Managing Hydrologic Data, Observations and Terrain Analysis

David Tarboton

Utah State University

http://www.engineering.usu.edu/dtarb/

Outline

- The CUAHSI Hydrologic Information System

CUAHSI HIS *Sharing hydrologic data*

http://his.cuahsi.org/

- Terrain Analysis Using Digital Elevation Models

TauDEM Version 5

http://hydrology.usu.edu/taudem

Hydrologic Data Challenges

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

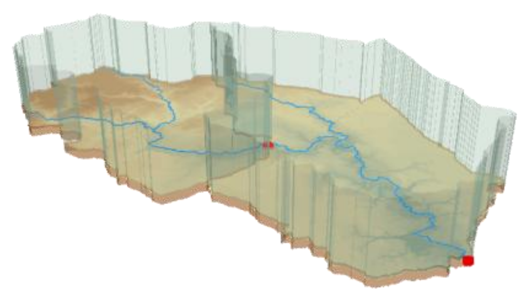
Data Heterogeneity


Water quality


Water quantity


Rainfall and Meteorology


Soil water


GIS


Groundwater

The way that data is organized can enhance or inhibit the analysis that can be done

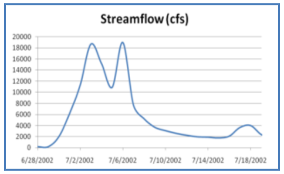
I have your information right here …

Picture from: http://initsspace.com/

Hydrologic Science

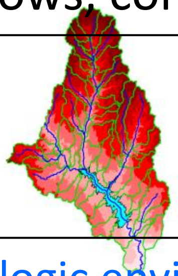
It is as important to represent *hydrologic environments* *precisely with data as it is to represent* *hydrologic processes* *with equations*

Physical laws and principles
(Mass, momentum, energy, chemistry)



Hydrologic Process Science
(Equations, simulation models, prediction)

Hydrologic conditions
(Fluxes, flows, concentrations)

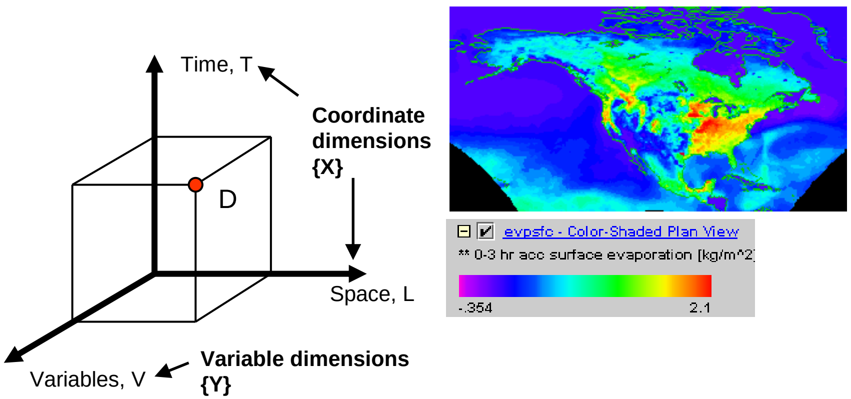


Hydrologic Information Science
(Observations, data models, visualization

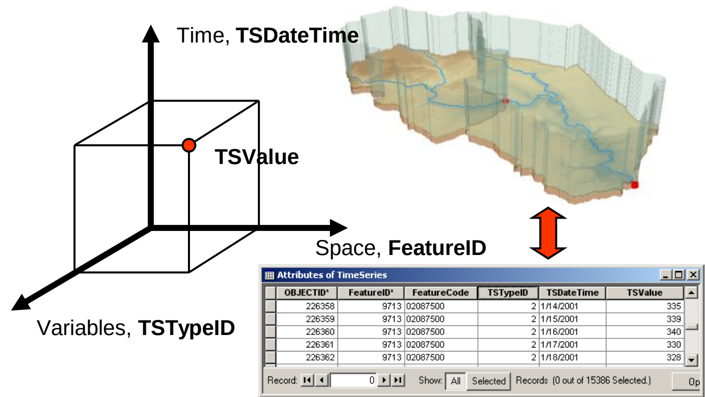
Hydrologic environment
(Physical earth)

Data models capture the complexity of natural systems

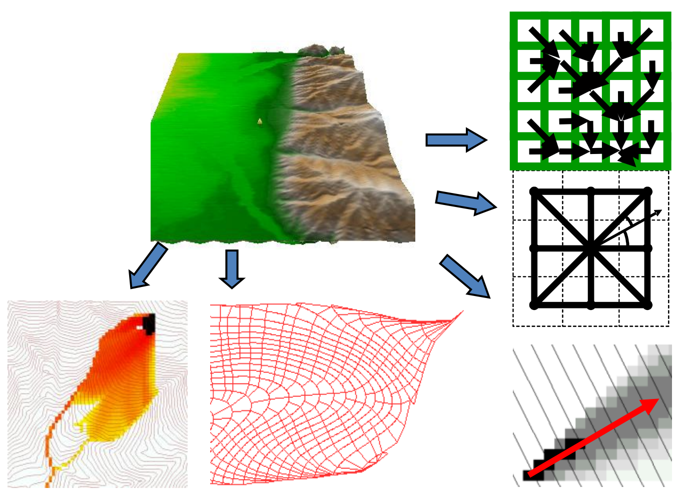
NetCDF (Unidata) - A model for Continuous Space-Time data



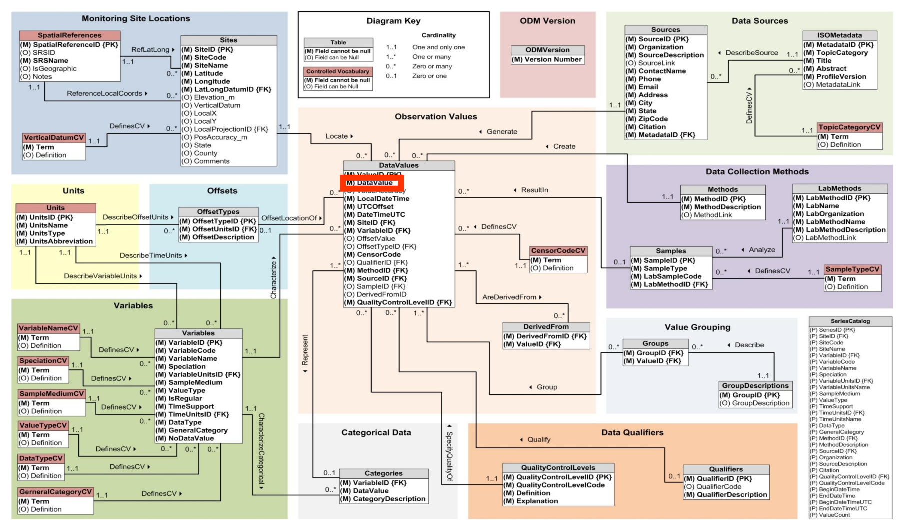
ArcHydro – A model for Discrete Space-Time Data



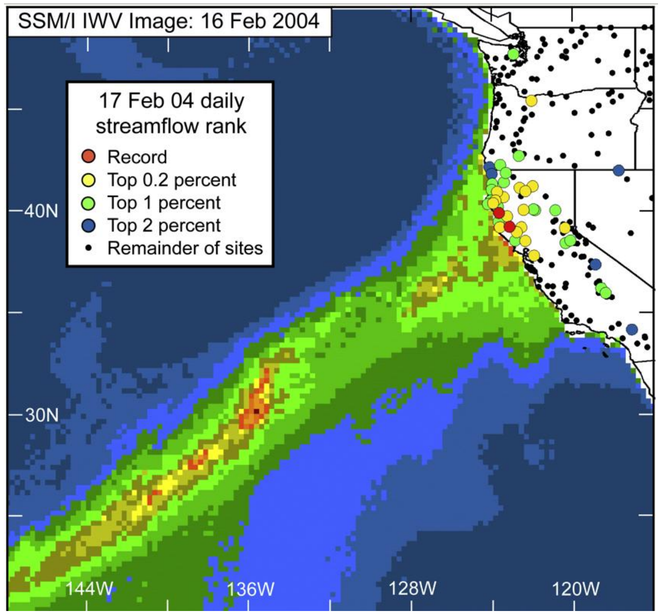
Terrain Flow Data Model used to enrich the information content of a digital elevation model

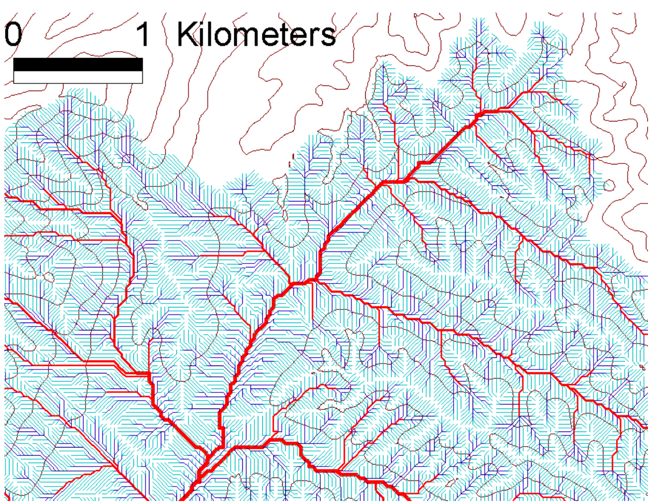


CUAHSI Observations Data Model: What are the basic attributes to be associated with each single data value and how can these best be organized?



Data intensive science synthesizes large quantities of information (Hey et al., 2009).

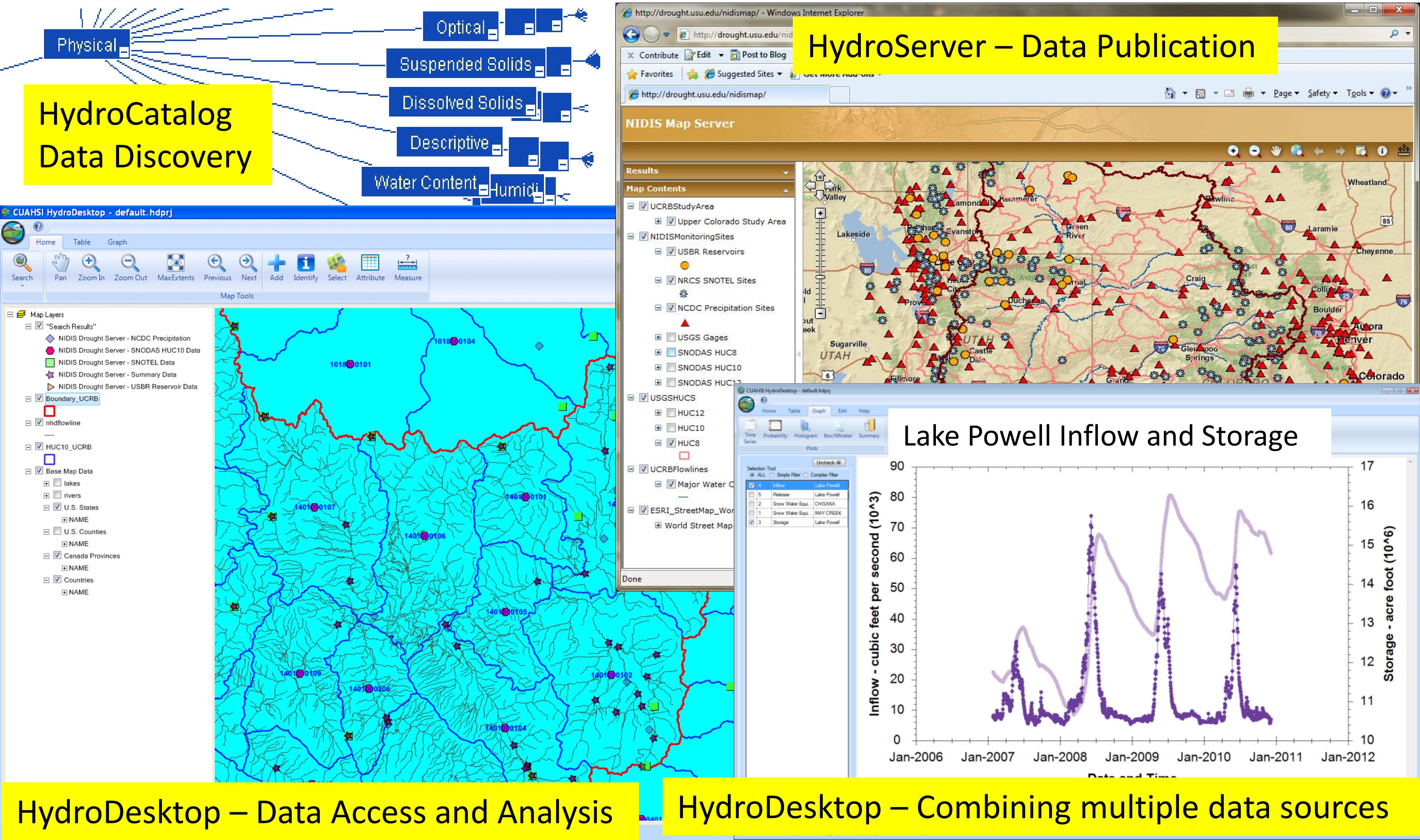




- exploiting advanced computational capability for the analysis and integration of large new datasets to elucidate complex and emergent behavior
- In hydrology, the image at left (Ralph et al., 2006) illustrates connection between extreme floods recorded in USGS stream gages and atmospheric water vapor from space based sensors
- Satellite remote sensing and massive datasets enhance understanding of multi-scale complexity in processes such as rainfall and river networks

CUAHSI HIS

The CUAHSI Hydrologic Information System (HIS) is an internet based system to support the sharing of hydrologic data. It is comprised of hydrologic databases and servers connected through web services as well as software for data publication, discovery and access.



The CUAHSI Community Hydrologic Information System

- University of Texas at Austin – David Maidment, Tim Whiteaker, James Seppi, Fernando Salas, Jingqi Dong, Harish Sangireddy
- San Diego Supercomputer Center – Ilya Zaslavsky, David Valentine, Tom Whitenack, Matt Rodriguez
- Utah State University – Jeff Horsburgh, Kim Schreuders, Stephanie Reeder, Edward Wai Tsui, Ravichand Vegiraju, Ketan Patil
- University of South Carolina – Jon Goodall, Anthony Castronova
- Idaho State University – Dan Ames, Ted Dunsford, Jiří Kadlec, Yang Cao, Dinesh Grover
- Drexel University/CUNY – Michael Piasecki
- WATERS Network – Testbed Data Managers
- CUAHSI Program Office – Rick Hooper, Yoori Choi, Conrad Matiuk
- ESRI – Dean Djokic, Zichuan Ye

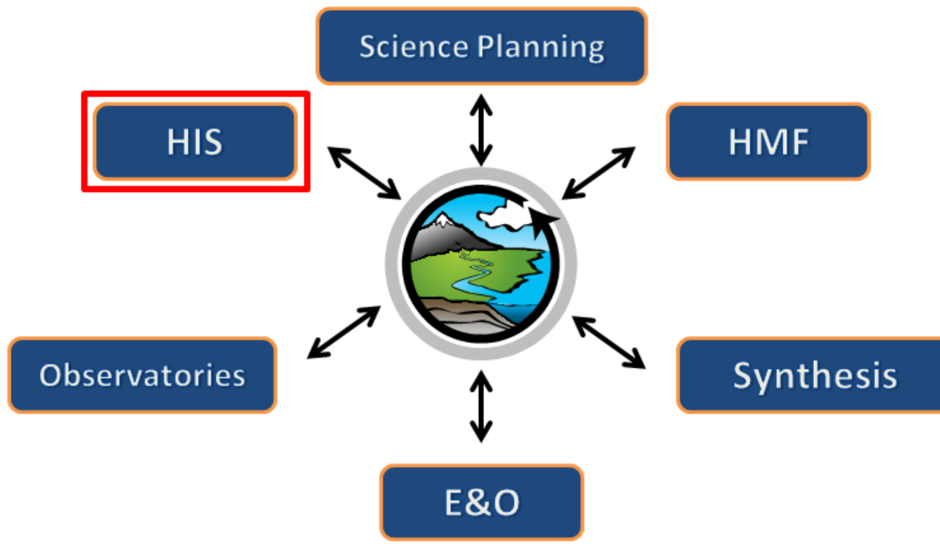
CUAHSI
HIS
Sharing hydrologic data

http://his.cuahsi.org/

Support
EAR 0622374

What is CUAHSI?

Consortium of Universities for the Advancement of Hydrologic Science, Inc.



- 110 US University members
- 6 affiliate members
- 12 International affiliate members

(as of March 2009)

Infrastructure and services for the advancement of hydrologic science and education in the U.S.

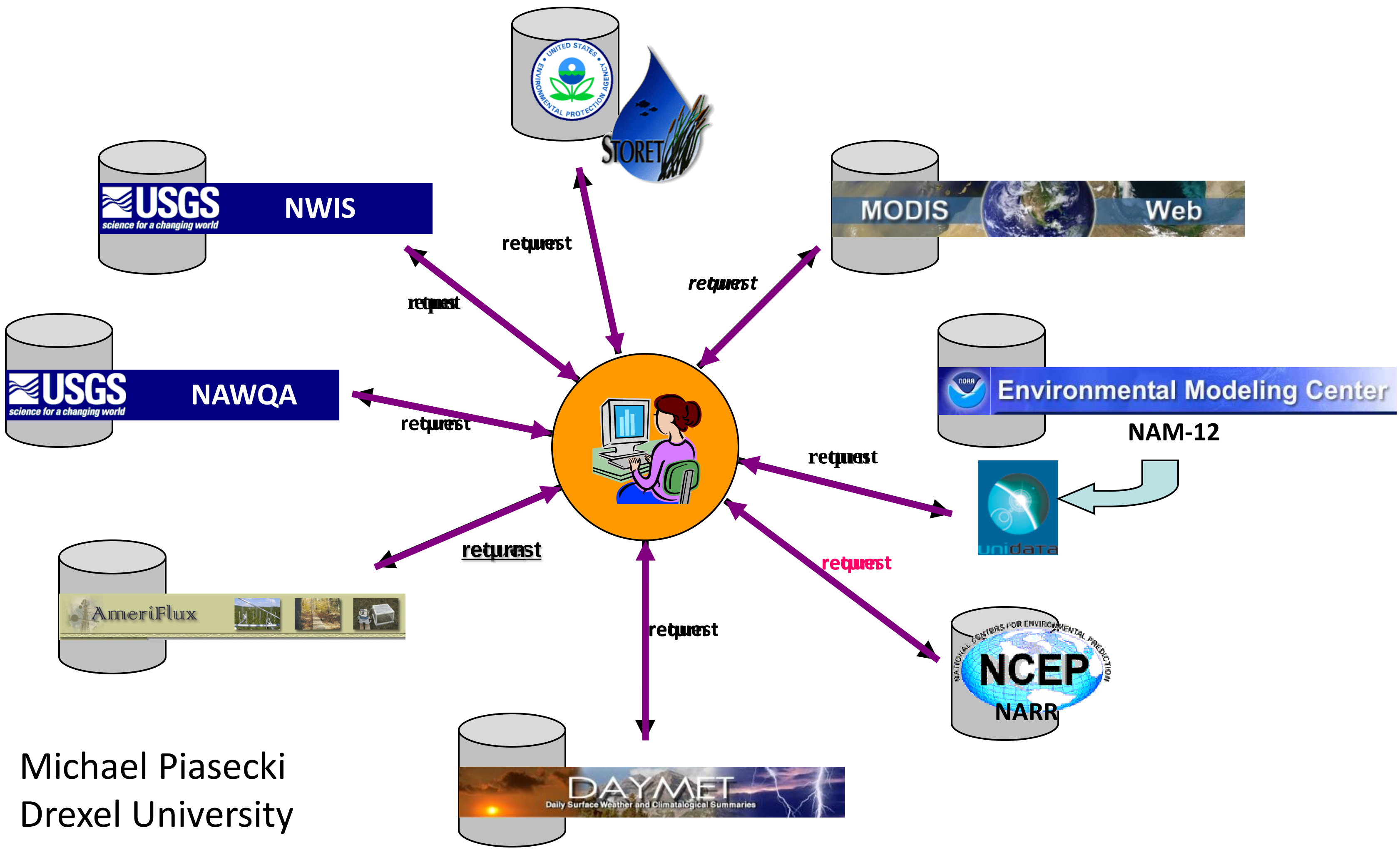
CUAHSI
universities allied for water research

http://www.cuahsi.org/

Support
EAR 0753521

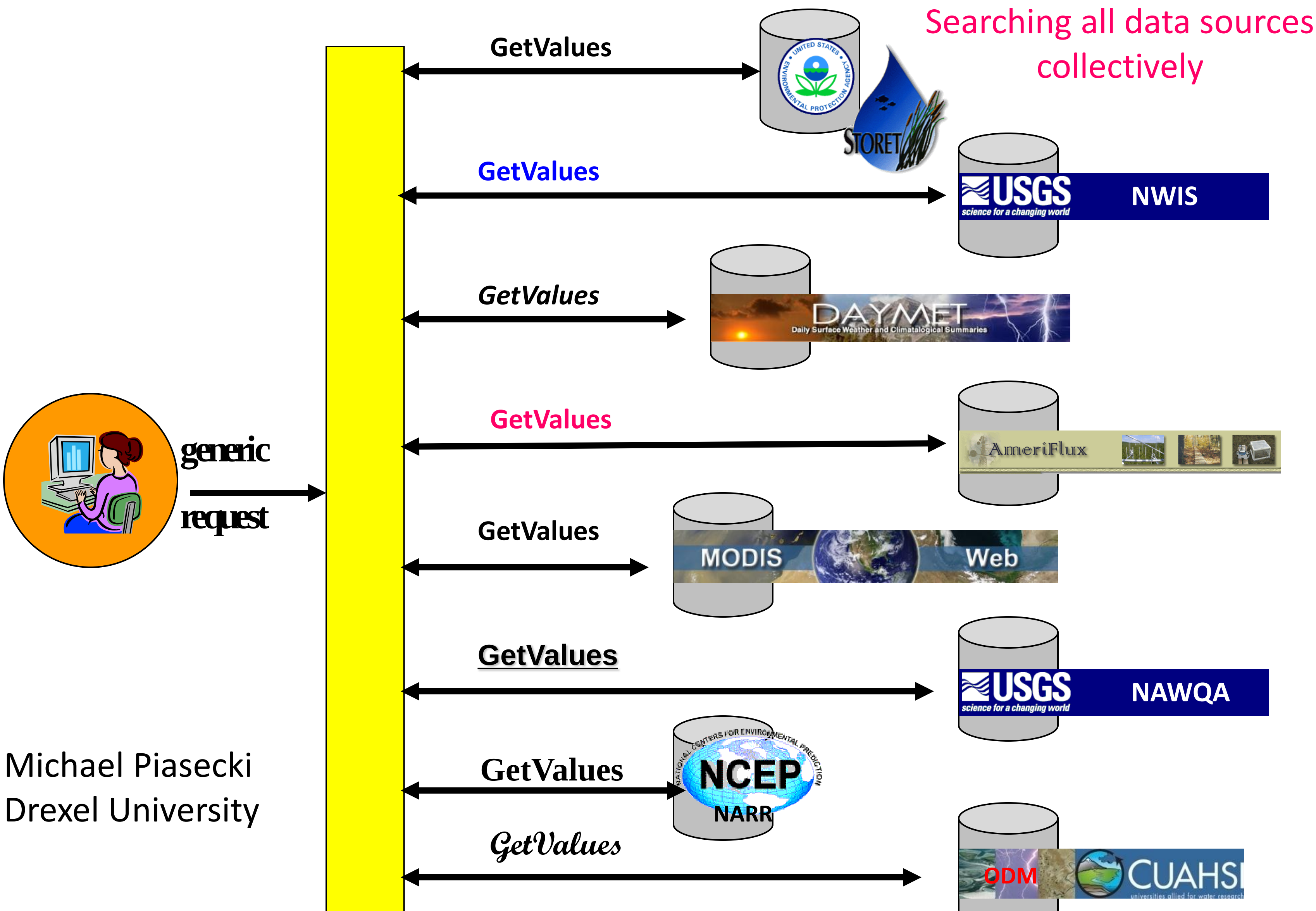
Data Searching – What we used to have to do

