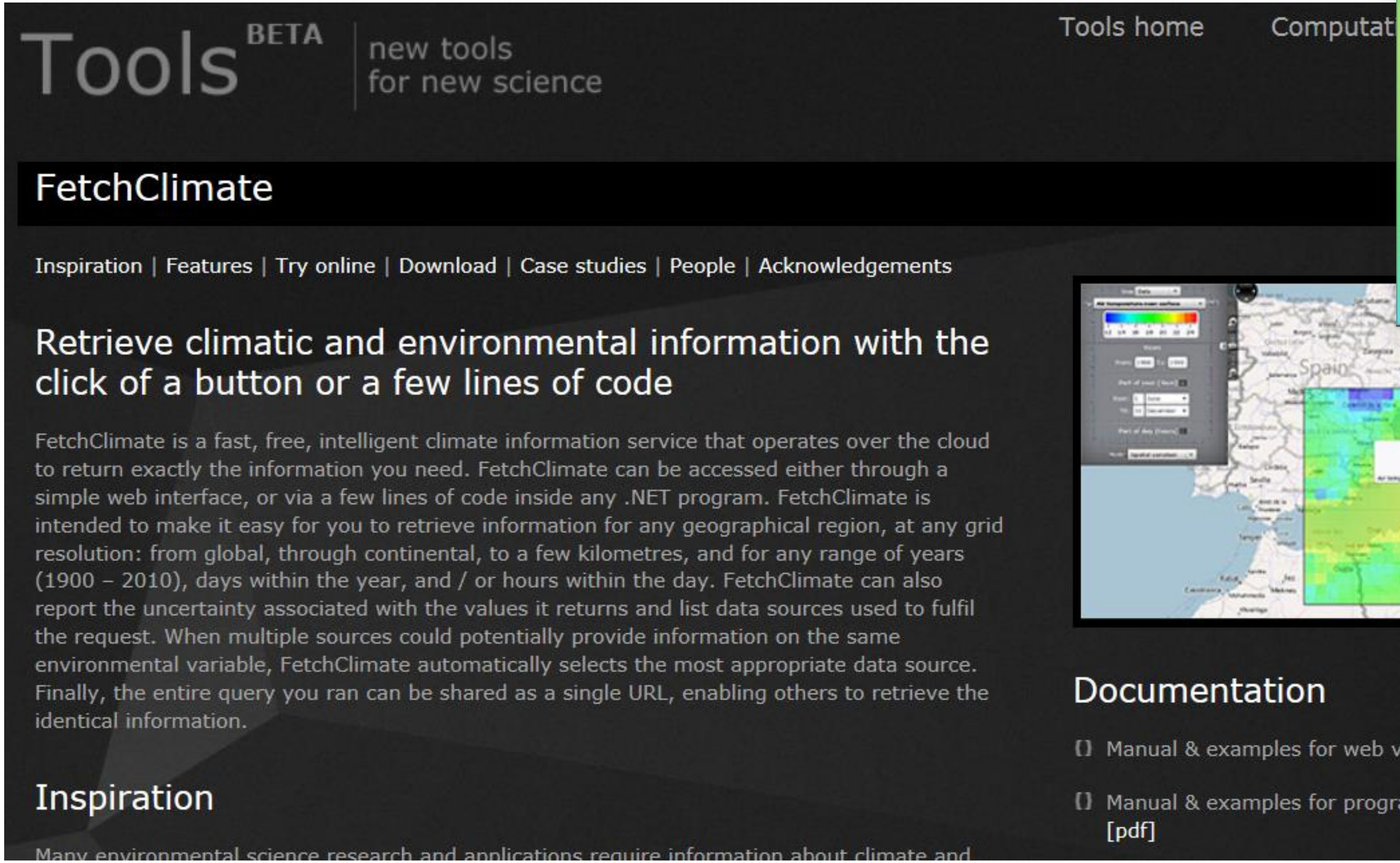
data encryption services for cloud applications so that they can roam encryption keys in a secure way.

Discoverability

Accessibility

Consumeability

# Fetch Climate



Tools BETA | new tools for new science

Tools home    Computat

## FetchClimate

Inspiration | Features | Try online | Download | Case studies | People | Acknowledgements

### Retrieve climatic and environmental information with the click of a button or a few lines of code

FetchClimate is a fast, free, intelligent climate information service that operates over the cloud to return exactly the information you need. FetchClimate can be accessed either through a simple web interface, or via a few lines of code inside any .NET program. FetchClimate is intended to make it easy for you to retrieve information for any geographical region, at any grid resolution: from global, through continental, to a few kilometres, and for any range of years (1900 – 2010), days within the year, and / or hours within the day. FetchClimate can also report the uncertainty associated with the values it returns and list data sources used to fulfil the request. When multiple sources could potentially provide information on the same environmental variable, FetchClimate automatically selects the most appropriate data source. Finally, the entire query you ran can be shared as a single URL, enabling others to retrieve the identical information.

