

FACE-IT: Earth science workflows made easy with Globus and Galaxy technologies

(Provide more capability for more people at lower cost by delivering “Science as a Service”)

Raffaele Montella^{1,2}, Alison Brizius², Joshua Elliott², David Kelly², Ravi Madduri^{2,3}, Ketan Maheshwari³, Cheryl Porter⁴, Peter Vilter², Michael Wilde², Wei Xiong⁴, Meng Zhang⁴ and Ian Foster^{2,3,5}

¹Department of Science and Technologies, University of Naples Parthenope, Naples, ITALY;

²Computation Institute, Argonne National Laboratory and University of Chicago, Chicago, Illinois, USA;

³Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, USA;

⁴University of Florida, Department of Agricultural and Biological Engineering, Gainesville, Florida, USA;

⁵Department of Computer Science, University of Chicago, Chicago, Illinois, USA;



...less magic wands,
more magic minds...

Facing real problems with Information Technology

What's in a name...

Science Gateways

No buzzword
Real things!
An open playground for
the next generation of
earth system scientists

The user profile...

Scientists
Experts of their fields
Limited programming skills
Complex experiments

Data + Workflows = Results

faceit-portal.org

Information
Technology

Effective and efficient
solutions to real problems
Experts in design and
abstraction

Development-experts (in wizardry)...

Built on
widely used Galaxy,
Globus, and
Swift systems

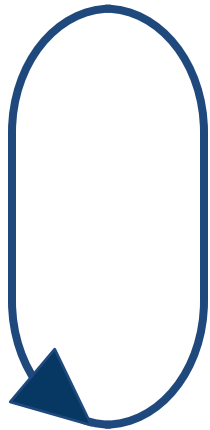


FACE-IT: A Framework to Advance Climate, Economic, and Impact Investigations with Information Technology

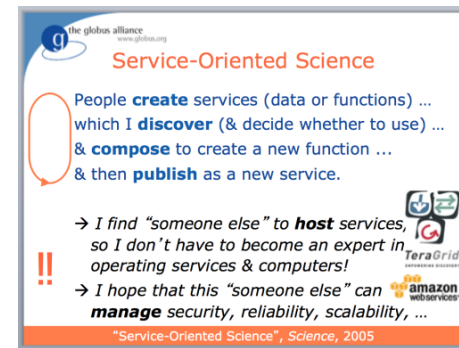


BACK 2005

Service Oriented Science



- People create services (data or functions) ...
- which I discover (& decide whether to use) ...
- & compose to create a new function ...
- & then publish as a new service.



→ I find “someone else” to **host** services, so I don’t have to become an expert in operating services & computers!

→ I hope that this “someone else” can **manage** security, reliability, scalability, ...