Searching each data source separately

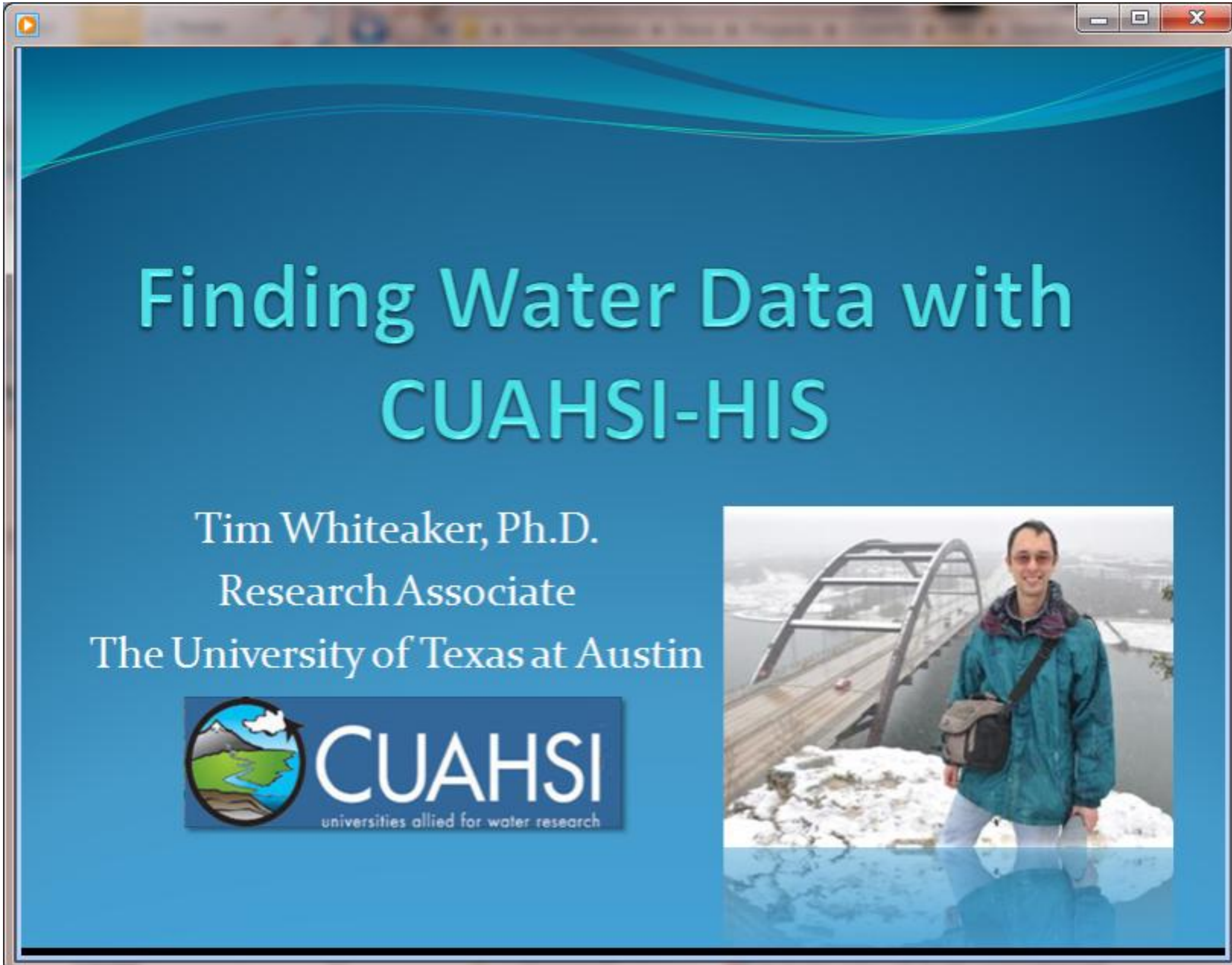


Michael Piasecki
Drexel University

What HIS enables

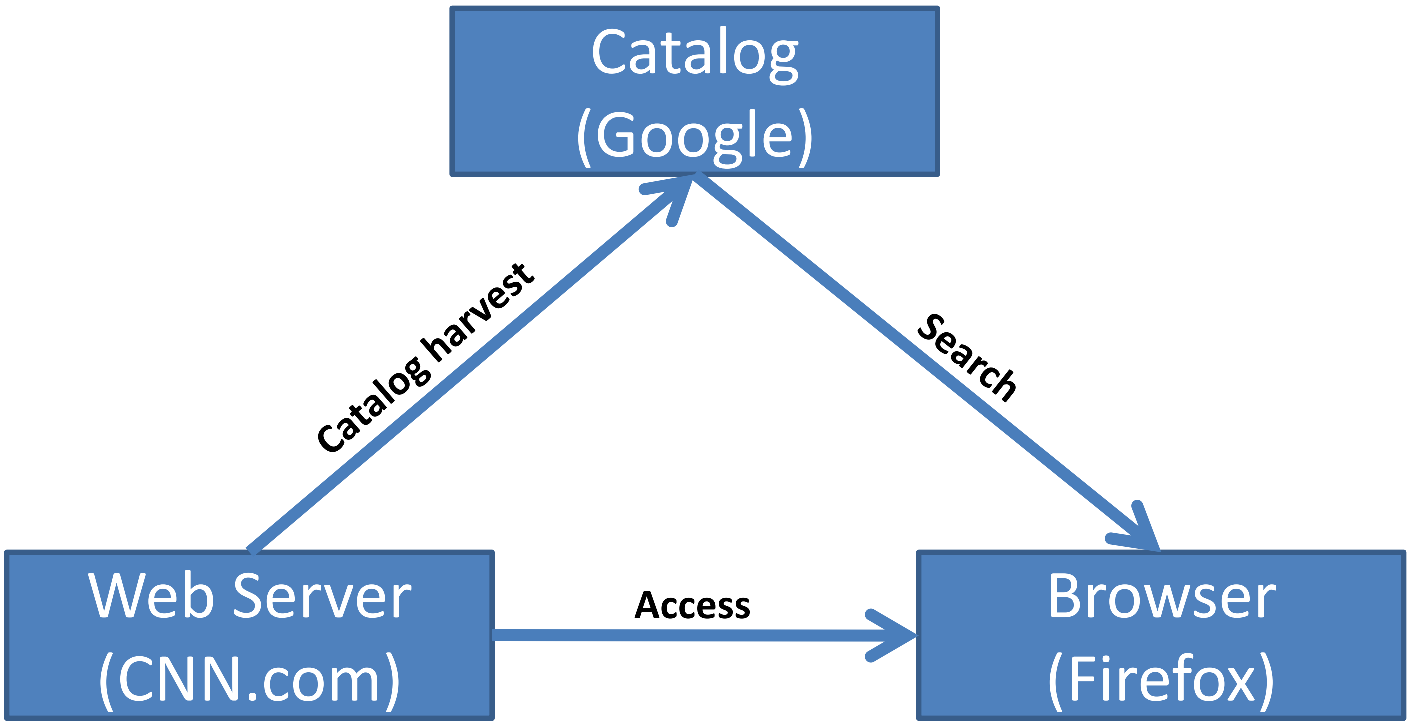


Michael Piasecki
Drexel University

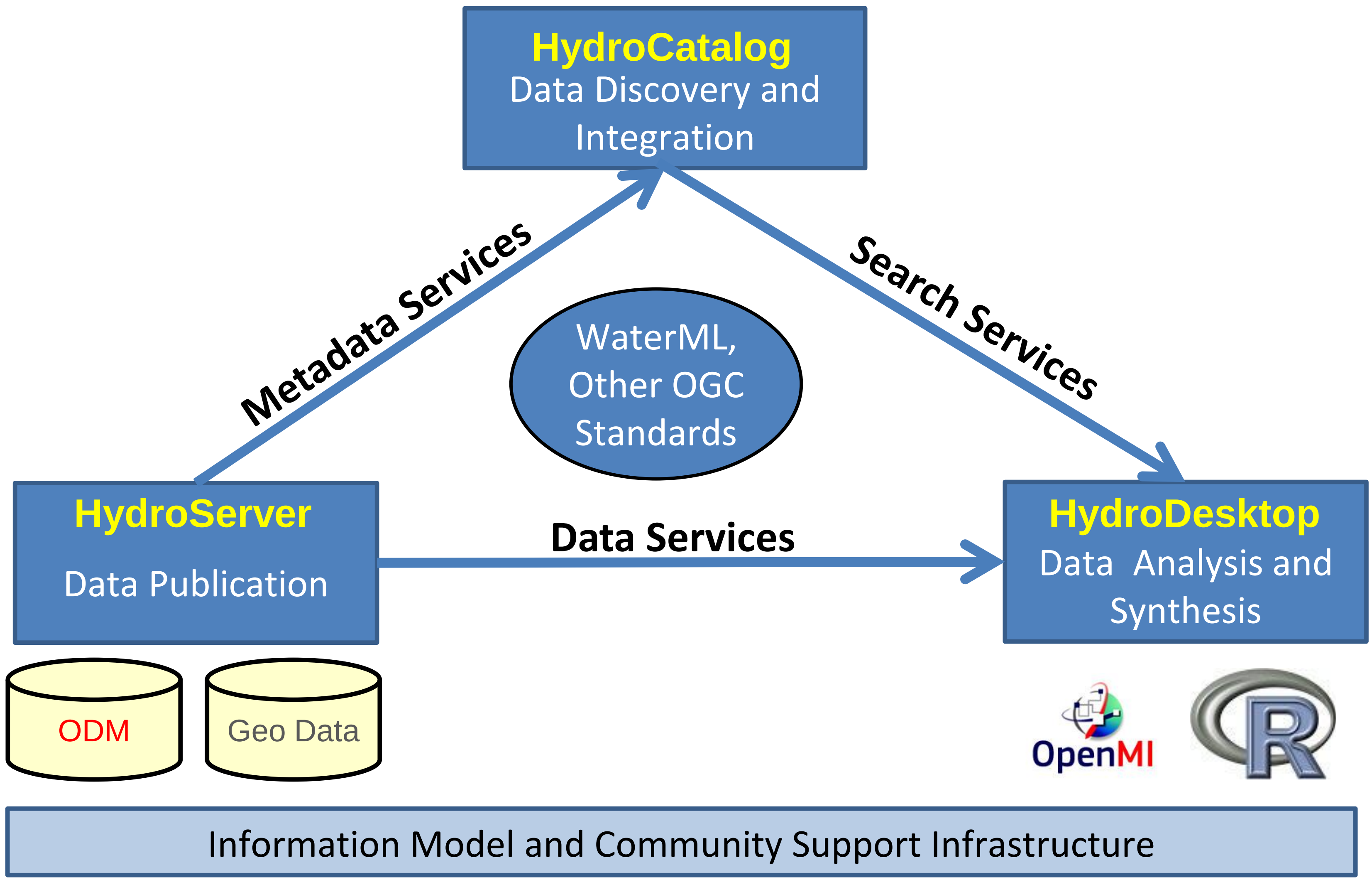


http://his.cuahsi.org/movies/JacobsWellSpring/JacobsWellSpring.html

Web Paradigm



CUAHSI Hydrologic Information System Services-Oriented Architecture



What are the basic attributes to be associated with each single data value and how can these best be organized?

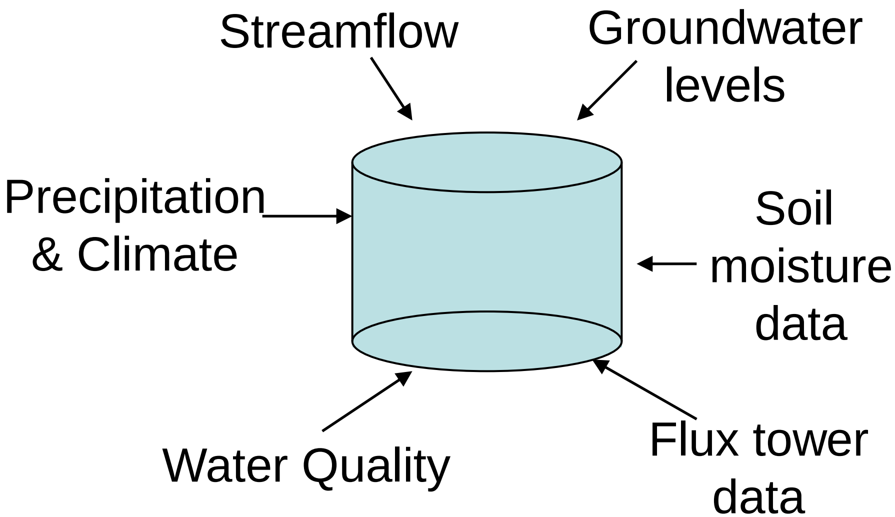
Value
DateTime
Variable
Location
Units
Interval (support)
Accuracy

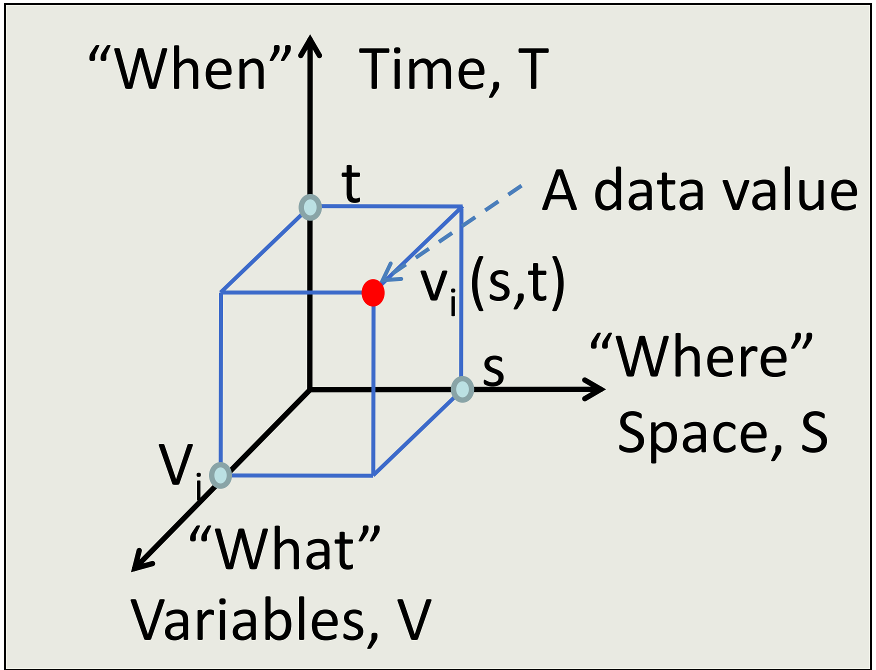
Offset
OffsetType/ Reference Point
Source/Organization
Censoring
Data Qualifying Comments

Method
Quality Control Level
Sample Medium
Value Type
Data Type

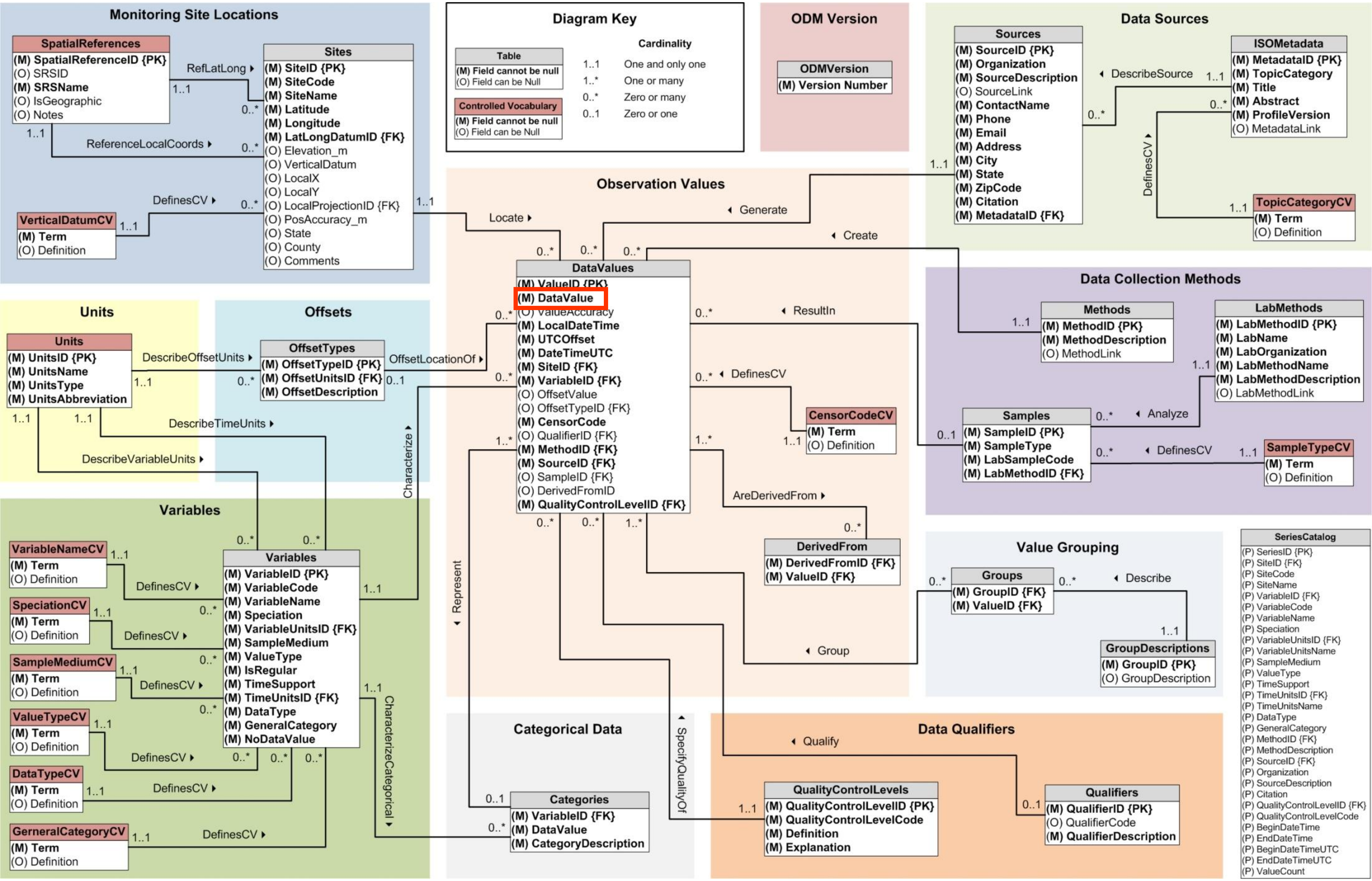
CUAHSI Observations Data Model

- A relational database at the single observation level (atomic model)
- Stores observation data made at points
- Metadata for unambiguous interpretation
- Traceable heritage from raw measurements to usable information
- Standard format for data sharing
- Cross dimension retrieval and analysis





CUAHSI Observations Data Model http://his.cuahsi.org/odmdatabases.html



Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2008), A Relational Model for Environmental and Water Resources Data, *Water Resour. Res.,* 44: W05406, doi:10.1029/2007WR006392.

Stage and Streamflow Example

DataValues : Table

ValueID	DataValue	ValueAccuracy	LocalDateTime	UTCOffset	SiteID	VariableID	MethodID	DerivedFromID
1	4.18		05/01/2006 0:00:00.000	-7	1	1	1	
97	748		05/01/2006 0:00:00.000	-7	1	2	2	1
193	722	22.89831642	05/01/2006 0:00:00.000	-7	1	3	3	100
2	4.18		05/01/2006 0:15:00.000	-7	1	1	1	
98	748		05/01/2006 0:15:00.000	-7	1	2	2	2
3	4.17		05/01/2006 0:30:00.000	-7	1	1	1	
99	742		05/01/2006 0:30:00.000	-7	1	2	2	3
4	4.17		05/01/2006 0:45:00.000	-7	1	1	1	
100	742		05/01/2006 0:45:00.000	-7	1	2	2	4
5	4.17		05/01/2006 1:00:00.000	-7	1	1	1	
101	742		05/01/2006 1:00:00.000	-7	1	2	2	5
6	4.17		05/01/2006 1:15:00.000	-7	1	1	1	
102	742		05/01/2006 1:15:00.000	-7	1	2	2	6

Record: 1 of 415

DerivedFrom : Table

DerivedFromID	ValueID
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17

Record: 18

Variables : Table

VariableID	VariableCode	VariableName	VariableUnitsID	SampleMedium	ValueType	IsRegular	TimeSupport	TimeUnitsID	DataType	GeneralCategory	NoDataValue
1	00065	Gage height	1	Water	Field Observation	☑	15	5	Continuous	Hydrologic	-9999
2	00060	Discharge	2	Water	Derived Value	☑	15	5	Continuous	Hydrologic	-9999
3	00060	Discharge, daily average	2	Water	Derived Value	☑	24	6	Average	Hydrologic	-9999
4	00300	Dissolved oxygen concentration	3	Water	Field Observation	☐	0		Instantaneous	Water Quality	-9999

Record: 4 of 7

Units : Table

UnitsID	UnitsName	UnitsType	UnitsAbbreviation
1	Feet	Length	ft
2	Cubic feet per second	Flow	ft^3/s
3	Milligrams per liter	Concentration	mg/L
4	Meters	Length	m
5	Minutes	Time	min
6	Hours	Time	hr

Record: 8 of 8

Methods : Table

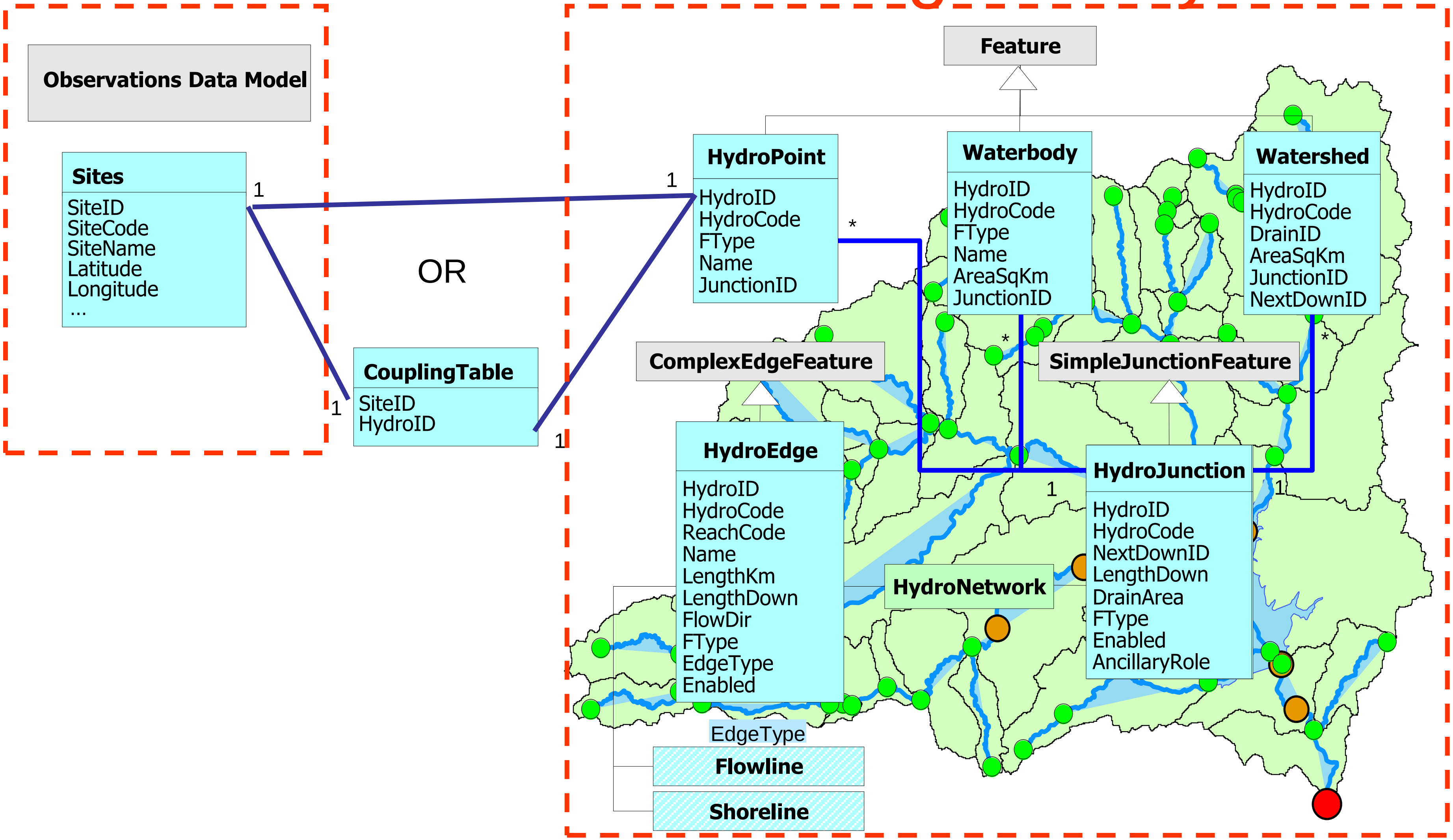
MethodID	MethodDescription
1	Gage height measured with continuous data logger
2	Discharge derived from water stage using site specific rating curve
3	Daily average discharge derived from 15 minute continuous discharge values
4	Dissolved oxygen measured with a Hydrolab multiprobe field instrument