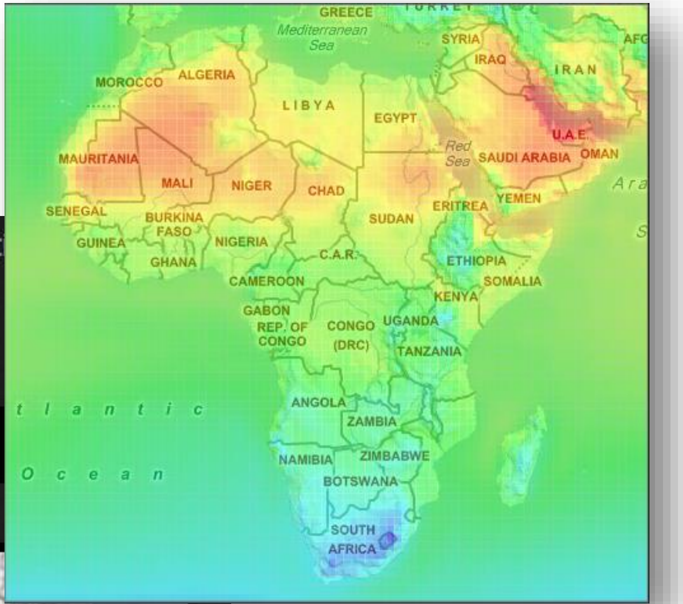
### Documentation

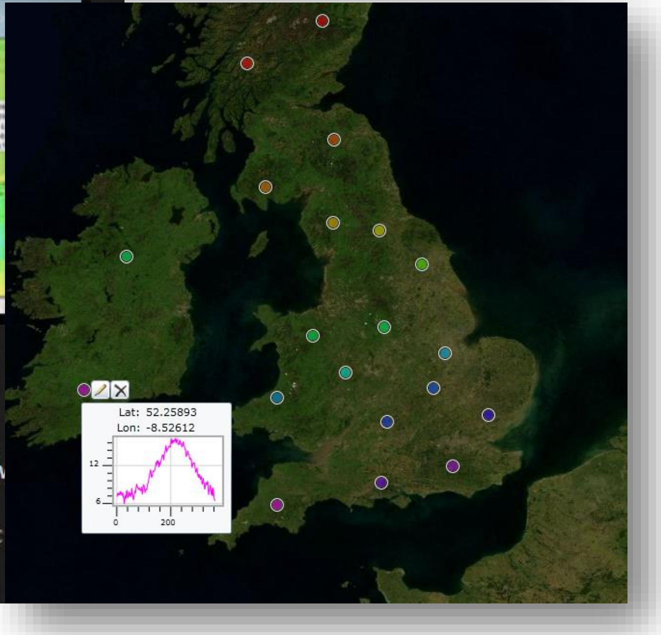
{} Manual & examples for web version [v

{} Manual & examples for programmatic [pdf]

### Inspiration

Many environmental science research and applications require information about climate and



