Microsoft Research
FacultySummit

Open Data for Open Science

Microsoft Environmental Informatics Framework

Yan Xu
Sr. Research Program Manager
Microsoft Corporation

FUTURE/WORLD 2011 2031

Microsoft Environmental Informatics

Since 2010

Vision: facilitate seamless access to environmental data and information

Focus: data discoverability, accessibility, and consumability

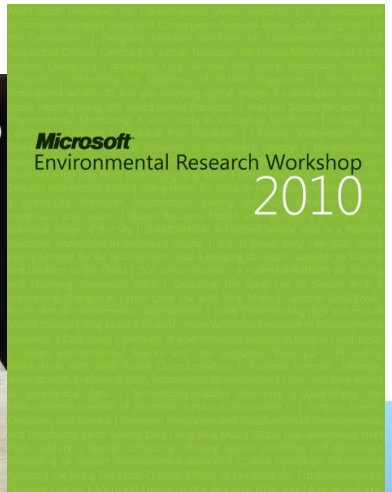
Objectives:

- advance the technology use in environmental research
- create design wins using Microsoft technologies to
  - Foster innovations in computational environmental research
  - Advance interoperability of data and information sharing
  - Facilitate citizen science for environmental research
- Enhance connections among multiple disciplines and stakeholders

Connect with Environmental Research

[Microsoft Environmental Research Workshop 2010]

- 50+ from academia, industry, government, and Microsoft
- Focus:
  - Mind-swap between Environmental scientists and Microsoft technologists
  - Computational challenges in environmental research
  - Mutually beneficial collaborations



Initial Findings

Grand Challenge: vast amount of heterogeneous data

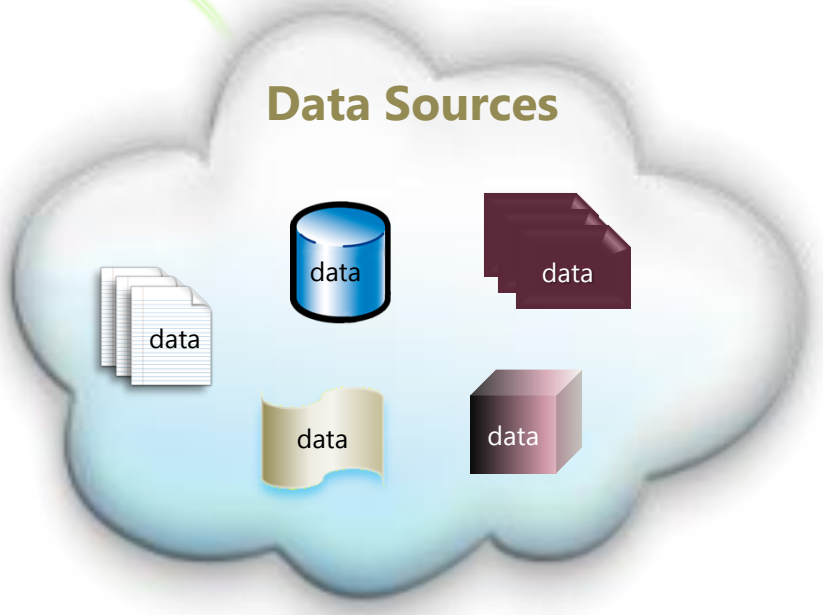
- Necessary to use data from difference sources
- Not easy (little incentive) to share data
- Need tools to enable producers & consumers to adopt standards

Unique Challenge: it doesn't end with scientific publications

- Compelling presentation of knowledge is critical to
  - Influence policy makers and the general public
  - Enable citizen science to scale the effort.

