Clearing your Desk!
Software and Data Services for Collaborative Web Based GIS Analysis

David Tarboton, Ray Idaszak, Jeffery Horsburgh, Dan Ames, Jon Goodall, Larry Band, Venkatesh Merwade, Alva Couch, Rick Hooper, David Maidment, Pabitra Dash, Michael Stealey, Hong Yi, Tian Gan, Tseganeh Gichamo, Ahmet Yildirim, Yan Liu

USU, RENCI, BYU, UNC, UVA, CUAHSI, Tufts, Texas, Purdue, Caktus

http://www.hydroshare.org
Outline

• Data and computational challenges
• HydroShare
  – Goals
  – Resource data model
  – Architecture
• Terrain analysis and TauDEM in CyberGIS
• Data services for hydrologic modeling
• Summary
Do you have the access or know how to take advantage of advanced computing capability?

A digital divide
The challenge of increasing Digital Elevation Model (DEM) resolution

1980’s DMA 90 m
$10^2$ cells/km$^2$

1990’s USGS DEM 30 m
$10^3$ cells/km$^2$

2000’s NED 10 m
$10^4$ cells/km$^2$

2010’s LIDAR ~1 m
$10^6$ cells/km$^2$

e.g. 50,000 km$^2$
Watershed
27 MB

240 MB

2 GB

200 GB
HydroShare web based collaboration environment

- Share your data and models with colleagues
- Manage who has access to the content that you share
- Share, access, visualize and manipulate a broad set of hydrologic data types
- Sharing and execution of models
- Web services API to facilitate automated and client access to almost all functionality
- Access to and use of high performance computing
- Publication of data and models with a DOI

Enable more rapid advances in hydrologic understanding through collaborative data sharing, analysis and modeling.

www.hydroshare.org
Types of data supported as resources

- Generic ✓
- Geographic Raster (GeoTIFF) ✓
- Time Series ✓
- Multidimensional (netCDF) ✓
- Model program ✓
  - SWAT Model Instance ✓
- Web App ✓
- Geographic Feature (Shapefile) ✓
- Referenced Time Series (CUAHSI HIS web service link) ▲
- River Geometry
- Sample based observations (ODM2 and CZO)
- Composite resources (Collections of resources)
Can this web based (cloud) computing approach deliver GIS and hydrologic analysis and modeling functionality as services over the web? and address

- Platform independence
- Big data
- Reproducibility
- Reduce needs for software installation and configuration
Can this web based (cloud) computing approach deliver GIS and hydrologic analysis and modeling functionality as services over the web and address:

- Platform independence
- Big data
- Reproducibility
- Reduce needs for software installation and configuration

Anyone can set up a server/app platform (software service) to operate on HydroShare resources through iRODS and API.
Terrain Analysis as a use case

- Topography is fundamental to hydrology
- Watersheds are the most basic hydrologic landscape elements

TauDEM [http://hydrology.usu.edu/taudem/](http://hydrology.usu.edu/taudem/) is software for deriving hydrologically useful information from Digital Elevation Models

- Stream and watershed delineation
- Multiple flow direction flow field
- Calculation of flow based derivative surfaces
- MPI Parallel Implementation for speed up and large problems
- Open source platform independent C++ command line executables for each function
- Deployed as an ArcGIS Toolbox with python scripts that drive command line executables
Using TauDEM today requires

• Expertise in Hydrologic DEM analysis
• The software
  – ArcGIS licenses (for ArcGIS plugin)
  – The ability to install software
  – TauDEM command functions with MPI installation
  – Compilation for other platforms
• Sufficient Hardware (RAM and Disk)
• The data (GDAL formatted rasters with consistent grid size and spatial reference)
Moving TauDEM to the cloud, CyberGIS TauDEM App
http://gateway.cigi.illinois.edu/
Moving TauDEM to the cloud, CyberGIS TauDEM App

http://gateway.cigi.illinois.edu/
http://cybergis.cigi.uiuc.edu/
Select the products you want

The wizard configures the sequence of functions to run to get the result
Results displayed in browser

And saved in HydroShare (in progress)
CyberGIS HydroShare Coupling

- CyberGIS data holdings shared as HydroShare resources
- HydroShare user resources staged at CyberGIS for input to computation
- Results persisted as resources in HydroShare to support
  - Collaboration
  - Input for further analysis and models
Advancing Data Services for Modeling and Analysis

Assumptions

1. GIS and hydrologic modelers have to learn and become comfortable using a modern scientific programming language (e.g. Python or R)
2. Modeling is increasingly data intensive (large datasets from a range of sources) demanding more data and computing resources than is in most PC’s
3. Reproducibly installing and configuring models on different platforms is a challenge
4. Should not have to become expert in HPC systems and learning them is a barrier to using HPC and research with big data and computationally intensive models
Computation via Python Client calling Data and Modeling Services

Input

```python
# Illustration of Watershed Delineation using CI-WATER data services
HDS = HydroDS(username='settings.USER_NAME', password='settings.PASSWORD')
subsetDEM = HDS.subset_raster(input_raster='nedWesternUS.tif', left=-111.97, top=41.629,
right=-111.48, bottom=41.36, output_raster='MyDEM.tif')
projectDEM = HDS.project_resample_raster(input_raster_url_path='output_raster',
cell_size_dx=30, cell_size_dy=30, epsg_code=26912,
output_raster='MyHydroProj.tif', resample='bicubic')

# Create outlet point
outlet = HDS.create_outlet_shapefile(point_x=-111.855, point_y=41.586,
output_shape_file_name='Outlet.shp')

outletProj = HDS.project_shapefile(outlet={'output_shape_file_name': 'OutletProj.shp'},
epsg_code=26912)

Watershed = HDS.delineate_watershed(projectDEM['output_raster'],
input_outlet_shapefile_url_path='OutletProj.shp',
output_threshold=1000, epsg_code=26912,
output_raster='Watershed.tif',
output_outlet_shapefile_url_path='OutletProj.shp')
ListMyFiles()
```

Result

Details in [https://www.hydroshare.org/resource/cfb8d71b7f1f4e75a44f5e634f4730d4/](https://www.hydroshare.org/resource/cfb8d71b7f1f4e75a44f5e634f4730d4/) or search for HydroGate in HydroShare.

Python session on desktop but data and analysis on server with results pushed to iRODS/HydroShare for storage and collaboration.
Summary

1. Web based Cyberinfrastructure for GIS and Hydrologic Data and Modeling is emerging to support
   • Large datasets
   • Collaboration
   • Reproducible workflows and results
   • And reduce software installation and configuration limitations.

2. HydroShare and CyberGIS are part of this.
   • Interoperability is key to leveraging full potential of multiple emerging cloud cyberinfrastructure systems
Thanks to the HydroShare, CyberGIS and CI-WATER teams!

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

http://www.hydroshare.org