NASA Earth Science Information Systems Capability Vision

Prepared by the Earth Science Data Systems Working Group
on Technology Infusion

Why a Capability Vision for Information Systems?

- Helps us focus our efforts
  - What capabilities are needed to achieve the Earth science goals?
  - What technologies need to be infused most?
  - What standards are needed most?
  - What reusable components are needed most?
- Helps us measure progress
  - What is the roadmap for deploying new capabilities?
  - How much progress have we made toward achieving the vision?

Earth Science Provides Important Information to Individuals, Organizations, and Societies

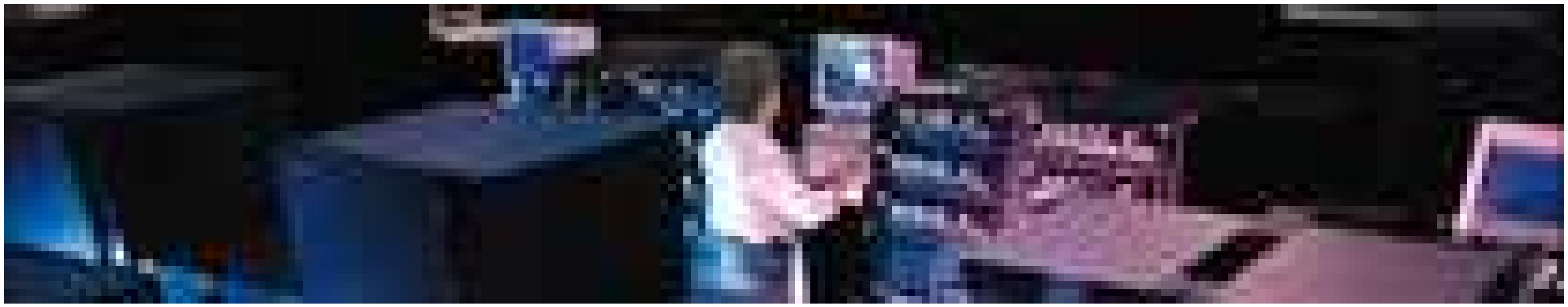
- Global observations from Earth observing satellites provide useful data on weather, climate, and natural hazards
- Knowledge gained through Earth science research has improved our understanding of Earth systems and global change
- NASA’s focus in the future will be on improving modeling and prediction capabilities

Improved Observation and Information Systems are Needed

- New observational capabilities will provide better resolution & coincident coverage

- New information system capabilities will provide the ability to quickly distill petabytes of data into usable information and knowledge

New Information System Capabilities: The Top Ten



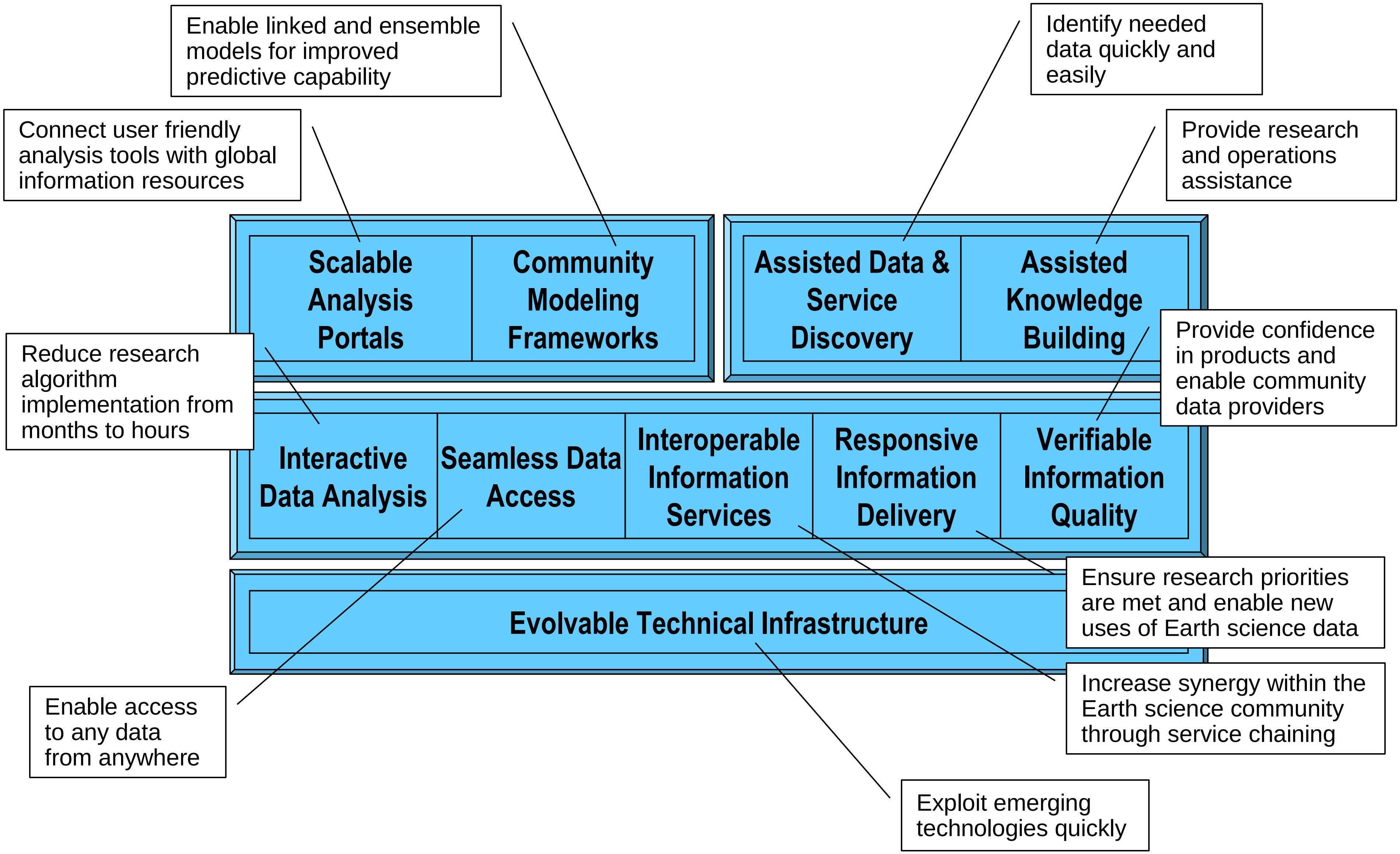
Scalable Analysis Portals	Community Modeling Frameworks

Assisted Data & Service Discovery	Assisted Knowledge Building

Interactive Data Analysis	Seamless Data Access	Interoperable Information Services	Responsive Information Delivery	Verifiable Information Quality

Evolvable Technical Infrastructure

New Information System Capabilities: The Top Ten



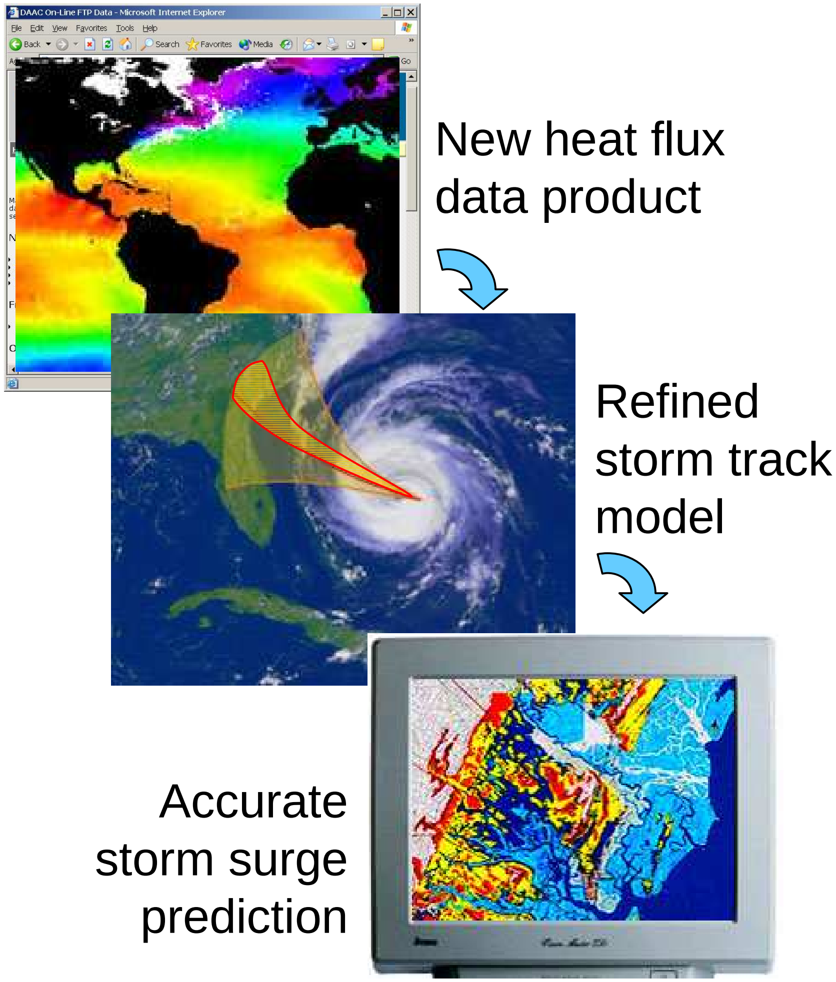
How Will New Information System Capabilities Help?