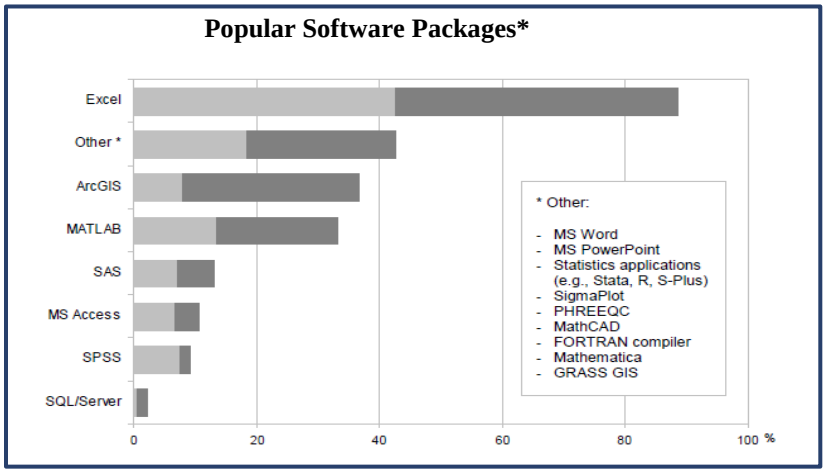
Common Problems with Data

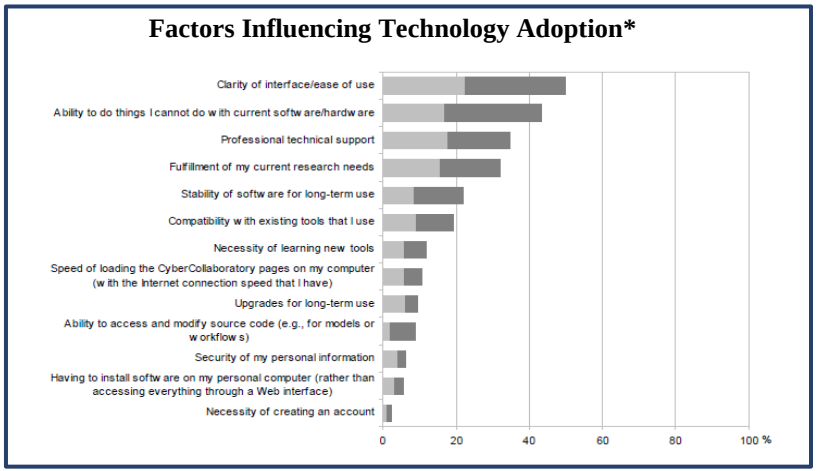
- To use data from difference sources
  - Non-standard formats, scales, and units
  - Lack of data quality control
  - Lack of metadata
  - Difficult to repurpose data for different (my) tools
- To share data
  - Lack of incentive
  - Expect no credit
  - Need extra resource and tools
- To properly maintain data, need
  - Versioning
  - Provenance
  - Curation



The Norms of Technology Use

*Cyberinfrastructure for the waters networks: a Survey of AEESP and CUSHAI Members, K.A. Lawrence et al, May 2006





The Lowest Common Denominator

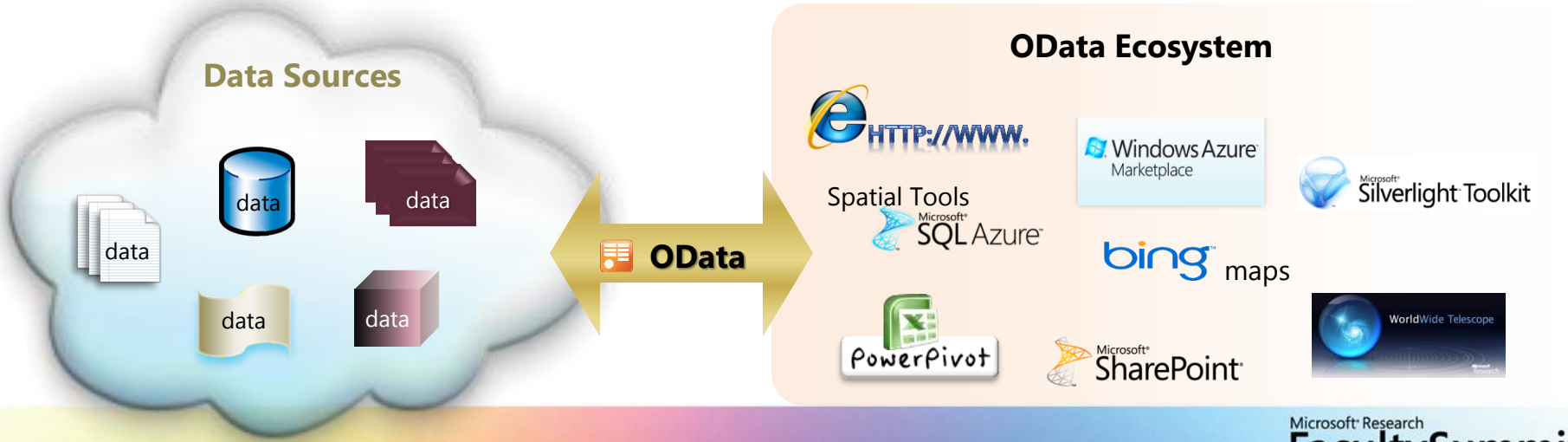
Requirements for adopting new technologies:

- Easy of use
- Additional value
- Professional technical support

Advance the Norms

Microsoft Environmental Informatics Framework -

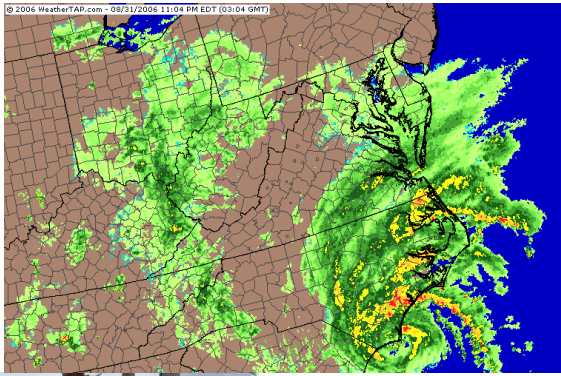
- a solution to engage Microsoft technologies with environmental research
- a strategy to amplify social impact of environmental research
- a common agenda among internal stakeholders



Example – Digital Urban Informatics

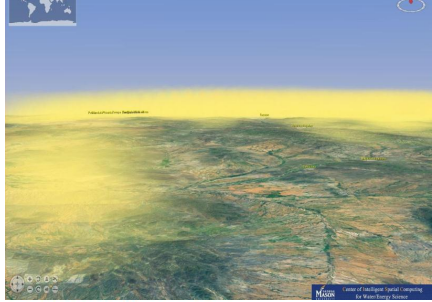
- **Objectives**
  - Research - Apply innovative computational thinking to environmental applications by developing a new "Digital Urban Informatics" computational framework addressing both long-term sustainability issues and short-term situational awareness
  - Education - Promote citizen science
- **Key technologies**
  - Microsoft WWT Earth,
  - Azure Cloud Computing,
  - Windows Phone 7 app + cloud, s
  - OData
  - Modeling in Azure, on-demand, in near-real-time

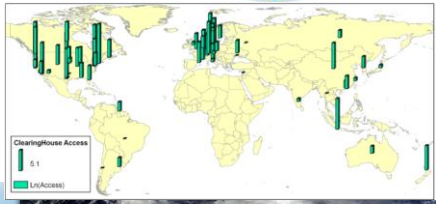
    e.g. Modflow parallel ensemble runs for groundwater sustainability

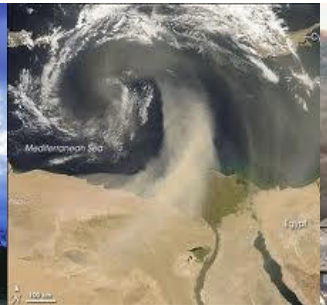


Example – GEOSS Clearinghouse

- **Objectives**
  - Share Global Earth Observation Data Among 140+ Countries to Address Global Challenges on Natural Hazards and Emergency Responses
  - Support Global End Users to Discover, Access, and Utilize EO Data
  - Provide Responses to End Users in Seconds
- **Key technologies**
  - Windows Azure - Responding to Massive Concurrent End Users
  - SQLAzure and SQL Spatial - Managing Millions to Billions of Metadata Records
  - Bing Maps and World Wide Telescope - Visualizing EO Data