Record: 5 of 5

Independent of, but can be coupled to Geographic Representation

ODM

e.g. Arc Hydro



Importance of the Observations Data Model

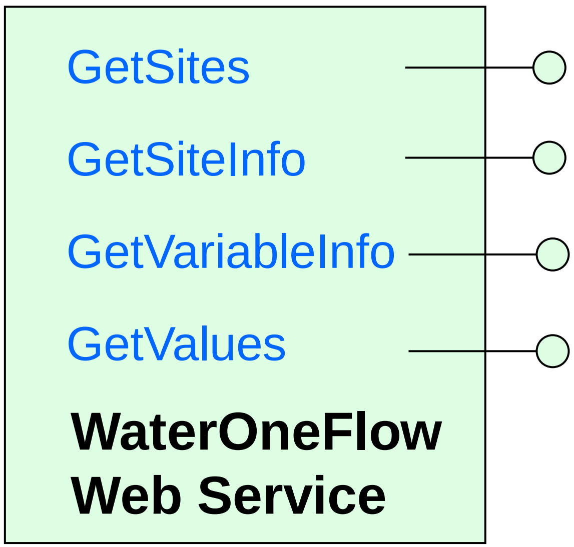
- Provides a common persistence model for observations data

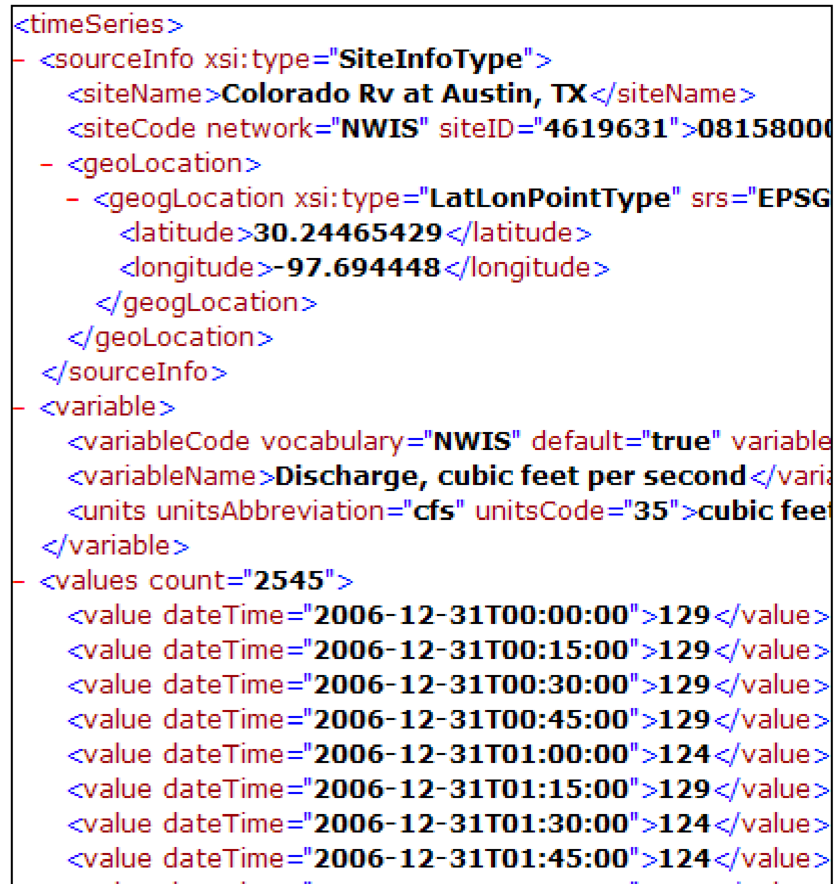
- Syntactic heterogeneity (File types and formats)
- Semantic heterogeneity
  - Language for observation attributes (structural)
  - Language to encode observation attribute values (contextual)
- Publishing and sharing research data
- Metadata to facilitate unambiguous interpretation
- Enhance analysis capability

WaterML and WaterOneFlow

WaterML is an XML language for communicating water data
WaterOneFlow is a set of web services based on WaterML

- Set of query functions
- Returns data in WaterML



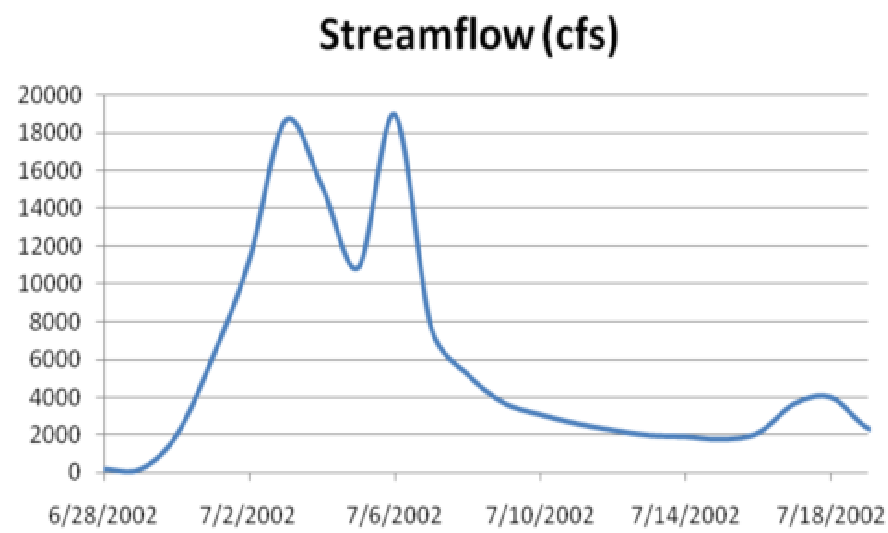


WaterML as a Web Language

USGS Streamflow data in WaterML language

Discharge of the San Marcos River at Luling, TX
June 28 - July 18, 2002


```
- <values count="21">
  <value qualifiers="A" dateTime="2002-06-28T00:00:00">203</value>
  <value qualifiers="A" dateTime="2002-06-29T00:00:00">195</value>
  <value qualifiers="A" dateTime="2002-06-30T00:00:00">2010</value>
  <value qualifiers="A" dateTime="2002-07-01T00:00:00">6170</value>
  <value qualifiers="A" dateTime="2002-07-02T00:00:00">11300</value>
  <value qualifiers="A" dateTime="2002-07-03T00:00:00">18700</value>
  <value qualifiers="A" dateTime="2002-07-04T00:00:00">15200</value>
  <value qualifiers="A" dateTime="2002-07-05T00:00:00">10900</value>
  <value qualifiers="A" dateTime="2002-07-06T00:00:00">19000</value>
  <value qualifiers="A" dateTime="2002-07-07T00:00:00">7720</value>
  <value qualifiers="A" dateTime="2002-07-08T00:00:00">5230</value>
  <value qualifiers="A" dateTime="2002-07-09T00:00:00">3710</value>
  <value qualifiers="A" dateTime="2002-07-10T00:00:00">3090</value>
  <value qualifiers="A" dateTime="2002-07-11T00:00:00">2610</value>
  <value qualifiers="A" dateTime="2002-07-12T00:00:00">2260</value>
  <value qualifiers="A" dateTime="2002-07-13T00:00:00">1990</value>
  <value qualifiers="A" dateTime="2002-07-14T00:00:00">1920</value>
  <value qualifiers="A" dateTime="2002-07-15T00:00:00">1780</value>
  <value qualifiers="A" dateTime="2002-07-16T00:00:00">2120</value>
  <value qualifiers="A" dateTime="2002-07-17T00:00:00">3680</value>
  <value qualifiers="A" dateTime="2002-07-18T00:00:00">4010</value>
  <qualifier qualifierCode="A" network="USGS" vocabulary="dv_rmk_cd">Approved for publication -- Processing and review completed.</qualifier
</values>
```

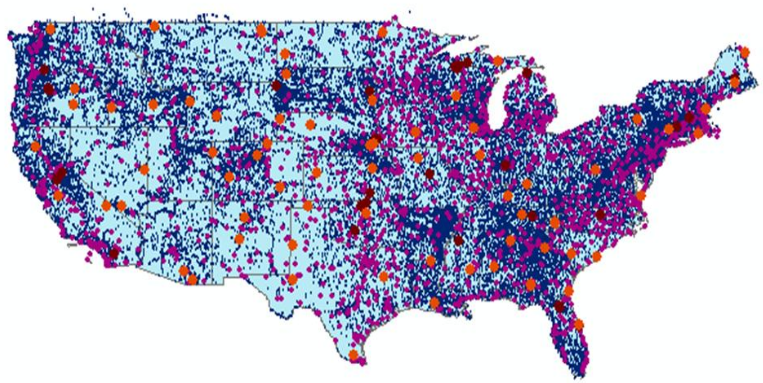



This is the WaterML GetValues response from NWIS Daily Values

Open Geospatial Consortium Web Service Standards

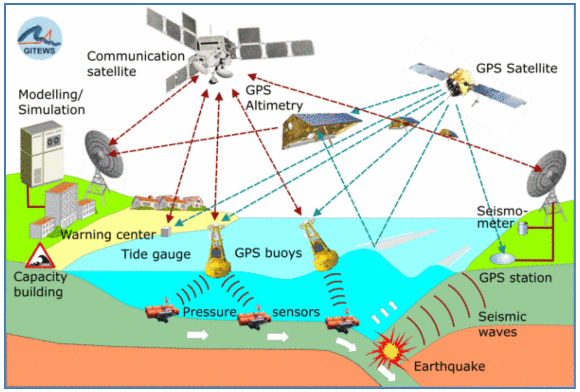
These standards have been developed over the past 10 years …. …. by 400 companies and agencies working within the OGC

- Map Services



- Web Map Service (WMS)
- Web Feature Service (WFS)
- Web Coverage Service (WCS)
- Catalog Services for the Web (CS/W)

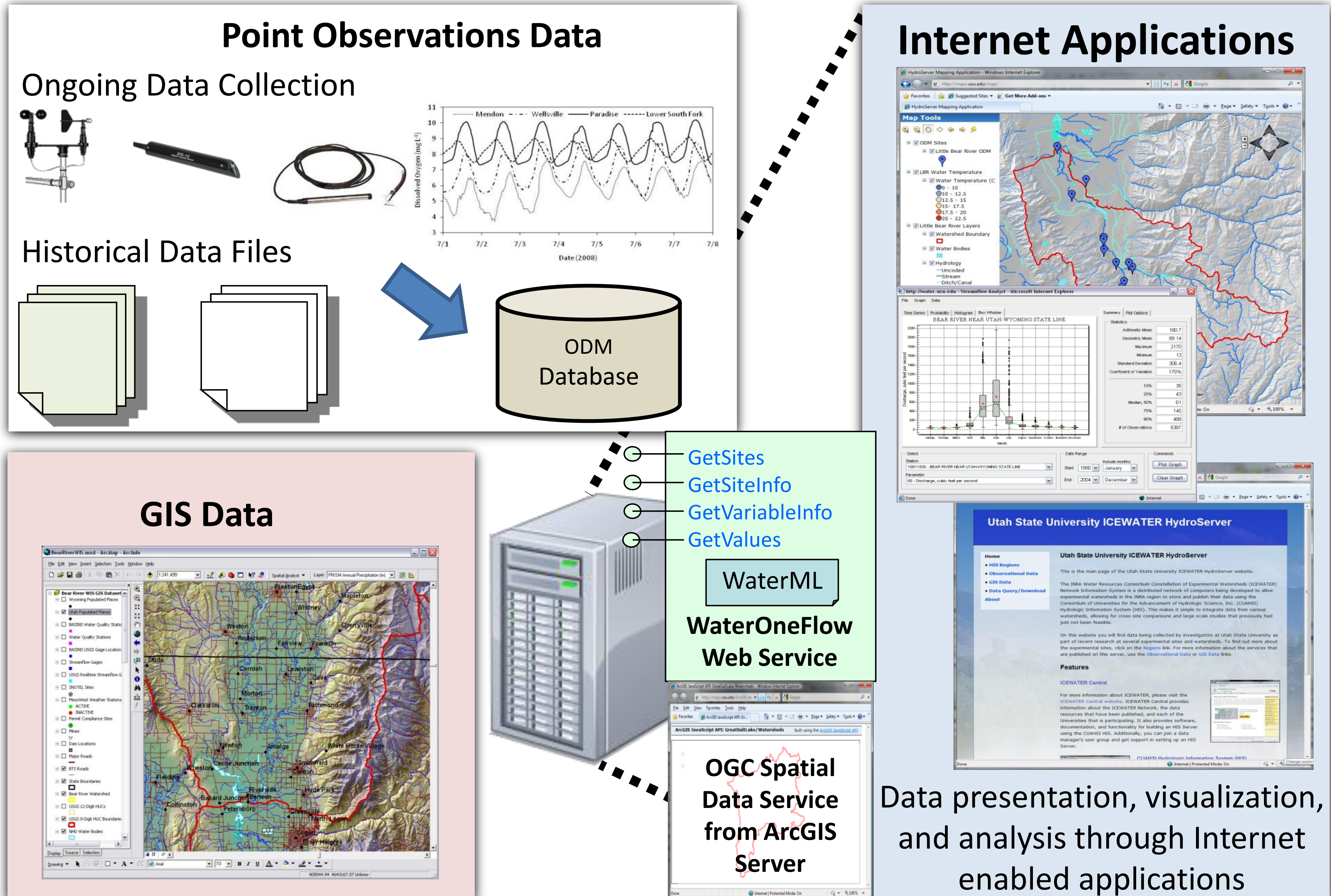
- Observation Services



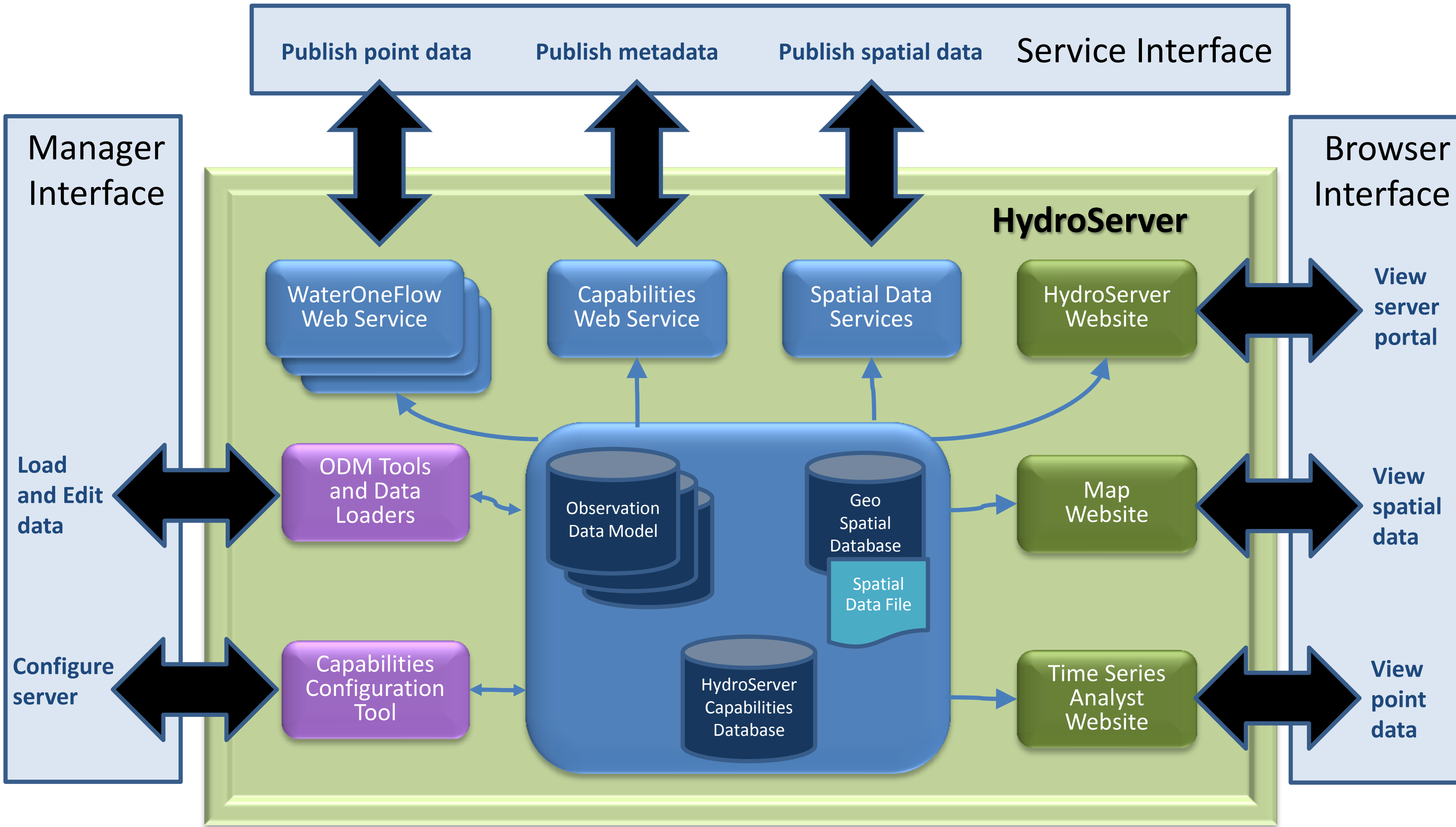
- Observations and Measurements Model
- Sensor Web Enablement (SWE)
- Sensor Observation Service (SOS)

OGC Hydrology Domain Working Group evolving WaterML into an International Standard
http://www.opengeospatial.org/projects/groups/waterml2.0swg

HydroServer – Data Publication

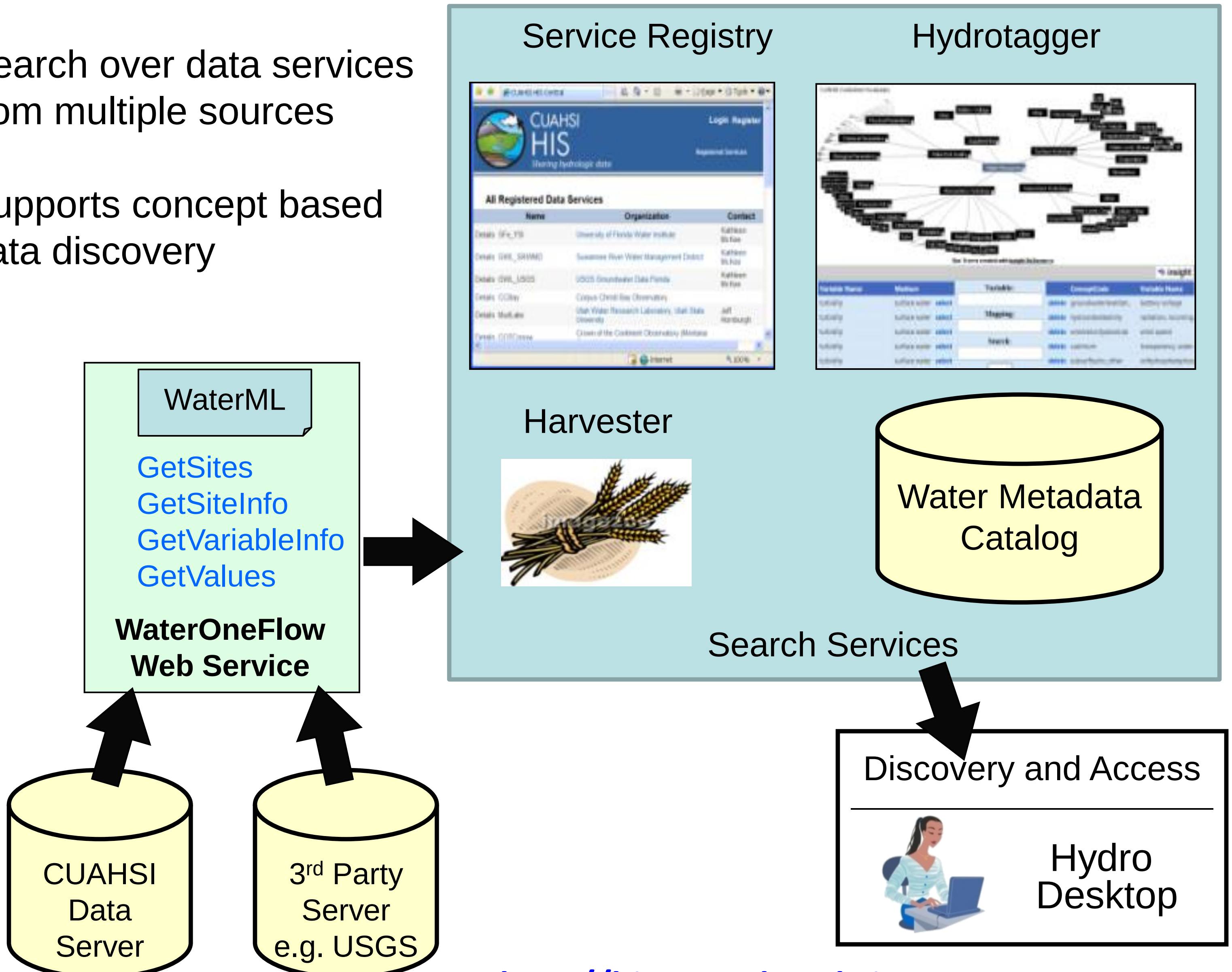


HydroServer



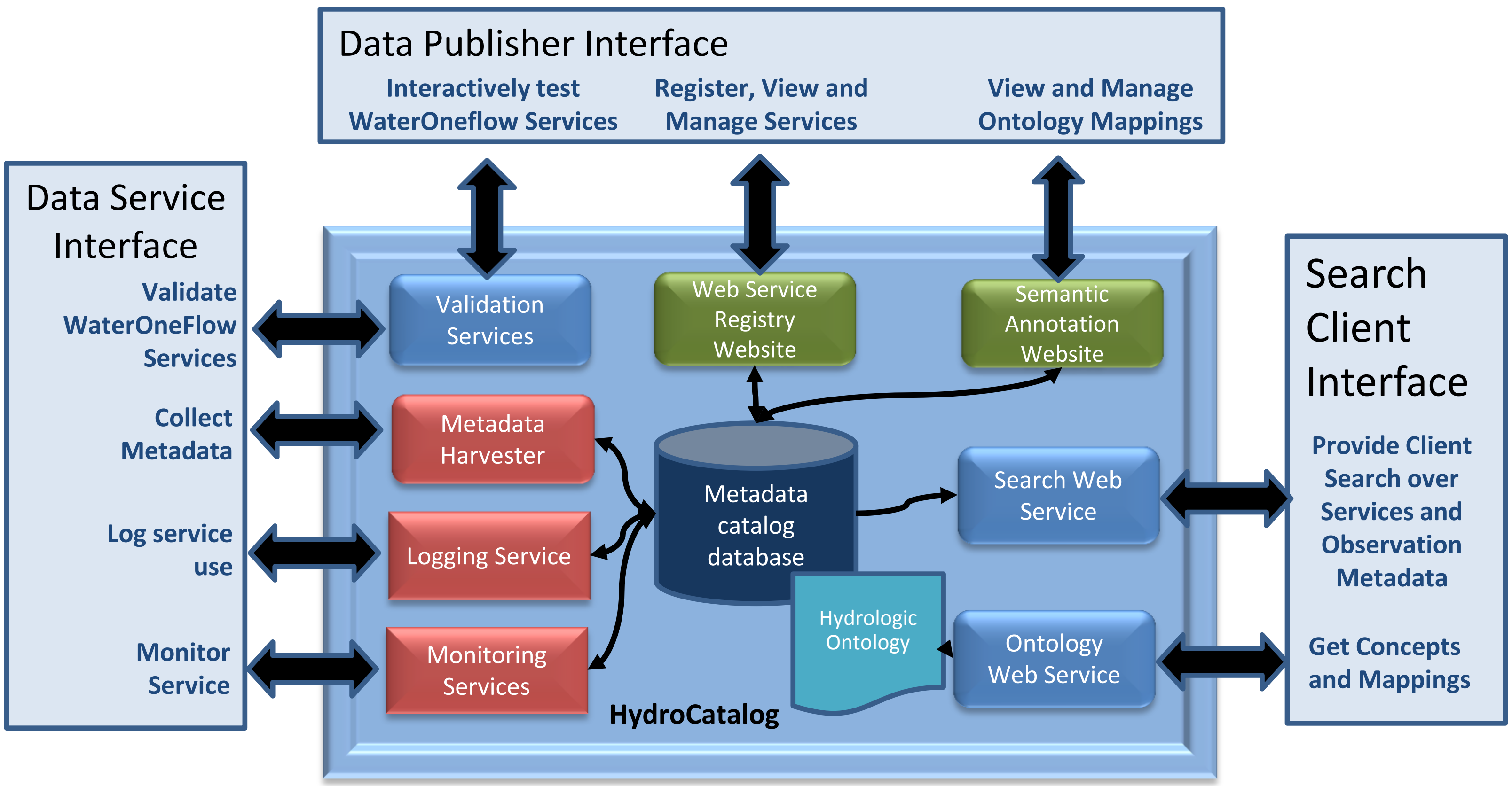
HydroCatalog

- Search over data services from multiple sources
- Supports concept based data discovery