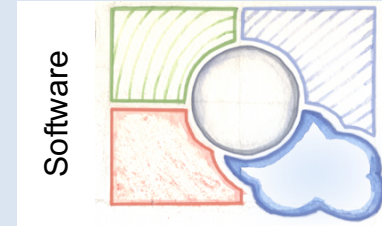


Making complex things straightforward

The Science Stack

- **Galaxy**

- Interactive execution
- Creation, Execution, Sharing, Discovering Workflows



SaaS

- **Globus**

- Data management
- Identity Management



PaaS

- **AWS**

- HTCondor, Chef, EC2, EBS, S3, SNS, NEWT
- Spot, Route 53, Cloud Formation



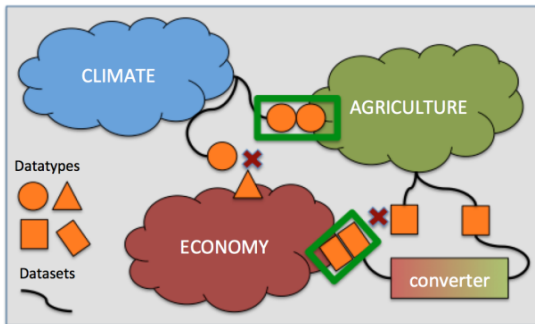
IaaS

Teaching Galaxy to speak Earth Science

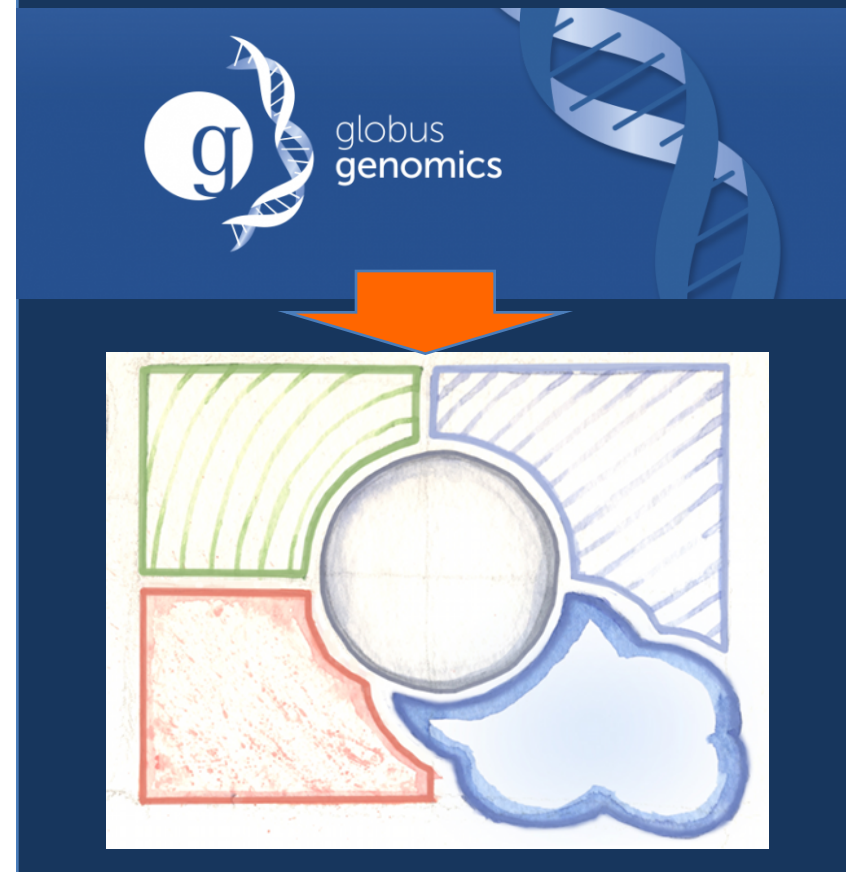
From genes to Earth in 6 steps

- Datatypes
- Tools
- Tool parameters

- Aggregated datatypes
- Data providers
- Visualizers



Data	NetCDF	GCM
file_ext mime-type ... Metadata ...	file_ext="nc" ... Metadata: NCML Metadata: WMS ...	file_ext="gcm.nc" ... schema= "GCM.ncxsd" ...
set_meta() sniff() display_peek() display_data() ...	set_meta() Sniff() display_peek() display_data() ...	