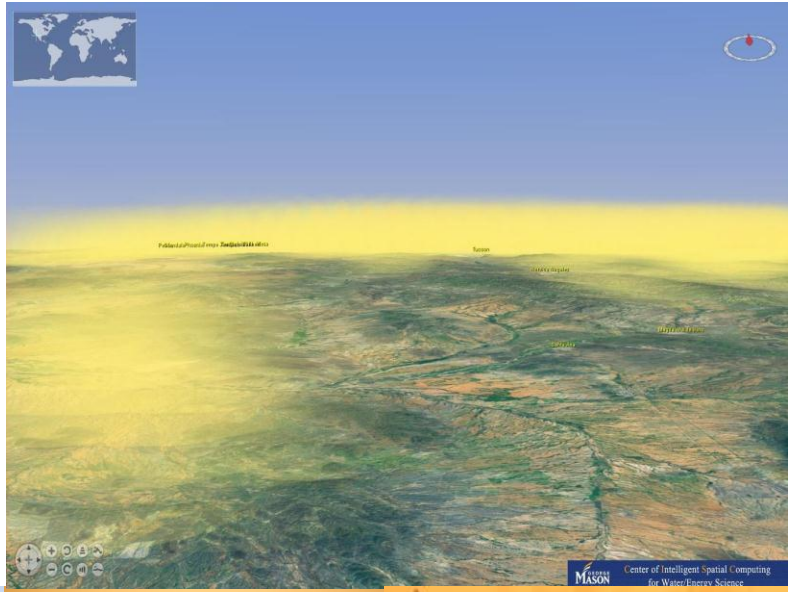


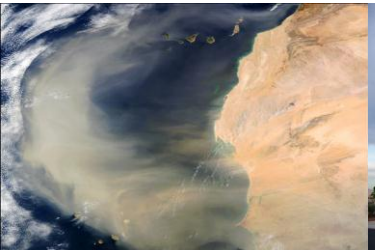




Example – Forecasting Dust Storms

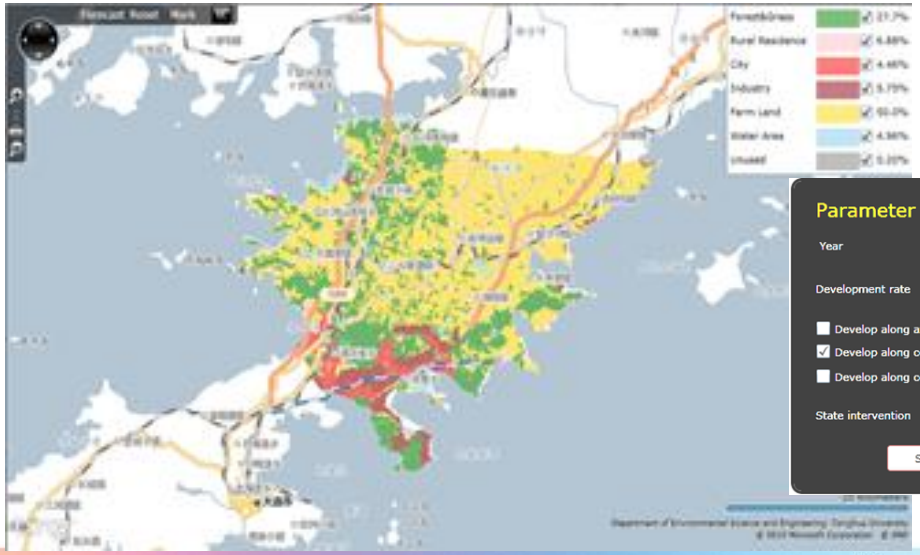
- **Objectives**
  - Provide timely forecasting of dust storm for public health emergency responses
  - Provide an intuitive interface for decision makers
- **Key technologies**
  - SQL Azure as an data management to access large volumetric data
  - World Wide Telescope(WWT) as an interactive 3D/4D visualization framework to render the data
  - Azure as an advanced cloud computing platform to support forecasting.





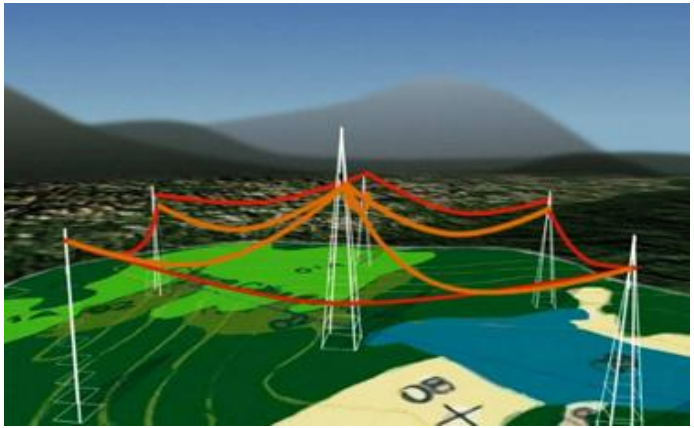
Example – Forecasting Urban Land Use Changes