- Severe weather prediction improvement scenario
  - Hypothetical science scenario to illustrate the envisioned capabilities in a practical context
  - Only one of many possible scenarios
  - Based on one of six science focus areas in NASA's Earth science strategy

Severe Weather Prediction Improvement

- Motivation
  - Hurricanes periodically hit the East Coast of the U.S., each causing up to $25B damage and dozens of deaths
- Goal
  - Improve 5 day track prediction from +/- 400km to +/-100km by 2014
  - Accurately predict secondary effects like tidal surge
- Impact
  - Better predictions allow preparations to be focused where needed, saving money and lives
  - Note: +/-400km covers about 25% of the East Coast, while +/-100km is about 6%
- Note
  - Emphasis is on the science behind the application

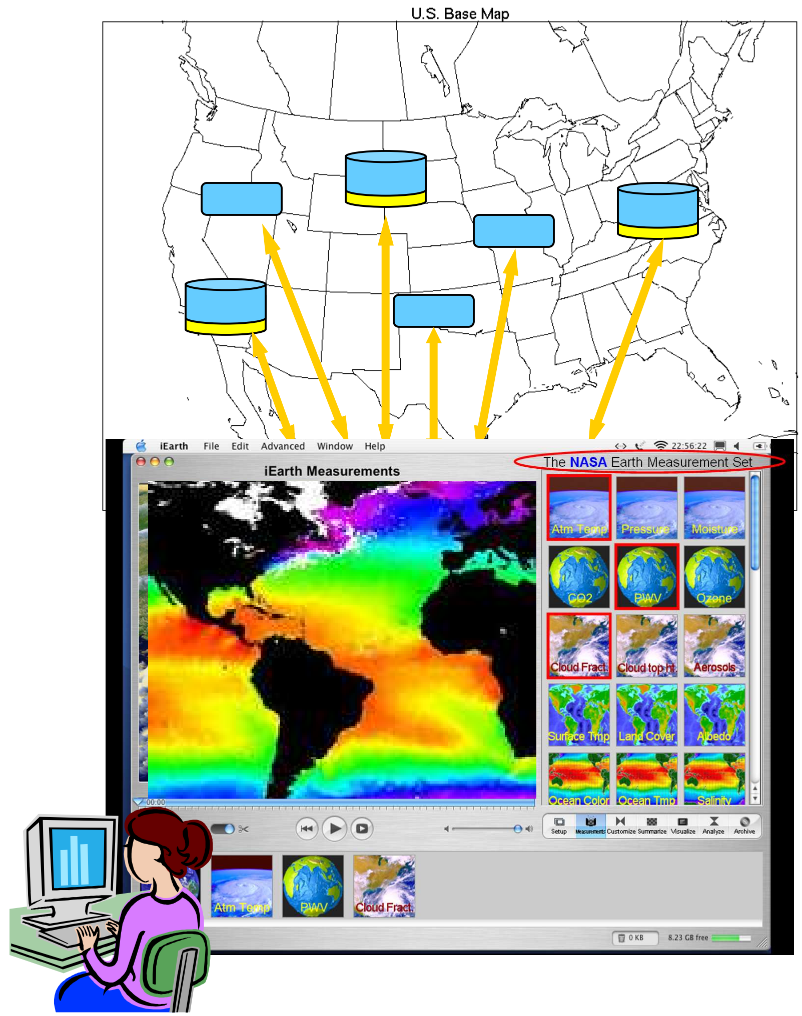
Severe Weather Prediction Improvement: How Envisioned Capabilities Would Help



- Scalable analysis portals
  - Researcher can quickly create a new ocean heat flux data product for use in severe storm models
- Community modeling frameworks
  - Several models are coupled together to create an accurate forecast the hurricane’s track and associated tidal surge
- Supporting capabilities
  - Ensure ease-of-use, quality, and timeliness

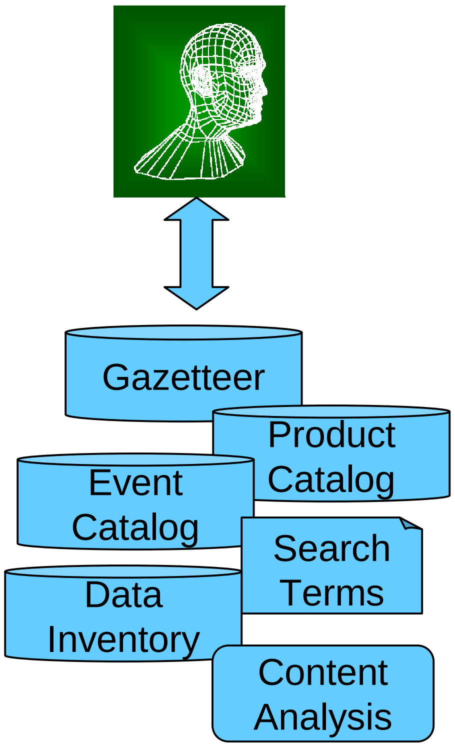
Scalable Analysis Portals

- Need
  - Researcher needs to combine a variety of local and remote data products and services to produce a new data product of estimated heat flux at ocean surface boundary
  - (Ocean heat is known to be the primary fuel of hurricanes but no heat flux product currently exists for use in severe storm models)
- Vision
  - Connect user friendly analysis tools with global information resources using common semantics
- Supporting capabilities
  - Assisted data & service discovery
  - Interactive data analysis
  - Seamless data access
  - Interoperable information services
  - Responsive information delivery
  - Verifiable information quality

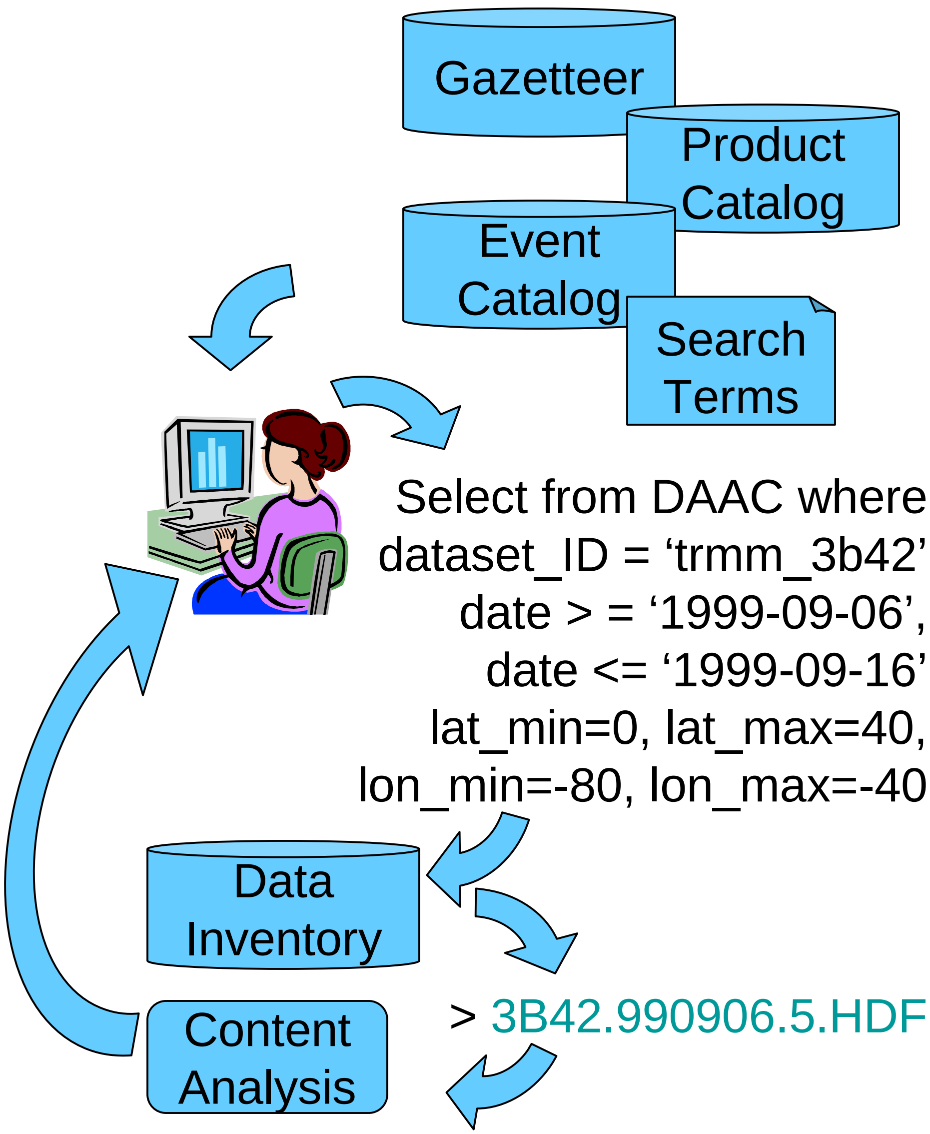


Assisted Data & Service Discovery

- Need
  - Researcher needs to identify datasets and information services required for heat flux calculations
- Vision
  - Identify needed information quickly and easily
- *Enabling technologies*
  - *Data and service description standards (XML, WSDL, RDF, OWL, OWL-S, DAML), web service directories (UDDI), syndication services (RSS), topic maps*
  - *Rule-based logic systems*
  - *Established directory services (GCMD, ECHO, THREDDS)*

