



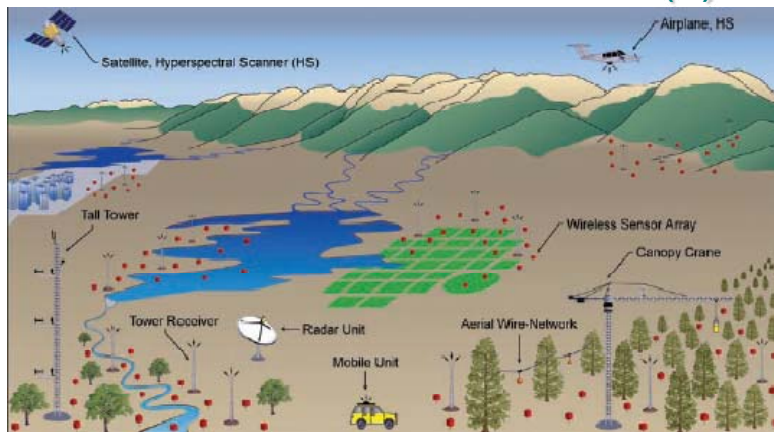
# Science Data Life Cycle

IST400/600

Instructor: Jian Qin



## Scenarios of science data (1)



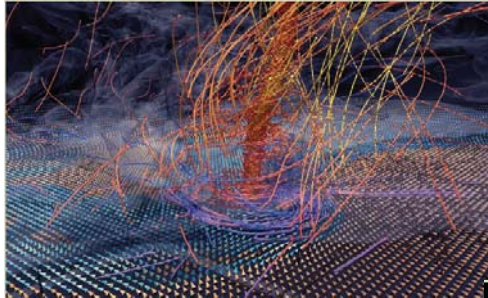
An artist's conception (above) depicts fundamental NEON observatory instrumentation and systems as well as potential spatial organization of the environmental measurements made by these instruments and systems.

3/23/2009

NSF. (2007). *Cyberinfrastructure Vision for 21st Century Discovery*.  
<http://www.nsf.gov/pubs/2007/nsf0728/index.jsp>

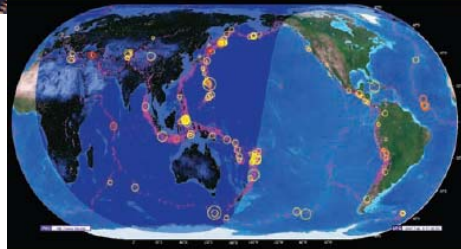
2

## Scenarios of science data (2)



Visualization of volcano simulation created from data generated by a tornado simulation calculated on the NCSA computing cluster, shows the tornado by spheres colored according to pressure. Orange and blue tubes represent the rising and falling airflow around the tornado.

The IRIS Seismic Monitor System allows scientists and others to monitor global earthquakes in near real-time, visit seismic stations world-wide, and search the web for earthquake information.



3/23/2009

3

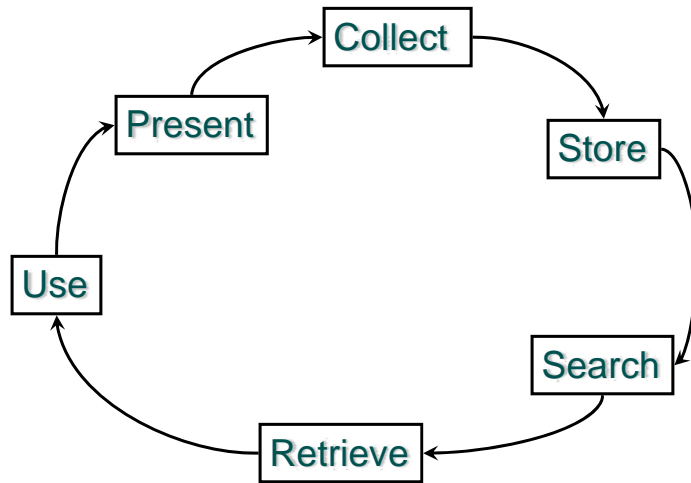
## What do the scenarios tell us?

