Community Approaches to Hydrologic Data and Model Sharing using Cloud Resources

David Tarboton, Ray Idaszak, Jeffery Horsburgh, Dan Ames, Jon Goodall, Larry Band, Venkatesh Merwade, Alva Couch, Jennifer Arrigo, Rick Hooper, David Valentine

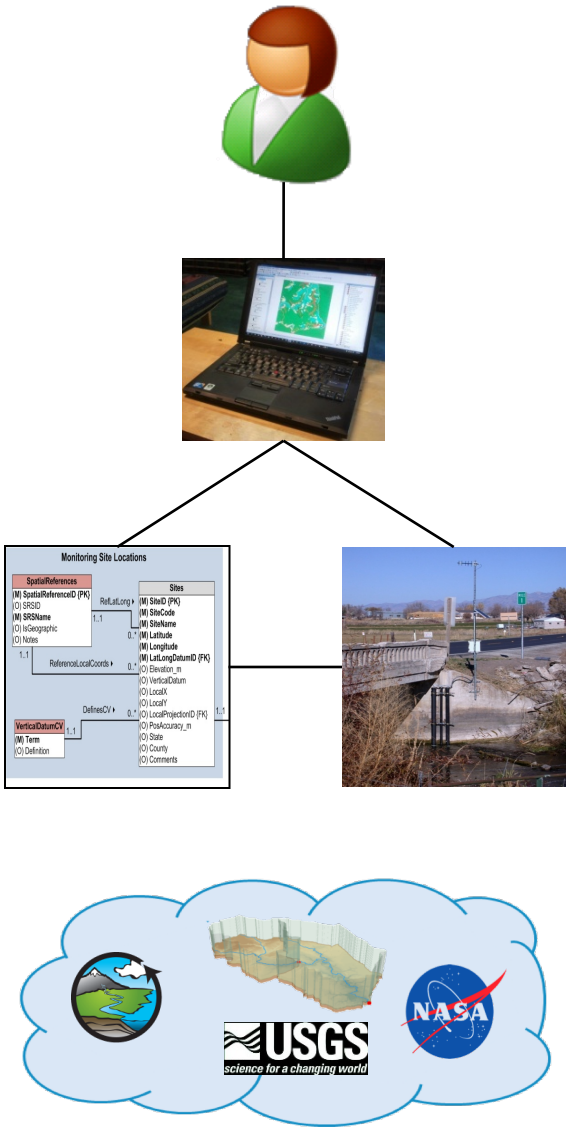
http://his.cuahsi.org
http://www.hydroshare.org

HYDROSHARE

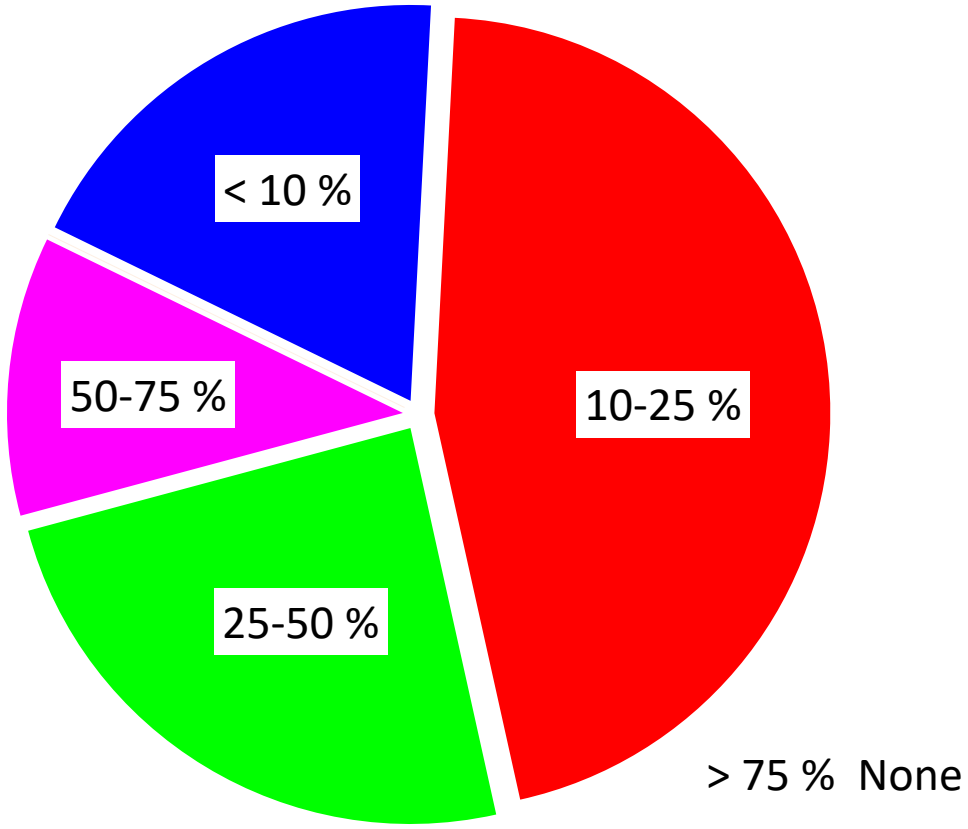
EAR 0622374 (2007-2012)
OCI-1148453 (2012-2017)
OCI-1148090 (2012-2017)

Outline

- Cloud Computing
- The CUAHSI HIS
  - A Services-Oriented Architecture Based System for Sharing Hydrologic Data
- HydroShare
  - A Web-Based Collaborative Environment for the Sharing of Hydrologic Data and Models



What proportion of your research time do you spend on preparing or preprocessing data into appropriate forms needed for research purposes?

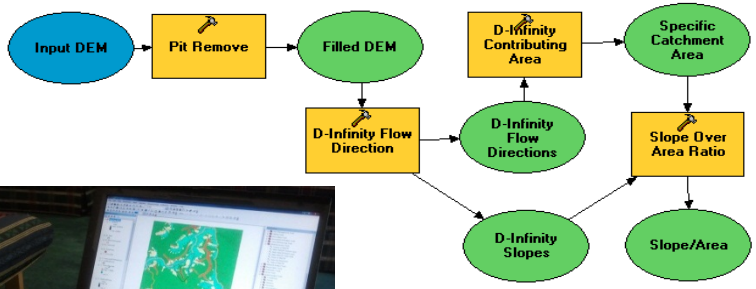


http://his.cuahsi.org/documents/HISStatusSept15.pdf

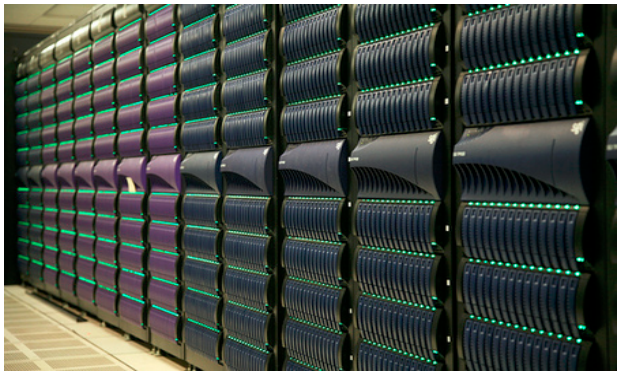
Do you have the access or know how to take advantage of advanced computing capability?

Researchers

- Experimentalists
- Modelers



HPC Specialists



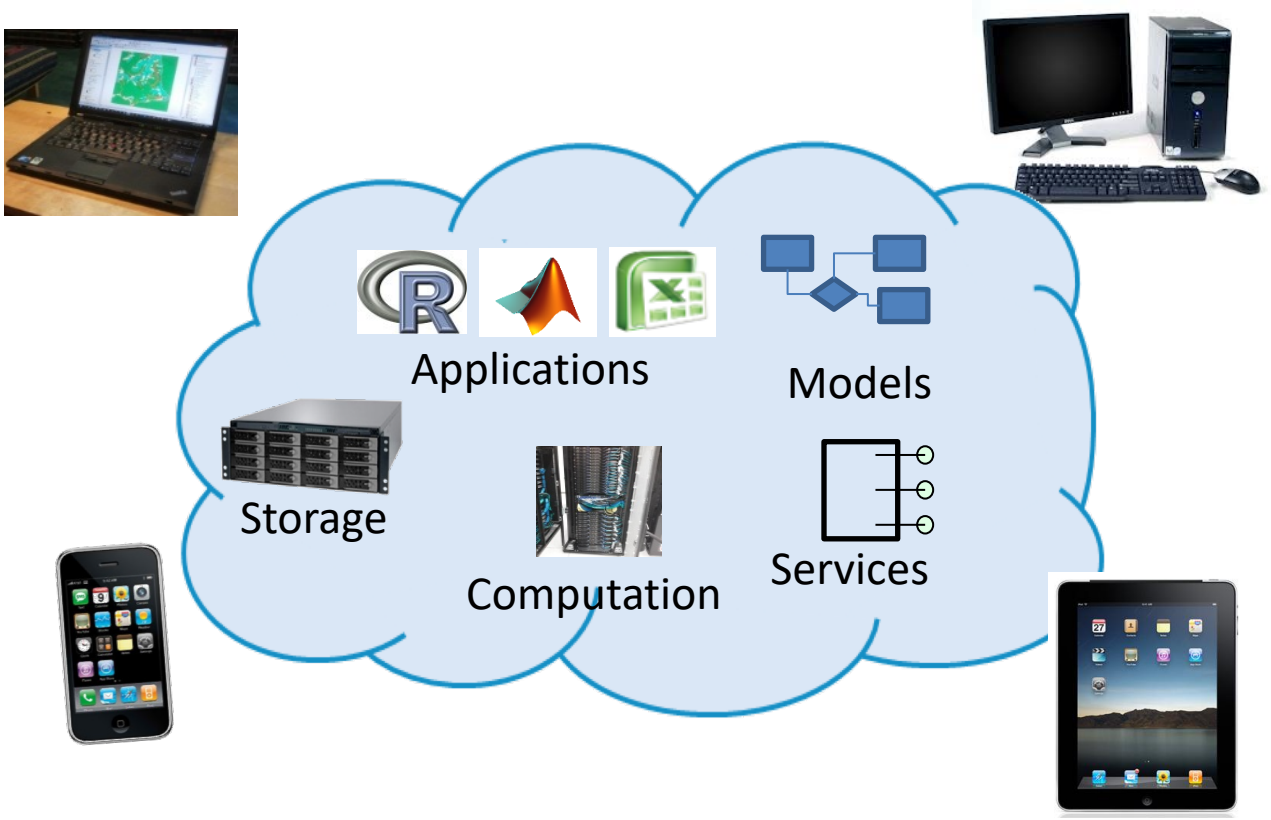
#!/bin/bash vi chmod

#PBS -l nodes=4:ppn=8

grep awk mpiexec

```
-bash-3.2$ ls tddata
logan               LoganOutlet.sbn  LoganOutlet.shp      LoganOutlet.shx
LoganOutlet.dbf  LoganOutlet.sbx  LoganOutlet.shp.xml
-bash-3.2$ ls tddata/logan
logan.tif
-bash-3.2$ ls
eric    logMPfel        run.bash        taudem.bash     taudem_submit.sh
logMP   run_all.bash    run_taudem.sh   taudem.o41959   tddata
-bash-3.2$ run_taudem.sh pitremove -z logan -fel loganfel
43058.ib-net
-bash-3.2$
```