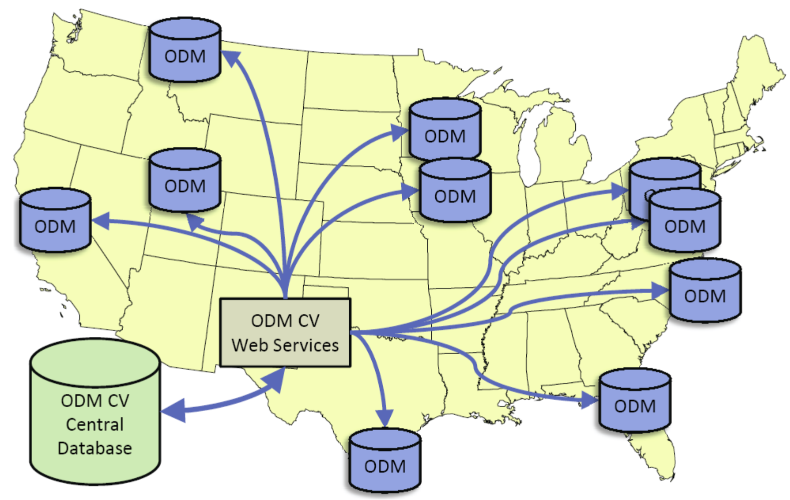
http://hiscentral.cuahsi.org

HydroCatalog



Overcoming Semantic Heterogeneity

- ODM Controlled Vocabulary System
  - ODM CV central database
  - Online submission and editing of CV terms
  - Web services for broadcasting CVs



ODM VariableNameCV

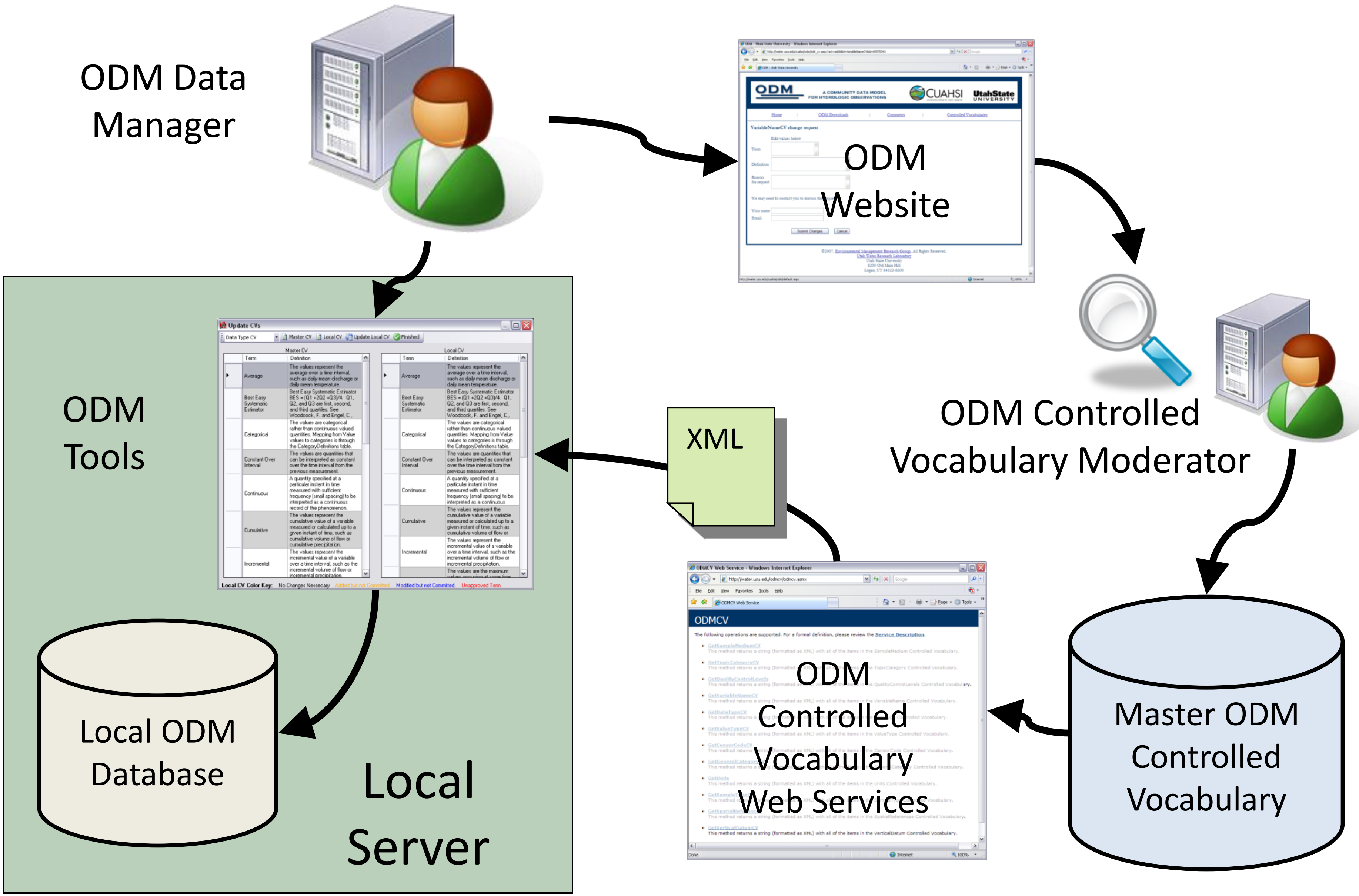
Variable Name

Investigator 1:	“Temperature, water”
Investigator 2:	“Water Temperature”
Investigator 3:	“Temperature”
Investigator 4:	“Temp.”

Term
…
Sunshine duration
Temperature
Turbidity
…

From Jeff Horsburgh

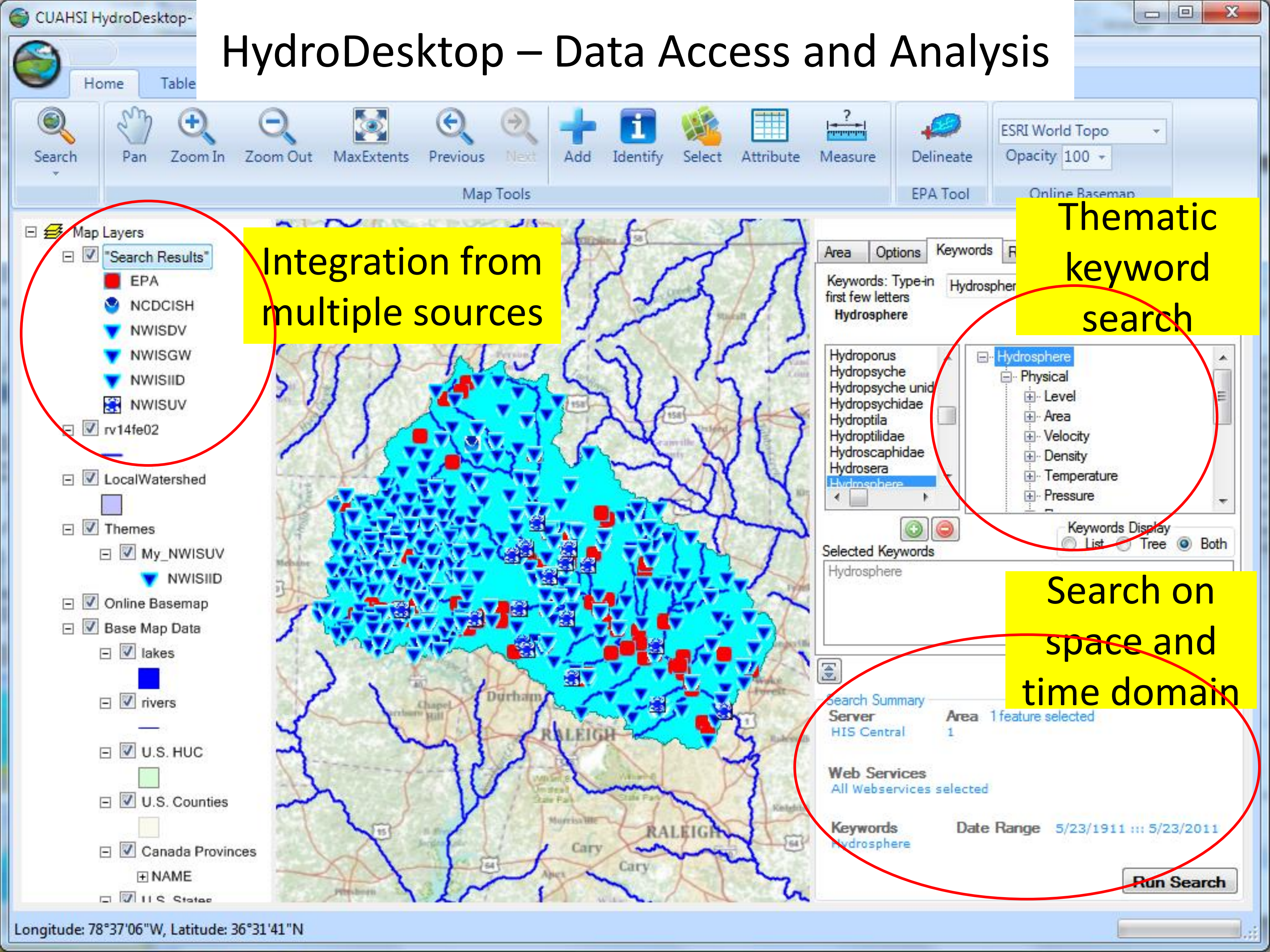
Dynamic controlled vocabulary moderation system



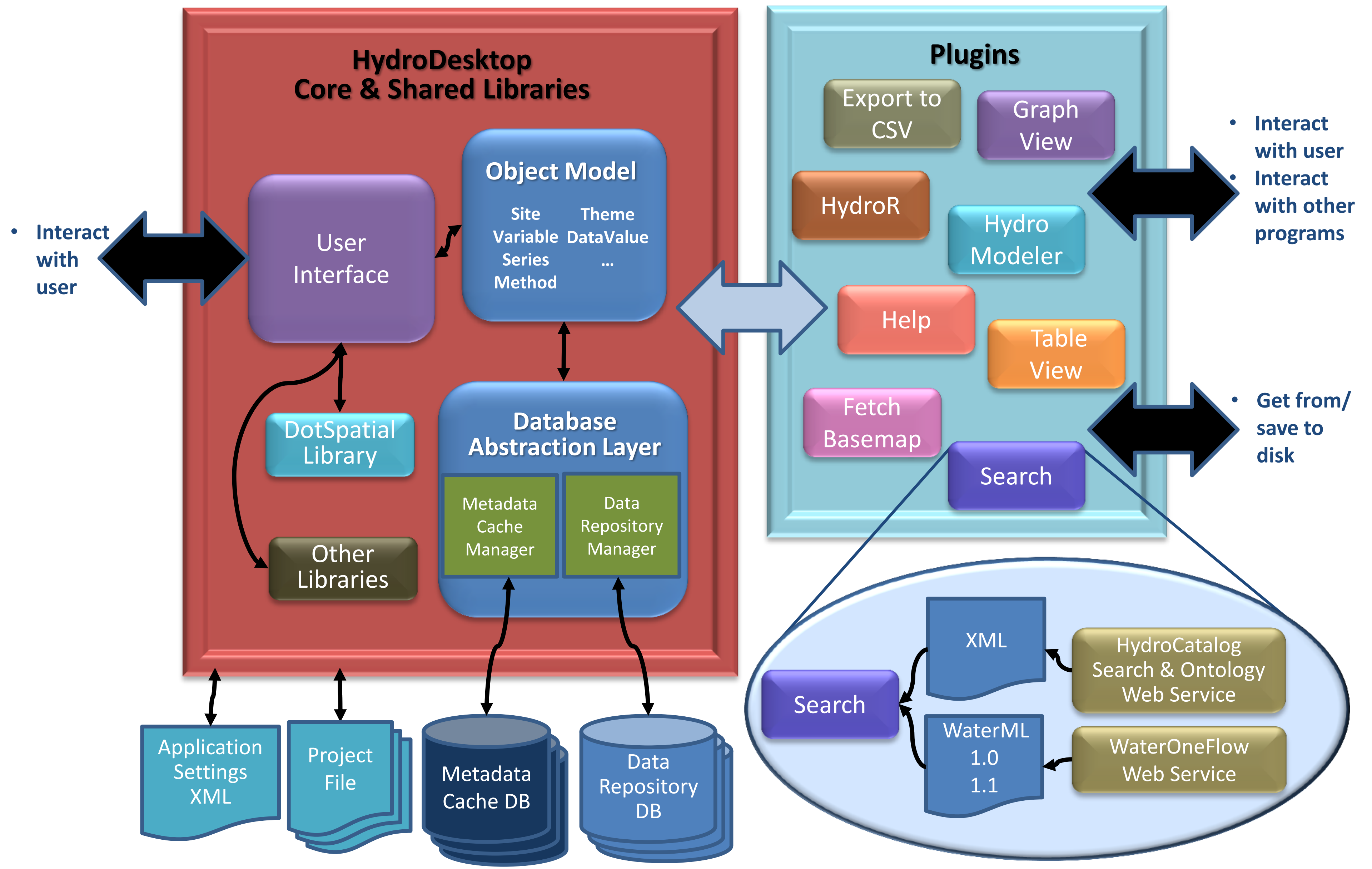
http://his.cuahsi.org/mastercvreg.html

From Jeff Horsburgh

HydroDesktop – Data Access and Analysis

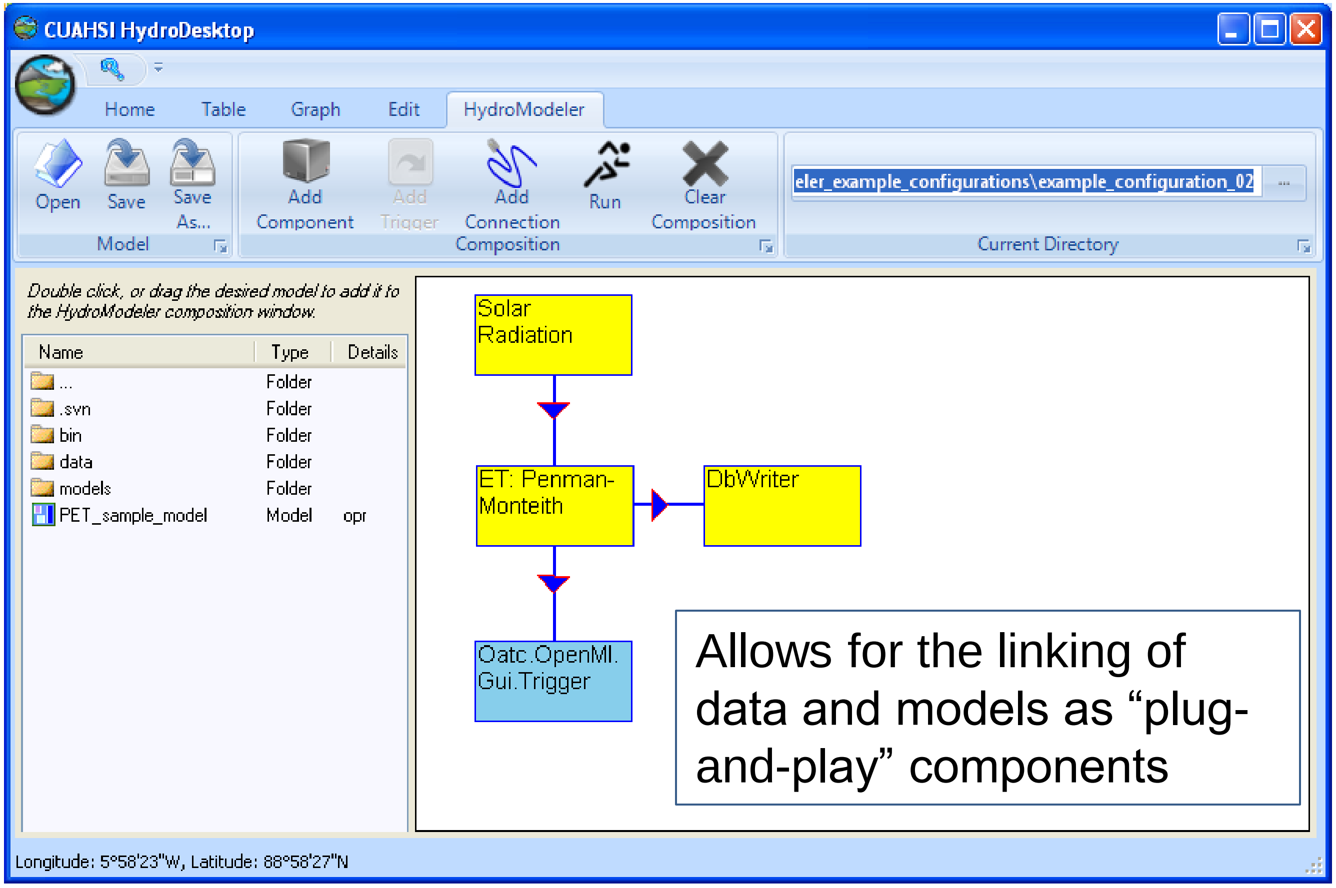


HydroDesktop



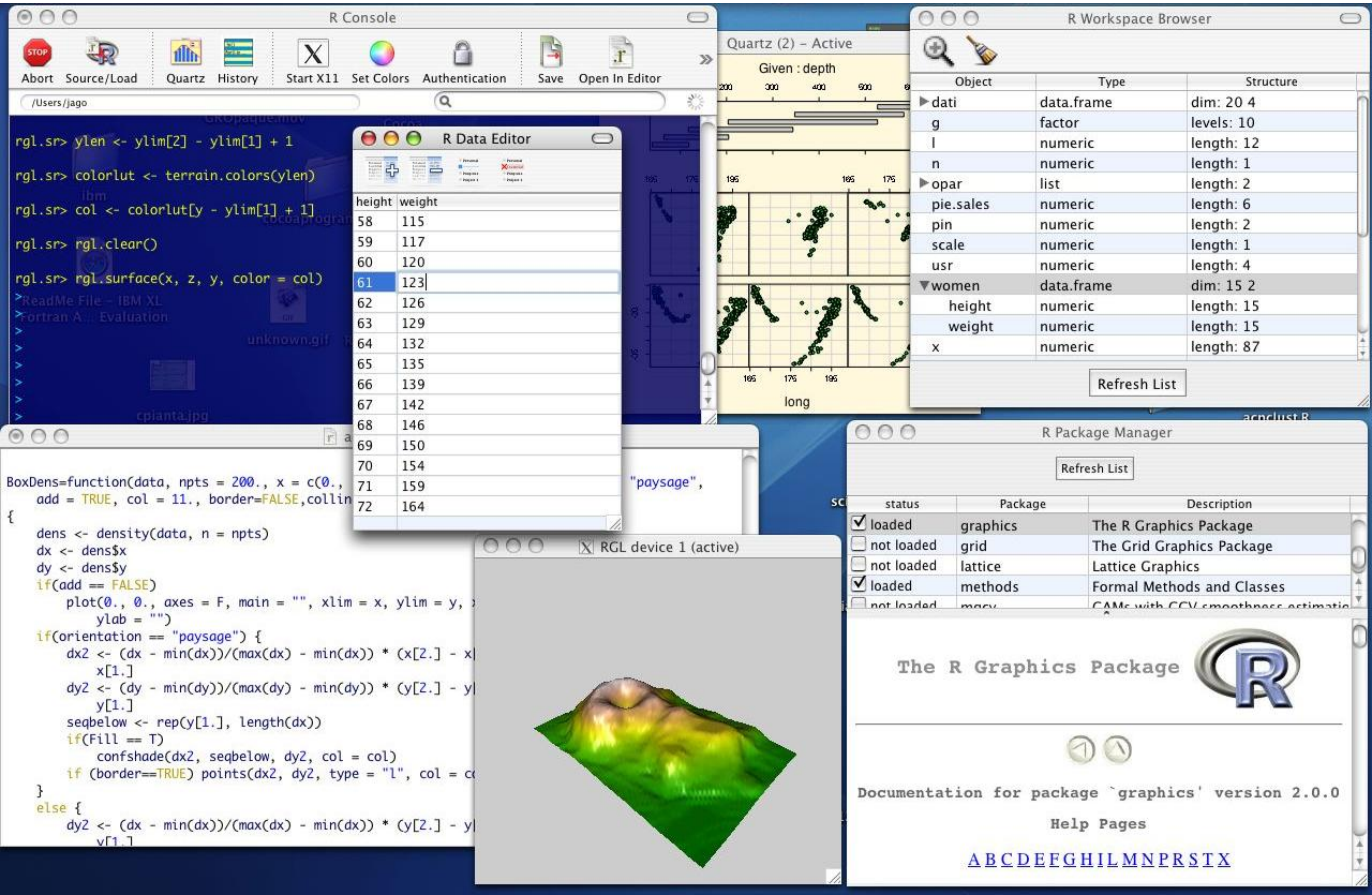
HydroModeler

An integrated modeling environment based on the Open Modeling Interface (OpenMI) standard and embedded within HydroDesktop



In development at the University of South Carolina by Jon Goodall, Tony Castronova, Mehmet Ercan, Mostafa Elag, and Shirani Fuller

Integration with “R” Statistics Package

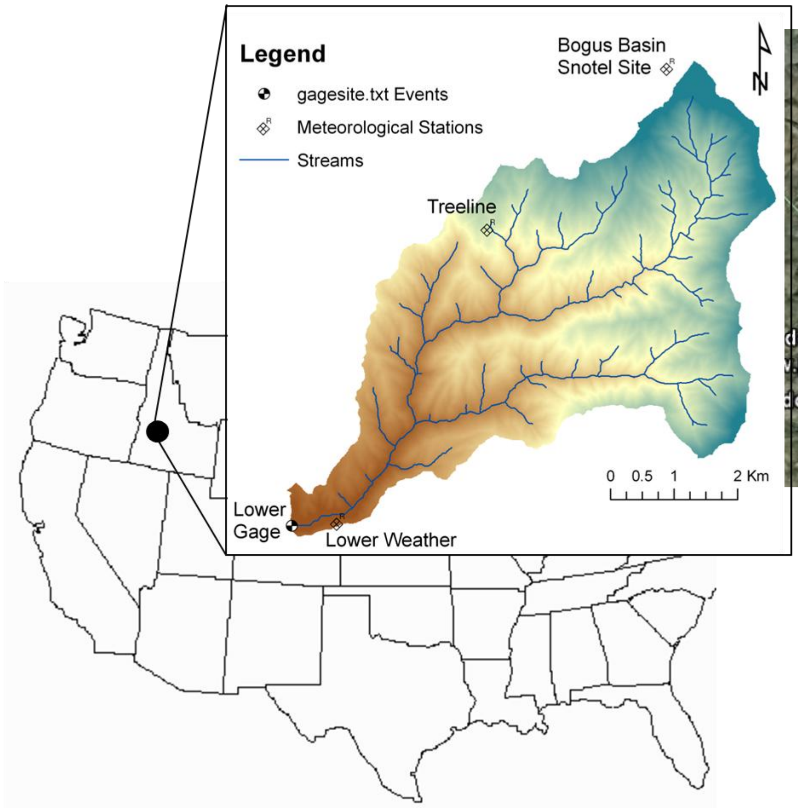


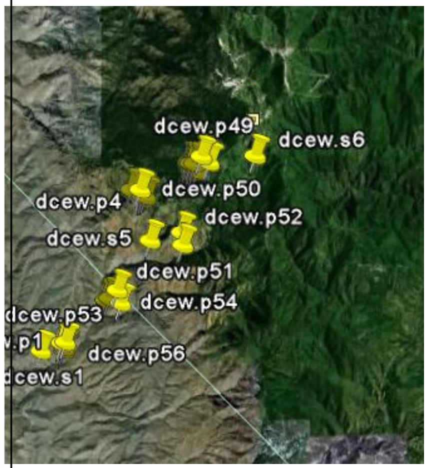
37 Water Data Services on HIS Central from 12 Universities

- University of Maryland, Baltimore County
- Montana State University
- University of Texas at Austin
- University of Iowa
- Utah State University
- University of Florida
- University of New Mexico
- University of Idaho
- Boise State University
- University of Texas at Arlington
- University of California, San Diego
- Idaho State University

Dry Creek Experimental Watershed (DCEW)

(28 km^2 semi-arid steep topography, Boise Front)