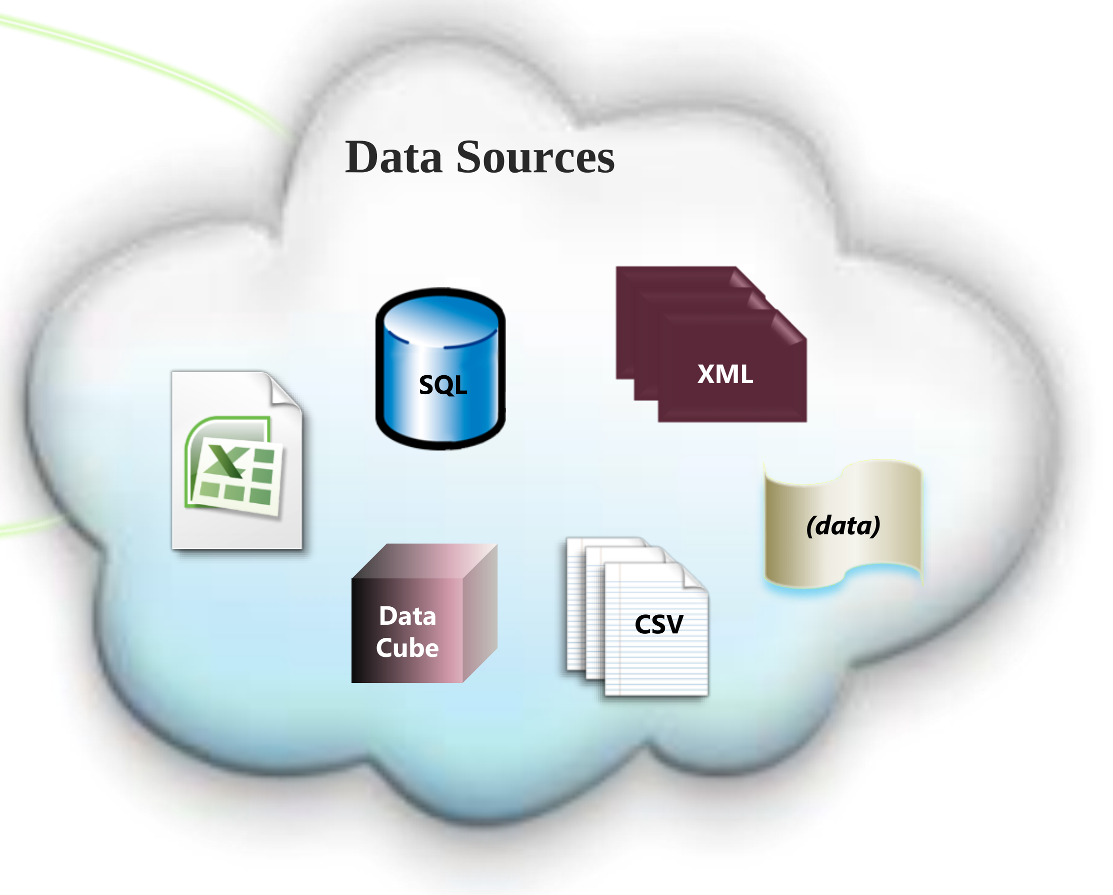


http://fetchclimate.cloudapp.net/

# Environmental Informatics Framework (EIF)

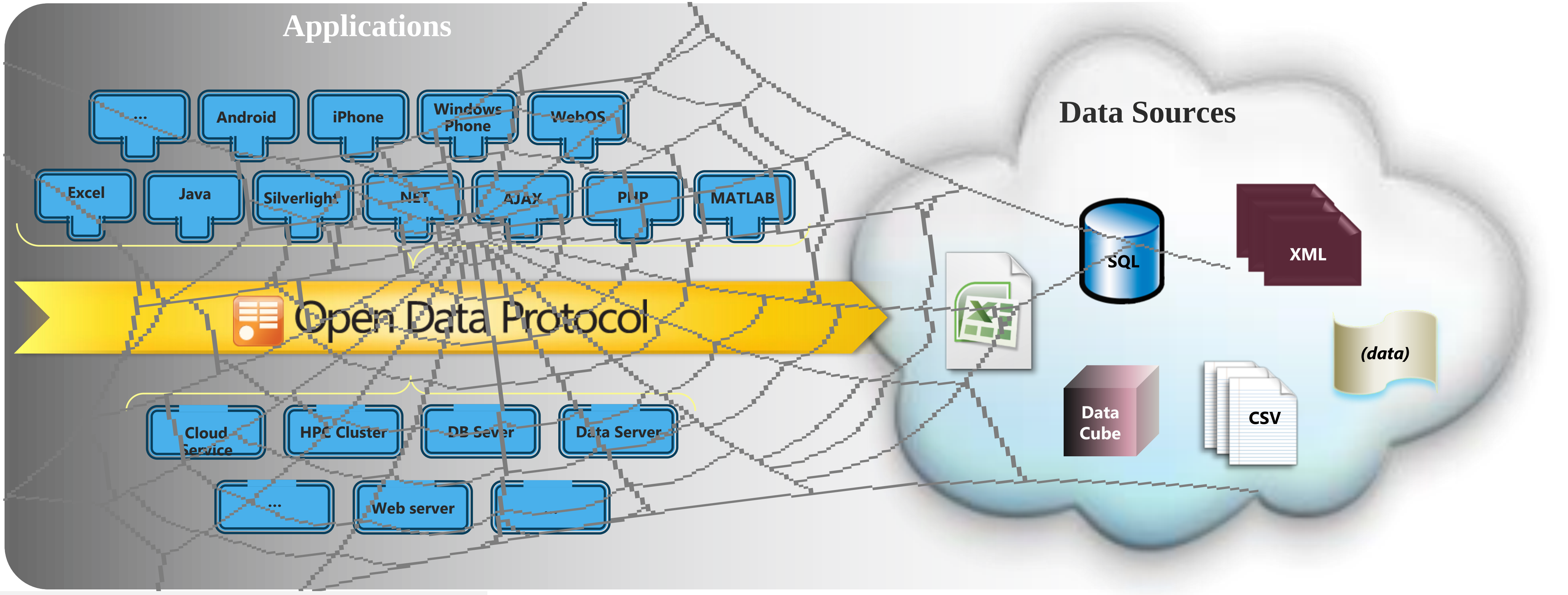
## Common Problems with Data

- To use data from different sources
  - Non-standard formats, scales, and units
  - Lack of data quality control
  - Lack of metadata
  - Difficult to repurpose data for different (my) tools
- To share data
  - Lack of incentive (no credit)
  - Need extra resources and tools
- Hidden problems, seldom addressed
  - Versioning
  - Provenance
  - Curation



# Environmental Informatics Framework (EIF)

## Current State of Data Ecosystem



# Environmental Informatics Framework (EIF)

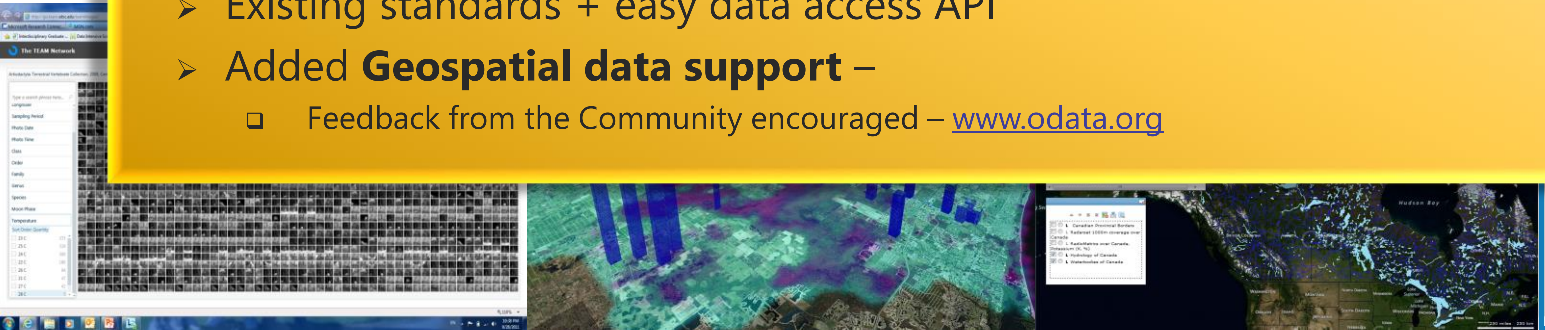
## Advance data discoverability, accessibility, and consumability

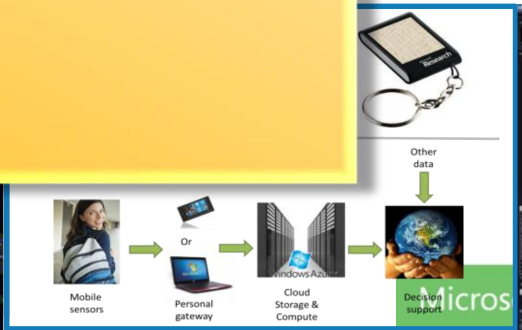
### Open Data Protocol (OData)

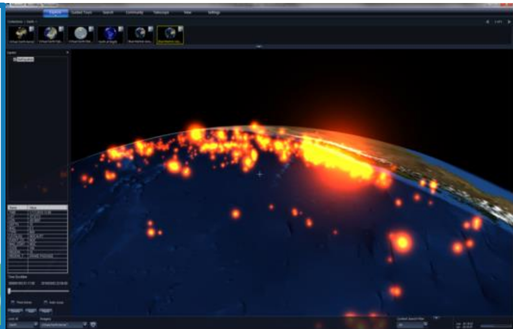
http://www.odata.org

**It allows you to form URLs based on what you know about the underlying data**

- A Web protocol for querying and updating data
  - provides a way to unlock your data and free it from data silos
  - does this by building upon Web technologies such as HTTP, Atom Publishing Protocol (AtomPub) and JSON to provide access to information from a variety of applications, services, and stores.
- In Open Source/Specifications Promise – being submitted to OASIS
- An application of a set of internet standards:
  - HTTP,
  - Atom (RFC 4287),
  - AtomPub (RFC 5023),
  - REST semantics
- Existing standards + easy data access API
- Added **Geospatial data support** –
  - Feedback from the Community encouraged – www.odata.org