Cloud Computing

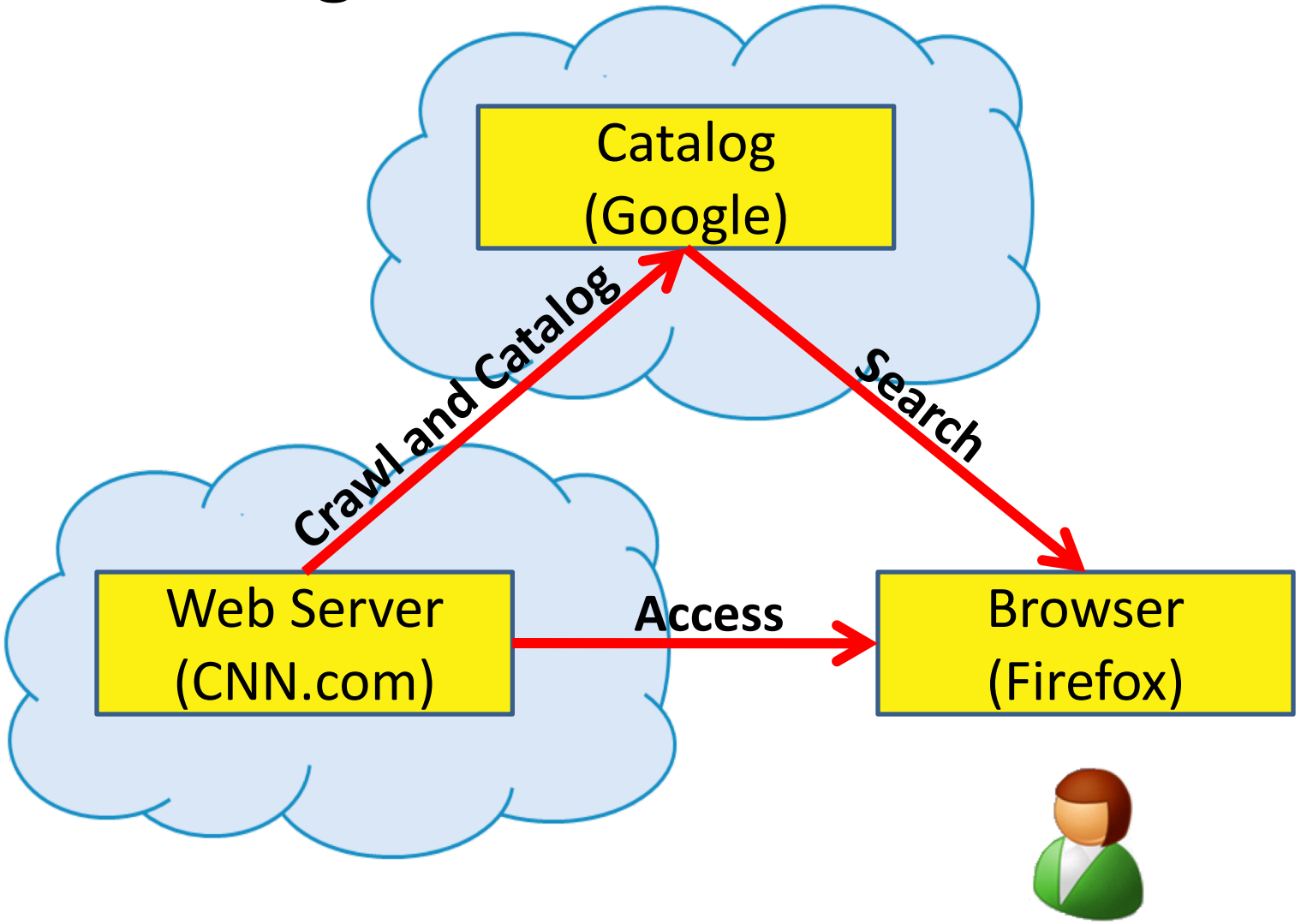


Wikipedia: Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet)

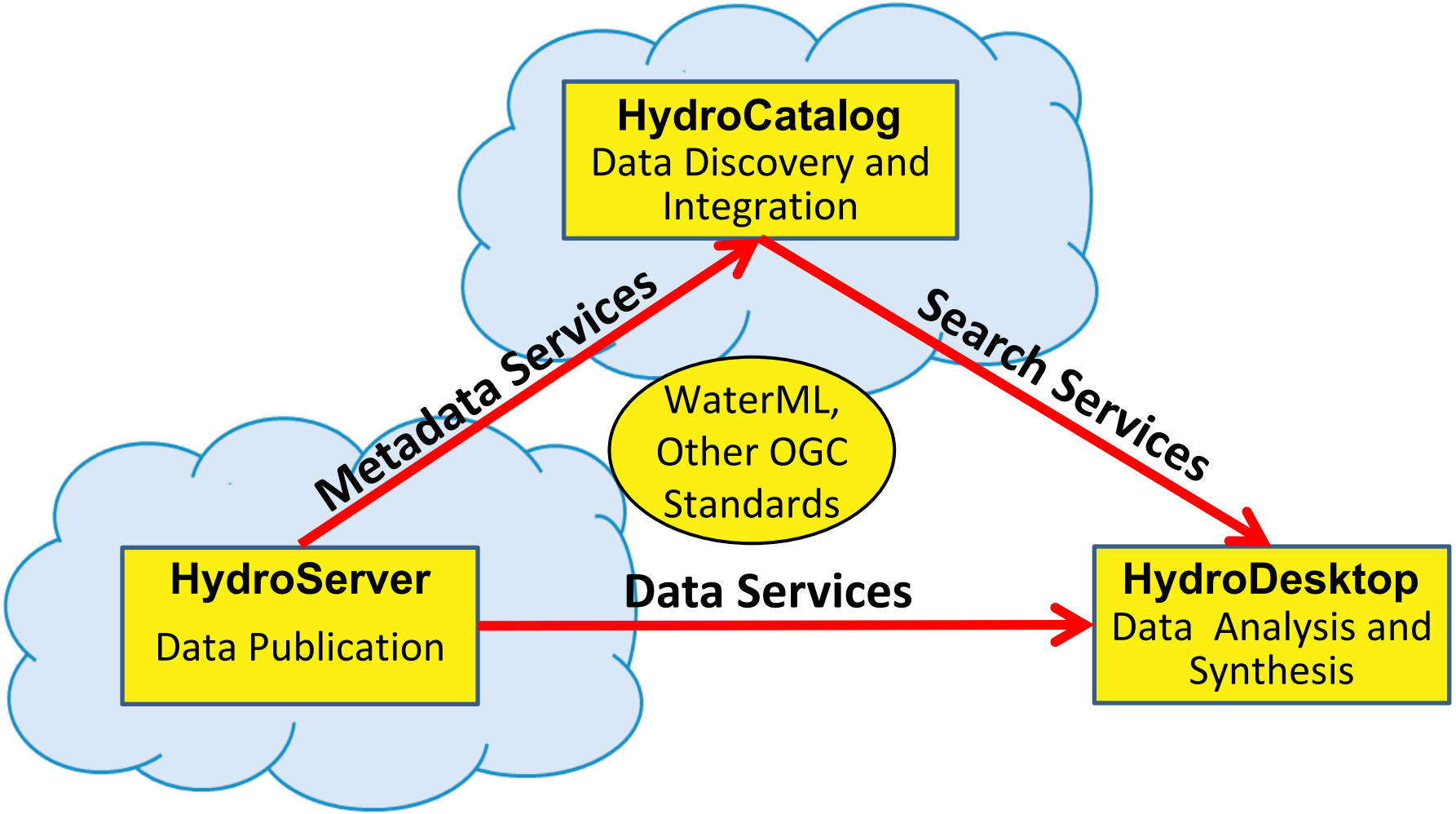
Google, Amazon, Microsoft, Apple, DropBox

XSEDE, Condor, BOINC

Cloud Services-Oriented Architecture Paradigm of the World Wide Web



CUAHSI Hydrologic Information System: A Services-Oriented Architecture Based System for Sharing Hydrologic Data



Hydrologic Data Challenges

- From dispersed federal agencies
- From investigators collected for different purposes
- Different formats
  - Points
  - Lines
  - Polygons
  - Fields
  - Time Series