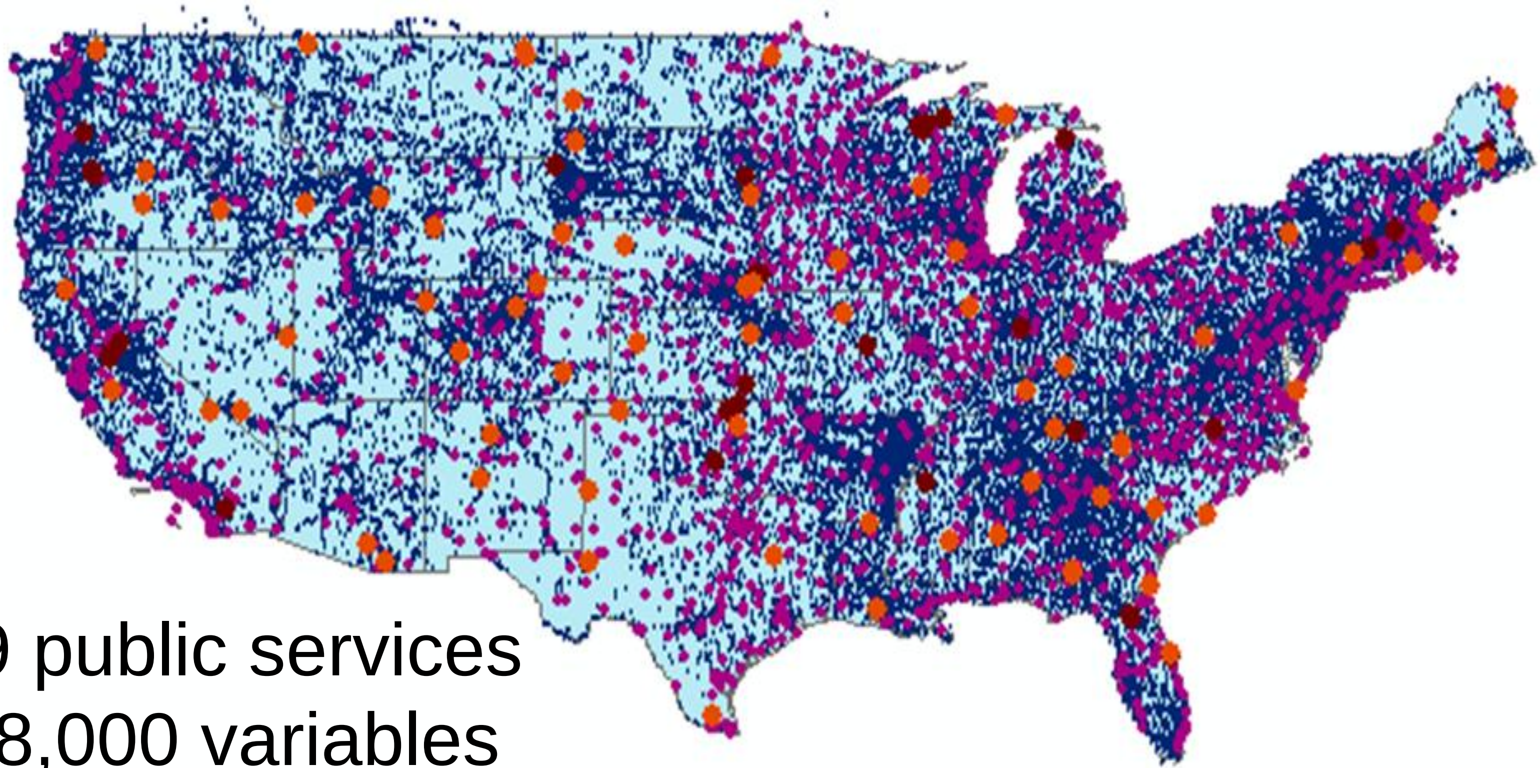
68 Sites
24 Variables
4,700,000+ values

Published by Jim McNamara, Boise State University

Water Agencies and Industry

- USGS, NCDC , Corps of Engineers publishing data using HIS WaterML
- OGC Hydrology Domain Working Group evaluating WaterML as OGC standard
- ESRI using CUAHSI model in ArcGIS.com GIS data collaboration portal
- Kisters WISKI support for WaterML data publication
- Australian Water Resources Information System Water Accounting System has adopted aspects of HIS
- NWS West Gulf River Forecast Center Multi-sensor Precipitation Estimate published from ODM using WaterML

CUAHSI Water Data Services Catalog



69 public services
18,000 variables
1.9 million sites
23 million series
5.1 billion data values
(as of June 2011)

The largest water data catalog in the world

maintained at the San Diego Supercomputer Center

Open Development Model

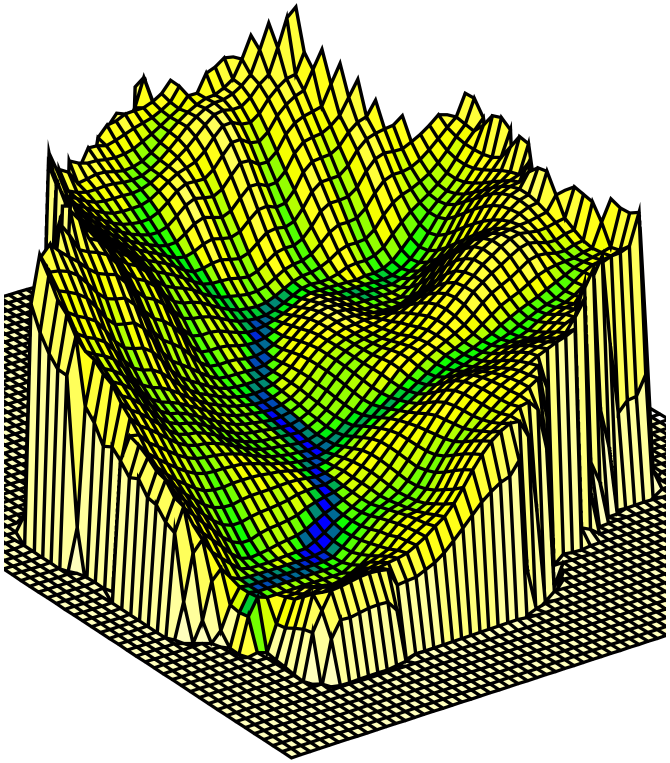
- http://hydrodesktop.codeplex.com
- http://hydroserver.codeplex.com
- http://hydrocatalog.codeplex.com

Summary

- **Data Storage** in an *Observations Data Model* (ODM) and publication through HydroServer
- **Data Access** through internet-based *Water Data Services* using a consistent data language, called WaterML from HydroDesktop
- **Data Discovery** through a *National Water Metadata Catalog* and thematic keyword search system at Central HydroCatalog (SDSC)
- **Integrated Modeling and Analysis** within HydroDesktop

The combination of these capabilities creates a common window on water observations data for the United States unlike any that has existed before.

Terrain Analysis Using Digital Elevation Models (TauDEM)



David Tarboton[1], Dan Watson[2], Rob Wallace,[3] Kim Schreuders[1], Jeremy Neff[1]

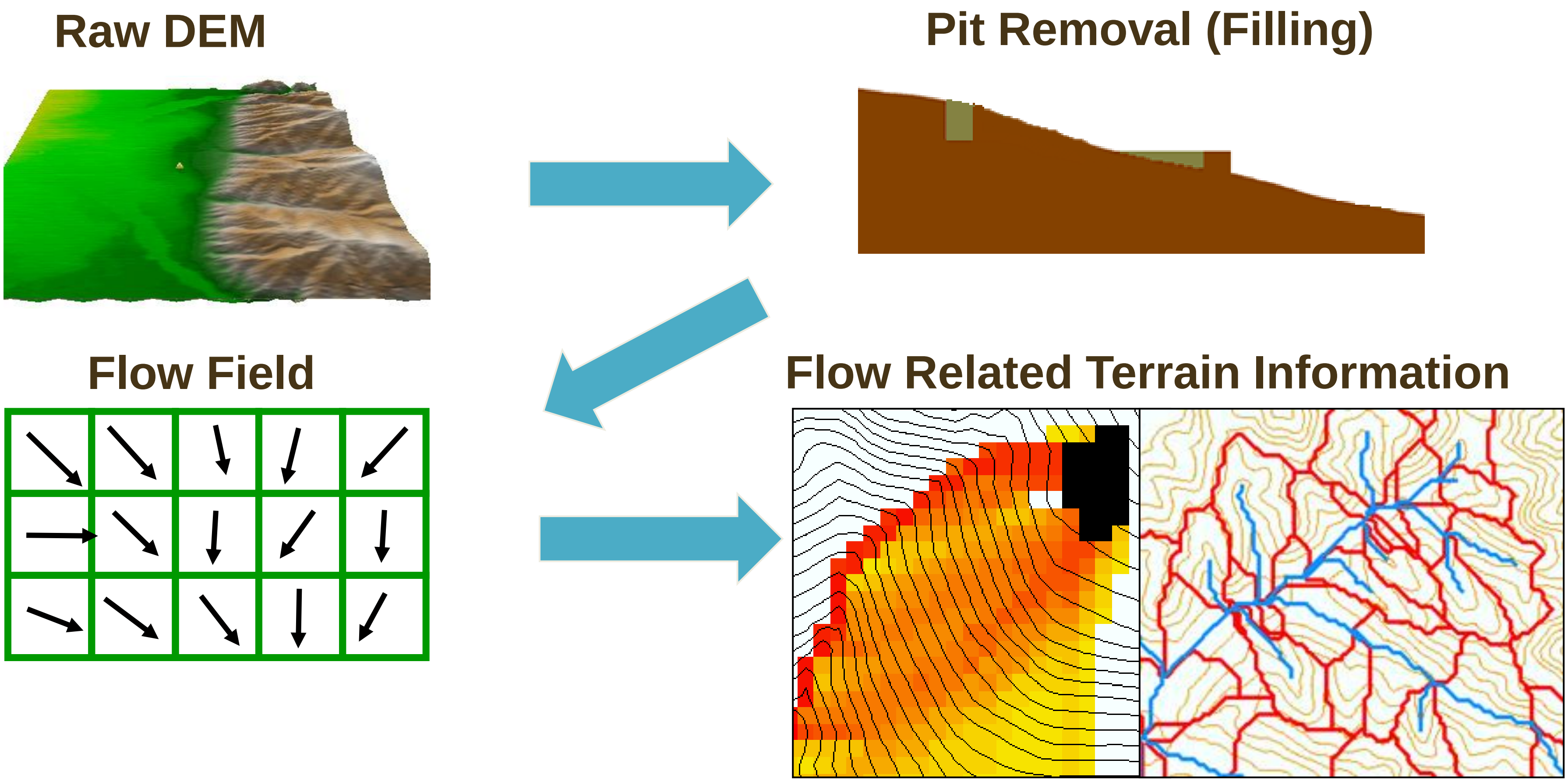
[1]Utah Water Research Laboratory, Utah State University, Logan, Utah

[2]Computer Science, Utah State University, Logan, Utah

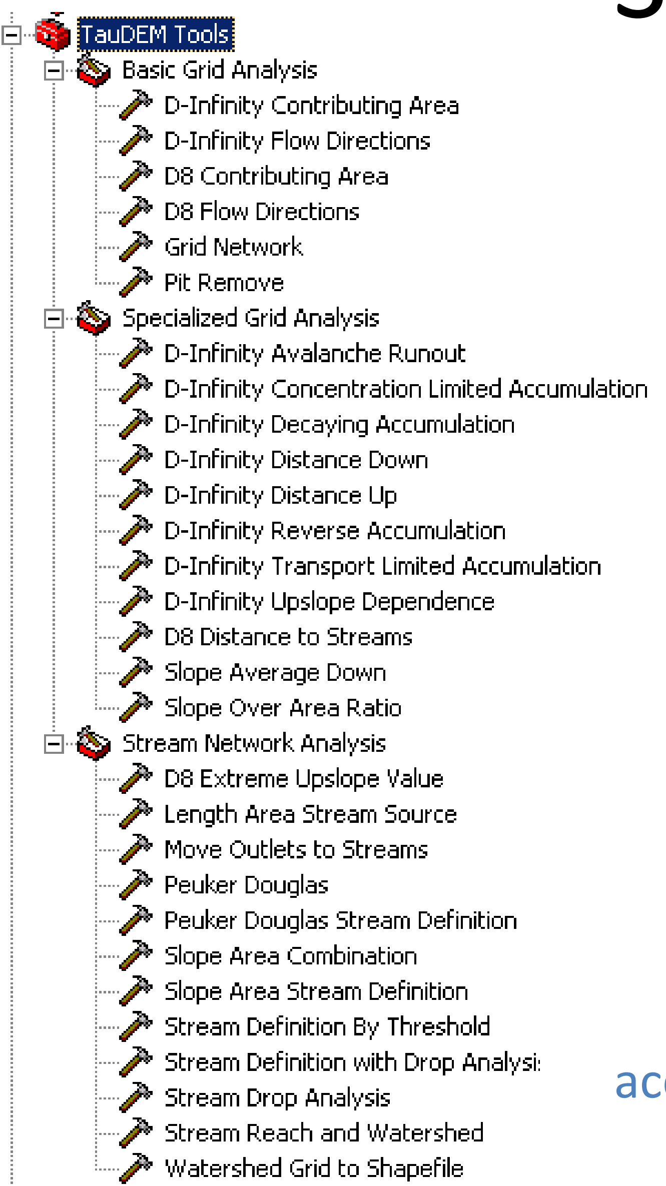
[3]US Army Engineer Research and Development Center, Information Technology Lab, Vicksburg, Mississippi

This research was funded by the US Army Research and Development Center under contract number W9124Z-08-P-0420

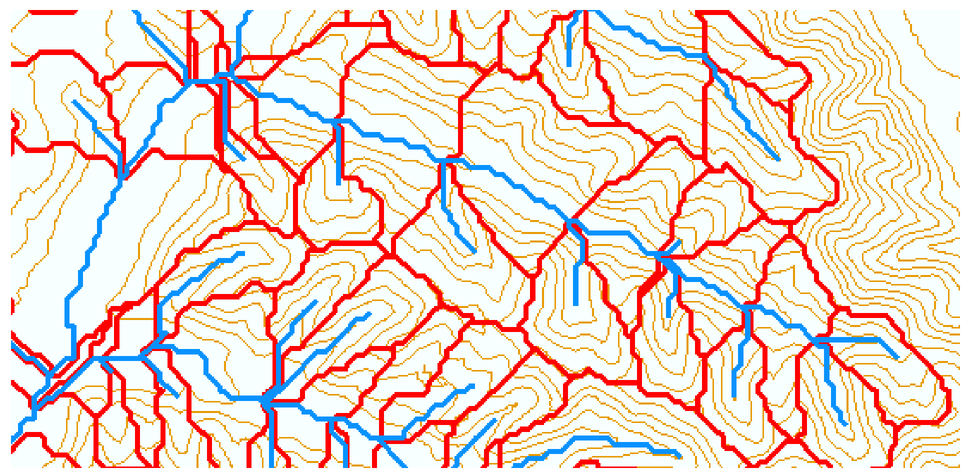
Deriving hydrologically useful information from Digital Elevation Models



A parallel version of the TauDEM Software Tools



- Improved runtime efficiency
- Capability to run larger problems
- Platform independence of core functionality



Deployed as an ArcGIS Toolbox with tools that drive accompanying command line executables, available from http://hydrology.usu.edu/taudem/

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 10-30 m

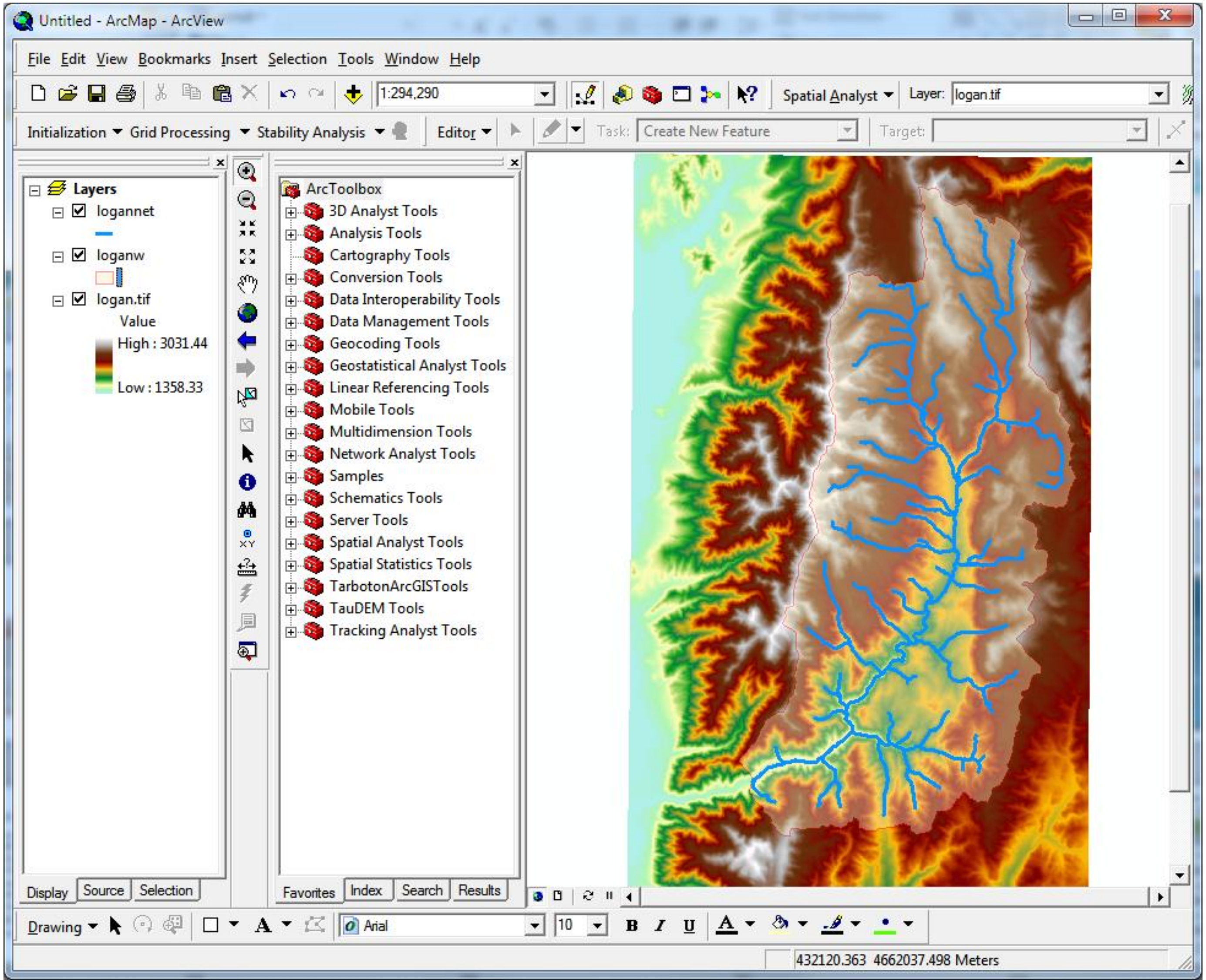
10^4 cells/km^2

2010's LIDAR ~1 m

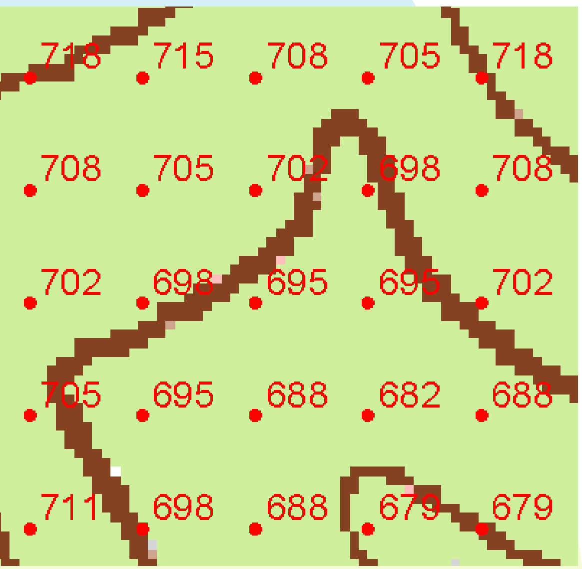
10^6 cells/km^2

Website and Demo

- http://hydrology.usu.edu/taudem



Grid Data Format Assumptions



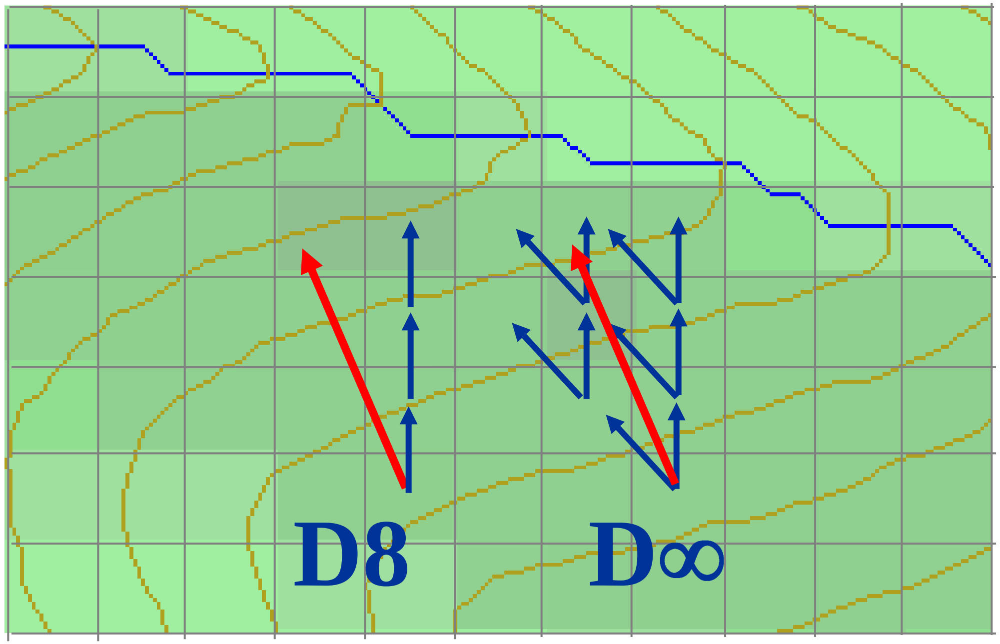
- Input and output grids are uncompressed GeoTIFF
- Maximum size 4 GB
- GDAL Nodata tag preferred (if not present, a missing value is assumed)
- Grids are square ($\Delta x = \Delta y$)
- Grids have identical extents, cell size and spatial reference
- Spatial reference information is not used (no projection on the fly)

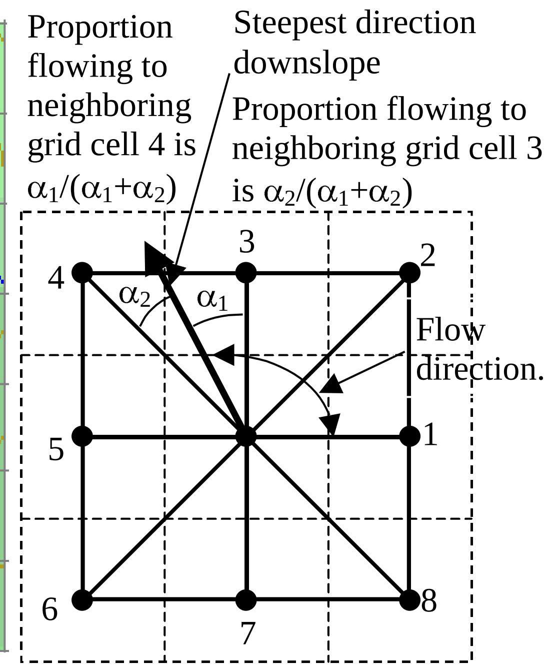
Representation of Flow Field

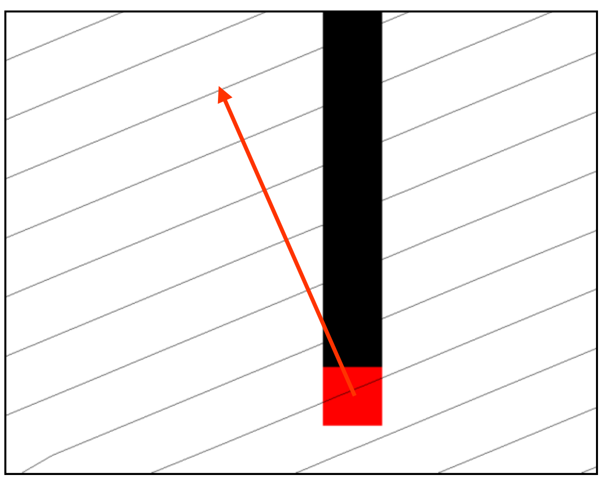
Steepest single direction

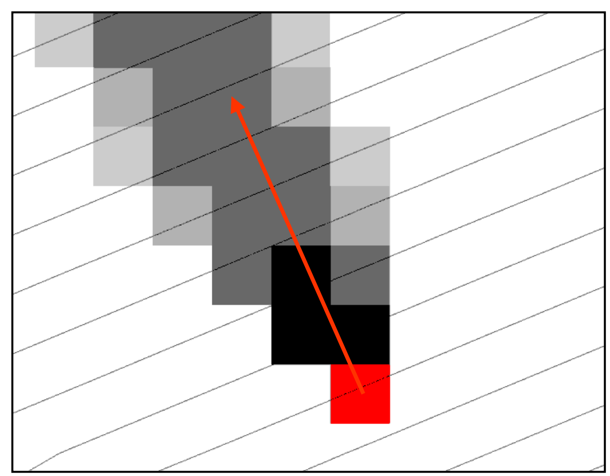
48	**52**
56	**67**

$$\frac{67-52}{30} = 0.50$$





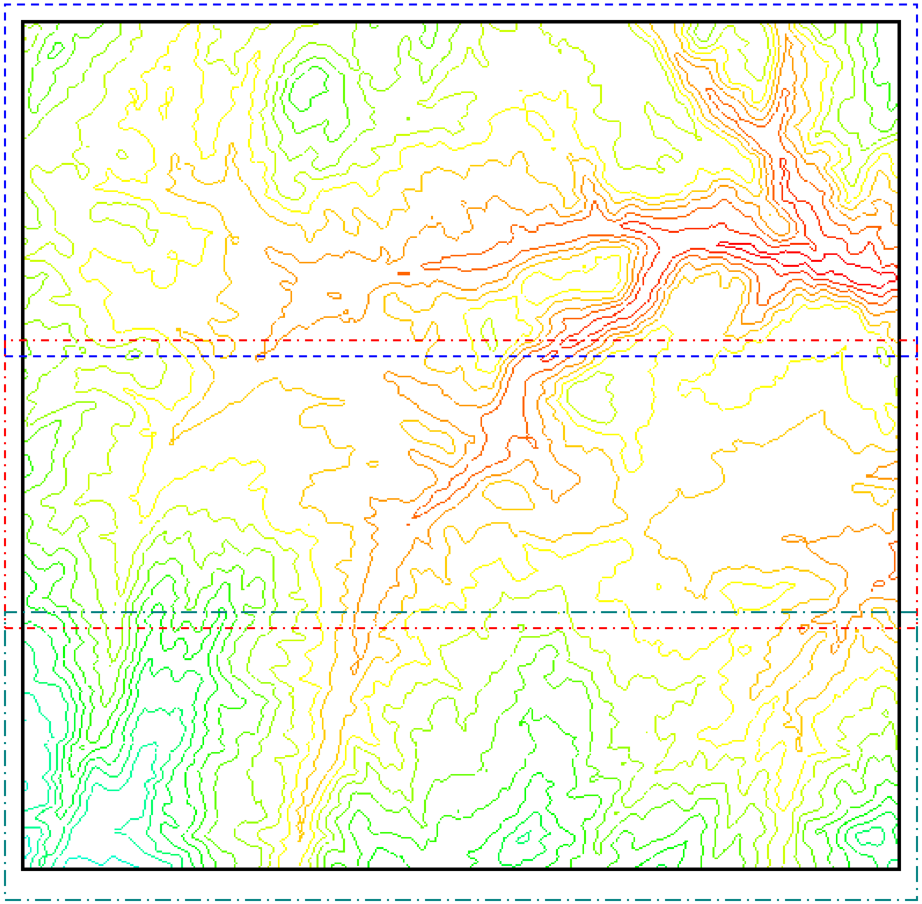




Tarboton, D. G., (1997), "A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models," Water Resources Research, 33(2): 309-319.)