# New ways to analyze and communicate data

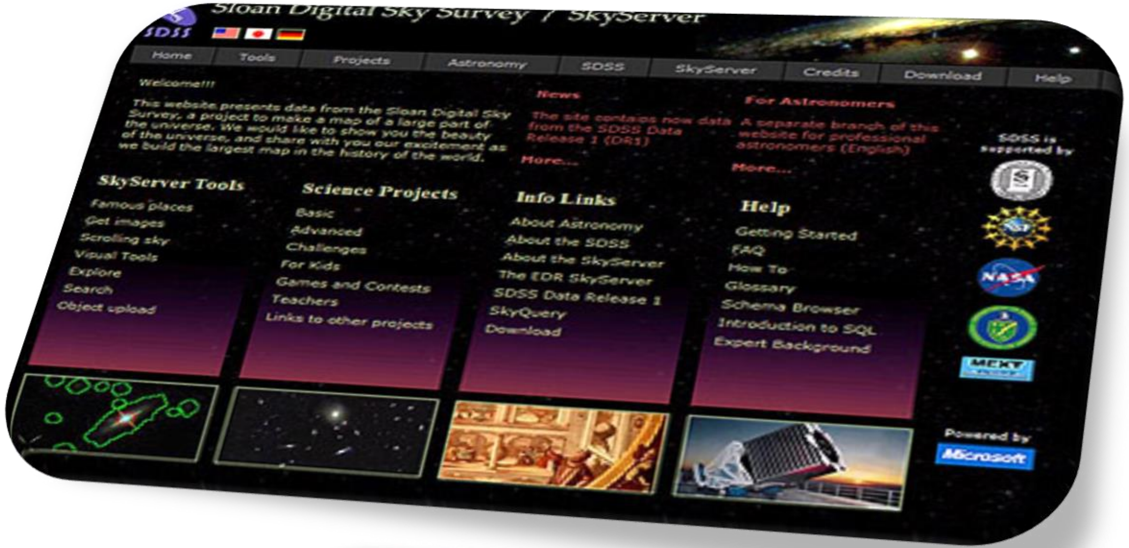
# The 'Cosmic Genome' Project

- The Sloan Digital Sky Survey is the first major astronomical survey project:
  - 5 color images of ¼ of the sky
  - Pictures of 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
  - Have to publish the data before astronomers publish their analysis

SDSS

# Public Use of the SkyServer

- **Posterchild in 21st century data publishing**
  - 380 million web hits in 6 years
  - 930,000 distinct users
    vs 10,000 astronomers
  - 1600 scientific papers
  - Delivered 50,000 hours
    of lectures to high schools
  - Delivered 100B rows of data
- **Citizen Science: GalaxyZoo**
  - Goal of 1 million visual galaxy classifications by the public
  - Allows general public to search for photographs and classify different types of galaxies



Welcome!!!

This website presents data from the Sloan Digital Sky Survey, a project to make a map of a large part of the universe. We would like to show you the beauty of the universe, and share with you our excitement as we build the largest map in the history of the world.

News

The site contains now data from the SDSS Data Release 1 (DR1)

More...

For Astronomers

A separate branch of this website for professional astronomers (English)

More...

SkyServer Tools

Famous places
Get images
Scrolling sky
Visual Tools
Explore
Search
Object upload

Science Projects

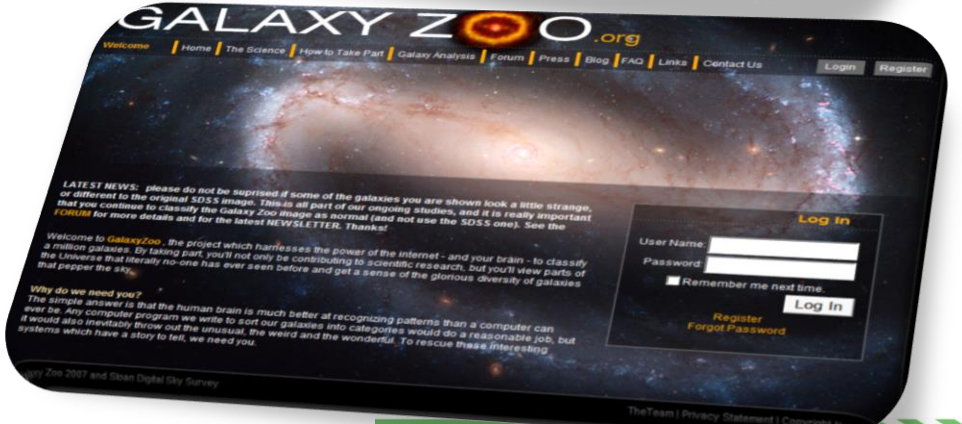
Basic
Advanced
Challenges
For Kids
Games and Contests
Teachers
Links to other projects

Info Links

About Astronomy
About the SDSS
About the SkyServer
The EDR SkyServer
SDSS Data Release 1
SkyQuery
Download

Help

Getting Started
FAQ
How To
Glossary
Schema Browser
Introduction to SQL
Expert Background



LATEST NEWS: please do not be suprised if some of the galaxies you are shown look a little strange, or different to the original SDSS image. This is all part of our ongoing studies, and it is really important that you continue to classify the Galaxy Zoo image as normal (and not use the SDSS one). See the FORUM for more details and for the latest NEWSLETTER. Thanks!

Welcome to GalaxyZoo , the project which harnesses the power of the internet - and your brain - to classify a million galaxies. By taking part, you'll not only be contributing to scientific research, but you'll view parts of the Universe that literally no-one has ever seen before and get a sense of the glorious diversity of galaxies that pepper the sky.

Why do we need you?

The simple answer is that the human brain is much better at recognizing patterns than a computer can ever be. Any computer program we write to sort our galaxies into categories would do a reasonable job, but it would also inevitably throw out the unusual, the weird and the wonderful. To rescue these interesting systems which have a story to tell, we need you.

Log In

User Name:

Password:

□ Remember me next time.