Data Heterogeneity

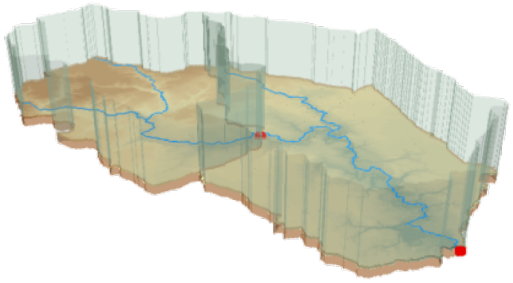
Water quality

Water quantity

Rainfall and Meteorology

Soil water

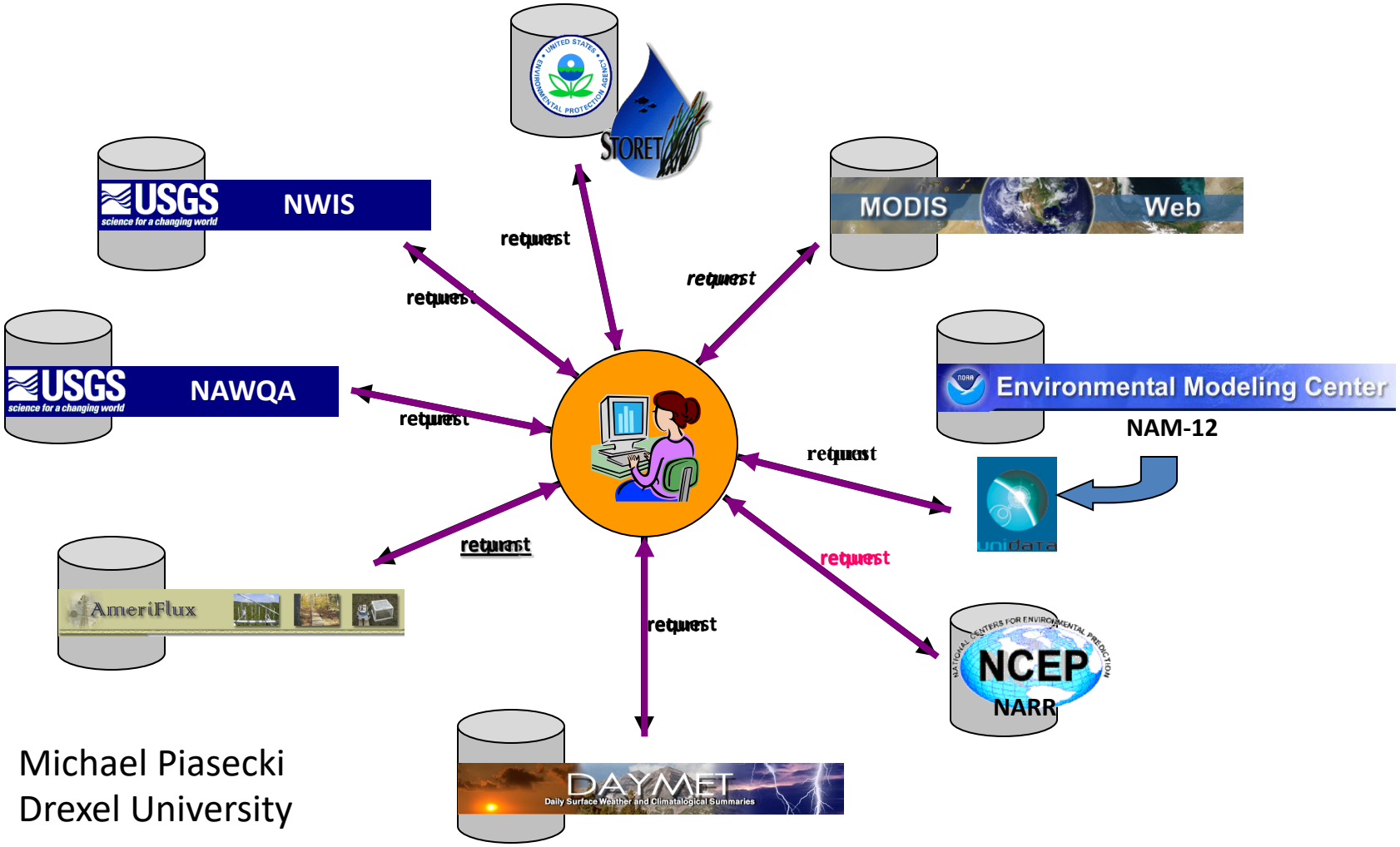
GIS



Groundwater

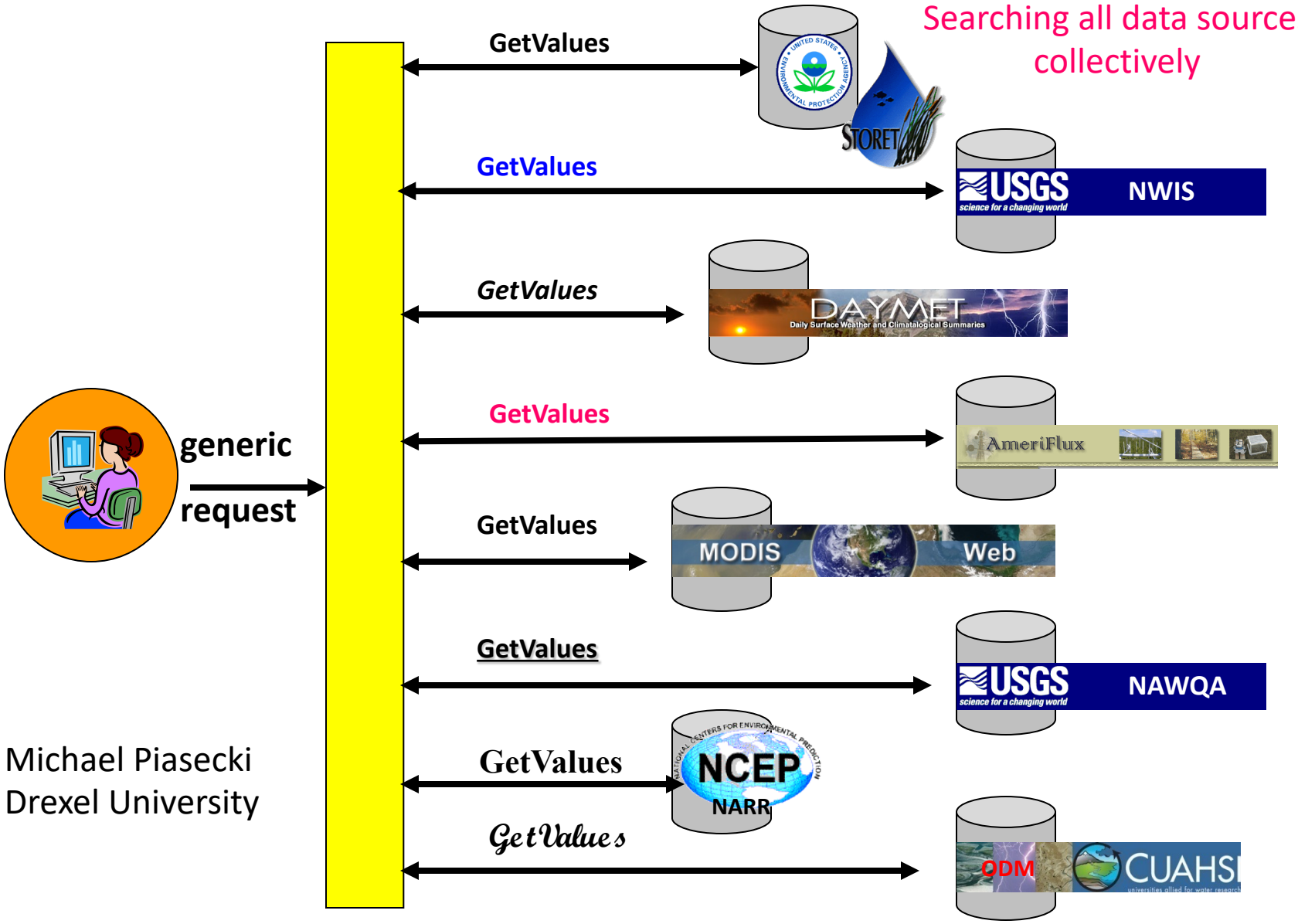
Data Searching – What we used to have to do