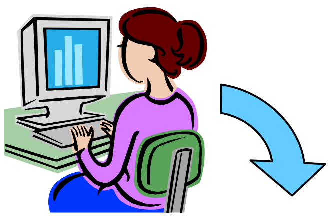


Assisted Data & Service Discovery: Current State

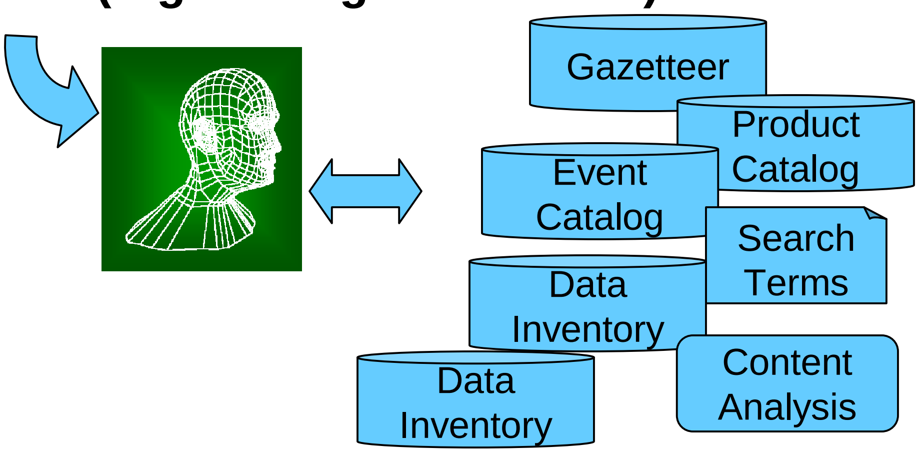


- Manual catalog searches result in dozens of similar datasets, many of which are unsuited to the intended use
- Inventory searches must be carefully constrained and user must know the exact data product needed, otherwise too much or too little data is returned
- Disparate catalog approaches impeded cross-catalog searches

Assisted Data & Service Discovery: Future Vision



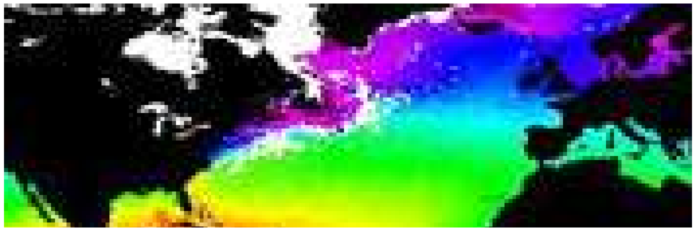
Select from
Semantic Web of Earth Data where
parameter=“esipfed:precipitation”
instrument=“gcmd:TRMM”
date=“between Sept 6 and Sept 16, 1996”
region=“ogc:South Atlantic”
phenomena = “esipfed:hurricane”
function= ‘rainfall(region=“ogc:Bermuda”) > 3’



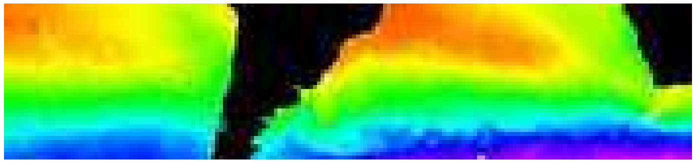
- Researcher uses semantic and content-based search to search for data using proper names, domain-specific jargon, and high-level specifications
- Researcher quickly finds data with the parameters, resolution, and coverage needed for the heat flux analysis

Interactive Data Analysis

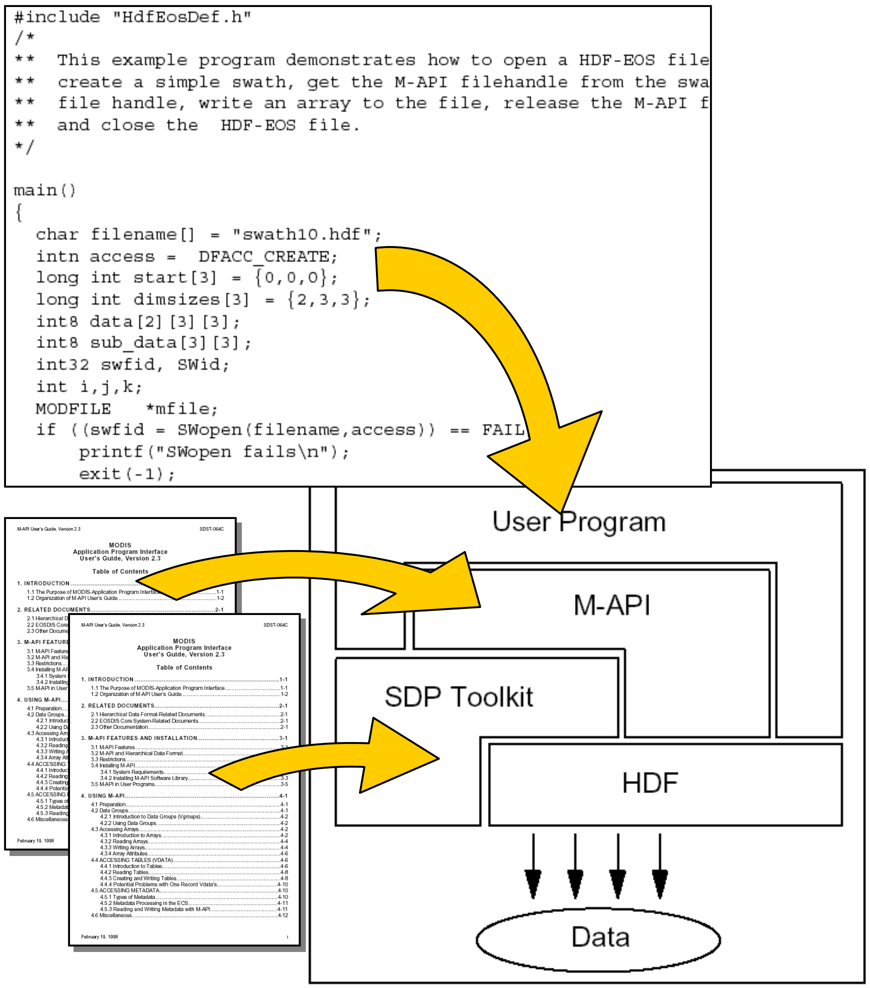
- Need
  - Researcher needs to implement a new algorithm in software to calculate ocean heat flux
- Vision
  - Reduce research algorithm implementation from months to hours
- *Enabling technologies*
  - *Visual grammars*
  - *Visual programming environments (Cantata, Triana, Grist/Viper, Wit)*
  - *High-level analysis tools (IDL, Matlab, Mathematica)*



$$\rho C_{p_g} u \frac{\partial T}{\partial x} = \lambda \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + G$$

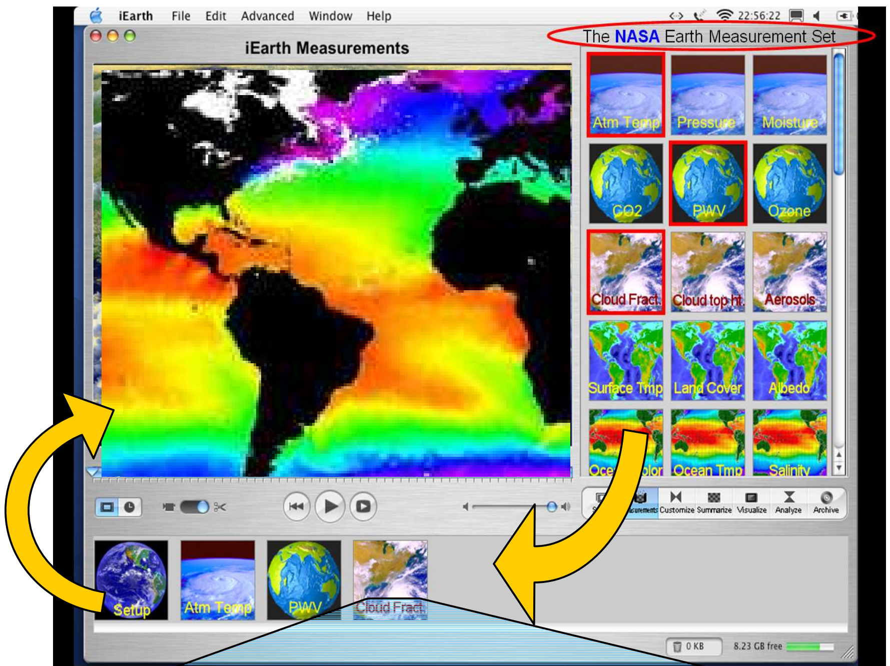


Interactive Data Analysis: Current State



- Coding, debugging, and deploying algorithms takes months of work
- Algorithms must be implemented by software engineers, not scientists, using custom procedural code
- Algorithm developers must learn complex application program interfaces for data manipulation and production control
- Monolithic programming & production environments do not support algorithm sharing