Teaching Galaxy to speak Earth Science

Step ONE of 6: earth system datatypes

- Color scheme

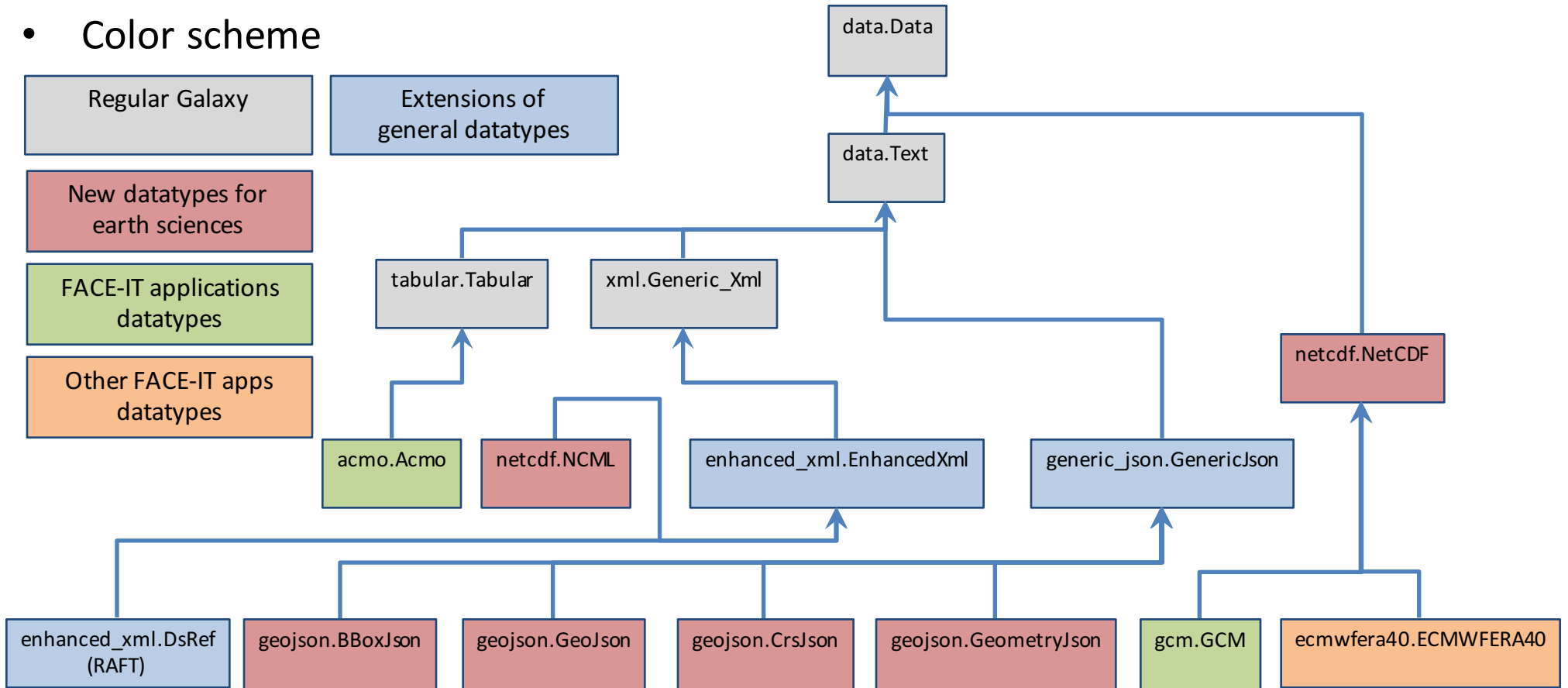
Regular Galaxy

Extensions of general datatypes

New datatypes for earth sciences

FACE-IT applications datatypes

Other FACE-IT apps datatypes





Teaching Galaxy to speak Earth Science (“Globusally”)

Step TWO of 6: new tools

FACE-IT | Galaxy Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data

- Upload File** Upload files from your computer, or transfer data from a URL
- Browse and Get Data via Globus** Interactively browse and transfer data from Globus Online into your FACE-IT history
- Download PSIMS Soil Data** Download PSIMS soil data by selecting latitude and longitude
- Download PSIMS Climate Data** Download PSIMS climate data by selecting latitude and longitude

Browse and Get Data via Globus (version 1.0.0)

Source Endpoint:

Search endpoint

fac

- fgf#genomicsfacilit
- galaxy#faceit
- globuspublish#artif
- jelliott#face-it
- jimdemo#ShareForl
- ratnayakesd#Projec
- talsaniaks#bustin_1
- u_w6egsyxa2ei6lb5
- fc93-11e5-a701-

any Globus Online endpoint to your Galaxy installation and creates a galaxy dataset from