- **Objectives**
  - Visually integrated urban land expansion module and environmental forecast module
  - Identify potential environmental impacts of various urban land layout over next 10-20 years
  - Promote smart decision making
- **Key technologies**
  - Bing Maps and SilverLight
  - HPC computing



Example – Wireless Sensor Networks in Brazilian Forest

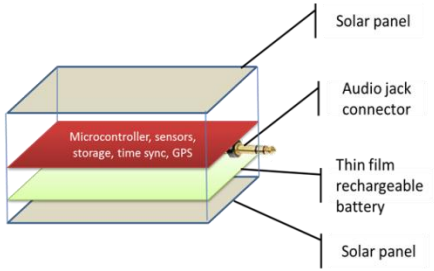
- **Hundreds of sensors to measure the vital signs of the rainforest**
  - Temperature
  - Water vapor
  - Solar radiation

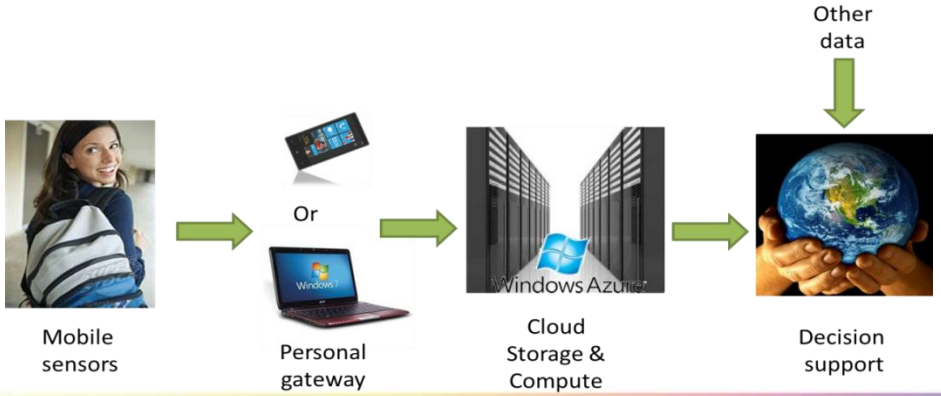


Deployment of towers with sensors (*in red*) placed on cables (*in orange*) above the rainforest canopy.

One of six towers equipped with sensors to measure temperature, water vapor, and solar radiation

Example – Participatory Environmental Monitoring Toolkit





- **Objectives**
  - Facilitate socially inclusive environmental observation
    - Time & GPS location
    - Temperature & Humidity
    - CO2
    - H2S
  - Leverage existing Microsoft technologies and user communities
  - Deliver a HW+SW toolkit in open source form
- **Key technologies**
  - Microsoft Research low energy GPS location sensing and mobile data collection services
  - OData
  - World Wide Telescope(WWT)
  - Windows Azure

Demo – WWT and Geo-Data Visualization

WorldWide Telescope (WWT)

- A visualization software environment
  - Enables a computer to function as a virtual telescope
    (astronomers call it "the best VO (virtual observatory) implementation")
  - Visualizes geo-data in 4D (space + time)
  - Integrated with Excel
  - Allows data sharing with controlled access – WWT Community
  - Empowers high-quality, intuitive, and interactive visual presentation via "WWT tour"
- Datasets under consideration
  - Seismic event distribution against sbuductionslab slab models (USGS NEIC)
  - Standardized-format datasets (OGC, WxS, NetCDF, Shapefile, CSV, HDF, …)
  - Dataset and model output concept: plugging data generators directly into WWT
  - Draped raster, e.g. MODIS ocean, land and atmospheric products
  - Alternatice topgraphy, e.g. ice sheet thickness and bathymetry
  - Climate change thematic datasets, e.g. monthly sea ice extent from NSIDC
- Free for research and education use

Demo – WWT and Dust Storm Simulation

A mutually beneficial casestudey

- Mind-swap, e.g. at [Open Data for Open Science Developers Training]
- Improve science modeling
- Improve computer engineering
- …