

DATA-DRIVEN COLLABORATION ENVIRONMENTS

INTEGRATING HUBZERO AND IRODS

Rajesh Kalyanam

Robert Campbell

Lan Zhao

Carol Song

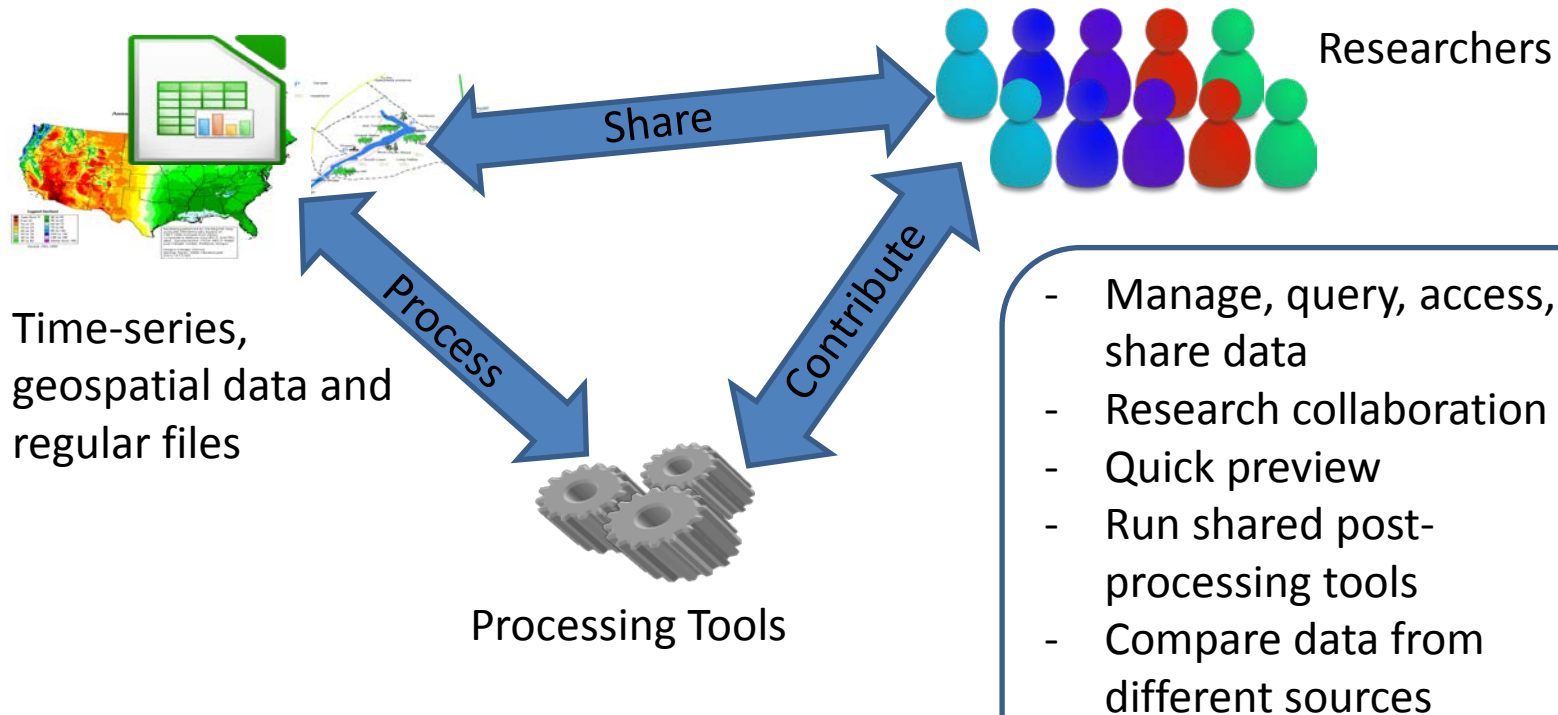
RCAC, Purdue University



G · A · B · B · S
geospatial data analysis building blocks

PURDUE
UNIVERSITY

HISTORY – GEOSHARE, DRINET HUBS



IDATA

- Hub component for scientific data management
- Supports familiar nested-folder file organization
- File sharing based on hub groups
- Support for metadata entry and metadata-based file query
- Preview, simple processing support for tabular, geospatial data

IDATA

iDATA Publish, Browse & Discover

Current Path: myprotection

Type	Name	Size	Date modified
Raster Image	spatial.nc	9.0 KB	9/7/2016 10:05 AM
File	hazard.tif	9.0 KB	9/7/2016 10:06 AM

mygeohub - mygeohub

https://mygeohub.org/odata/?controller=collection&action=browse&id=130&path=aster%20images&shared=0