Searching each data source separately



Michael Piasecki
Drexel University

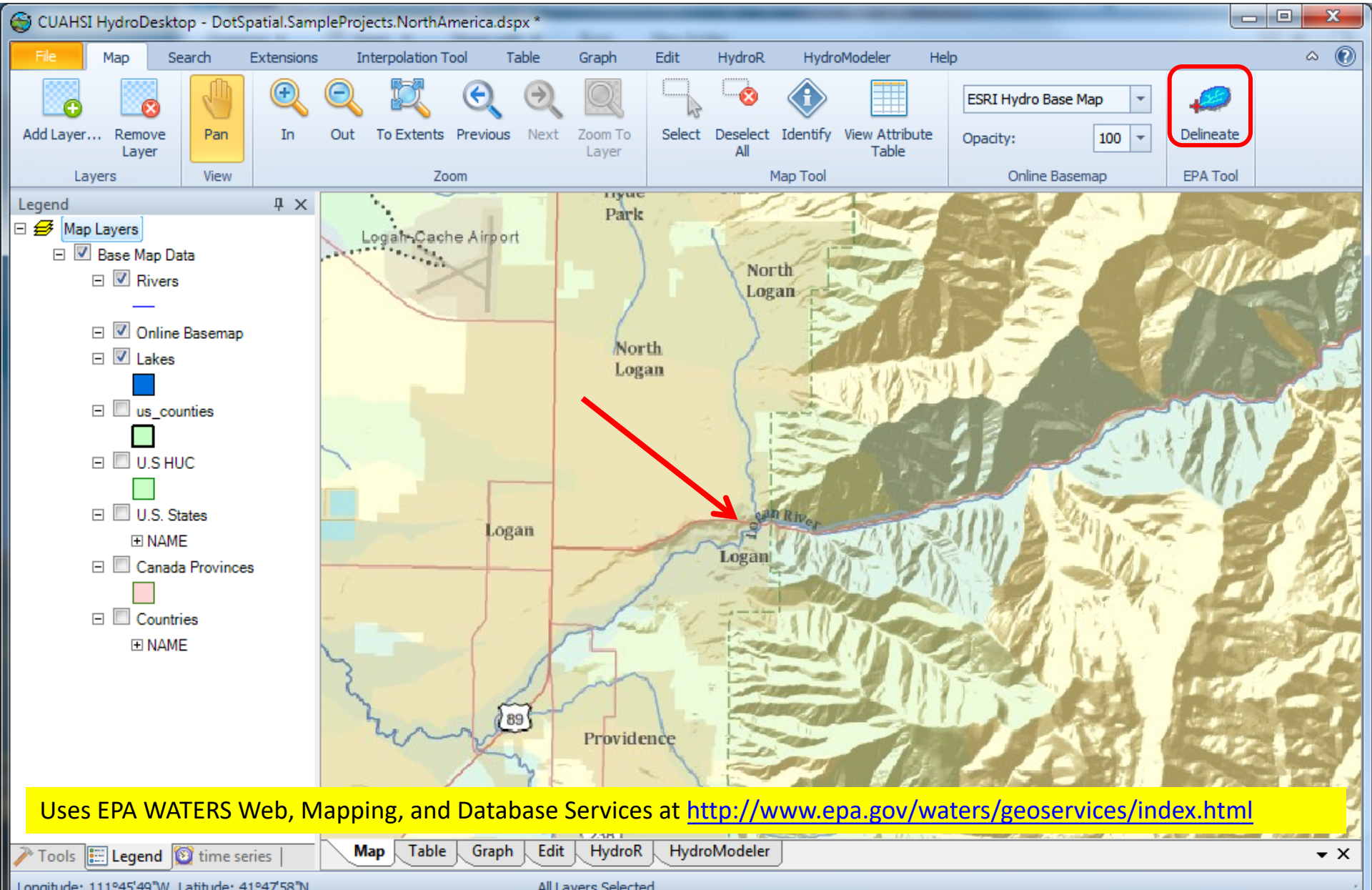
What CUAHSI HIS enables



Michael Piasecki
Drexel University

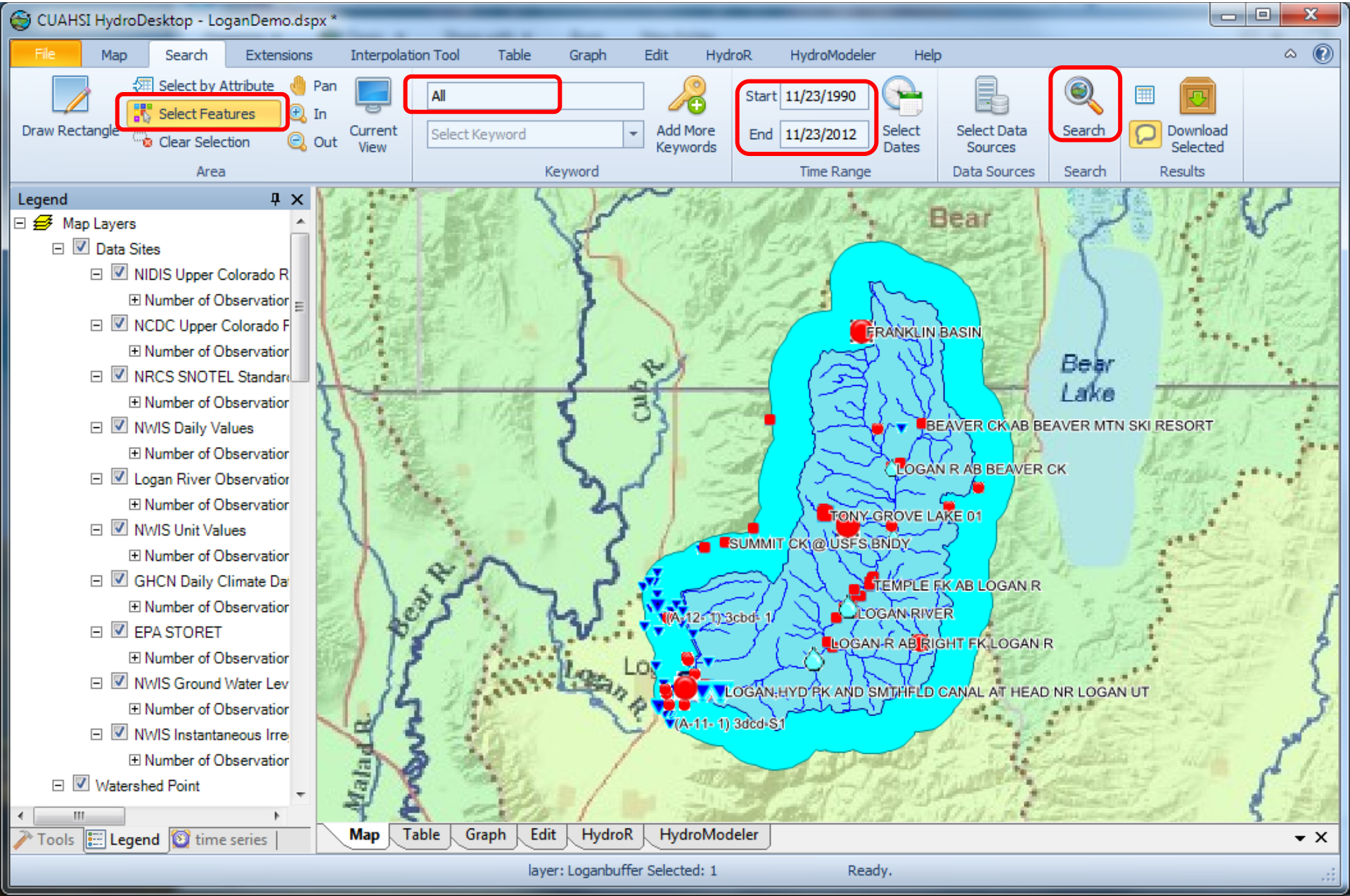
HIS Example.

1. Delineate Watershed using EPA Web Services

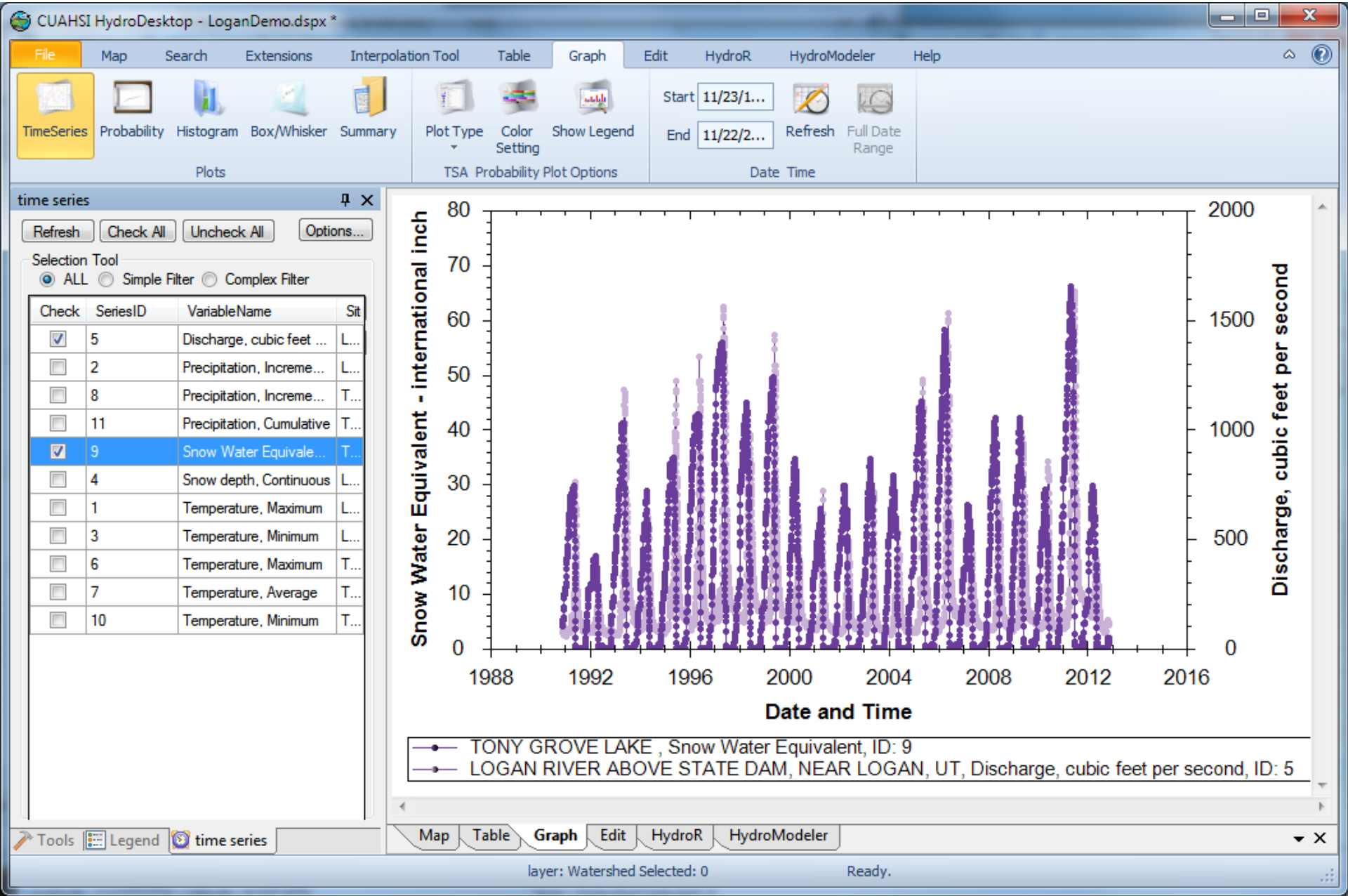


Uses EPA WATERS Web, Mapping, and Database Services at http://www.epa.gov/waters/geoservices/index.html

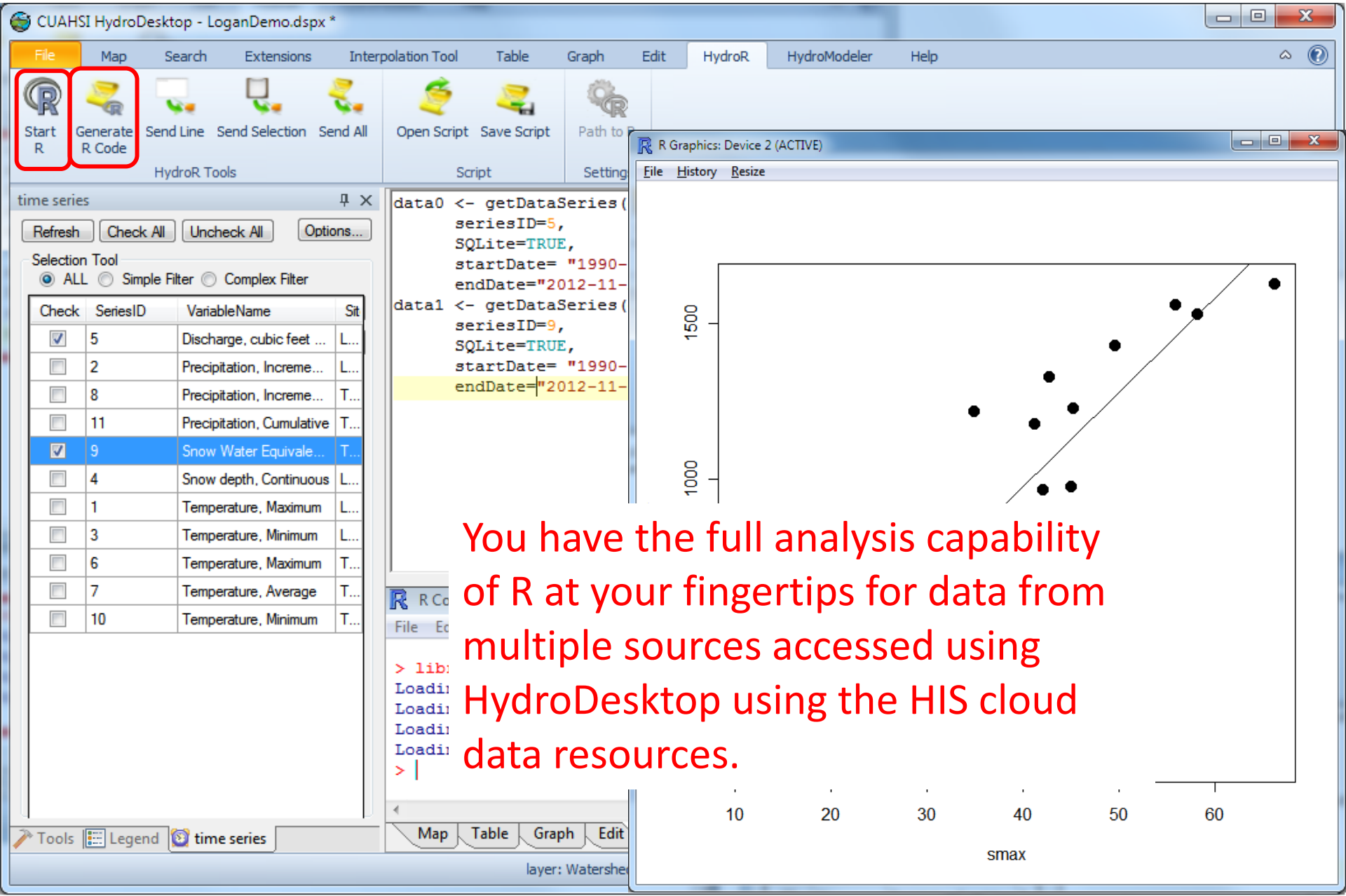
2. Search last 22 years for all data in buffer around watershed