Parallel Approach

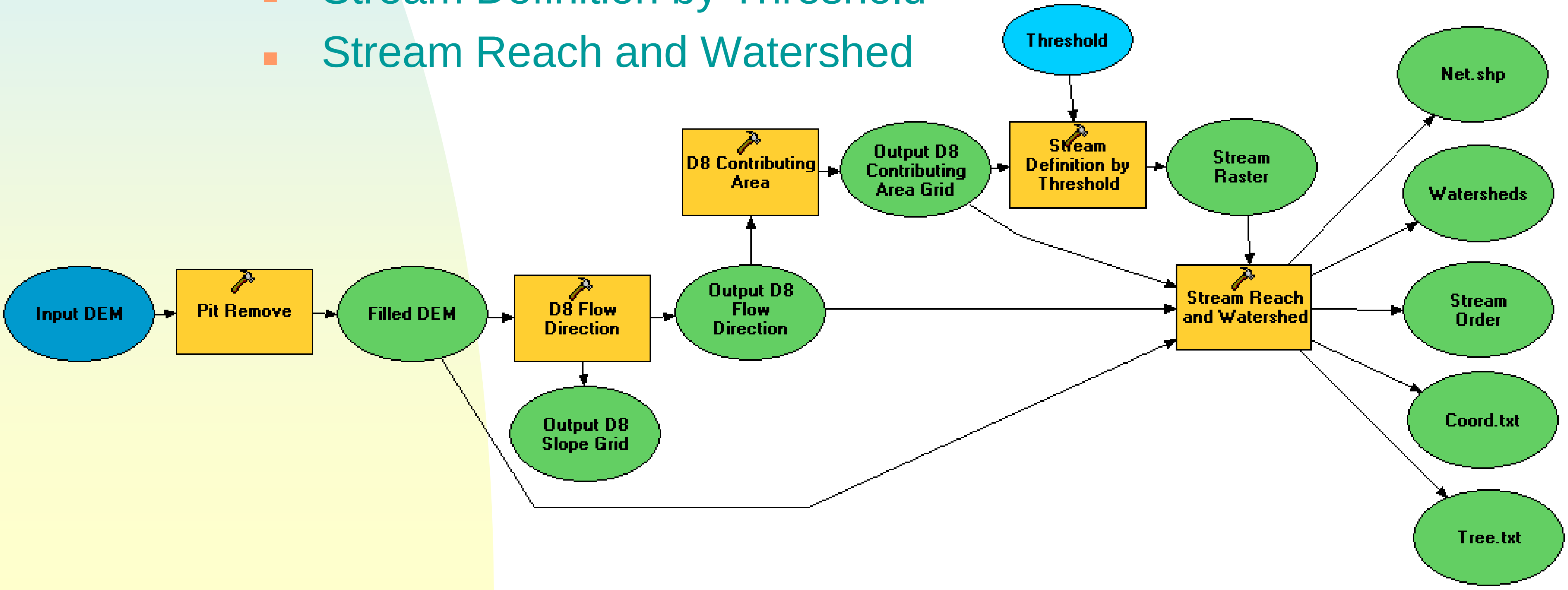
- MPI, distributed memory paradigm
- Row oriented slices
- Each process includes one buffer row on either side
- Each process does not change buffer row



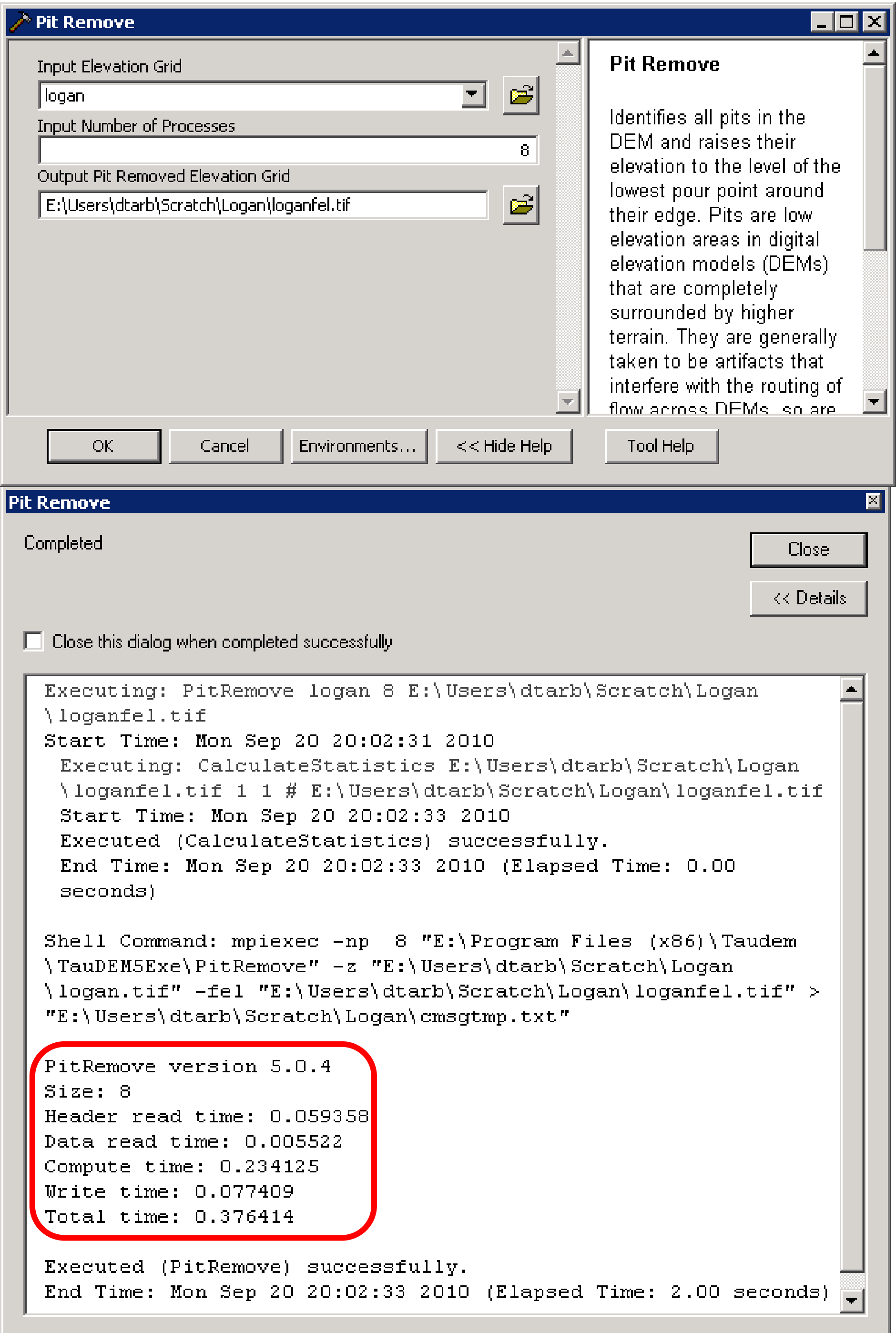
Illustrative Use Case: Delineation of channels and watersheds using a constant support area threshold

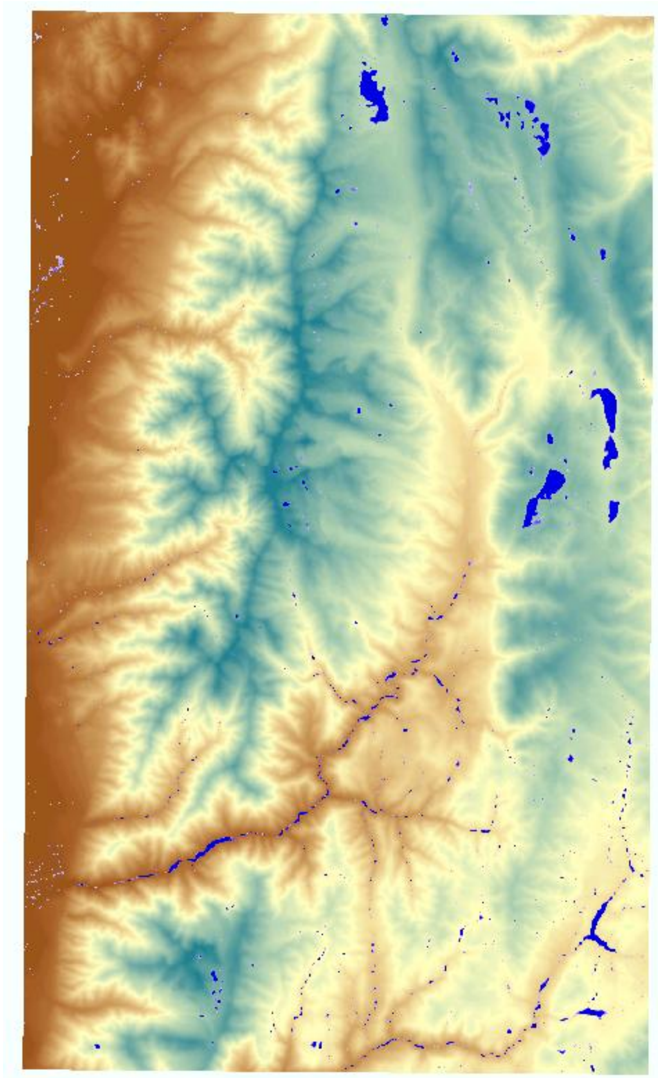
Steps

- Pit Remove
- D8 Flow Directions
- D8 Contributing Area
- Stream Definition by Threshold
- Stream Reach and Watershed



Pit Remove


Pit Remove
Input Elevation Grid
logan
Input Number of Processes
8
Output Pit Removed Elevation Grid
E:\Users\dtarb\Scratch\Logan\loganfel.tif
Pit Remove
Identifies all pits in the DEM and raises their elevation to the level of the lowest pour point around their edge. Pits are low elevation areas in digital elevation models (DEMs) that are completely surrounded by higher terrain. They are generally taken to be artifacts that interfere with the routing of flow across DEMs, so are
OK
Cancel
Environments...
<< Hide Help
Tool Help
Pit Remove
Completed
Close
<< Details
Close this dialog when completed successfully
Executing: PitRemove logan 8 E:\Users\dtarb\Scratch\Logan \loganfel.tif
Start Time: Mon Sep 20 20:02:31 2010
Executing: CalculateStatistics E:\Users\dtarb\Scratch\Logan \loganfel.tif 1 1 # E:\Users\dtarb\Scratch\Logan\loganfel.tif
Start Time: Mon Sep 20 20:02:33 2010
Executed (CalculateStatistics) successfully.
End Time: Mon Sep 20 20:02:33 2010 (Elapsed Time: 0.00 seconds)
Shell Command: mpiexec -np 8 "E:\Program Files (x86)\Taudem \TauDEM5Exe\PitRemove" -z "E:\Users\dtarb\Scratch\Logan \logan.tif" -fel "E:\Users\dtarb\Scratch\Logan\loganfel.tif" > "E:\Users\dtarb\Scratch\Logan\cmsgtmp.txt"
PitRemove version 5.0.4
Size: 8
Header read time: 0.059358
Data read time: 0.005522
Compute time: 0.234125
Write time: 0.077409
Total time: 0.376414
Executed (PitRemove) successfully.
End Time: Mon Sep 20 20:02:33 2010 (Elapsed Time: 2.00 seconds)



D8 Flow Direction (and Slope)

D8 Flow Direction

Input Pit Filled Elevation Grid

loganfel

Input Number of Processes

8

Output D8 Flow Direction Grid

E:\Users\dtarb\Scratch\Logan\loganp.tif

Output D8 Slope Grid

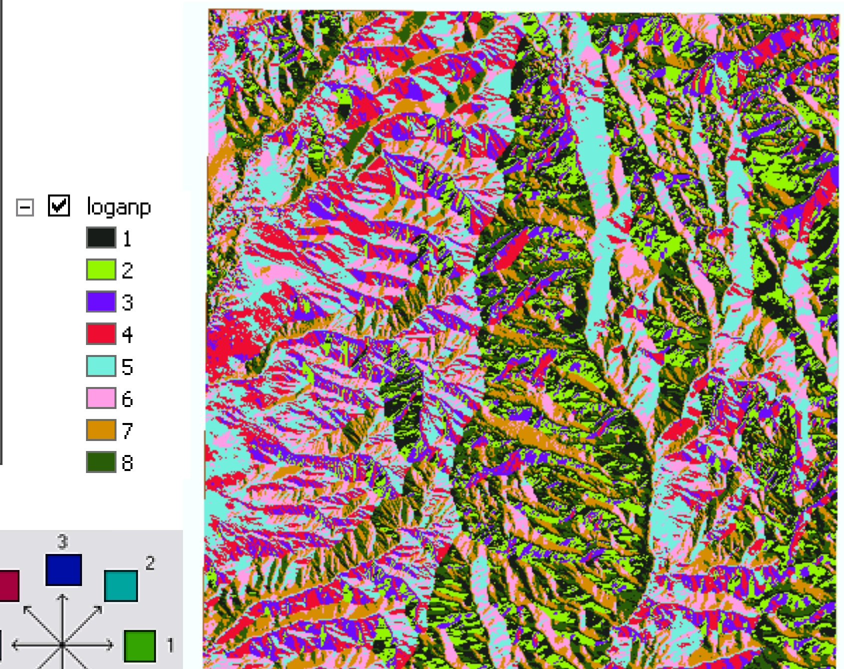
E:\Users\dtarb\Scratch\Logan\logansd8.tif

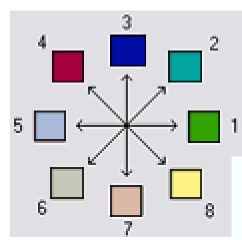
D8 Flow Direction

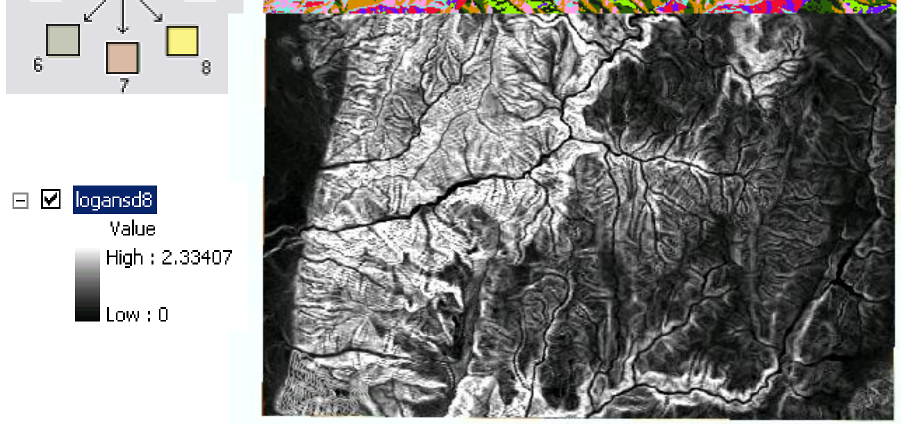
Calculates 2 grids. The first contains the D8 flow directions which are defined, for each cell, as the direction of the one of its eight adjacent or diagonal neighbors with the steepest downward slope. Flow Direction Coding: 1 - East, 2 - North East, 3 - North, 4 - North West , 5 - West, 6 - South West, 7 - South, 8 - South East. The

OK | Cancel | Environments... | << Hide Help | Tool Help

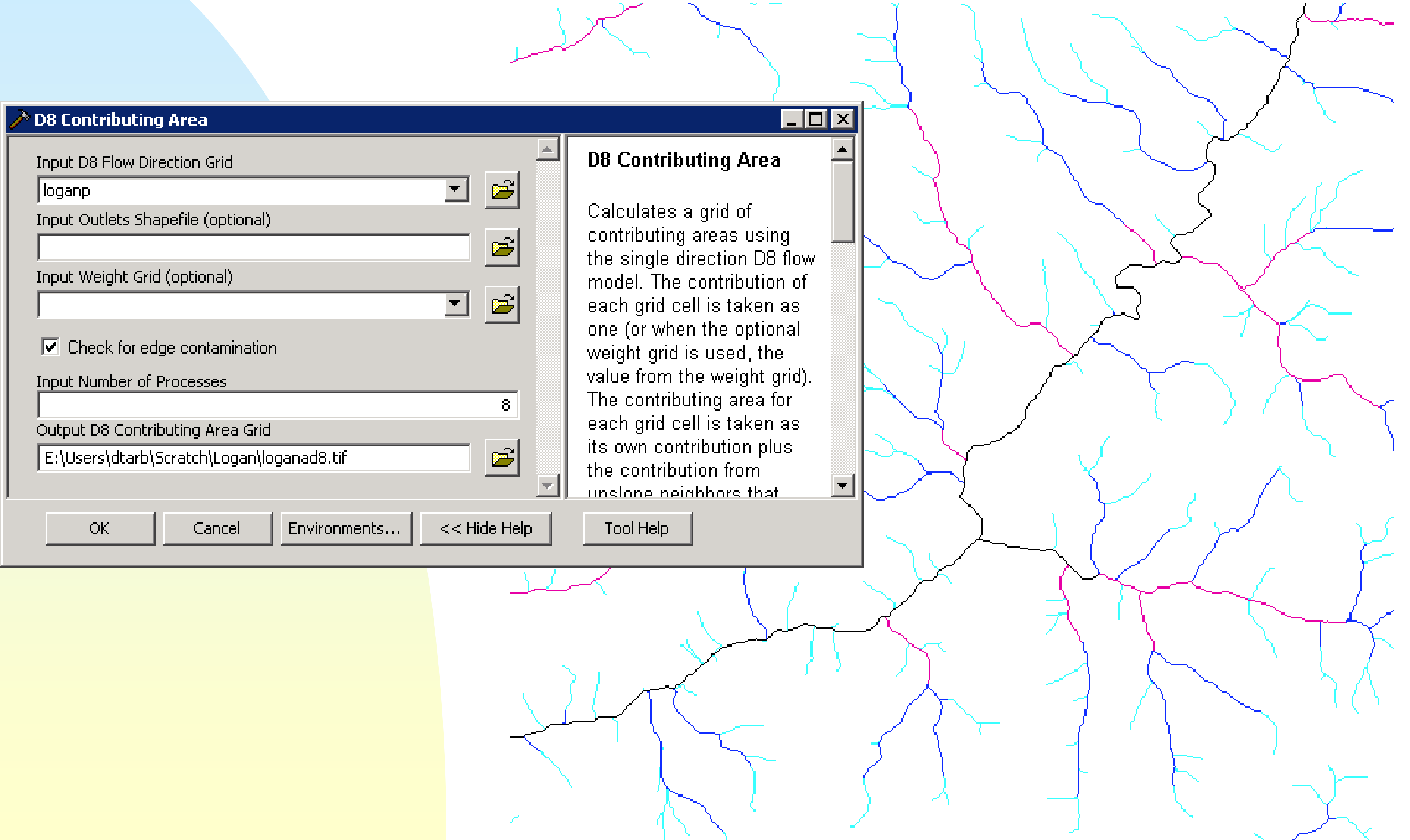
```
Executed (CalculateStatistics) successfully.
End Time: Mon Sep 20 20:11:18 2010 (Elapsed Time: 1.00
seconds)

Shell Command: mpiexec -np  8 "E:\Program Files (x86)\Taudem
\TauDEM5Exe\D8FlowDir" -fel "E:\Users\dtarb\Scratch\Logan
\loganfel.tif" -p "E:\Users\dtarb\Scratch\Logan\loganp.tif" -
sd8 "E:\Users\dtarb\Scratch\Logan\logansd8.tif" > "E:\Users
\dtarb\Scratch\Logan\cmsgtmp.txt"

D8FlowDir version 5.0.4
Size: 8
Header read time: 0.027037
Data read time: 0.009796
Compute Slope time: 0.102044
Write Slope time: 0.084885
Resolve Flat time: 0.353239
Write Flat time: 0.047082
Total time: 0.624084

Executed (D8FlowDirections) successfully.
End Time: Mon Sep 20 20:11:18 2010 (Elapsed Time: 3.00 seconds)
```