Log In

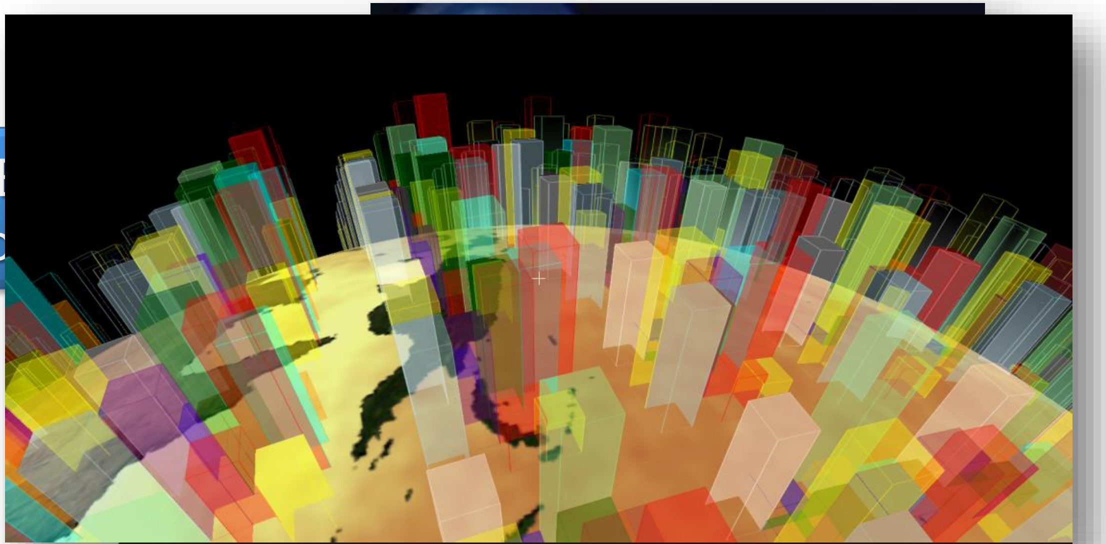
Register
Forgot Password

laxy Zoo 2007 and Sloan Digital Sky Survey

TheTeam | Privacy Statement | Copyright Noti

# World Wide Telescope

Seamless Rich Social Media Virtual Sky and E
Web application for science and educatio



Goals

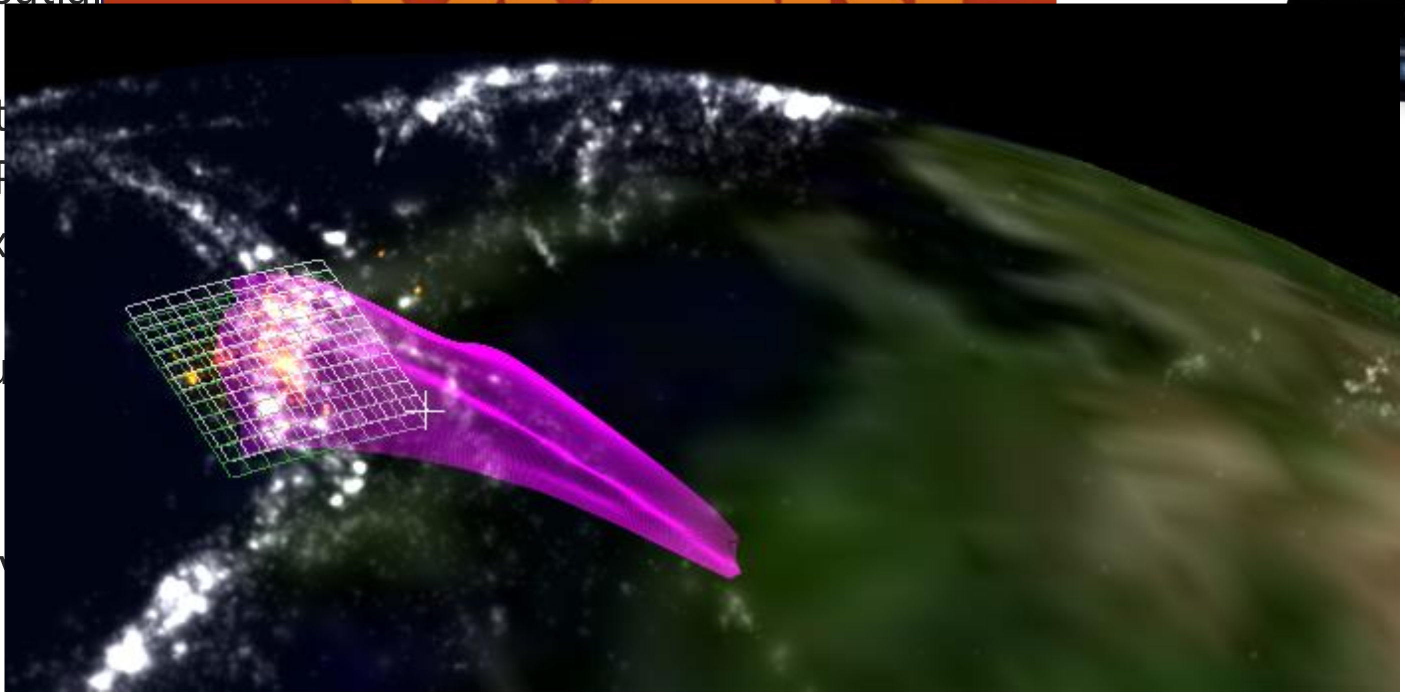
- Integra ... cess
- Easy ac
- Tours f
- Spatial