Interactive Data Analysis: Future Vision

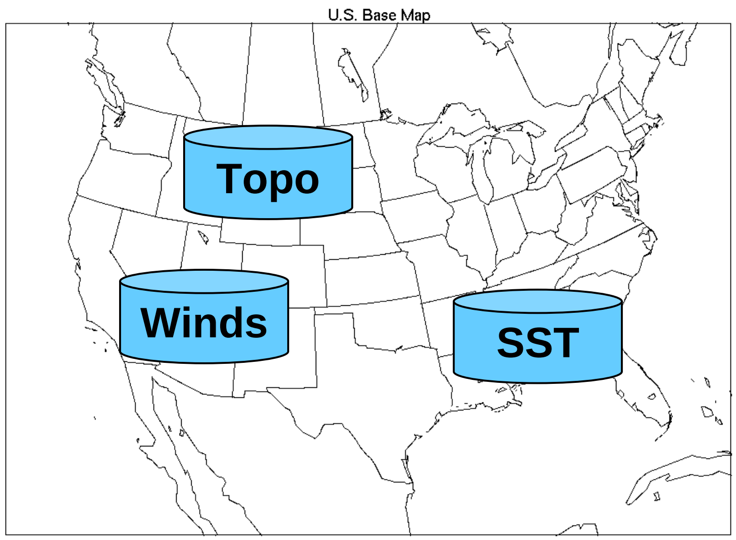


$$\rho C_{p_g} u \frac{\partial T}{\partial x} = \lambda \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + G$$

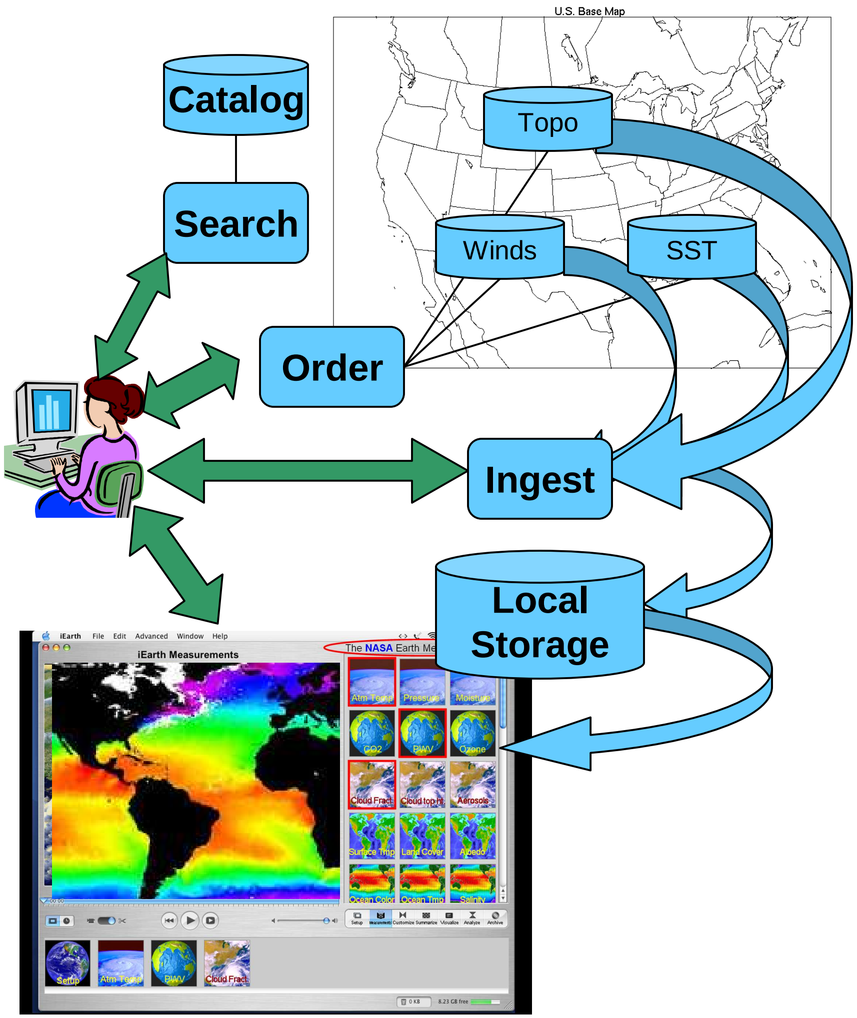
- Researcher uses a visual programming environment to create a new heat flux product in hours rather than months
- Researcher plugs useful transforms created by others into the visual programming environment as needed
- Researcher analyzes data with interactive tool to identify and quantify relationships between sea surface winds, temperature, topography, and heat transfer
- Researcher publishes analysis results as a data product for use in hurricane models

Seamless Data Access

- Need
  - Researcher needs to incorporate a variety of data such as sea winds, sea surface temperature, and ocean topography into the heat flux analysis
- Vision
  - Users can access current data from authoritative sources from any programming environment or analysis tool regardless of the data's physical location
- *Enabling technologies*
  - *Network data access protocols (OpenDAP, WMS/WCS, WebDAV, GridFTP)*
  - *Established data server tools (MapServer, DODS/LAS, ArcWeb)*
  - *Semantic metadata (OWL-S)*

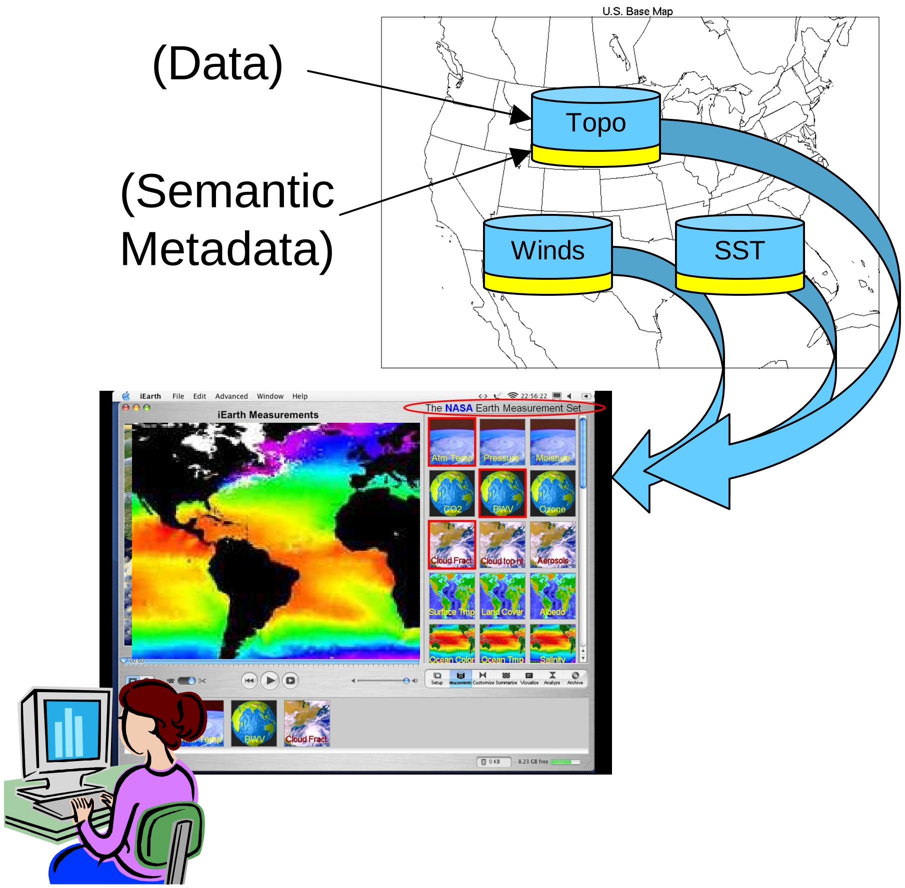


Seamless Data Access: Current State



- Data access is broken into separate search, order, and ingest processes
- Remote data products must first be imported into local storage systems before they can be accessed by analysis tools
- Different logins are required to access each data product
- Information on file format and data semantics is not bound to the data and must be manually interpreted