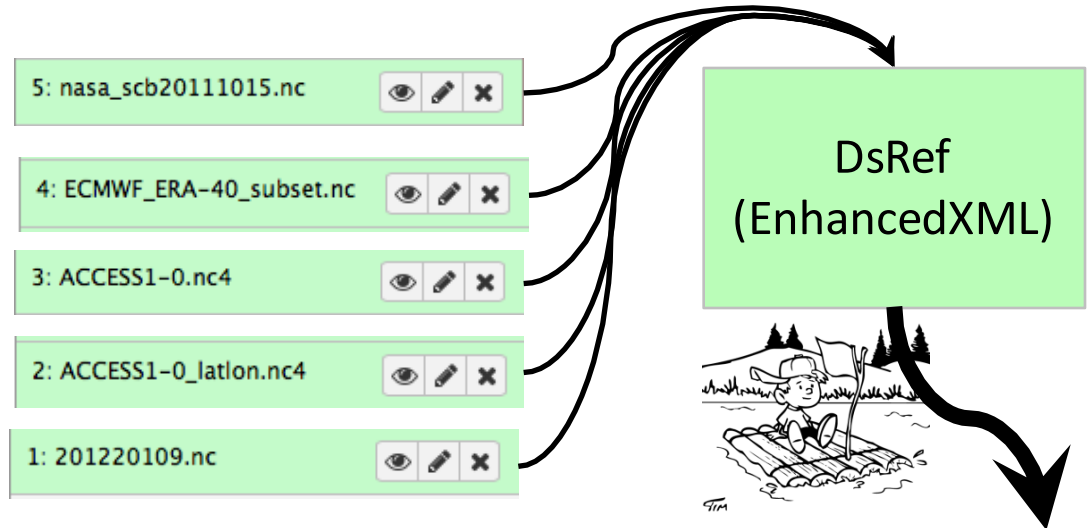
- **Tool parameters:**
Define the user interface elements for a tool
- Regular tool parameters wrap text fields, radio buttons and drop down lists.
- Custom tool parameters for Globus, OpenDap, date peaking and feature selection on maps.

```
<inputs>
  <param format="nc" name="format" value="nc" />
  <conditional name="mode" />
  <param name="mode" value="mode" />
</inputs>
```

The image shows two software interfaces. On the left is the OpenDap Browser (version 1.0.0) with a URL field containing 'http://meteo.uniparthenope.it/opens...'. Below the URL are buttons for 'Clean', 'All', and 'Prism', and an 'Ins' button with a regex pattern. On the right is the GeoJSON Generator (version 1.0.0) displaying a map of the Great Lakes region. A red bounding box is drawn around the Chicago area. Below the map, a JSON output is shown, representing a FeatureCollection with a single Feature. The JSON is: { "type": "FeatureCollection", "features": [{ "type": "Feature", "properties": { "_id": 41, "name": "Feature", "desc": "This is a feature" } }] }. At the bottom of the GeoJSON Generator window, the text 'Avg minimum air temperature (TMINA)' is visible.

- **Dataset References:**
XML based datatype grouping references to different datasets in the same history.
- The regular Galaxy works on single file datasets or composite file datasets.
- Acts as a 'struct' or an 'array' or a mix of both.
- Supports schemas and translators.

Globus HTTPS Endpoints
(...future...)