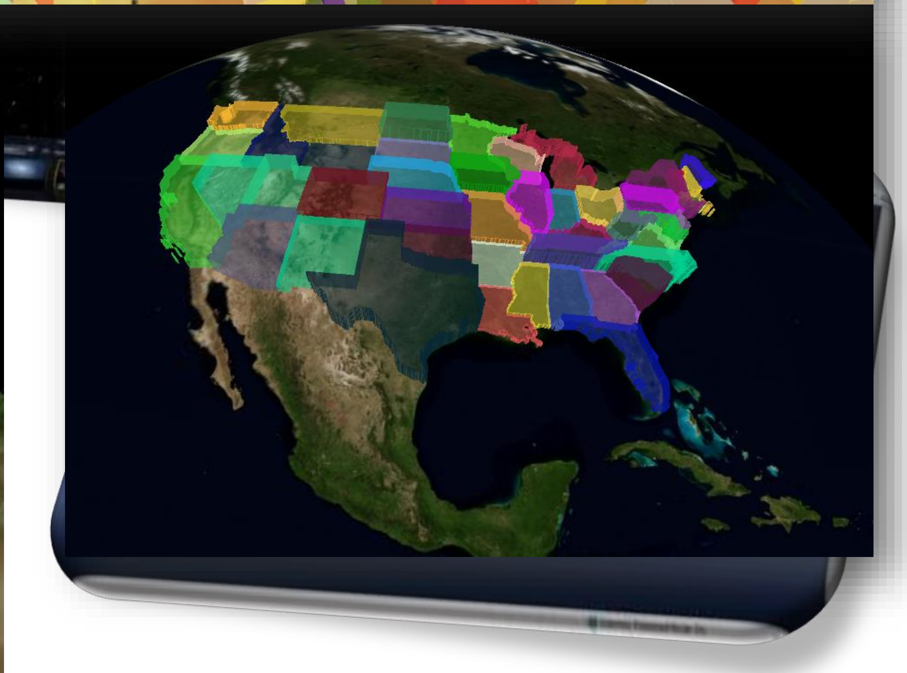
Updat

- AF
- Ex

Not ju
Being

We in



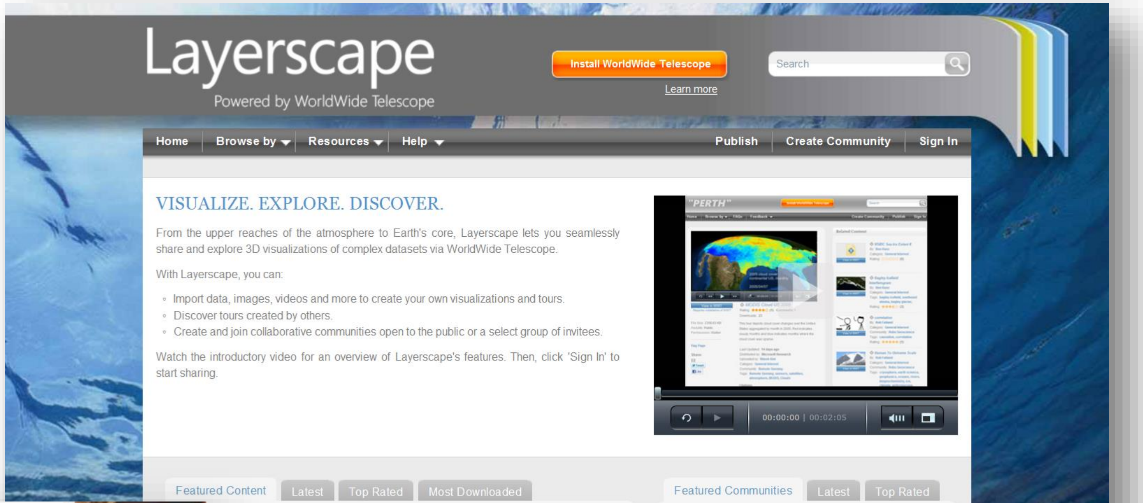


# Layerscape

Powered by WorldWide Telescope

www.layerscape.org

- Community Site for WWT Tours and Layers (Data)
- Sharing by groups/individuals



VISUALIZE. EXPLORE. DISCOVER.

From the upper reaches of the atmosphere to Earth's core, Layerscape lets you seamlessly share and explore 3D visualizations of complex datasets via WorldWide Telescope.

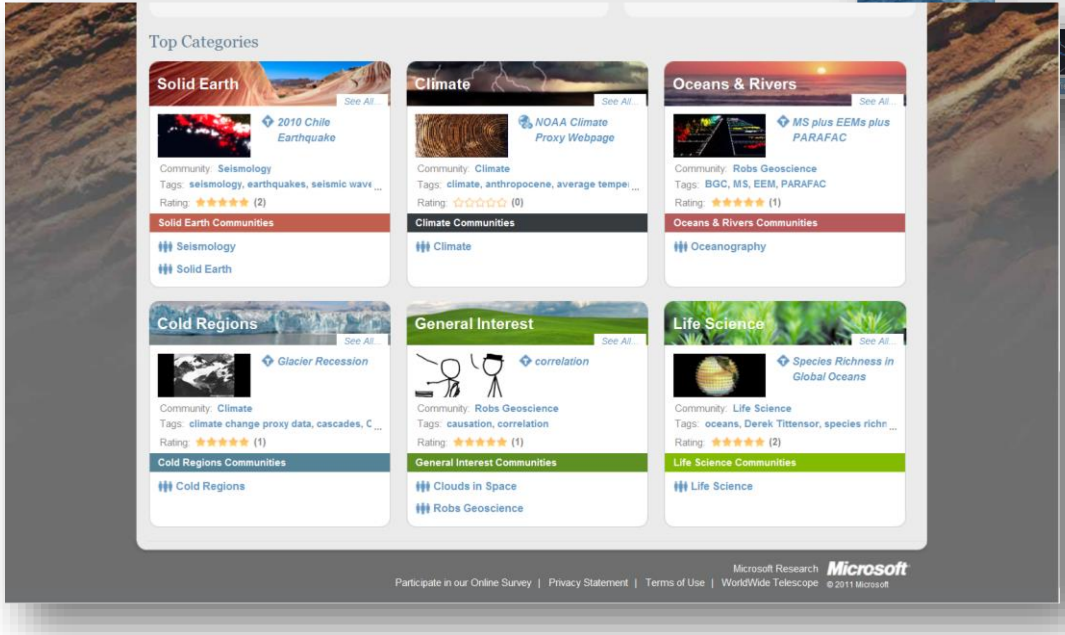
With Layerscape, you can:

- Import data, images, videos and more to create your own visualizations and tours.
- Discover tours created by others.
- Create and join collaborative communities open to the public or a select group of invitees.