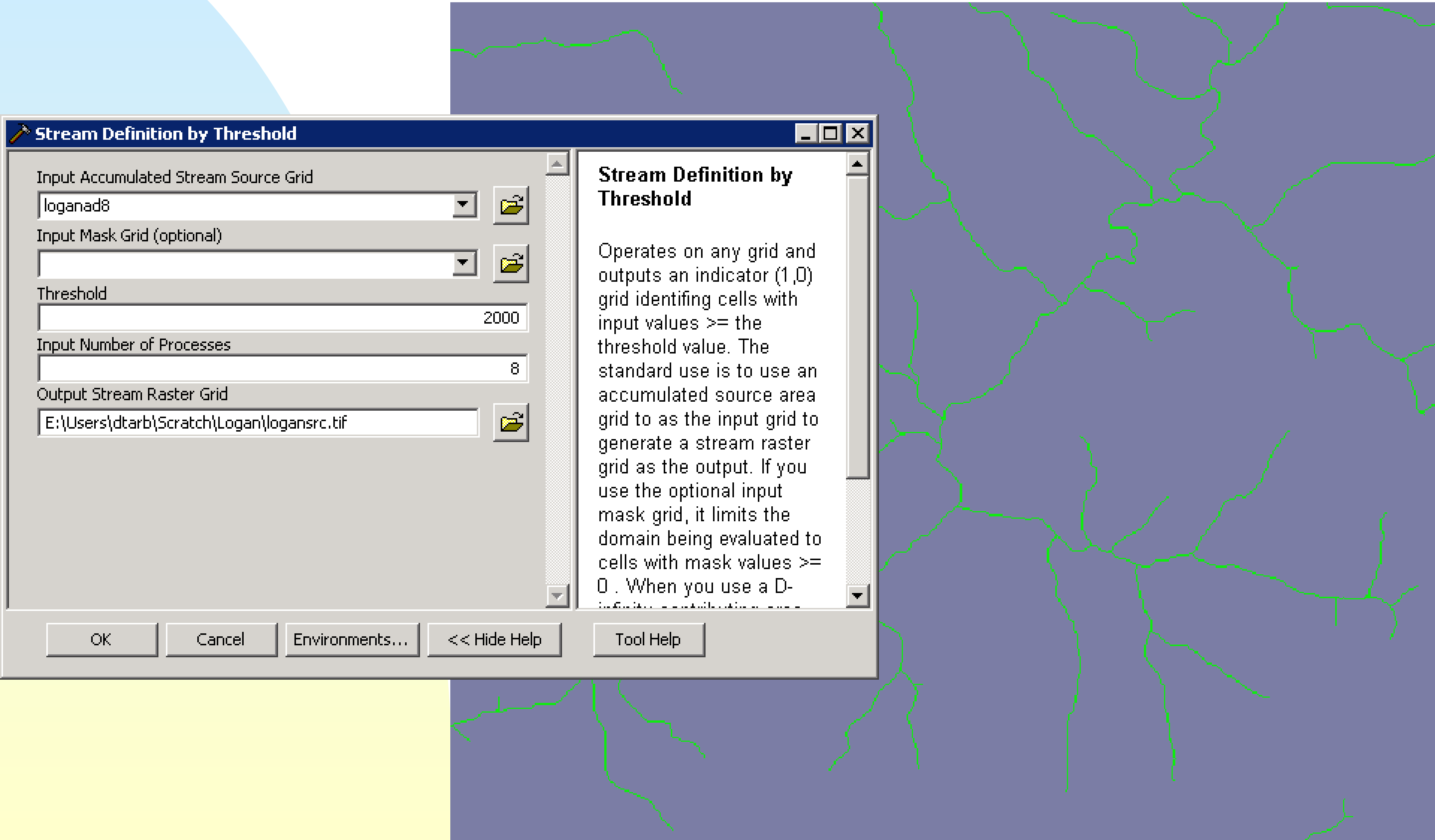




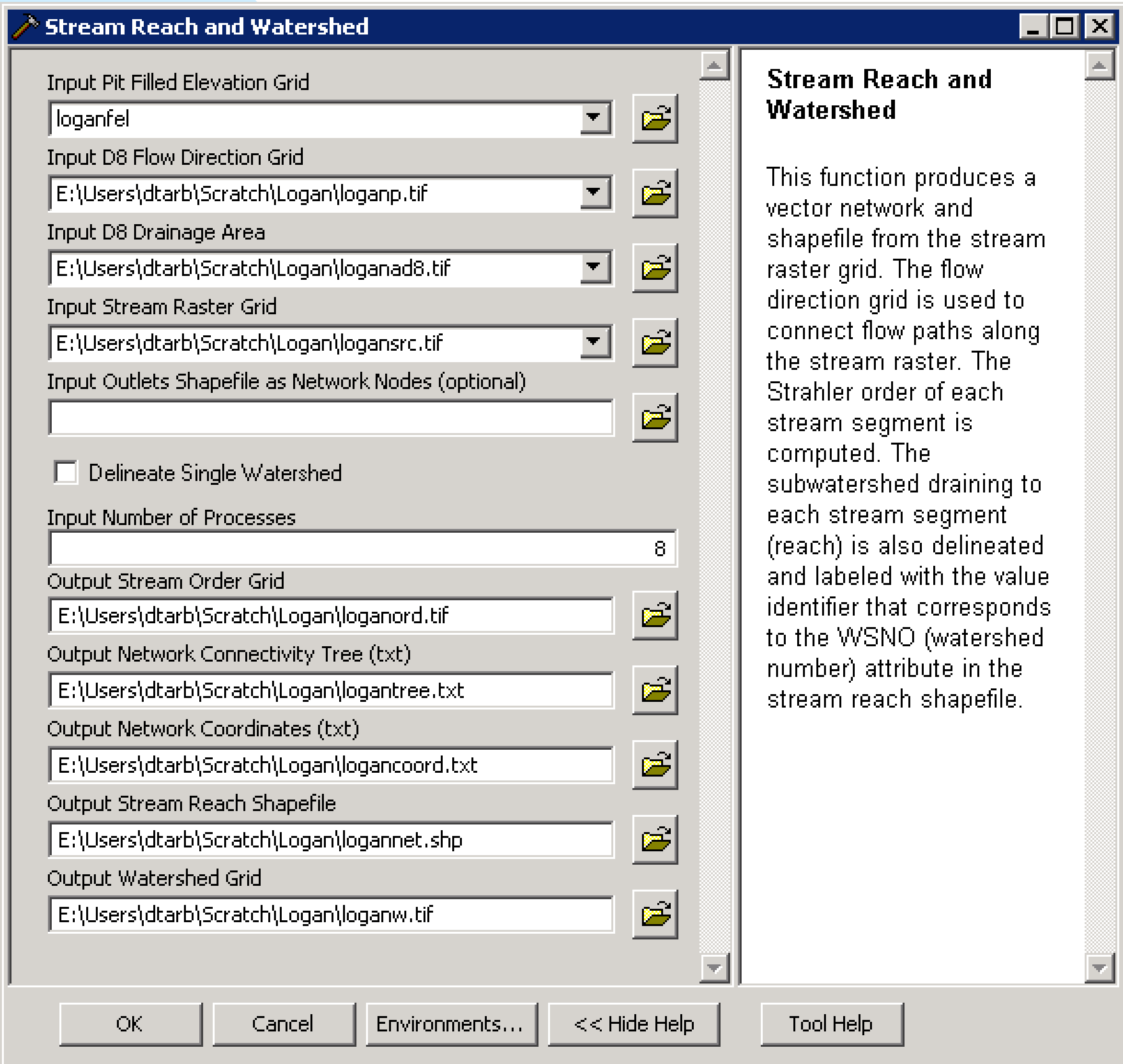
D8 Contributing Area

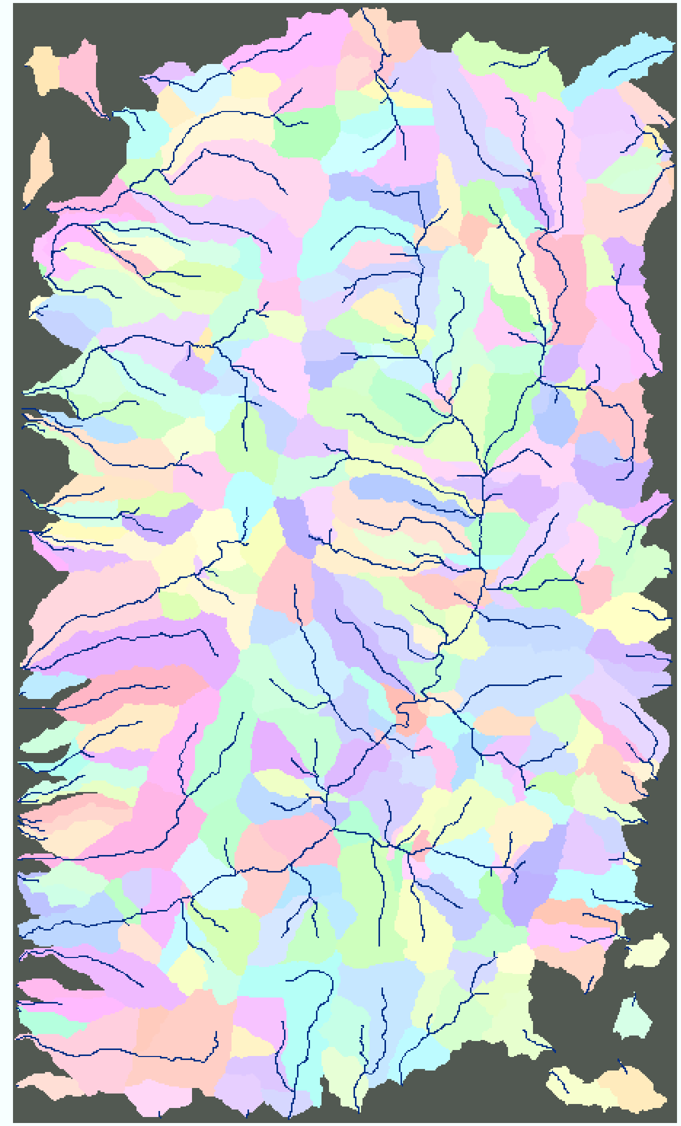


Stream Definition by Threshold



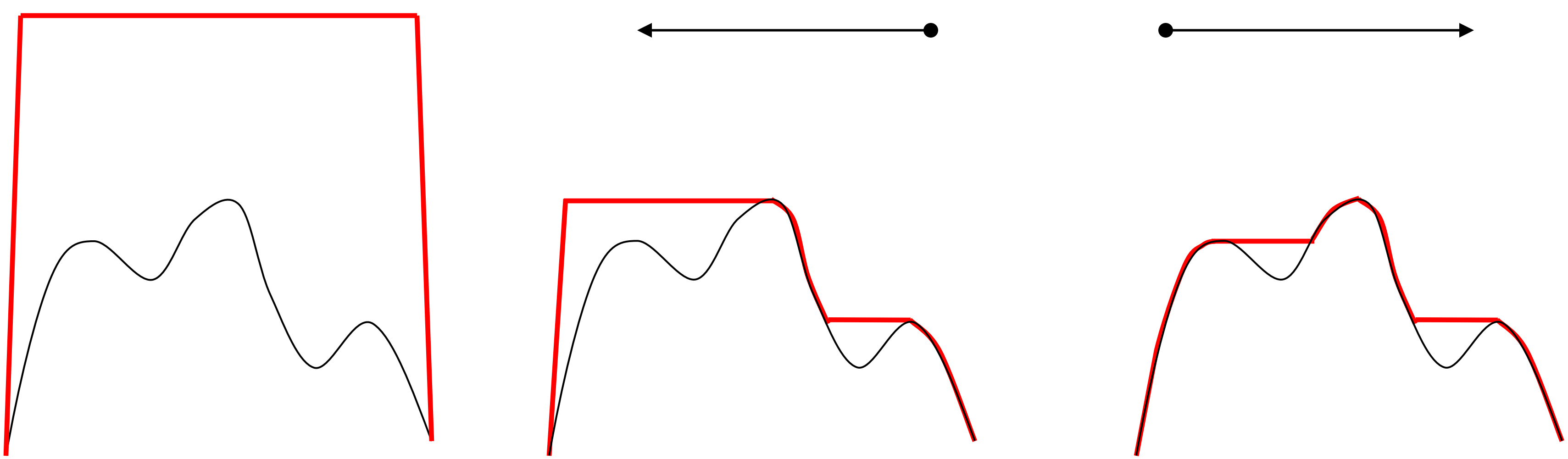
Stream Reach and Watershed





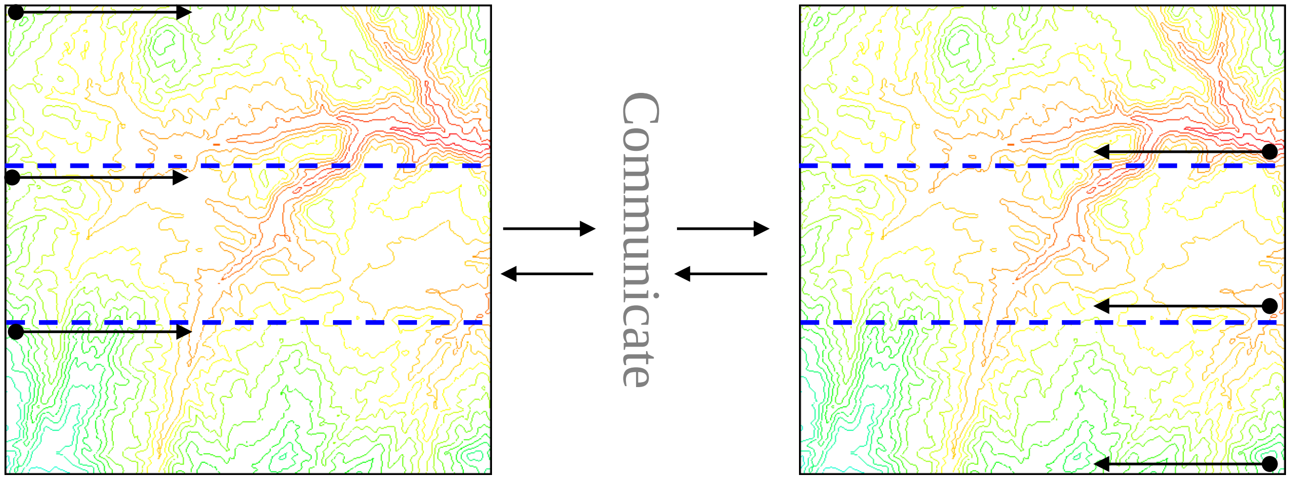
Some Algorithm Details
Pit Removal: Planchon Fill Algorithm

Initialization | 1st Pass | 2nd Pass



Planchon, O., and F. Darboux (2001), A fast, simple and versatile algorithm to fill the depressions of digital elevation models, *Catena*(46), 159-176.

Parallel Scheme




```
Initialize( Z,F)
Do
    for all grid cells i
        if Z(i) > n
            F(i) ← Z(i)
        Else
            F(i) ← n
            i on stack for next pass
    endfor
    Send( topRow, rank-1 )
    Send( bottomRow, rank+1 )
    Recv( rowBelow, rank+1 )
    Recv( rowAbove, rank-1 )
Until F is not modified
```

Z denotes the original elevation.
F denotes the pit filled elevation.
n denotes lowest neighboring elevation
i denotes the cell being evaluated

Iterate only over stack of changeable cells

Parallelization of Contributing Area/Flow Algebra

1. Dependency grid

Executed by every process with grid flow field P, grid dependencies D initialized to 0 and an empty queue Q.

FindDependencies(P,Q,D)

for all i
  for all k neighbors of i
    if $P_{ki}>0$ D(i)=D(i)+1
  if D(i)=0 add i to Q
next

2. Flow algebra function

Executed by every process with D and Q initialized from FindDependencies.

FlowAlgebra(P,Q,D,θ,γ)

while Q isn't empty
  get i from Q
  $\underline{\theta}_i = FA(\underline{\gamma}_i, \underline{P}_{ki}, \underline{\theta}_k, \underline{\gamma}_k)$
  for each downslope neighbor n of i
    if $P_{in}>0$
      D(n)=D(n)-1
      if D(n)=0
        add n to Q
  next n
end while
swap process buffers and repeat

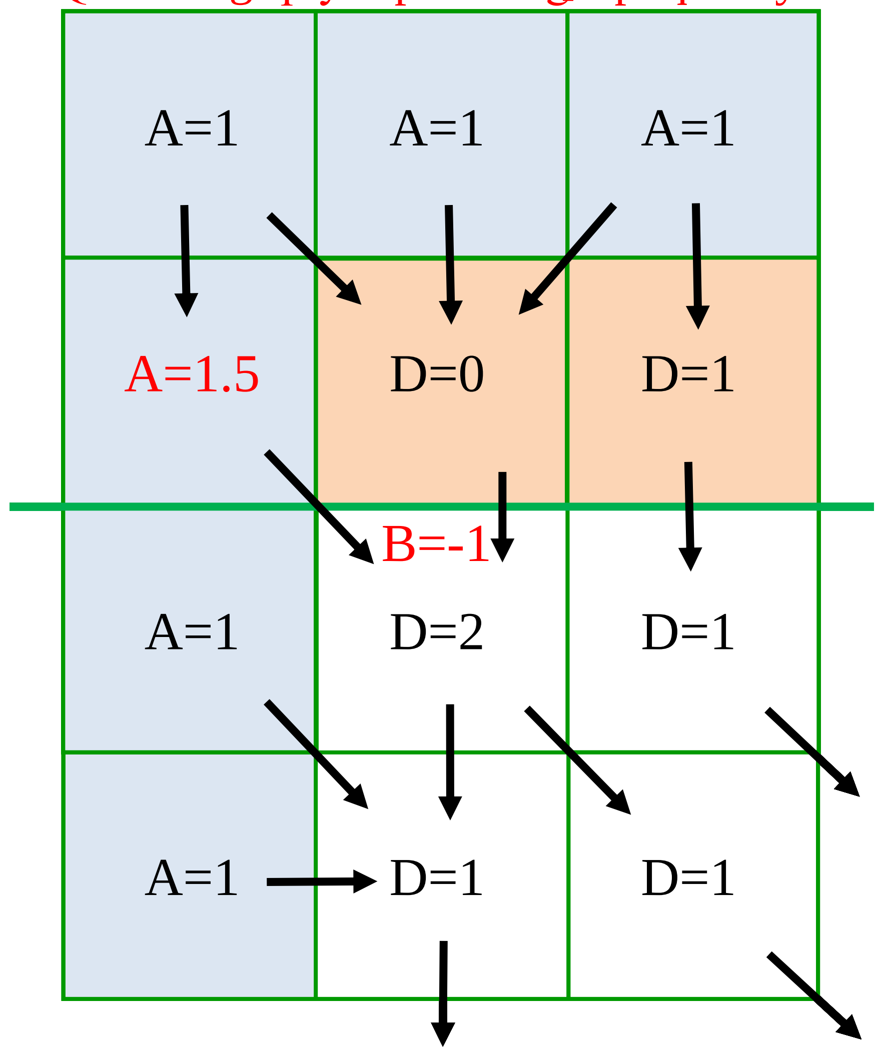
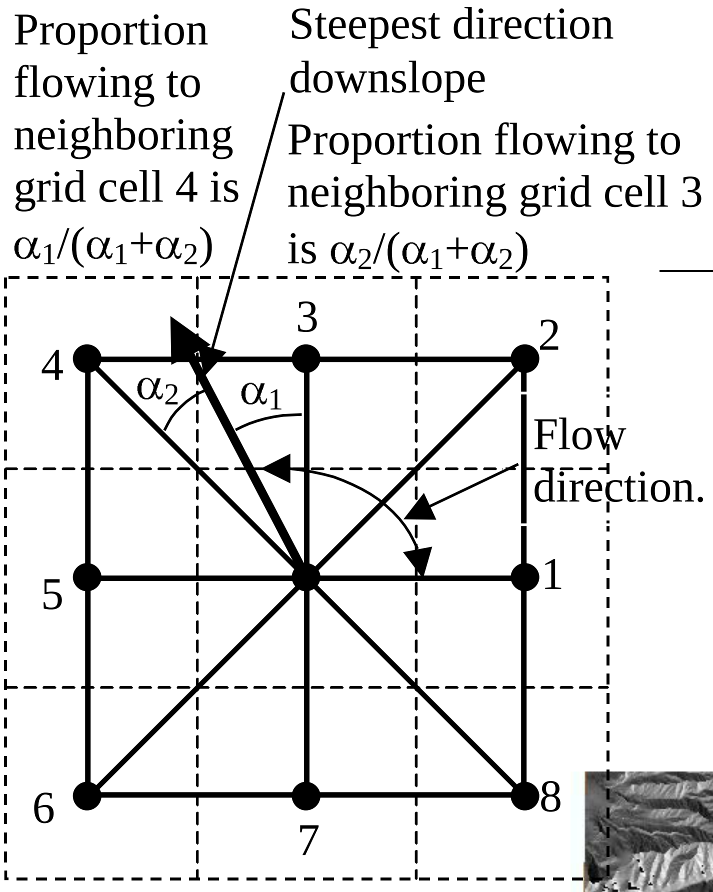
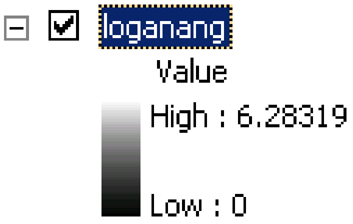
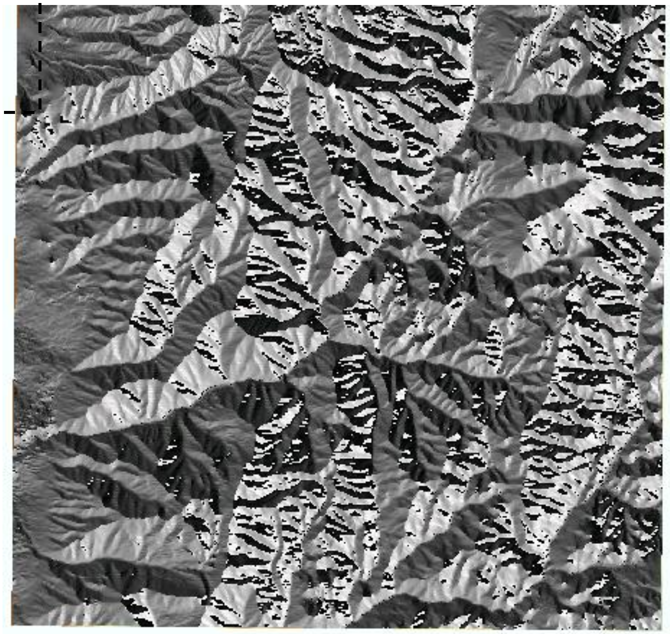