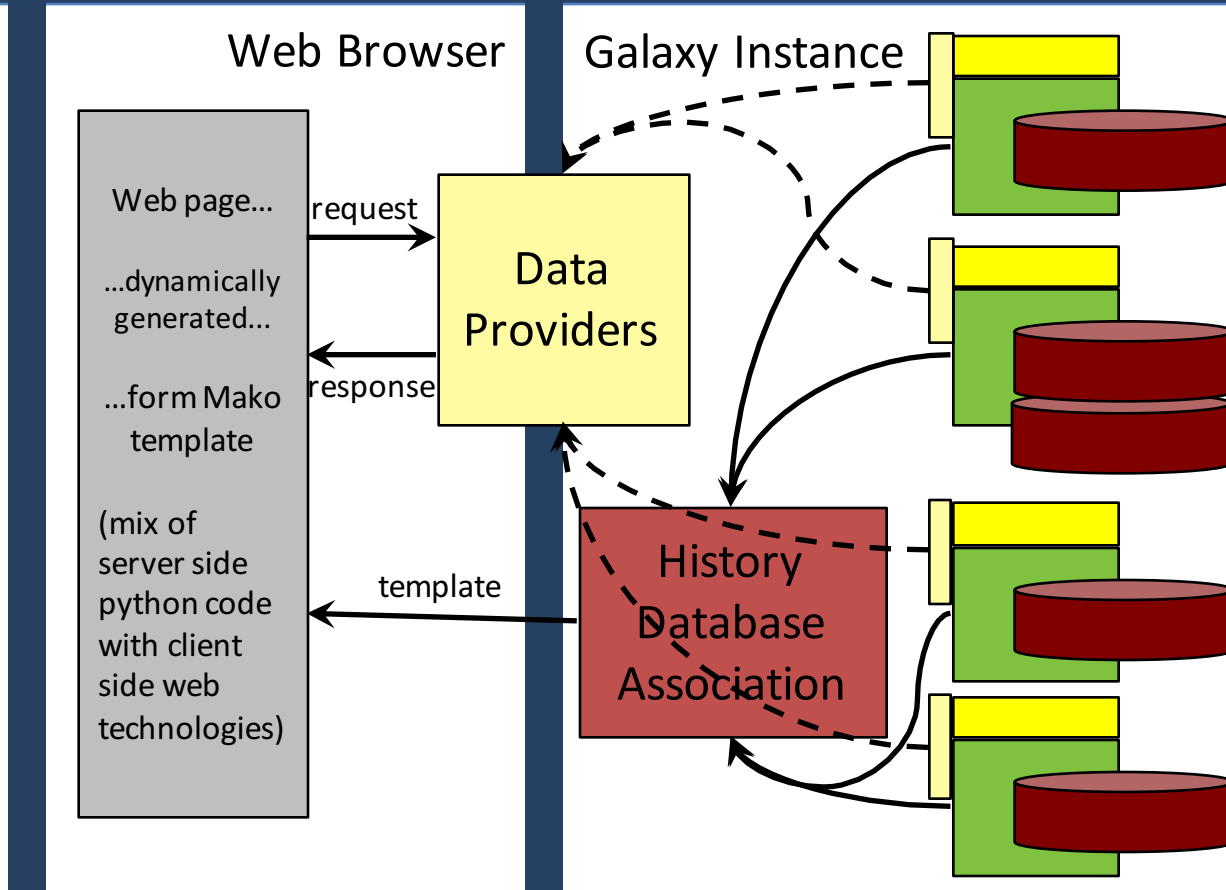
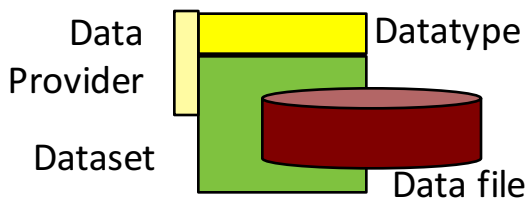
Used when:

- A tool consumes and/or produces a variable number of datasets
- The tool is implemented using a Swift script working in parallel