Map Properties: Google Roadmap, Google Satellite, Google Terrain, Google Hybrid

Layers: yield_geospatial_example_1016_31, Selected Features

Click on map to get feature info for location.

iDATA Publish, Browse & Discover

Current Path: mytab1

date/time	level	temp
2010-10-01 00:0	0.324	15.6
2010-10-01 00:3	0.326	15.76
2010-10-01 01:0	0.325	15.09
2010-10-01 01:3	0.326	15.64
2010-10-01 02:0	0.325	15.06
2010-10-01 02:3	0.325	15.6
2010-10-01 03:0	0.325	15.49
2010-10-01 03:3	0.324	15.33
2010-10-01 04:0	0.323	16.23
2010-10-01 04:3	0.322	15.14
2010-10-01 05:0	0.321	15.04
2010-10-01 05:3	0.32	14.95
2010-10-01 06:0	0.32	14.88
2010-10-01 06:3	0.319	14.75
2010-10-01 07:0	0.317	14.64
2010-10-01 07:3	0.317	14.53
2010-10-01 08:0	0.315	14.4
2010-10-01 08:3	0.315	14.27
2010-10-01 09:0	0.319	14.15
2010-10-01 09:3	0.312	14.05

View: 20

File View Raster Map View Resample Raster **Reclassify Raster** Subtract Rasters

2852, -29.89261

Layers

nodata 0 thru 0.4 = 0.4

Save raster map

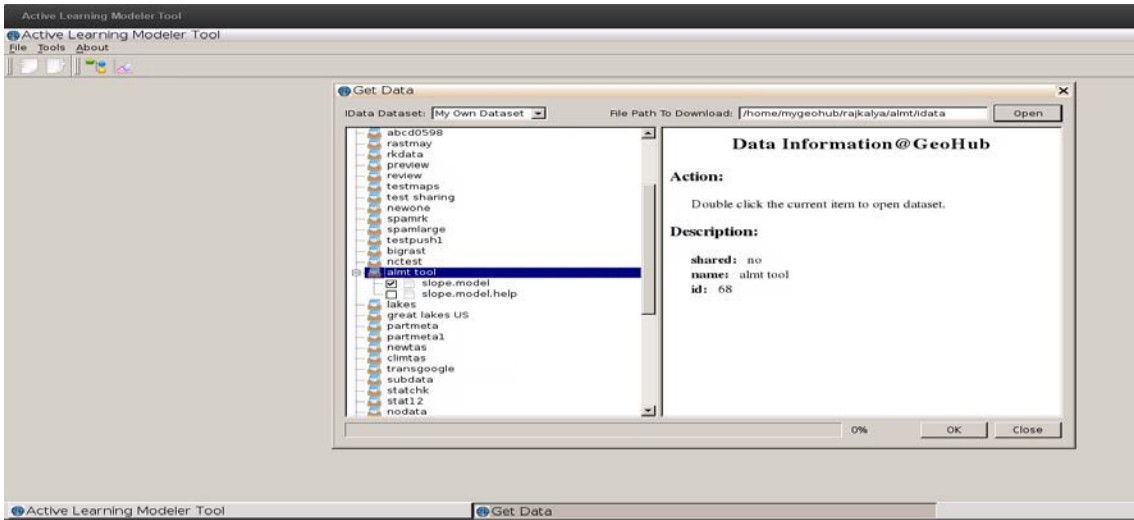
DATA FILE MANAGEMENT - IRODS

- Abstracts physical storage in iRODS
 - Distributed data management system
 - Developed at RENCi, DICE - UNC
- Supports easy storage expandability
- File metadata stored as attribute-value pairs
- iCAT enables file query based on metadata
- PHP client APIs enable easy integration with Hub component code

DATA IN GABBS

- Geospatial Data Analysis Building Blocks
 - Reusable blocks for geo-augmented tools, geospatial file management, visualization
- Enable interoperability between hub data management, tools, publications
- Enable tool input, output, discovery, provenance
 - display list of applicable tools for a file
 - support tool result reproducibility
- Support richer keyword (include geospatial) search

IDATA - HUB TOOL INTEGRATION



- Hub API controller exposes service interface to data management operations
- Tools invoke service methods for data listing, file upload, download and metadata
- Tool invocation URL arguments specify file input when launched from iData



iData - a web-based data management tool. Upload, share, process, and archive data collections including spreadsheet, raster and vector datasets.