Watch the introductory video for an overview of Layerscape's features. Then, click 'Sign In' to start sharing.

MBARI Dorado Drifter Tour 1 CTD Salinity
By: Rob Fatland
Category: Oceans & Rivers
Community: Oceanography

Robs Geoscience
By: Rob Fatland
Rating: ★★★★☆ (2)



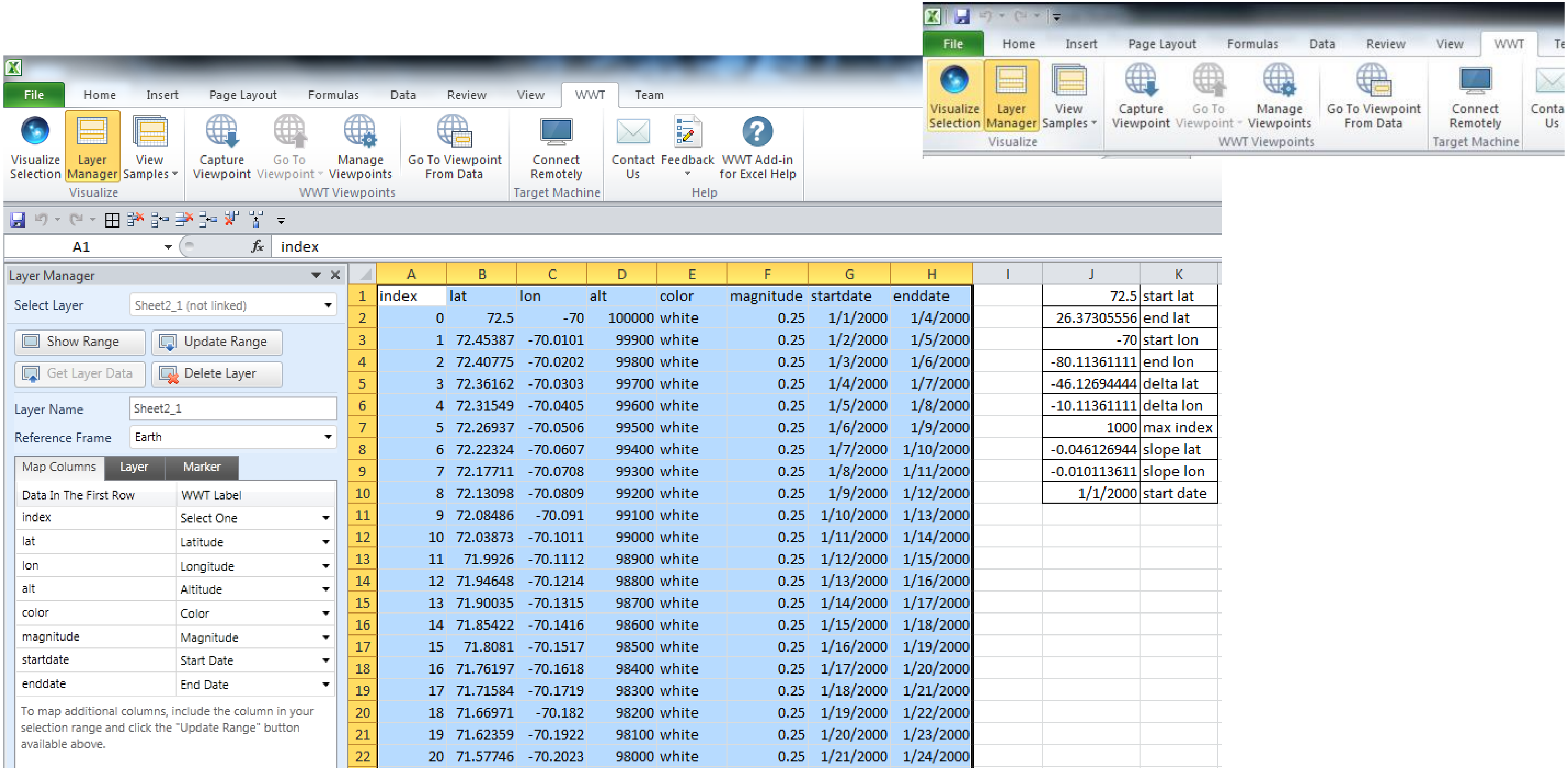
Top Categories

| Solid Earth | Climate | Oceans & Rivers |
|---|---|---|
| 2010 Chile Earthquake<br>Community: Seismology<br>Tags: seismology, earthquakes, seismic wave ...<br>Rating: ★★★★★ (2) | NOAA Climate Proxy Webpage<br>Community: Climate<br>Tags: climate, anthropocene, average tempe ...<br>Rating: ☆☆☆☆☆ (0) | MS plus EEMs plus PARAFAC<br>Community: Robs Geoscience<br>Tags: BGC, MS, EEM, PARAFAC<br>Rating: ★★★★★ (1) |
| Solid Earth Communities<br>Seismology<br>Solid Earth | Climate Communities<br>Climate | Oceans & Rivers Communities<br>Oceanography |

| Cold Regions | General Interest | Life Science |
|---|---|---|
| Glacier Recession<br>Community: Climate<br>Tags: climate change proxy data, cascades, C ...<br>Rating: ★★★★★ (1) | correlation<br>Community: Robs Geoscience<br>Tags: causation, correlation<br>Rating: ★★★★★ (1) | Species Richness in Global Oceans<br>Community: Life Science<br>Tags: oceans, Derek Tittensor, species richn ...<br>Rating: ★★★★★ (2) |
| Cold Regions Communities<br>Cold Regions | General Interest Communities<br>Clouds in Space<br>Robs Geoscience | Life Science Communities<br>Life Science |

Participate in our Online Survey | Privacy Statement | Terms of Use | WorldWide Telescope
Microsoft Research Microsoft
© 2011 Microsoft