3. Download and plot data from multiple sources



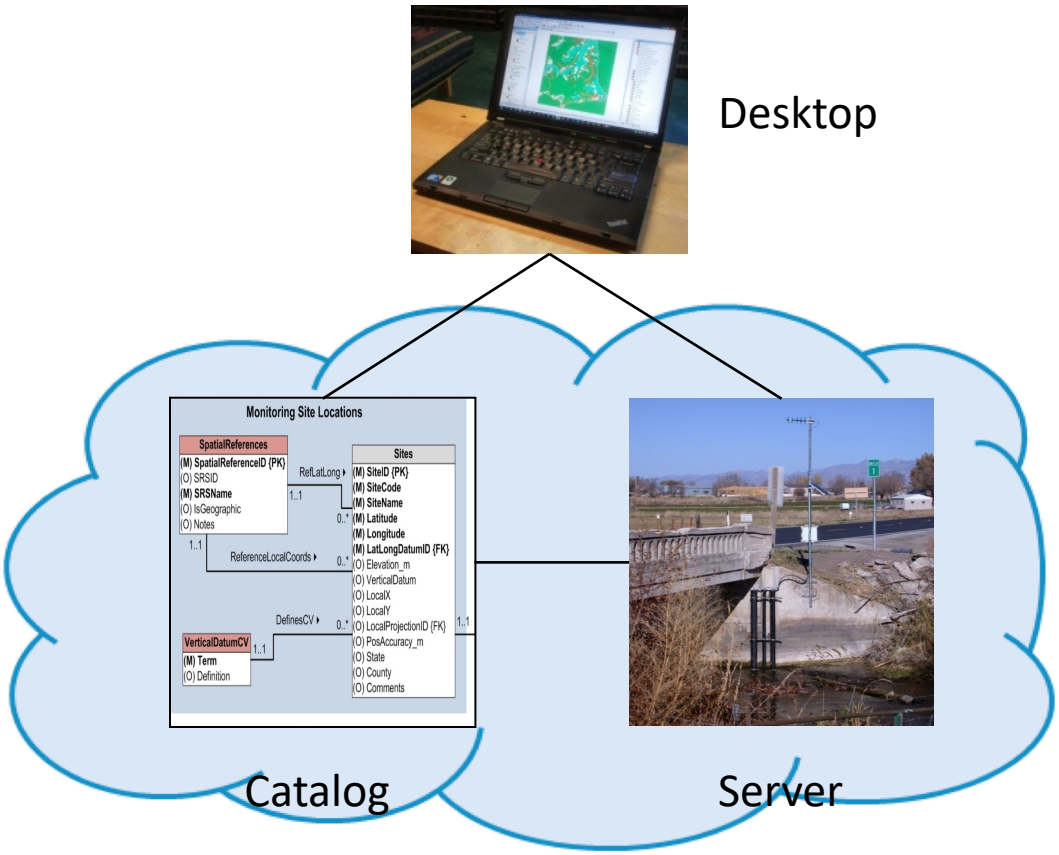
4. Analyze with R using the HydroR plugin



You have the full analysis capability of R at your fingertips for data from multiple sources accessed using HydroDesktop using the HIS cloud data resources.

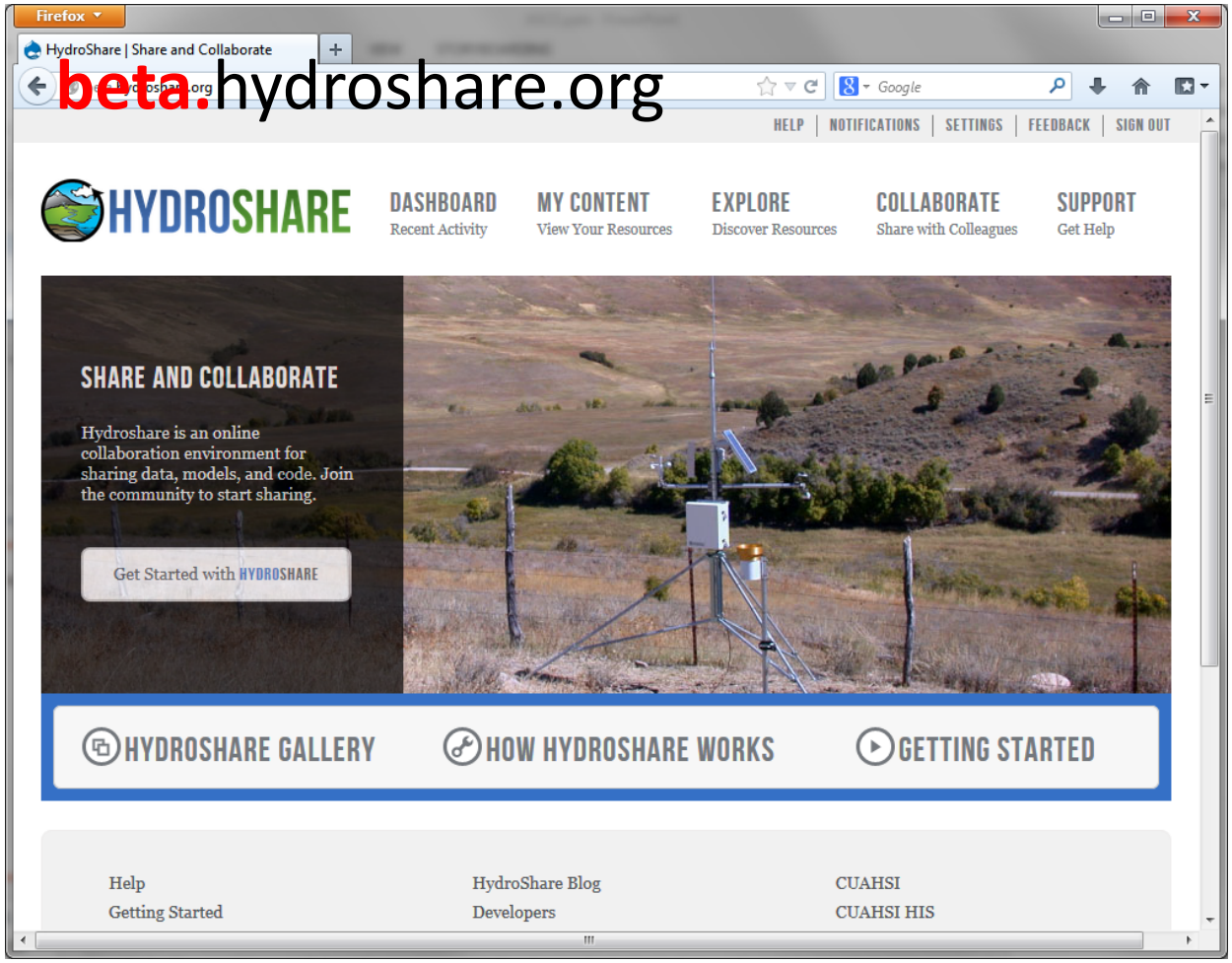
CUAHSI HIS: A common window on water observations data for the United States unlike any that has existed before

- Storage in a community data model
- Publication from a server
- Data access through internet-based services using consistent language and format
- Tools for access and analysis
- Discovery through thematic and geographic search functionality
- Integrated modeling and analysis combining information from multiple sources



Looking to the Future

HydroShare - A web-based collaborative environment for the sharing of hydrologic data and models



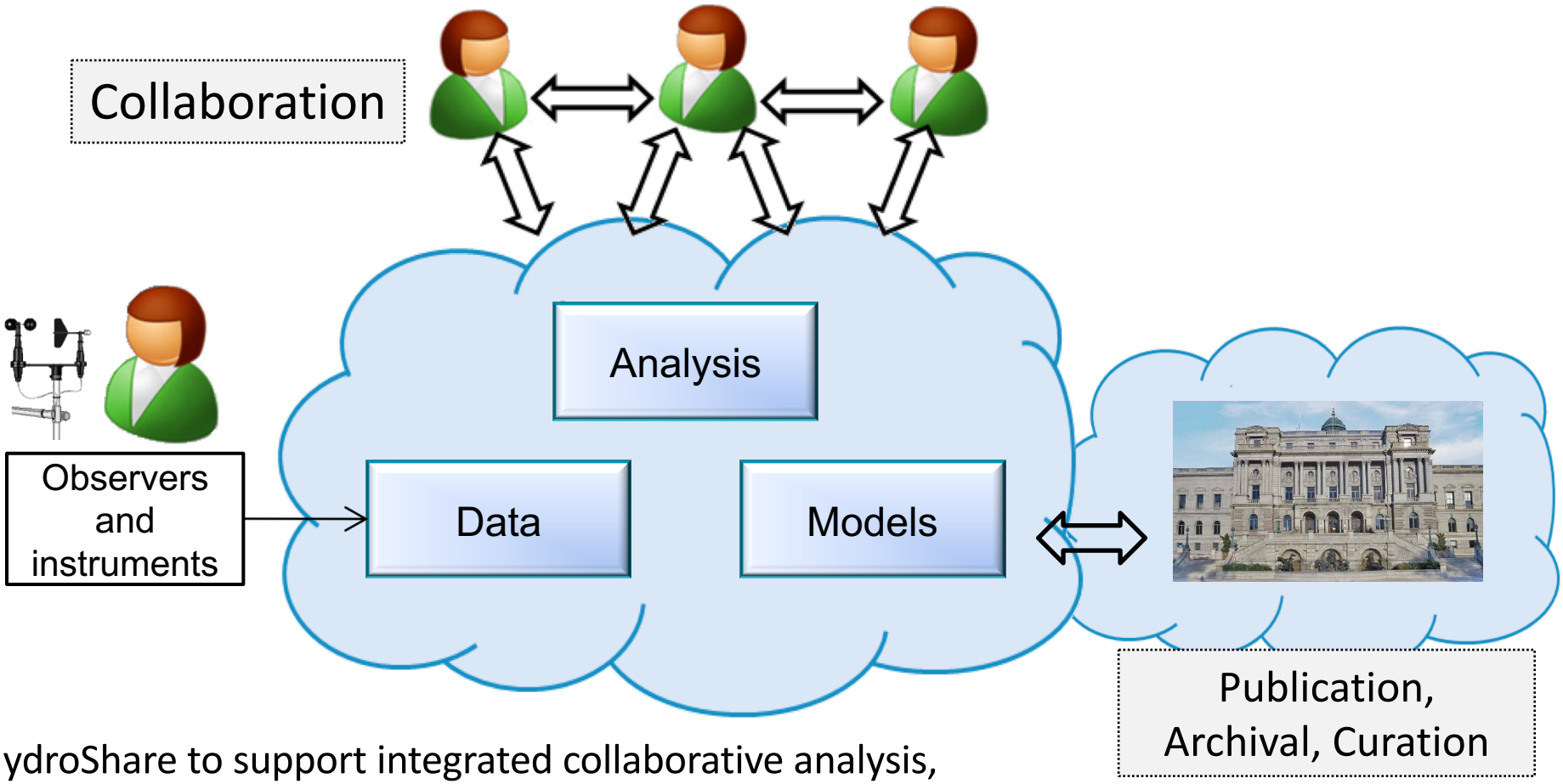
Can sharing data and models be as easy as sharing photos on Facebook or videos on YouTube?

Can finding data and models be as easy as shopping on Amazon?