- Science data are collected or generated
  - From
    - observations of natural world
    - simulations of models
    - experiments in laboratories
  - By using
    - Instruments and devices
    - Computing resources

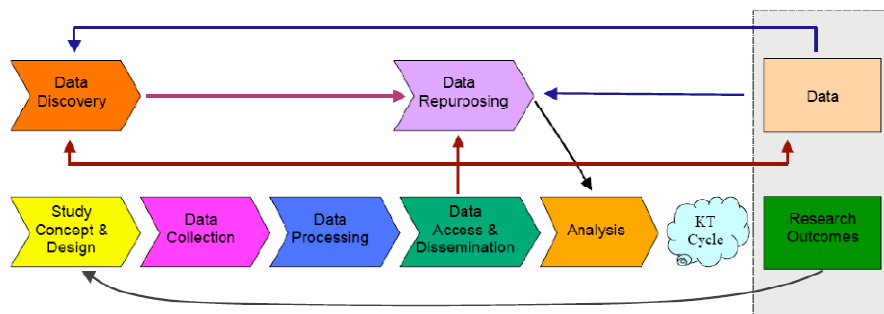
3/23/2009

4

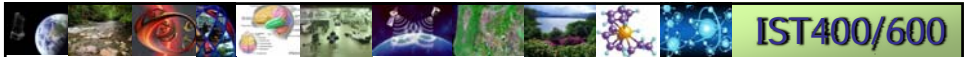
### Science data life cycle



### E-Science and the life cycle of research




The "KT Cycle" in the diagram represents the processes of knowledge transfer. This life cycle diagram comes from Charles Humphrey, "E-Science and the Life Cycle of Research" (2006) available online at <http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc>




## Four categories of issues

- Science-based issues
- Data management issues
- Policy issues
- Technical issues

3/23/2009
7



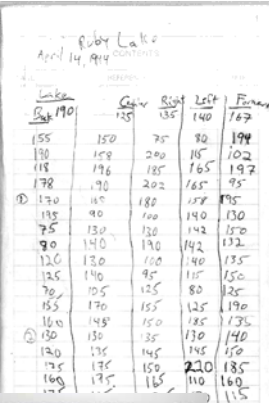
## Collection




Satellite imagery

- Protocols
  - Manual records
  - Sampling strategies
- Physical formats
  - Digital
  - Analog
- Metadata capture
  - Context and connections

Field notebooks




| Lat | Long | Depth | Temp | Salinity |
|-----|------|-------|------|----------|
| 155 | 150  | 75    | 80   | 194      |
| 130 | 158  | 200   | 165  | 102      |
| 118 | 176  | 185   | 165  | 197      |
| 178 | 170  | 202   | 165  | 95       |
| 170 | 165  | 180   | 158  | 195      |
| 135 | 140  | 100   | 140  | 130      |
| 75  | 130  | 130   | 142  | 150      |
| 90  | 140  | 190   | 142  | 132      |
| 120 | 130  | 100   | 140  | 135      |
| 125 | 140  | 95    | 135  | 150      |
| 70  | 105  | 125   | 80   | 128      |
| 155 | 170  | 155   | 125  | 190      |
| 160 | 145  | 150   | 125  | 135      |
| 130 | 130  | 135   | 130  | 140      |
| 120 | 135  | 145   | 145  | 150      |
| 175 | 175  | 150   | 210  | 185      |
| 160 | 175  | 165   | 110  | 160      |
| 100 | 175  | 165   | 110  | 115      |
| 100 | 175  | 165   | 110  | 145      |



Water survey

3/23/2009
8



IST400/600

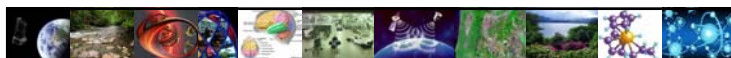
## Storage

- Archiving
  - Security / stability
  - Maintenance
- Accessibility
  - Online vs. offline
  - Multiple formats
- Metadata capture
  - Context and connections

Example:  
- Abandoned data (see handout)

3/23/2009

9



IST400/600

## Search


- Discovery
  - Existence
  - Characteristics
- Directory services
  - Scope and custody
- Metadata
  - interoperability

Examples

- [Global Change Master Directory](#)
- Data product handbooks
  - [Digital Elevation Model Standards](#)
- [THREDDS](#)

3/23/2009

10




## Retrieval

- Queries
  - Attributes
  - Conditions
- Subsetting
  - Projection
  - Subsampling

- SQL
- URL

3/23/2009 11




## Use

- Analysis and synthesis
  - Types, e.g. discrete vs. continuous data
  - Formats
  - Auditing, e.g. tracking process steps

### Examples

- Data visualization & analysis platform: IDL (<http://www.itvis.com/idl/>)
- MATLAB – the language of technical computing (<http://www.mathworks.com/products/matlab/>)
- Models
- ad hoc programs

3/23/2009 12




## Presentation and publication

- Visualization
  - Media
  - Metaphors
- Formats
  - Displayability
- Results vs. intermediate steps
  - Data vs. conclusions

Examples

- Poster
- Web page
- Article
- CD-ROM
- Video
- Web service
- Web site

3/23/2009 13



## Summary

- Each stage in the science data life cycle involves some issues
  - Science-based
  - Data management
  - Policy
  - Technical
- Goal of science data management
  - Access in short and long term
  - Use for various purposes by various groups of users

3/23/2009 14