# Excel to Visualize in seconds



A1 fx index

Select Layer: Sheet2_1 (not linked)

Layer Name: Sheet2_1

Reference Frame: Earth

| Data In The First Row | WWT Label |
|---|---|
| index | Select One |
| lat | Latitude |
| lon | Longitude |
| alt | Altitude |
| color | Color |
| magnitude | Magnitude |
| startdate | Start Date |
| enddate | End Date |

To map additional columns, include the column in your selection range and click the "Update Range" button available above.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | index | lat | lon | alt | color | magnitude | startdate | enddate |
| 2 | 0 | 72.5 | -70 | 100000 | white | 0.25 | 1/1/2000 | 1/4/2000 |
| 3 | 1 | 72.45387 | -70.0101 | 99900 | white | 0.25 | 1/2/2000 | 1/5/2000 |
| 4 | 2 | 72.40775 | -70.0202 | 99800 | white | 0.25 | 1/3/2000 | 1/6/2000 |
| 5 | 3 | 72.36162 | -70.0303 | 99700 | white | 0.25 | 1/4/2000 | 1/7/2000 |
| 6 | 4 | 72.31549 | -70.0405 | 99600 | white | 0.25 | 1/5/2000 | 1/8/2000 |
| 7 | 5 | 72.26937 | -70.0506 | 99500 | white | 0.25 | 1/6/2000 | 1/9/2000 |
| 8 | 6 | 72.22324 | -70.0607 | 99400 | white | 0.25 | 1/7/2000 | 1/10/2000 |
| 9 | 7 | 72.17711 | -70.0708 | 99300 | white | 0.25 | 1/8/2000 | 1/11/2000 |
| 10 | 8 | 72.13098 | -70.0809 | 99200 | white | 0.25 | 1/9/2000 | 1/12/2000 |
| 11 | 9 | 72.08486 | -70.091 | 99100 | white | 0.25 | 1/10/2000 | 1/13/2000 |
| 12 | 10 | 72.03873 | -70.1011 | 99000 | white | 0.25 | 1/11/2000 | 1/14/2000 |
| 13 | 11 | 71.9926 | -70.1112 | 98900 | white | 0.25 | 1/12/2000 | 1/15/2000 |
| 14 | 12 | 71.94648 | -70.1214 | 98800 | white | 0.25 | 1/13/2000 | 1/16/2000 |
| 15 | 13 | 71.90035 | -70.1315 | 98700 | white | 0.25 | 1/14/2000 | 1/17/2000 |
| 16 | 14 | 71.85422 | -70.1416 | 98600 | white | 0.25 | 1/15/2000 | 1/18/2000 |
| 17 | 15 | 71.8081 | -70.1517 | 98500 | white | 0.25 | 1/16/2000 | 1/19/2000 |
| 18 | 16 | 71.76197 | -70.1618 | 98400 | white | 0.25 | 1/17/2000 | 1/20/2000 |
| 19 | 17 | 71.71584 | -70.1719 | 98300 | white | 0.25 | 1/18/2000 | 1/21/2000 |
| 20 | 18 | 71.66971 | -70.182 | 98200 | white | 0.25 | 1/19/2000 | 1/22/2000 |
| 21 | 19 | 71.62359 | -70.1922 | 98100 | white | 0.25 | 1/20/2000 | 1/23/2000 |
| 22 | 20 | 71.57746 | -70.2023 | 98000 | white | 0.25 | 1/21/2000 | 1/24/2000 |

| J | K |
|---|---|
| 72.5 | start lat |
| 26.37305556 | end lat |
| -70 | start lon |
| -80.11361111 | end lon |
| -46.12694444 | delta lat |
| -10.11361111 | delta lon |
| 1000 | max index |
| -0.046126944 | slope lat |
| -0.010113611 | slope lon |
| 1/1/2000 | start date |

# Natural User Interfaces (NUI) Kinect SDK and WWT

KINECT™

for Windows®

- Rethinking ways in which people will interact with computers/technologies of the future
- Re-evaluating everything from their (non-) physical design to the human needs and interaction models
- Revolutionize the way we think about technology and what it can do on our behalf

# Our Next Steps

- Continue to work with Scientists
  - Implement features from feedback
    - Ie. Netcdf support
  - New clients and User Interface
  - Make it easier to use/create tours

- Cloud: Windows Azure integration - Write, Run, or Use Software
  - Write: Platform as a Service
    - Web Sites
    - Storage
    - Languages - .Net, Java, Python, PHP, Ruby, NodeJS, C++
  - Run: Infrastructure as a Service
    - Virtual Machines – Windows Servers & Linux
    - Connect to on-premises
  - Use: Software as a Service
    - Azure Media Services
    - HPC and Big Data – Hadoop, etc
    - Windows Azure Marketplace

# eScience in Action

# Microsoft eScience Workshop 2012

**October 8–9, 2012 | Chicago, Illinois**

http://research.microsoft.com/events/escience2012

Microsoft Research Connections

# Data Storage Sustainability?

- Digital Data can be open – who should pay the cost?
- Spinning Disks, Bandwidth, Cooling, etc

Why you need a TV Licence

**To watch or record TV programmes you must have a TV Licence**

You only need 1 TV Licence for your home no matter how many TVs you have.
The licence is for the place, not the TV.

**You can watch TV programmes in lots of different ways**

On a TV set
On a computer or laptop
On a mobile phone

**You might also have**

A digital box (sometimes these are called digi boxes)

# No Silver Bullet - What is needed?

- Algorithms that scale
- Data Management from the Start
- Automatic Ancillary Data capture
- Thinking about the Data, and retention
- Data sharing is natural from the start
- Visualization for everyone
- Best practices – insights and challenges shared amongst domains
    - Ie. eScience Workshop, etc

# Challenges

- Balancing

  Data Acquisition | Bandwidth | Storage/Processing

- Cross Discipline Collaboration – Knowledge sharing
- The data deluge - How to manage and analyze information?
- New types of Scientists:
  - Data Collectors & Data Analysis
- Riding the commodity curve
- Technology/Computing in support of Science

Microsoft®

Research Connections