HydroShare Functionality to be Developed

1. A new, web-based system for advancing model and data sharing
2. Sharing features to HydroDesktop
3. Access more types of hydrologic data using standards compliant data formats and interfaces
4. Enhance catalog functionality that broadens discovery functionality to different data types
5. New model sharing and discovery functionality
6. Facilitate and ease access to use of high performance computing
7. New social media and collaboration functionality
8. Links to other data and modeling systems

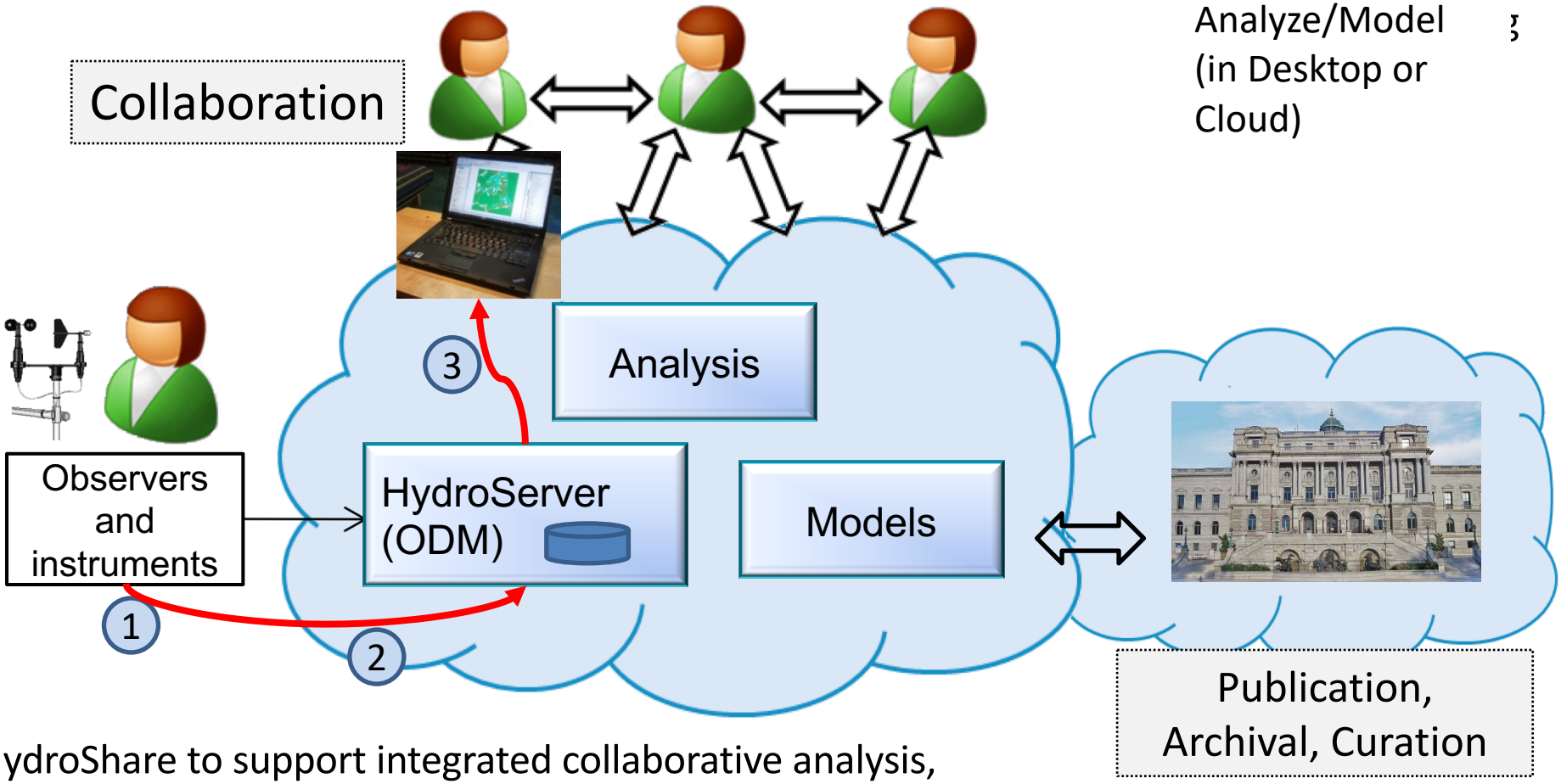
Imagine the Possibilities Enabled by Cloud-Based Systems



HydroShare to support integrated collaborative analysis, modeling and data publication

Imagine the Possibilities Enabled by Cloud-Based Systems

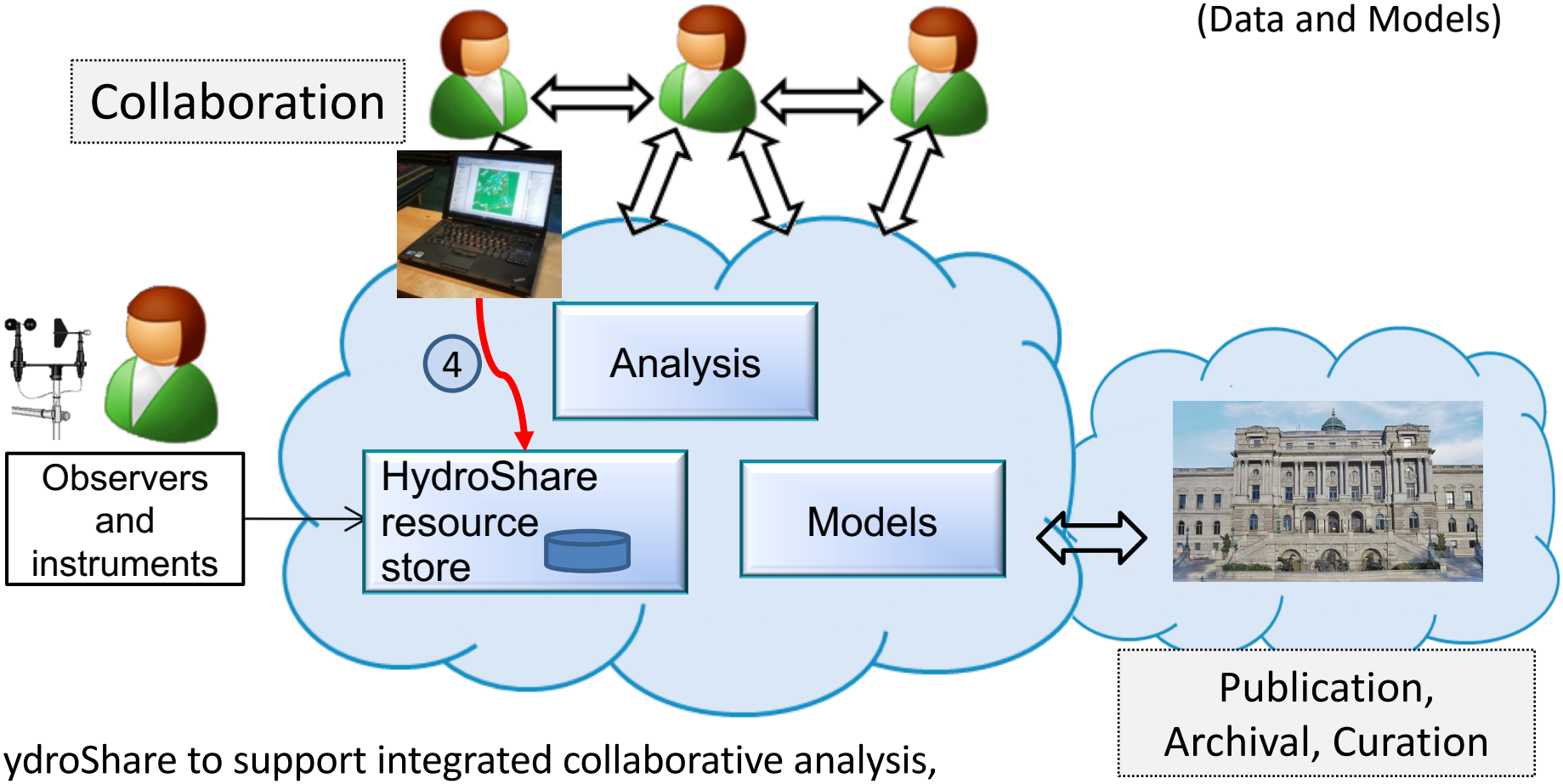
3. Discover and Analyze/Model (in Desktop or Cloud)



HydroShare to support integrated collaborative analysis, modeling and data publication

Imagine the Possibilities Enabled by Cloud-Based Systems

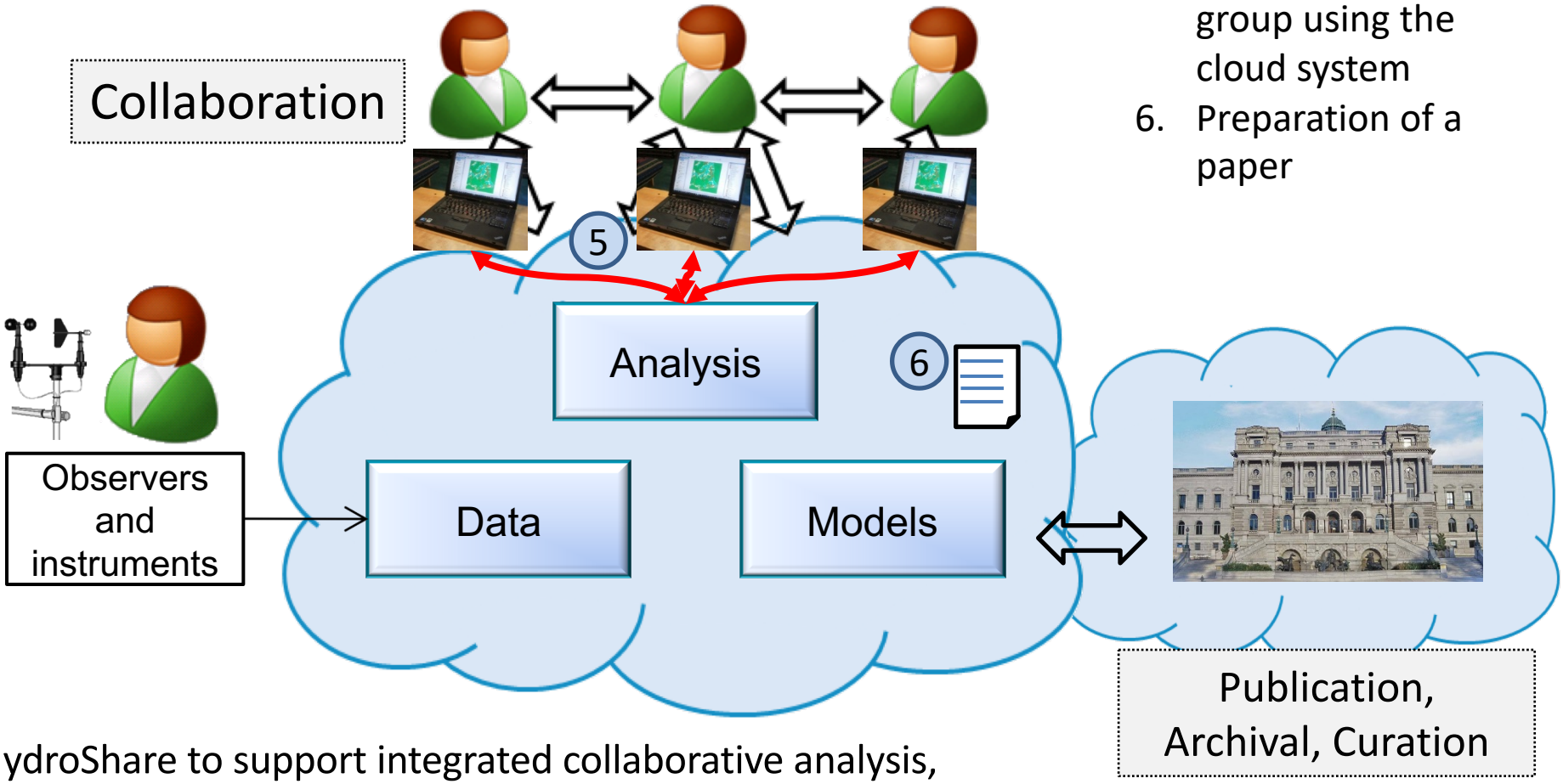
4. Share the results (Data and Models)