METADATA CAPTURE – IRODS MICROSERVICES

```
#acPostProcForPut { ON($objPath like ".*txt") {writeLine("serverLog","File $objP
#acPostProcForPut { }
acPostProcForPut { on($objPath like ".*tif") { msiExtractGeoMetaTif; } }
acPostProcForPut { on($objPath like ".*nc") { msiExtractGeoMetaNc; } }
acPostProcForPut { on($objPath like ".*shx") { msiExtractGeoMetaShp; } }
acPostProcForPut { on($objPath like ".*shp") { msiExtractGeoMetaShp; } }
acPostProcForPut { on($objPath like ".*prj") { msiExtractGeoMetaShp; } }
acPostProcForPut { on($objPath like ".*dbf") { msiExtractGeoMetaShp; } }
acPostProcForCopy { }
acPostProcForFilePathReg { }
acPostProcForCreate { }
# acPostProcForOpen {writeLine("serverLog",$objPath); }
acPostProcForOpen { }
acPostProcForPhymv { }
acPostProcForRepl { }
# 9) acSetNumThreads - Rule to set the number of threads for a data transfer
# This rule supports condition based on $rescName so that different
```

Event-based file processing

Microservice C-methods

```
/* Geo module microservices */
/* Service name          # args      Service function */
{ "msiExtractGeoMetaTif",      0,      (funcPtr) msiExtractGeoMetaTif },
{ "msiExtractGeoMetaNc",      0,      (funcPtr) msiExtractGeoMetaNc },
{ "msiExtractGeoMetaShp",     0,      (funcPtr) msiExtractGeoMetaShp }
```

```
OGRSpatialReferenceH hSpatialRef, tSpatialRef;
OGREnvelope hExtent;

hDS = OGROpen( path, FALSE, NULL );

if( hDS == NULL )
{
    exit( 1 );
}

hLayer = OGR_DS_GetLayer( hDS, 0 );

OGR_L_ResetReading(hLayer);

hSpatialRef = OGR_L_GetSpatialRef(hLayer);

snprintf( metaname, sizeof metaname, "projection" );
snprintf( metavalue, sizeof metavalue, "%s", (char *)OSRGetAttrValue( hSpatialRef, "PROJECTION", 0 ));
addKeyVal( &kvdriver, metaname, metavalue );
```

Shapefile processing with GDAL C APIs

INDEXED FILE SEARCH – APACHE SOLR

mygeohub - mygeohub
https://mygeohub.org/idata/?controller=collection&task=metadata&gid=130&path=StJosephWatershed/shapefiles&shared=0&obj=StJosephWatershed/shap

Title	boundary.shp
Subject	OBJECTID,GRIDCODE,Subbasin,Area,Sio1,Len1,Sil,Csl,Wid1,Dep1,Lat,Long,Elev,ElevMin,ElevMax,B.name,Shape_Leng,Shape_Area,Hydro
Contributor	Rajesh Kalyanam
Publisher	iData@MyGeoHub
Date	7/18/2015 3:20 AM
Identifier	StJosephWatershed/shapefiles/boundary.shp
Format	shp
Type	geospatial
Source	
Creator	

- DCMI metadata fields (+ geographic coverage)
- Captured by iRODS microservices
- Indexed into Solr

mygeohub - mygeohub
https://mygeohub.org/idata/?controller=collections&task=presearch&shared=0

Search Reset

Enter text for metadata search:

Or, highlight a search area on the map: Latitude: 29.0 to 49.0, Longitude: -103.0 to -72.0

65.742, 37.300

2000 km
1000 mi

bing

Copyright © 2013 Intelmap Enterprise Geographics, Inc. © 2013 Microsoft Corporation. © 2013 NAVTEQ. Terms of Use

6:29 AM
7/19/2015

- All DCMI text fields added to keyword search index
- Geographic search uses Solr coverage field

GOING FORWARD

TIGHTER IRODS – HUB INTEGRATION

- **Storage option for hub Project Files**
 - iDrop desktop client integration for large files
- **Explore iRODS Fuse mounts**
 - Can simplify file access, tool development
 - Will enable “local” file input/output in tool session (no service calls)
- **Faceted search for hub files**
 - iRODS microservice-based metadata capture
 - Import into Solr index and hooks for Solr faceted search
- **Post-process files for compiled preview**

GOING FORWARD

IRODS FEDERATION

- Federation enables file sharing across remote iRODS zones accessible as local collections
- Distinct hubs can support file sharing, collaboration across projects by federating iRODS storage
- Can enable remote file access in local hub tools
 - Potentially support tool workflows across hubs

GOING FORWARD

ENABLING RICHER TOOL WORKFLOWS

- Richer tool input filtering against iRODS-extracted metadata
 - For e.g. restricted to geographic region, require specific spreadsheet structure, images with 3 bands
- Tool output post-processing enabling subsequent tool input
 - For e.g. re-project geospatial data to standard projections
- Offline remote data download
- Implement data access protocols (OpenDAP) on top of iRODS for data subsetting

ACKNOWLEDGEMENTS



This work was supported by the NSF Award ACI - 1261727
CIF21 DIBBs : Integrating Geospatial Capabilities into
HUBzero