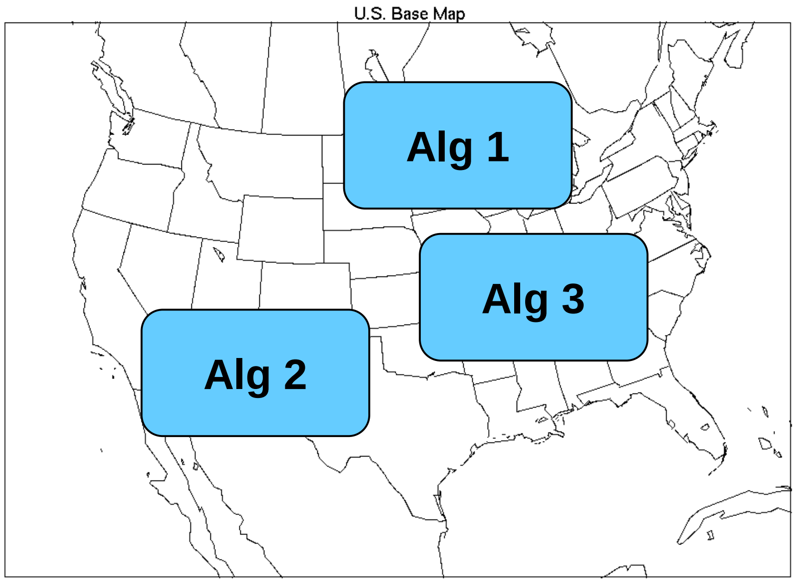
Seamless Data Access: Future Vision



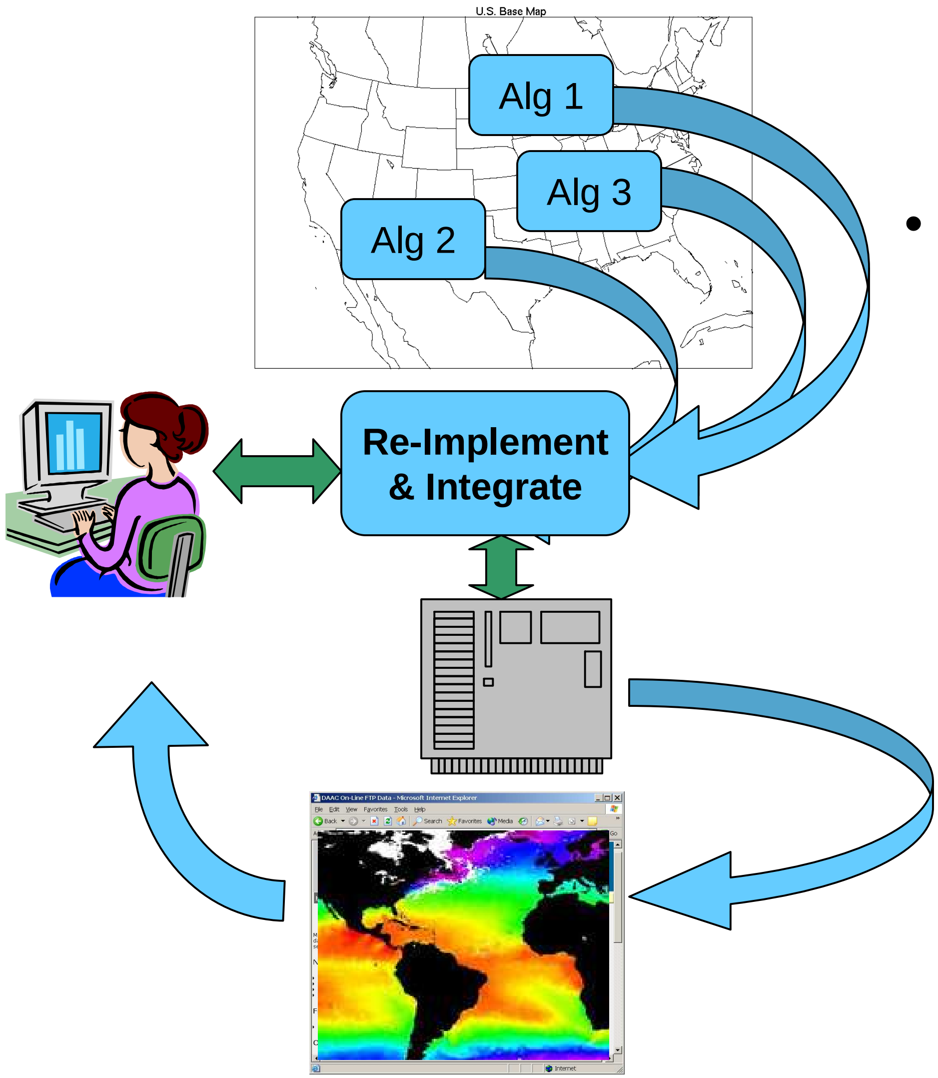
- Researcher simply opens remote datasets from within any analysis tool as if they were local
- Researcher obtains access to all datasets using single sign-on
- Sea winds, sea surface temperature, ocean topography, and other data are quickly incorporated into the heat flux analysis
- Data are correctly interpreted and automatically combined by the analysis tool using the associated semantic metadata

Interoperable Information Services

- Need
  - Researcher needs to incorporate algorithms available at remote locations into the local heat flux analysis
- Vision
  - Increase synergy in the Earth science community by leveraging in-place resources and expertise to provide information services on demand
- *Enabling technologies*
  - *Network service protocols (SOAP, Java RMI, OpenDAP, WS-*)*
  - *Grid toolkits (Globus)*
  - *Semantic metadata (OWL-S)*

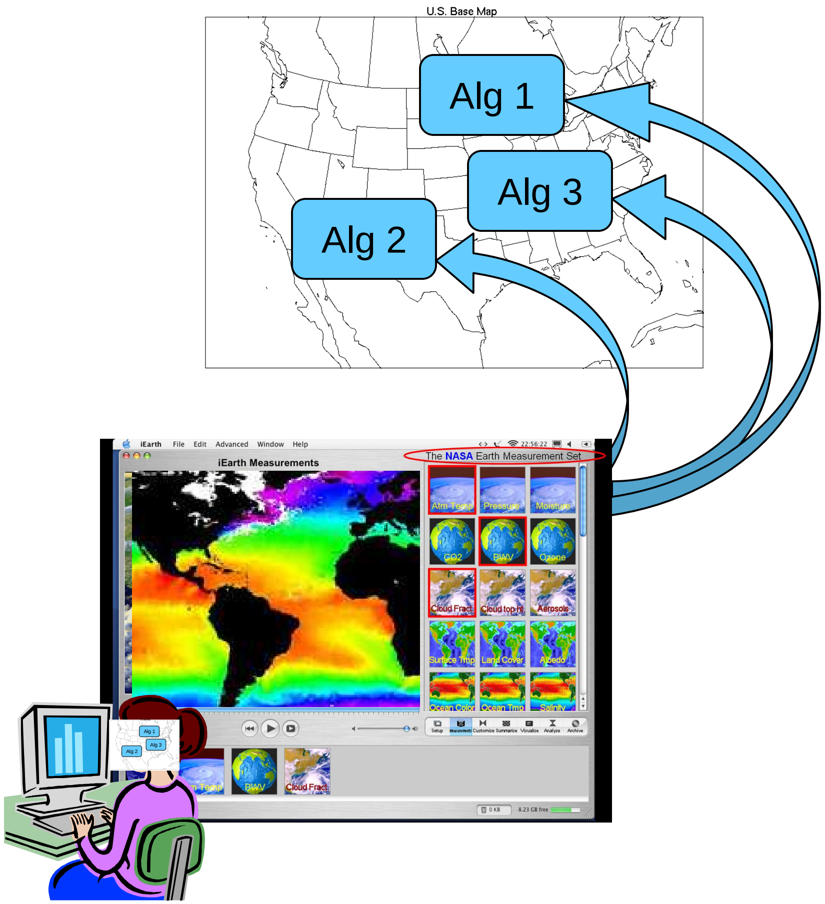


Interoperable Information Services: Current State



- Remote algorithms must first be ported to the local environment before they can be run
- Incompatibilities and dependencies sometimes result in recoding of the entire algorithm

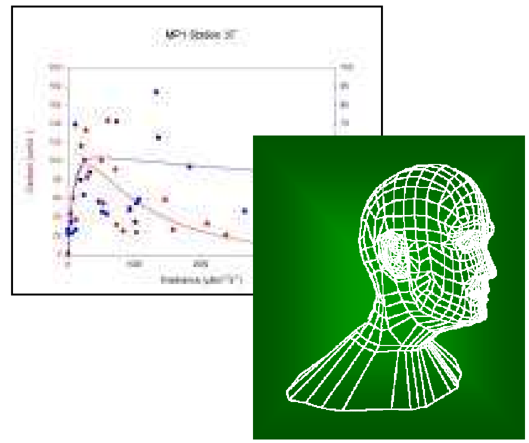
Interoperable Information Services: Future Vision



- Researcher simply invokes remote services from within the local analysis tool
- Ocean topography data is sent to proven services for sea roughness calculation and reprojection to enhance heat transfer calculation

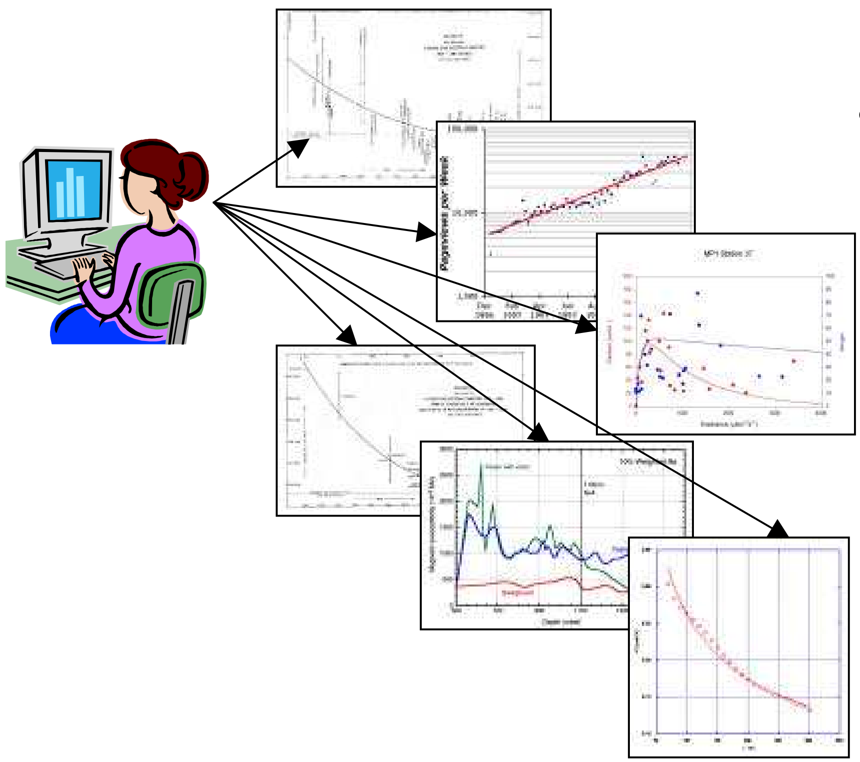
Assisted Knowledge Building

- Need
  - Researcher needs to determine how the storm track and other storm parameters affect storm surge
- Vision
  - Provide research and operations assistance using intelligent systems
- *Enabling technologies*
  - *Data mining algorithms (Support vector machines, independent component analysis, rule induction)*
  - *Data mining toolkits (Adam, D2K, Darwin)*
  - *Data mining plug-ins (IMAGINE, ENVI, ArcGIS)*