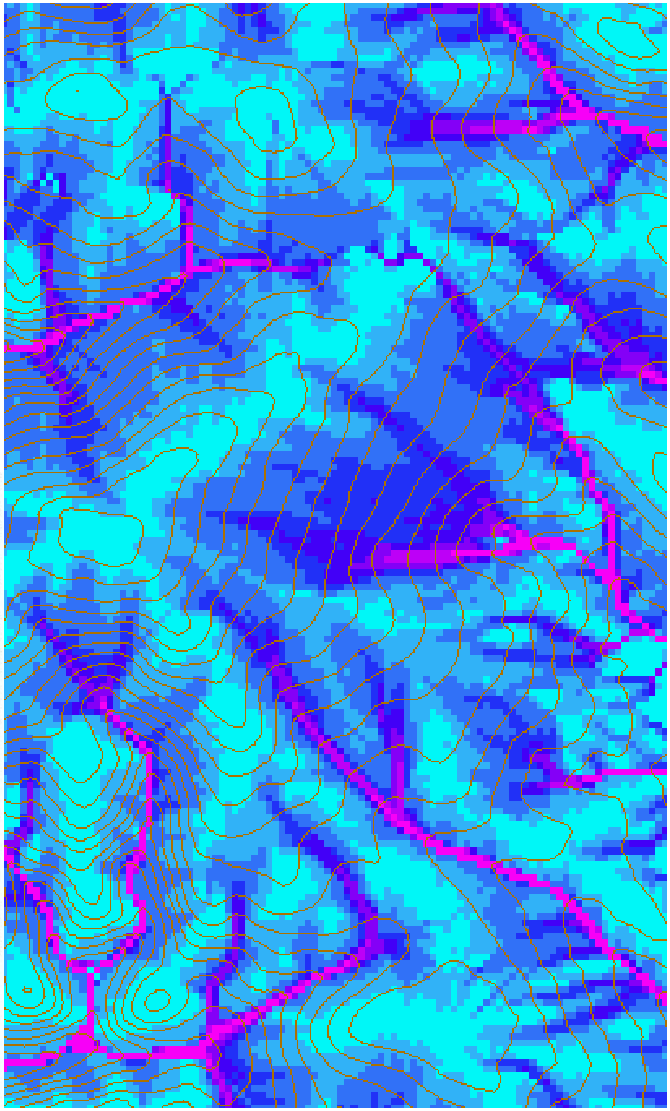
Illustration of some other functions



D-Infinity Slope and Contributing Area

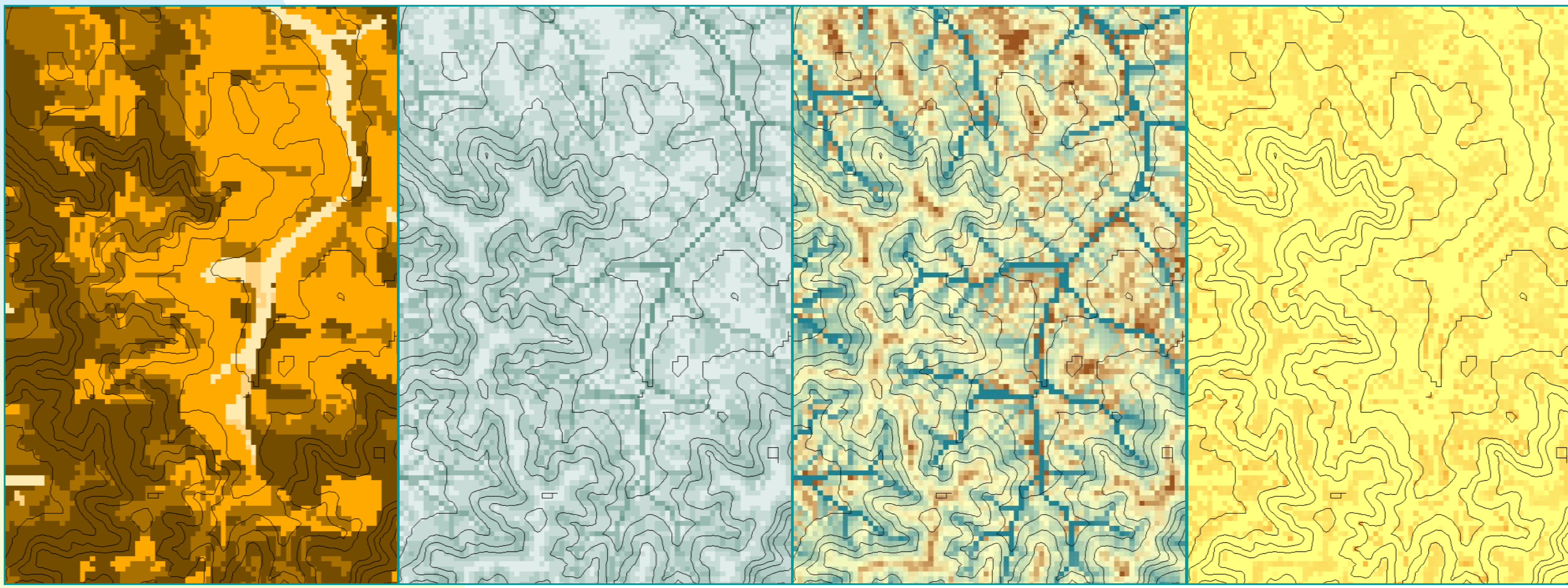






Transport limited accumulation

Supply | Capacity | Transport | Deposition



S

$T_{cap} = \chi a^2 \tan(b)^2$

$T_{out} = \min\{S + \sum T_{in}, T_{cap}\}$

$D = S + \sum T_{in} - T_{out}$

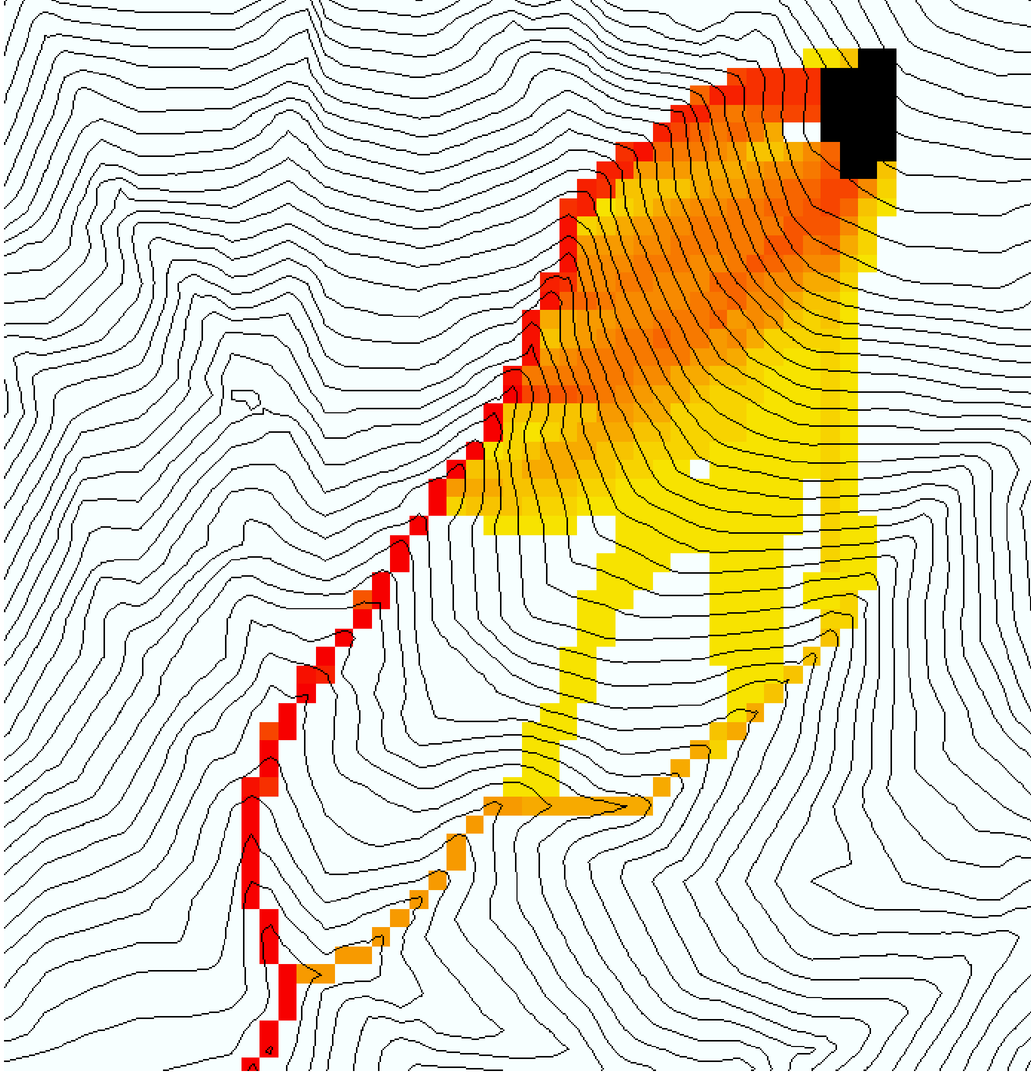
Useful for modeling erosion and sediment delivery, the spatial dependence of sediment delivery ratio and contaminant that adheres to sediment

Decaying Accumulation

A decayed accumulation operator DA[.] takes as input a mass loading field m(x) expressed at each grid location as m(i, j) that is assumed to move with the flow field but is subject to first order decay in moving from cell to cell. The output is the accumulated mass at each location DA(x). The accumulation of m at each grid cell can be numerically evaluated

$$DA[m(x)] = DA(i,j) = m(i,j)\Delta^2 + \sum_{k \text{ contributing neighbors}} p_k d(i_k, j_k) DA(i_k, j_k)$$

Here d(x) = d(i ,j) is a decay multiplier giving the fractional (first order) reduction in mass in moving from grid cell x to the next downslope cell. If travel (or residence) times t(x) associated with flow between cells are available d(x) may be evaluated as $\exp(-\lambda t(x))$ where λ is a first order decay parameter.



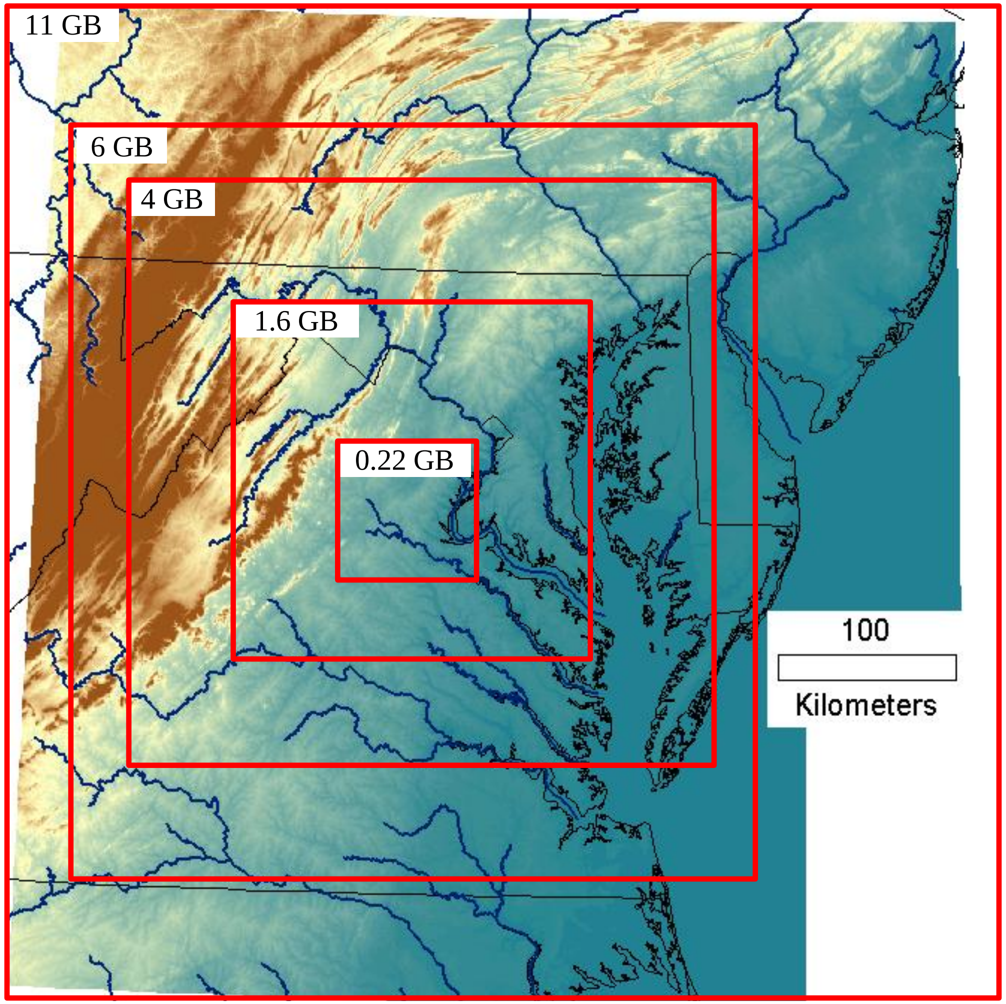
Useful for a tracking contaminant or compound subject to decay or attenuation

Capabilities Summary

Capability to run larger problems

		Processors used	Grid size	
			Theoretcal limit	Largest run
2008	TauDEM 4	1	0.22 GB	0.22 GB
Sept 2009	Partial implement-ation	8	4 GB	1.6 GB
June 2010	TauDEM 5	8	4 GB	4 GB
Sept 2010	Multifile on 48 GB RAM PC	4	Hardware limits	6 GB
Sept 2010	Multifile on cluster with 128 GB RAM	128	Hardware limits	11 GB

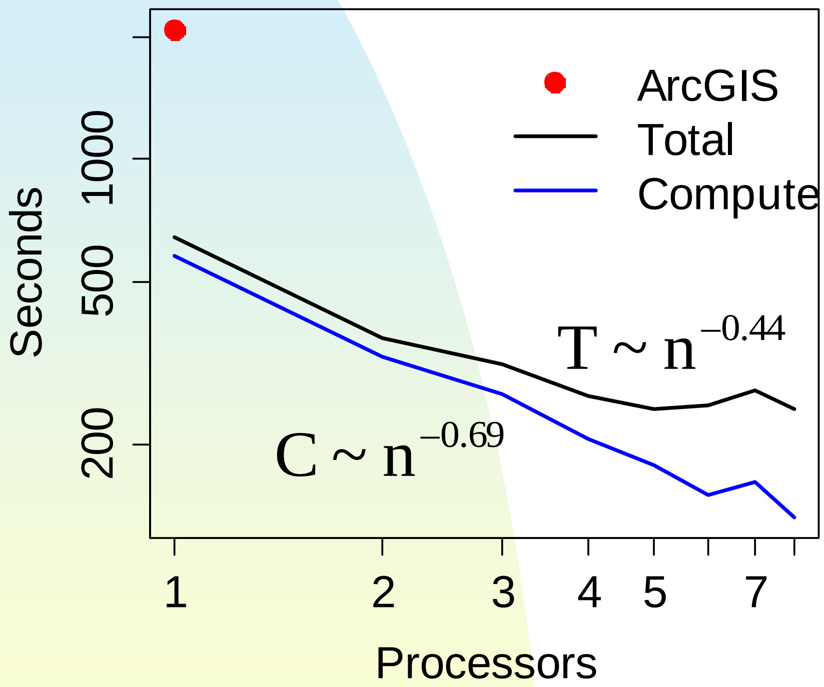
Single file size limit 4GB



At 10 m grid cell size

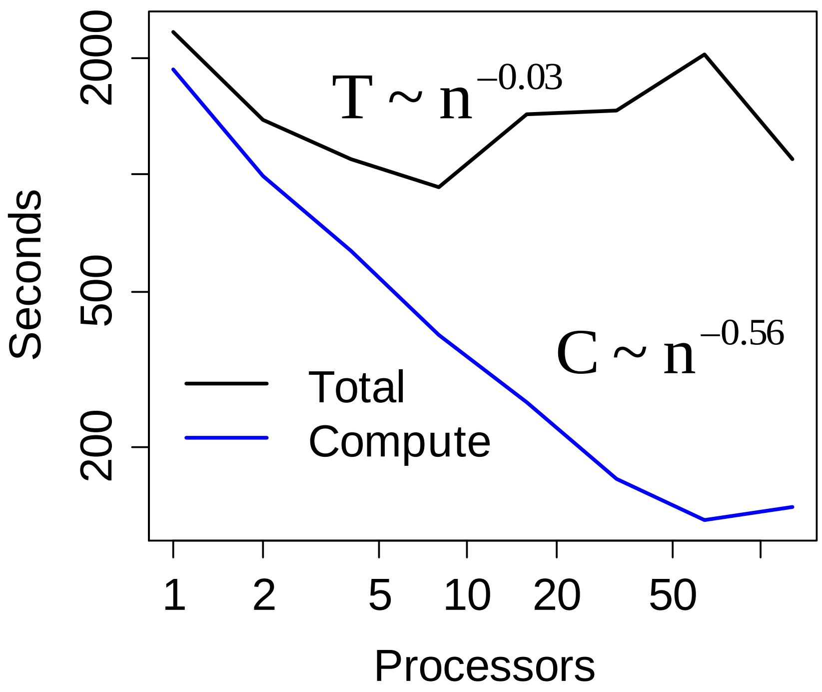
Improved runtime efficiency

Parallel Pit Remove timing for NEDB test dataset (14849 x 27174 cells ≈ 1.6 GB).



8 processor PC

Dual quad-core Xeon E5405 2.0GHz PC with 16GB RAM

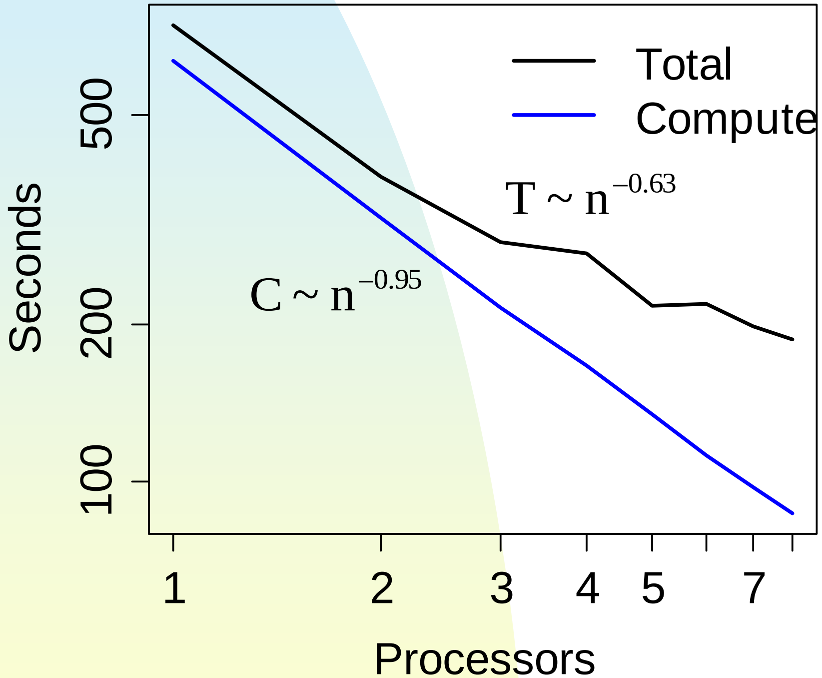


128 processor cluster

16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM

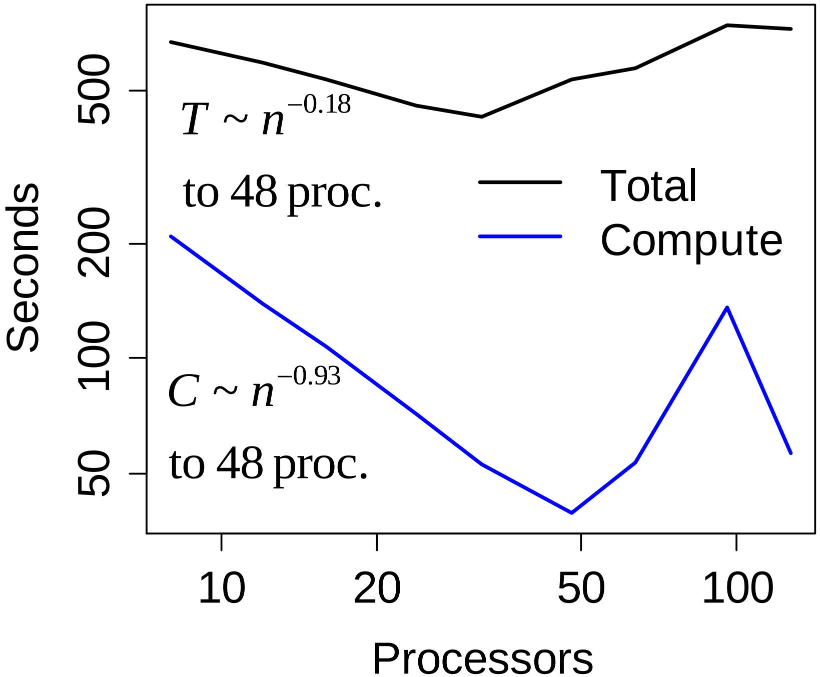
Improved runtime efficiency

Parallel D-Infinity Contributing Area Timing for Boise River dataset (24856 x 24000 cells ~ 2.4 GB)



8 processor PC

Dual quad-core Xeon E5405 2.0GHz PC with 16GB RAM



128 processor cluster

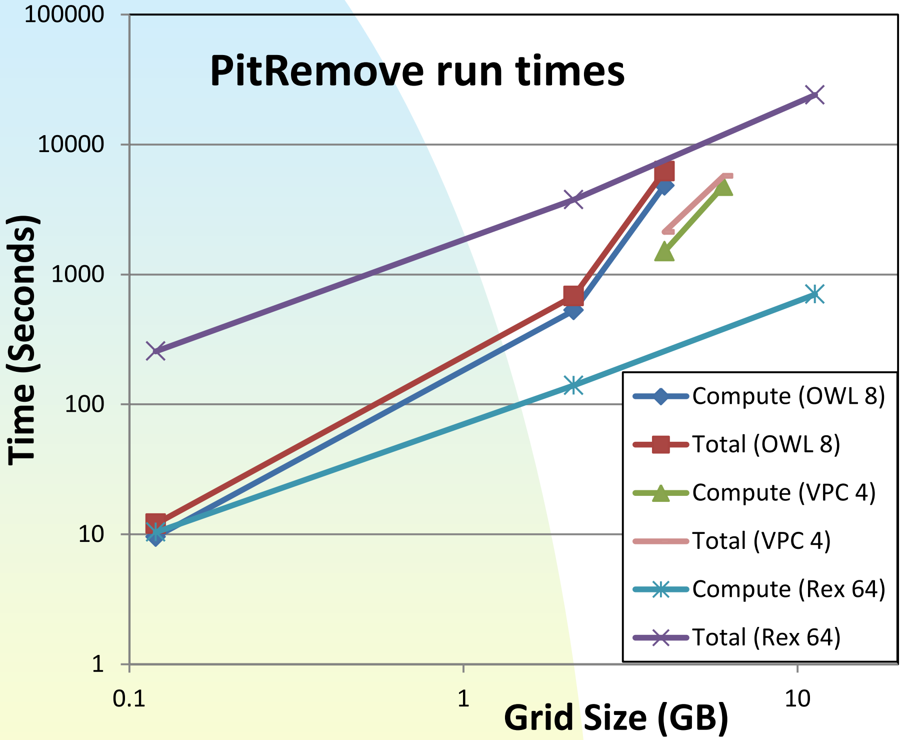
16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM

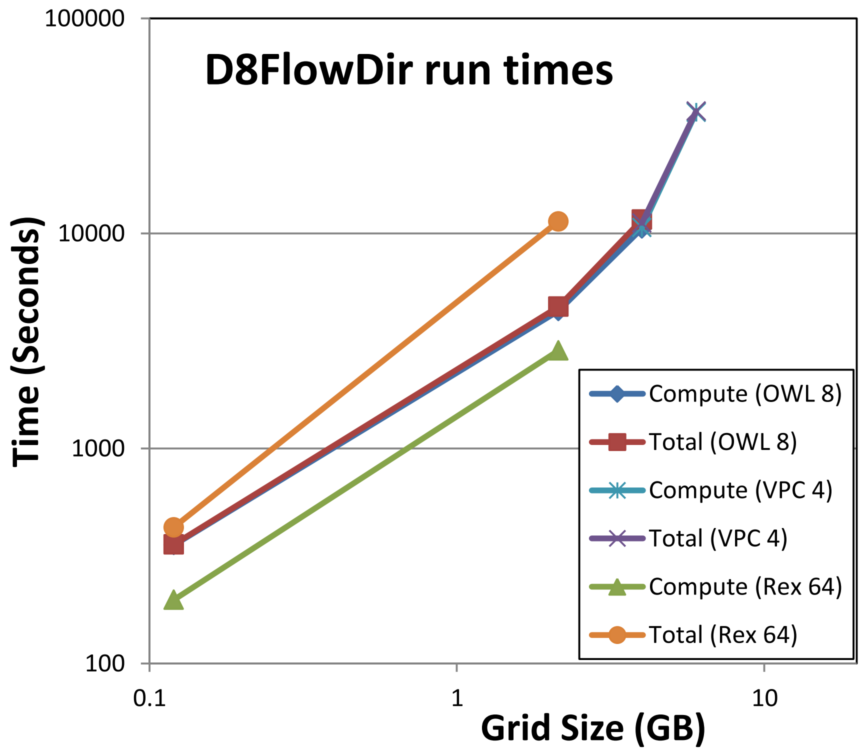
Scaling of run times to large grids

Dataset	Size	Hardware	Number of Processors	PitRemove (run time seconds)		D8FlowDir (run time seconds)	
	(GB)			Compute	Total	Compute	Total
GSL100	0.12	Owl (PC)	8	10	12	356	358
GSL100	0.12	Rex (Cluster)	8	28	360	1075	1323
GSL100	0.12	Rex (Cluster)	64	10	256	198	430
GSL100	0.12	Mac	8	20	20	803	806
YellowStone	2.14	Owl (PC)	8	529	681	4363	4571
YellowStone	2.14	Rex (Cluster)	64	140	3759	2855	11385
Boise River	4	Owl (PC)	8	4818	6225	10558	11599
Boise River	4	Virtual (PC)	4	1502	2120	10658	11191
Bear/Jordan/Weber	6	Virtual (PC)	4	4780	5695	36569	37098
Chesapeake	11.3	Rex (Cluster)	64	702	24045		

1. Owl is an 8 core PC (Dual quad-core Xeon E5405 2.0GHz) with 16GB RAM
2. Rex is a 128 core cluster of 16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM
3. Virtual is a virtual PC resourced with 48 GB RAM and 4 Intel Xeon E5450 3 GHz processors
4. Mac is an 8 core (Dual quad-core Intel Xeon E5620 2.26 GHz) with 16GB RAM

Scaling of run times to large grids





1. Owl is an 8 core PC (Dual quad-core Xeon E5405 2.0GHz) with 16GB RAM
2. Rex is a 128 core cluster of 16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM
3. Virtual is a virtual PC resourced with 48 GB RAM and 4 Intel Xeon E5450 3 GHz processors

Summary and Conclusions

- Parallelization speeds up processing and partitioned processing reduces size limitations
- Parallel logic developed for general recursive flow accumulation methodology (flow algebra)
- Documented ArcGIS Toolbox Graphical User Interface
- 32 and 64 bit versions (but 32 bit version limited by inherent 32 bit operating system memory limitations)
- PC, Mac and Linux/Unix capability
- Capability to process large grids efficiently increased from 0.22 GB upper limit pre-project to where < 4GB grids can be processed in the ArcGIS Toolbox version on a PC within a day and up to 11 GB has been processed on a distributed cluster (a 50 fold size increase)

Limitations and Dependencies

- Uses MPICH2 library from Argonne National Laboratory
  http://www.mcs.anl.gov/research/projects/mpich2/
- TIFF (GeoTIFF) 4 GB file size (for single file version)
- Capability to use multiple files to cover domain not yet on web site
- Processor memory
- Toolbox user interface only for ArcGIS 9.3.1

Are there any questions ?