HydroShare to support integrated collaborative analysis, modeling and data publication

Imagine the Possibilities Enabled by Cloud-Based Systems

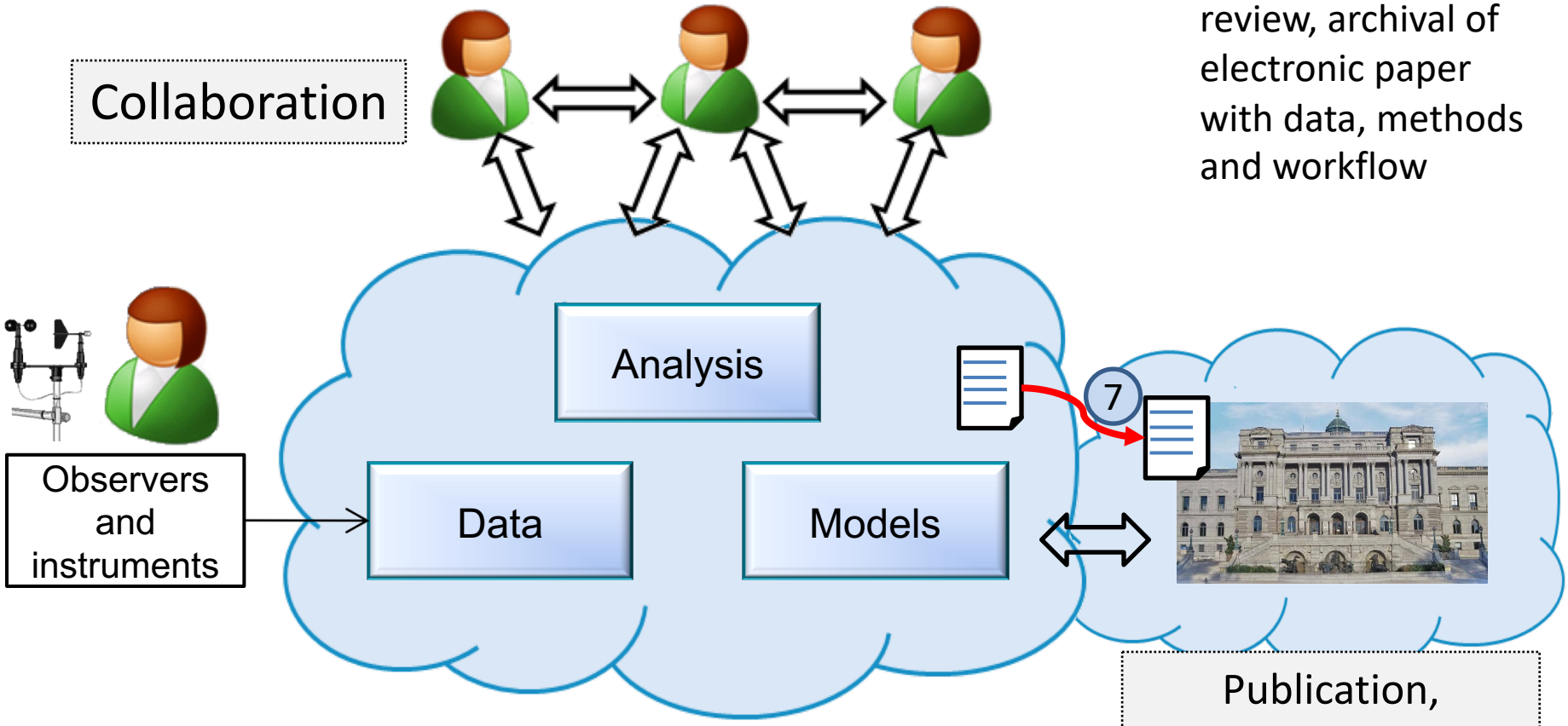
5. Collaboration with group using the cloud system
6. Preparation of a paper



HydroShare to support integrated collaborative analysis, modeling and data publication

Imagine the Possibilities Enabled by Cloud-Based Systems

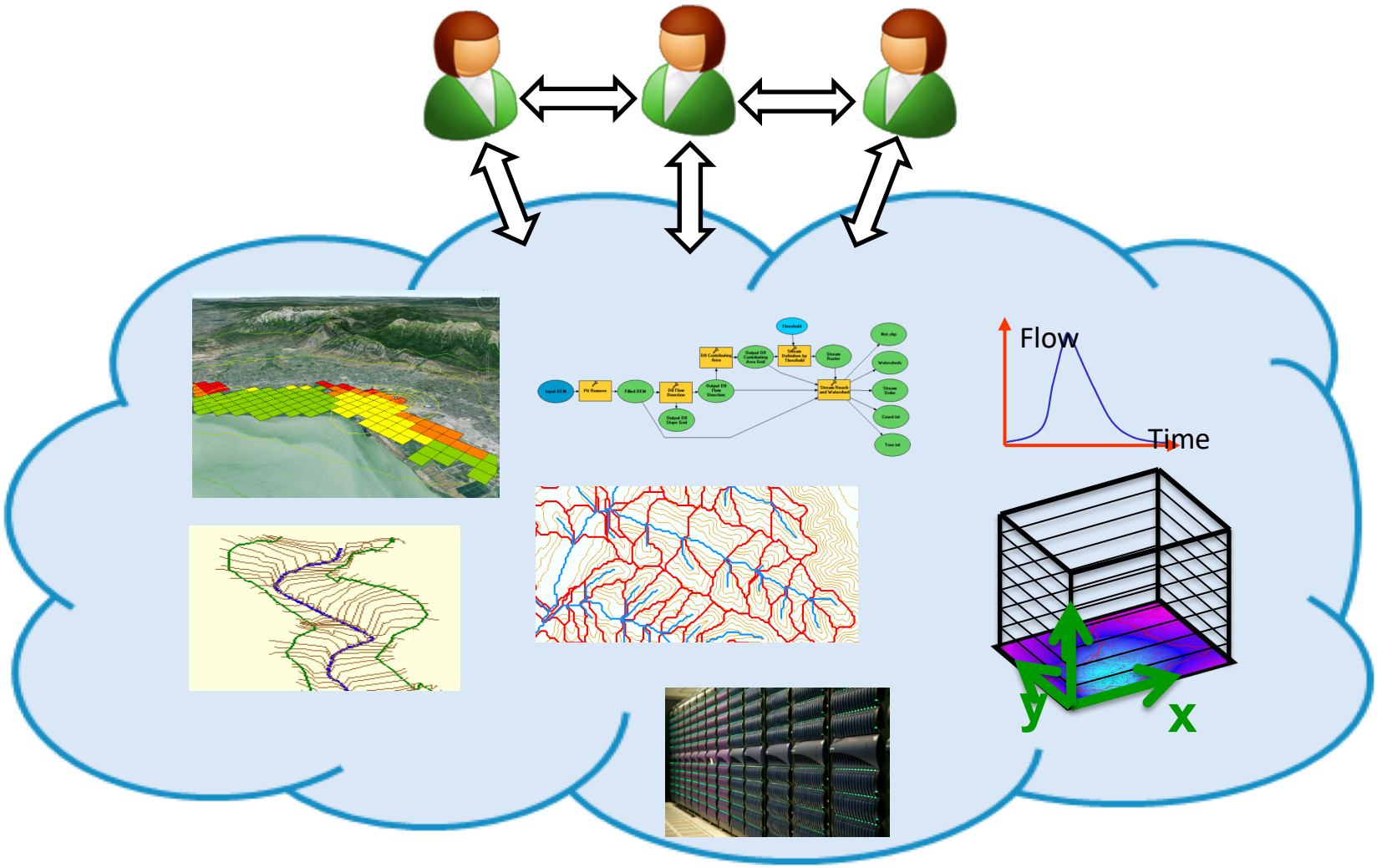
7. Submittal of paper, review, archival of electronic paper with data, methods and workflow



HydroShare to support integrated collaborative analysis, modeling and data publication

DataOne, EarthCube, …

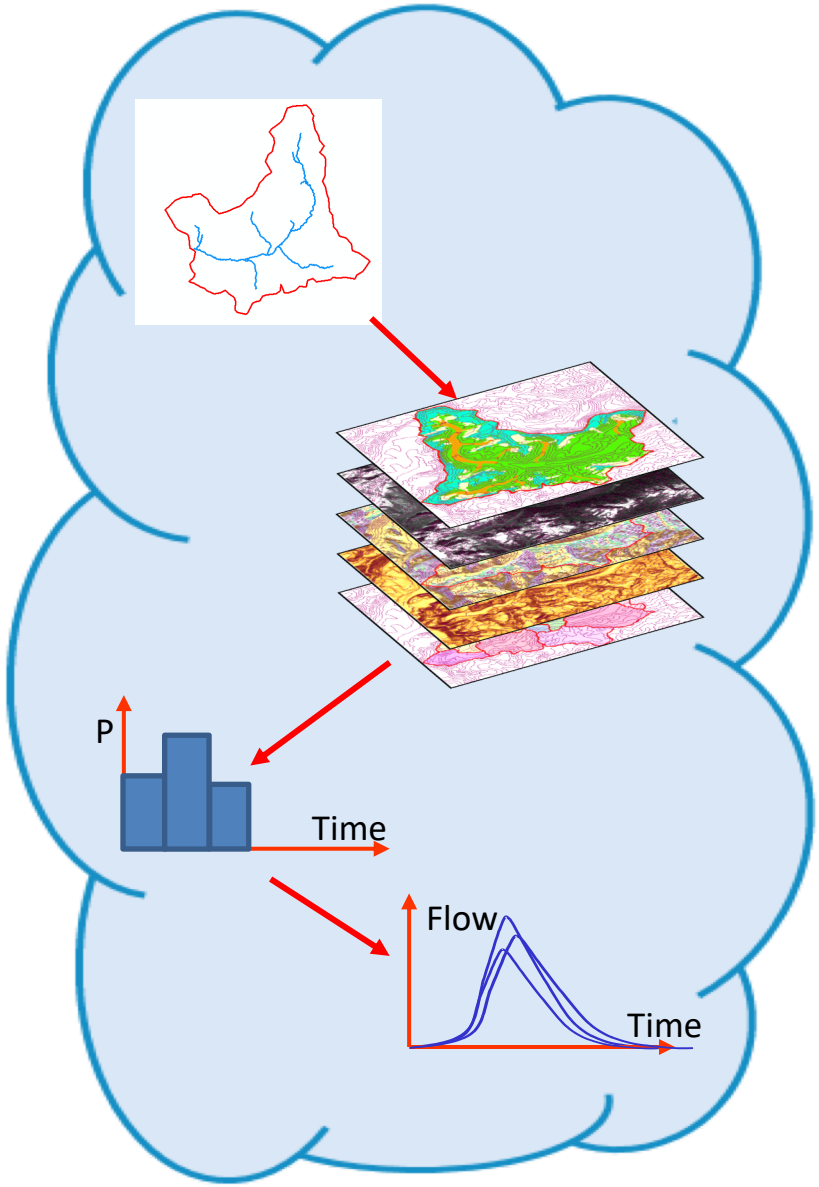
Imagine Cloud Modeling



- Automated reasoning to couple models based on purpose, context, data and resources