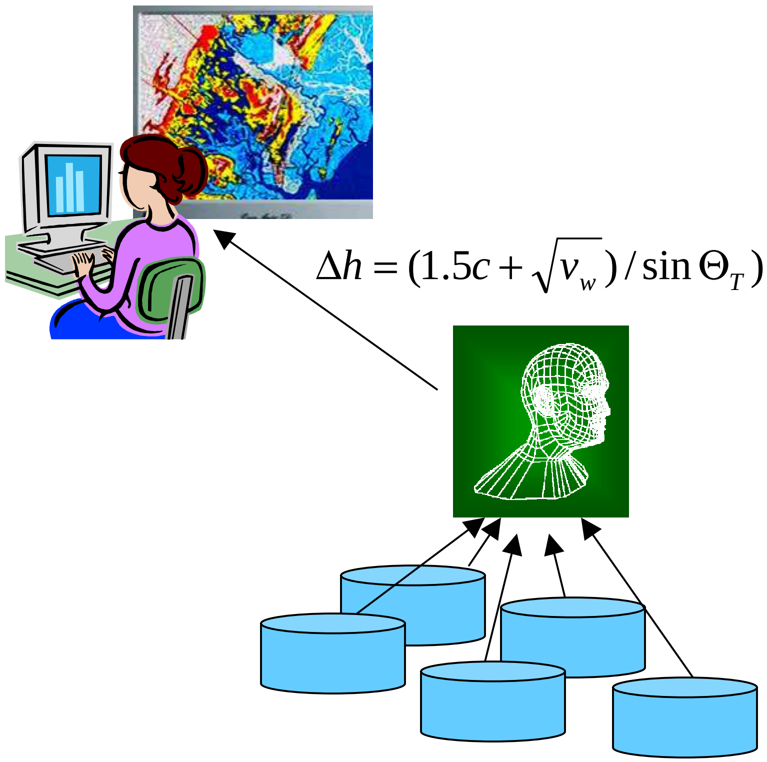
Assisted Knowledge Building: Current State

$$\Delta h = f(C, v_w, \Theta_T, \Phi, ?)$$



- Manual generation and testing of hypotheses regarding data interrelationships is time consuming and misses unexpected relationships.
- Manual analysis misses infrequent events and results in lost opportunities to collect additional data related to the event

Assisted Knowledge Building: Future Vision

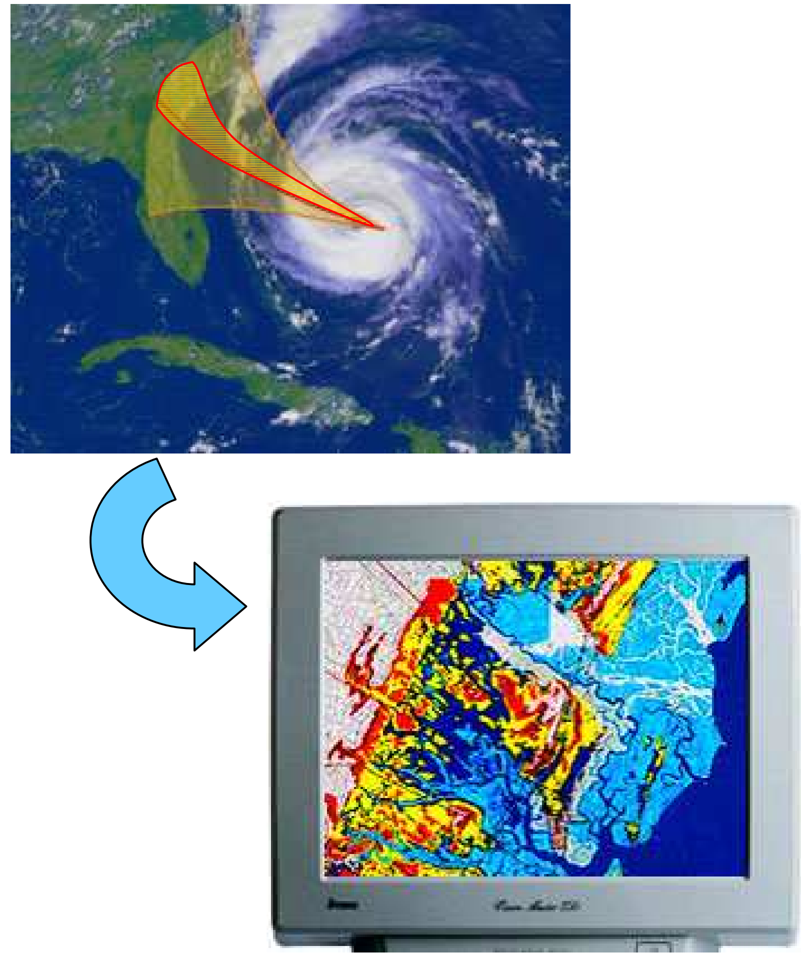


$$\Delta h = (1.5c + \sqrt{v_w}) / \sin \Theta_T)$$

- Data mining algorithms automatically infer a statistical model of storm surge based on storm size, angle of track, speed along track, wind speed, lunar phase, coastal shelf depth, and other parameters
- Researcher combines the inferred model and physical models to create a precision storm surge model

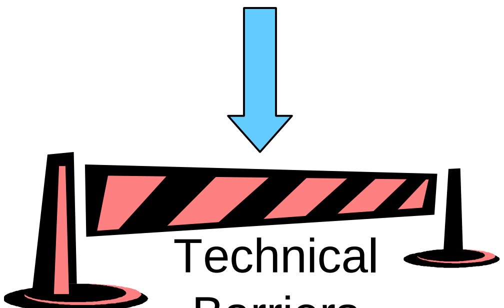
Community Modeling Frameworks

- Need
  - Researcher needs to couple hurricane forecast model to storm surge model to create more accurate predictions of coastal inundation
- Vision
  - Enable linked and ensemble models for improved predictive capability
- *Enabling technologies*
  - *Multi-model frameworks (ESMF, Tarsier, MCT, COCOLIB)*
  - *Model data exchange standards (BUFR, GRIB)*
  - *Semantic metadata (OWL-S)*



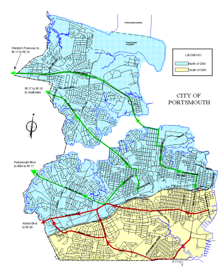
Community Modeling Frameworks: Current State

Storm Prediction Information



Technical Barriers

Inundation Model

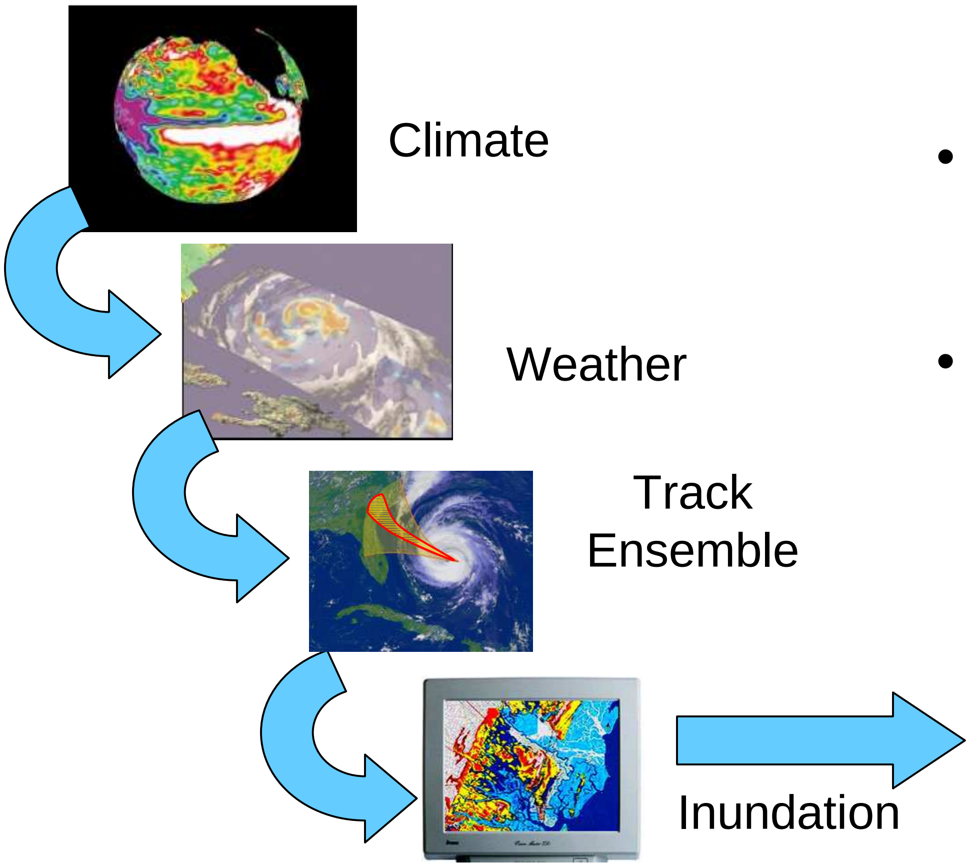


Evacuation Planning

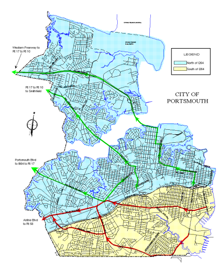
Relief Planning

- Disparate and non-interoperable modeling environments with language and OS dependencies
- Scientific models and remote sensing observations rarely connected directly to decision support systems
- Evacuation and relief planning based largely on historical averages and seat-of-the-pants estimates

Community Modeling Frameworks: Future Vision



- Researcher combines multiple models into an ensemble model to forecast the hurricane's track
- Researcher couples the storm track model to the storm surge model
- Analyst assesses property and transportation impact in decision support system fed by storm surge/inundation model

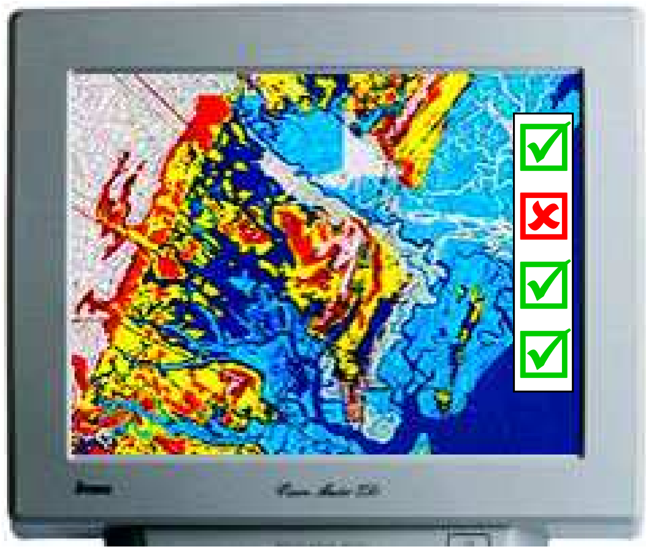


Evacuation Planning

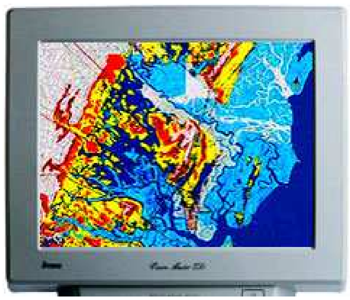
Relief Planning

Verifiable Information Quality

- Need
  - Relief and evacuation planners need to assess the quality of the coastal inundation prediction, which has been based on a long chain of calculations
- Vision
  - Provide confidence in information products and enable the community information provider marketplace
- *Enabling technologies*
  - *Data pedigree algorithms (Ellis)*
  - *Machine-readable formats (XML) and semantics (OWL-S)*



Verifiable Information Quality: Current State



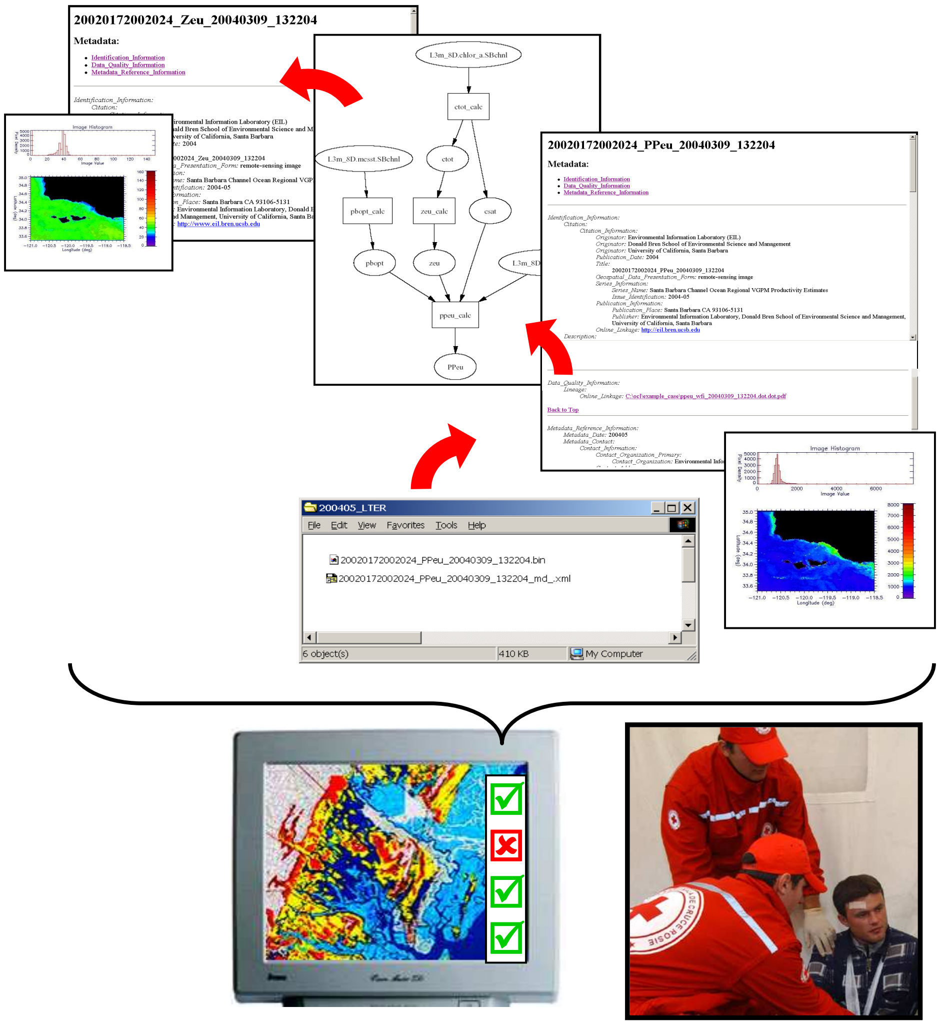
?

Inundation Prediction

Relief Planning

- End user has little insight into the quality of the analysis
- Data quality is sometimes implicit or assumed based on provider or dataset reputation
- Non-standard quality indicators cannot be automatically interpreted by COTS analysis software and are sometimes overlooked
- No machine-readable, standard representation of data lineage

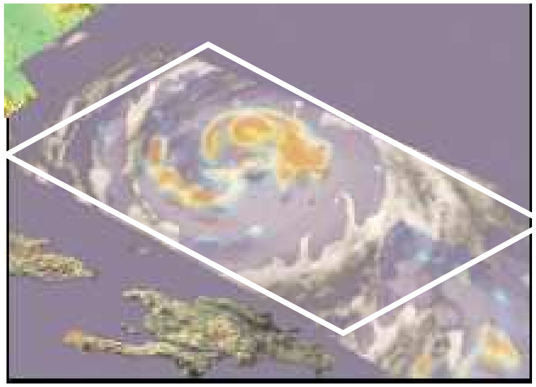
Verifiable Information Quality: Future Vision



- Users can easily explore data pedigree determine its reliability
- Commercial tools understand data quality flags and automatically handle issues such as missing data
- Researcher and end user can quantify the quality of the inundation prediction and use the results appropriately

Responsive Information Delivery

- Need
  - Researcher needs current storm data to update the storm track prediction
- Vision
  - Ensure research priorities are met and enable new uses of Earth science data
- *Enabling technologies*
  - *Optical networks (National LambdaRail)*
  - *Peer-to-peer networks with swarming (Modster)*
  - *Direct downlink (MODIS/AIRS DDL)*



Responsive Information Delivery: Current State

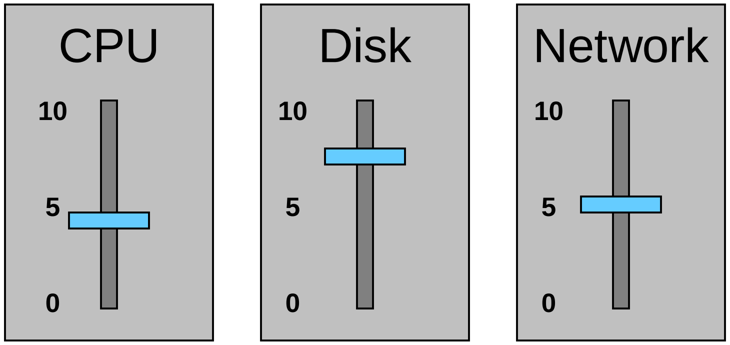
- Static products delivered weeks after collection
- Data is stored, cataloged, and delivered in granules that reflect processing and storage constraints more than end user needs
- Network delivery is slower and more expensive than physical media delivery
- First-come first-served data dissemination regardless of intended use