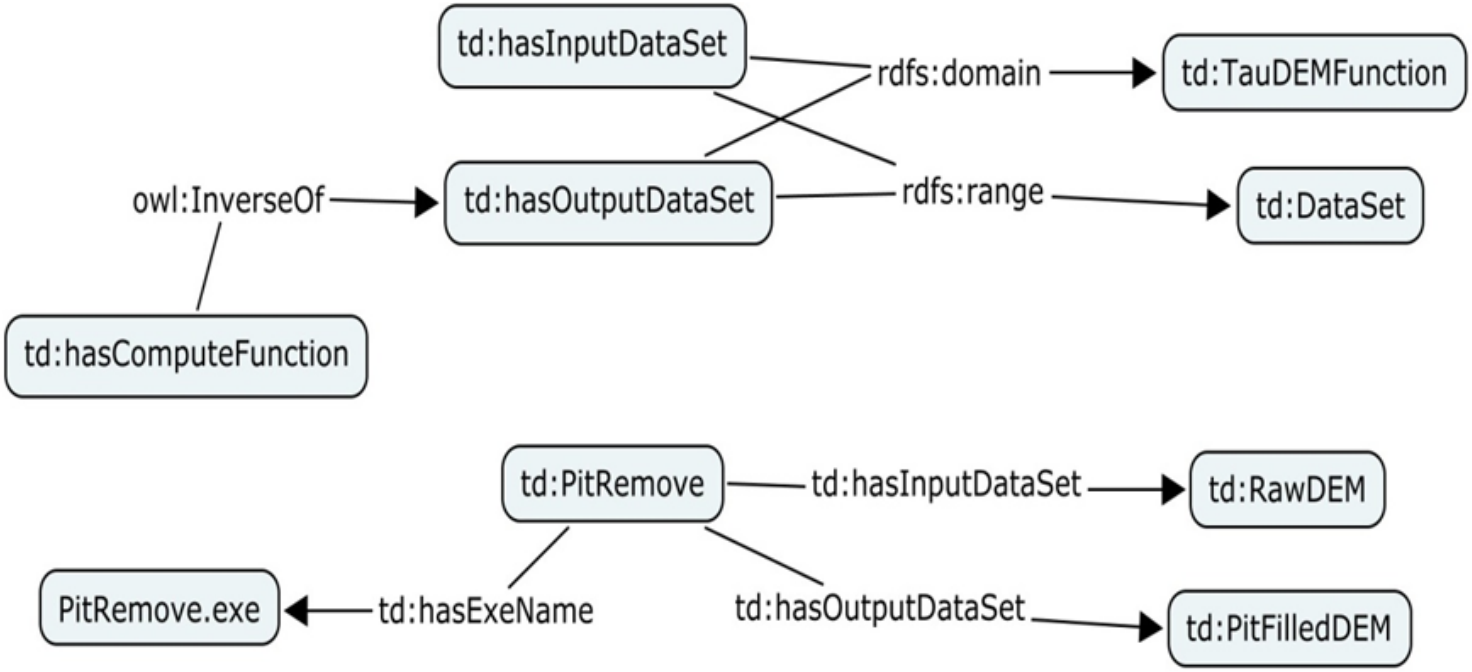
Another Example

- Big snow year
- Will my city flood?
- Click to delineate watershed (model domain)
- Generate model package from Essential Terrestrial Variables
- Generate suite of input scenarios
- Execute model and view results



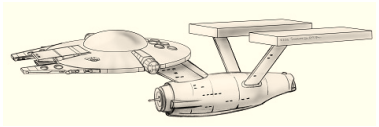
But there is more…

What if I could express my decision needs to the system and have it reason and deduce which models need to run, then configure and run them based on the inputs available, precision needs and resources and time available.



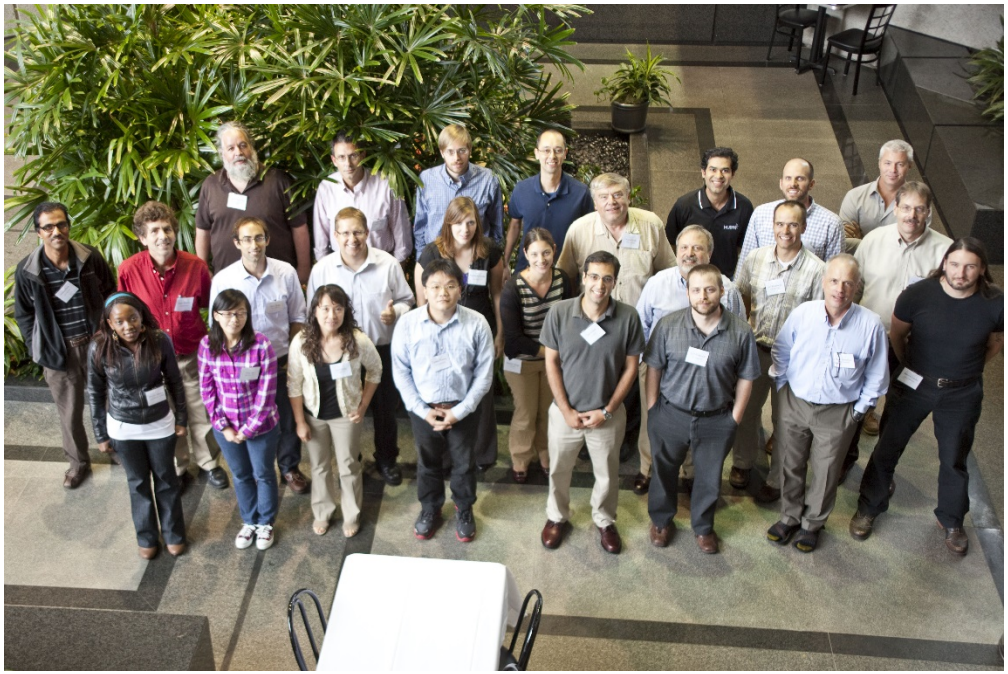
Summary and Conclusions

- Cloud Computing and Data Services
  - Community Participation
  - Interoperability
  - Standards
  - Open Development
- CUAHSI HIS
  - Enhanced Access to Hydrologic Data
  - Combining information from multiple sources
- HydroShare
  - A collaborative website for the sharing of hydrologic data and models
  - To expand data sharing capability of CUAHSI HIS
    - Additional data classes
    - Models, scripts, tools and workflows
  - To boldly go where no one has gone before



Thanks to a lot of people

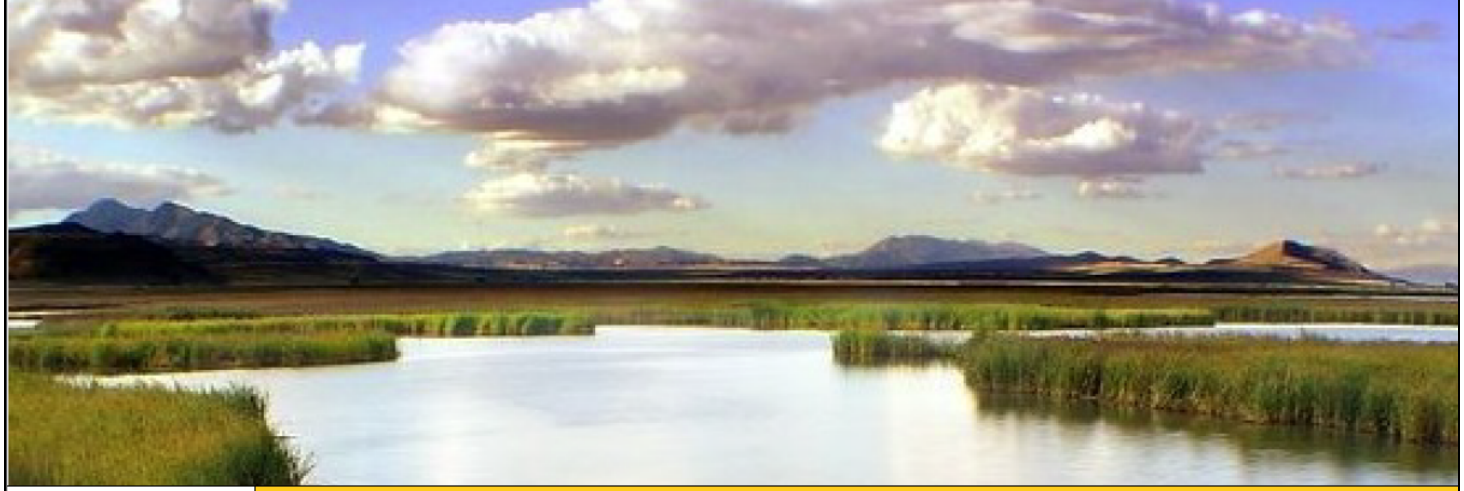
- USU
- RENCI/UNC
- CUAHSI
- BYU
- Tufts
- USC
- Texas
- Purdue
- SDSC



HydroShare team, CUAHSI HIS team, CI-WATER team, Students and Colleagues (Ray Idaszak, Dan Ames, Jeff Horsburgh, Jon Goodall, Larry Band, Venkatesh Merwade, Carol Song, Alva Couch, David Valentine, Rick Hooper, Jennifer Arrigo, David Maidment, Tim Whiteaker, Alex Bedig, Jason Coposky, Pabitra Dash, Tian Gan, Karl Gustafson, Taehee Hwang, Stephen Jackson, Harry Johnson, Yuri Kim, Phyllis Mbewe, Brian Miles, Jon Pollak, Stephanie Reeder, Terrell Russell, Tom Whitenack, Ilya Zaslavsky, Lan Zhao, Michael Piasecki, Kim Schreuders, Jiri Kadlec, Scott Peckham, Chris Duffy, Lorne Leonard, Tseganeh Gichamo, Norm Jones, Fred Ogden, Aaron Byrd)

HYDROSHARE http://www.cuahsi.org/hydroshare.aspx

EAR 0622374 (2007-2012)
OCI-1148453 (2012-2017)
OCI-1148090 (2012-2017)



CUAHSI Conference on Hydroinformatics and Modeling

July 17-19, 2013

Utah State University, Logan, Utah

NOW OPEN! Conference Registration and Abstract Submission

Abstracts Due: June 15, 2013
Registration Closes: June 30, 2013

For conference details, schedule, costs, presenter information, and logistics
http://www.cuahsi.org/WDCconf2013

Conference format includes

- Technical sessions on **hydrologic information systems and hydrologic modeling**
- Hands-on training and workshops to introduce community to services provided by newly established **Water Data Center (WDC)**
- Planned sessions on **using water data and information systems in the classroom**

Questions on conference or registration? Contact kberry@cuahsi.org