Responsive Information Delivery: Future Vision

- Automated data quality assurance and autonomous operations are used to expedite time-critical data
- Researcher obtains storm data within minutes of sensor overpass based on the application’s assigned priority
- Data are delivered in the preferred format specified in the researcher’s profile
- Data are delivered with the extents and parameter subsets specifically needed by the storm track model

Evolvable Technical Infrastructure

- Need
  - Researcher needs to take advantage of new processing, storage, and communications technologies to improve performance and reduce costs
- Vision
  - Exploit emerging technologies quickly
- *Enabling technologies*
  - *Processor & storage virtualization software (VMware, volume manager)*
  - *Scalable architectures (Beowolf, Grid)*
  - *Bandwidth-on-demand*



Evolvable Technical Infrastructure: Current State

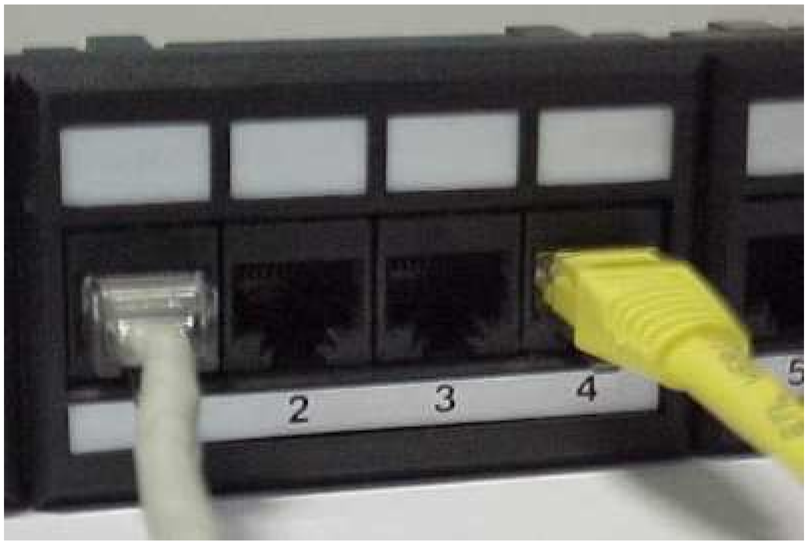
Old

Migration
•Data
•Software

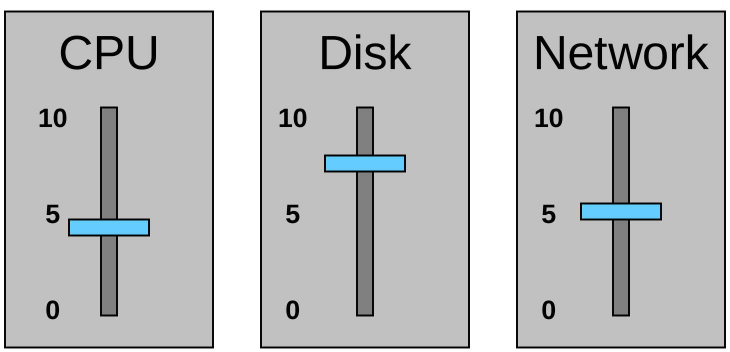
New

- Network capacity established early in mission and difficult to change
- Processing, storage, and communications upgrades are difficult and disruptive
  - Manual migration of data
  - Cutover is risky, and parallel operations are costly
  - Communication outages common during upgrades
- Non-standard interfaces impede introduction of new technologies

Evolvable Technical Infrastructure: Future Vision



Old New

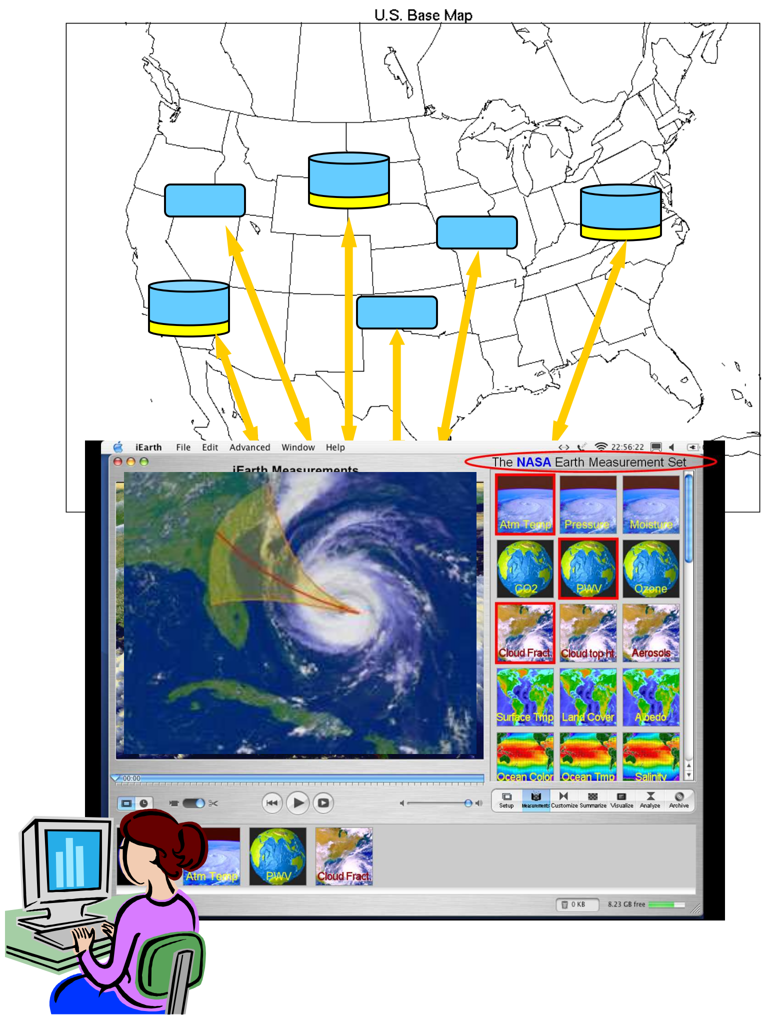


- Researcher simply plugs in new equipment to meet storm track model demands
- Researcher places on-line order for additional processing, storage, and communications capacity based on requirements and budget
- Additional capacity is obtained within minutes
- Data and processes automatically migrate to take advantage of new equipment or capacity

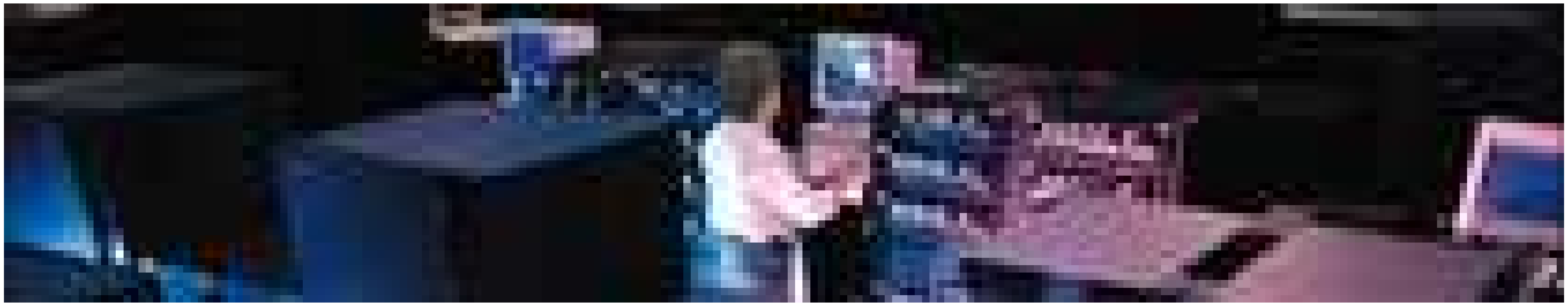
Focused Effort on Key Capabilities will Enhance Earth Science Community Capabilities

The envisioned capabilities empower researchers to...

- Quickly distill petabytes of data into usable information and knowledge
- Achieve new analysis & modeling results
- Build a community geospatial knowledge network that advances Earth science



Envisioned Capabilities Help Us Understand the Challenge In an Actionable Way



Scalable Analysis Portals	Community Modeling Frameworks

Assisted Data & Service Discovery	Assisted Knowledge Building

Interactive Data Analysis	Seamless Data Access	Interoperable Information Services	Responsive Information Delivery	Verifiable Information Quality

Evolvable Technical Infrastructure

Contributors

- Karen Moe
- Rob Raskin
- Peter Cornillon
- Tom Yunck
- Karl Benedict
- Liping Di
- Elaine Dobinson
- Jim Frew
- Kerry Handron
- Rudy Husar
- David Isaac
- Brian Wilson
- Oscar Casteneda
- Wenli Yang
- Other members of the Technology Infusion Working Group
- Many workshop participants