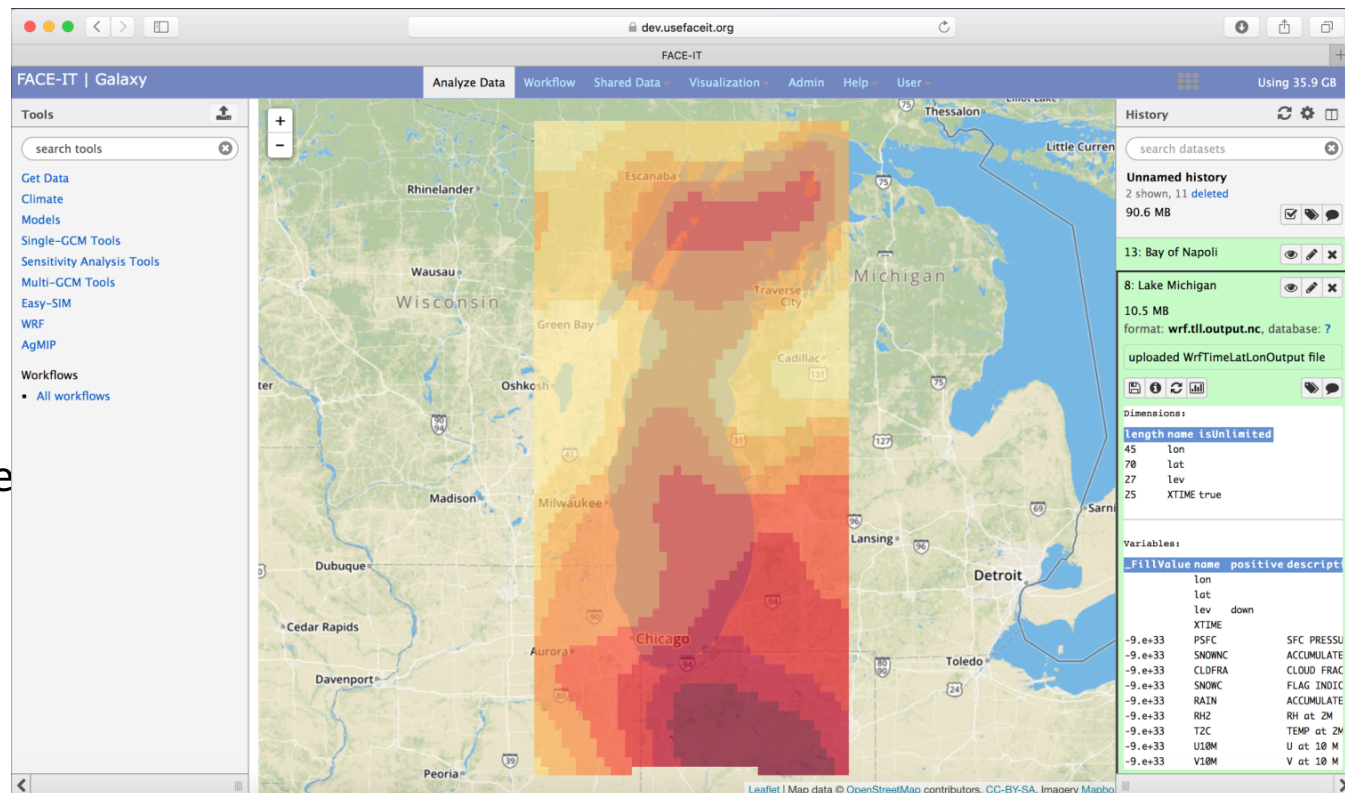
Teaching Galaxy to speak Earth Science

Step FIVE of 6: data providers

- **Data providers:** software components interfacing the datasets with the web browser.
- They provide data as array of JSON objects
- Key/Values, Columnar, custom
- Implemented in Datatype classes



- **Visualizers:**
client-side software
components for interactive
data visualization
- **Quasi-GIS!**
- **Map:**
Visualizes vector data produced
as GeoJSON objects by a data
provider
- **Wms (World Map Server):**
Visualizes raster data from
NetCDF datatypes.





Weather Research and Forecast @AWS: a (real) application.

Deliver on demand weather simulations

Canvas Generator (version 1.0.0)

Tools

search tools

- Get Data
- Climate
- Models
- Single-GCM Tools
- Sensitivity Analysis Tools
- Multi-GCM Tools
- WRF
- Easy-SIM
- AgMIP

Workflow control

Inputs

Workflow Canvas | Chicago HR MAR18h24

Details

Tool: ARWpost2grads

Version: 1.0.0

Wrf Model config file
Data input 'input' (wrf.model.xml)

Wrf Output
Data input 'wrfOutput' (wrf.output.nc)

Edit Step Actions

Rename Dataset
output Create

Add actions to this step; actions are applied when this workflow step completes.

Edit Step Attributes

Annotation / Notes:

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

This tool is a Face-IT Galaxy application for extreme weather simulations

- wrf-model.org
- +30k people community
- 150 countries
- high computing demanding

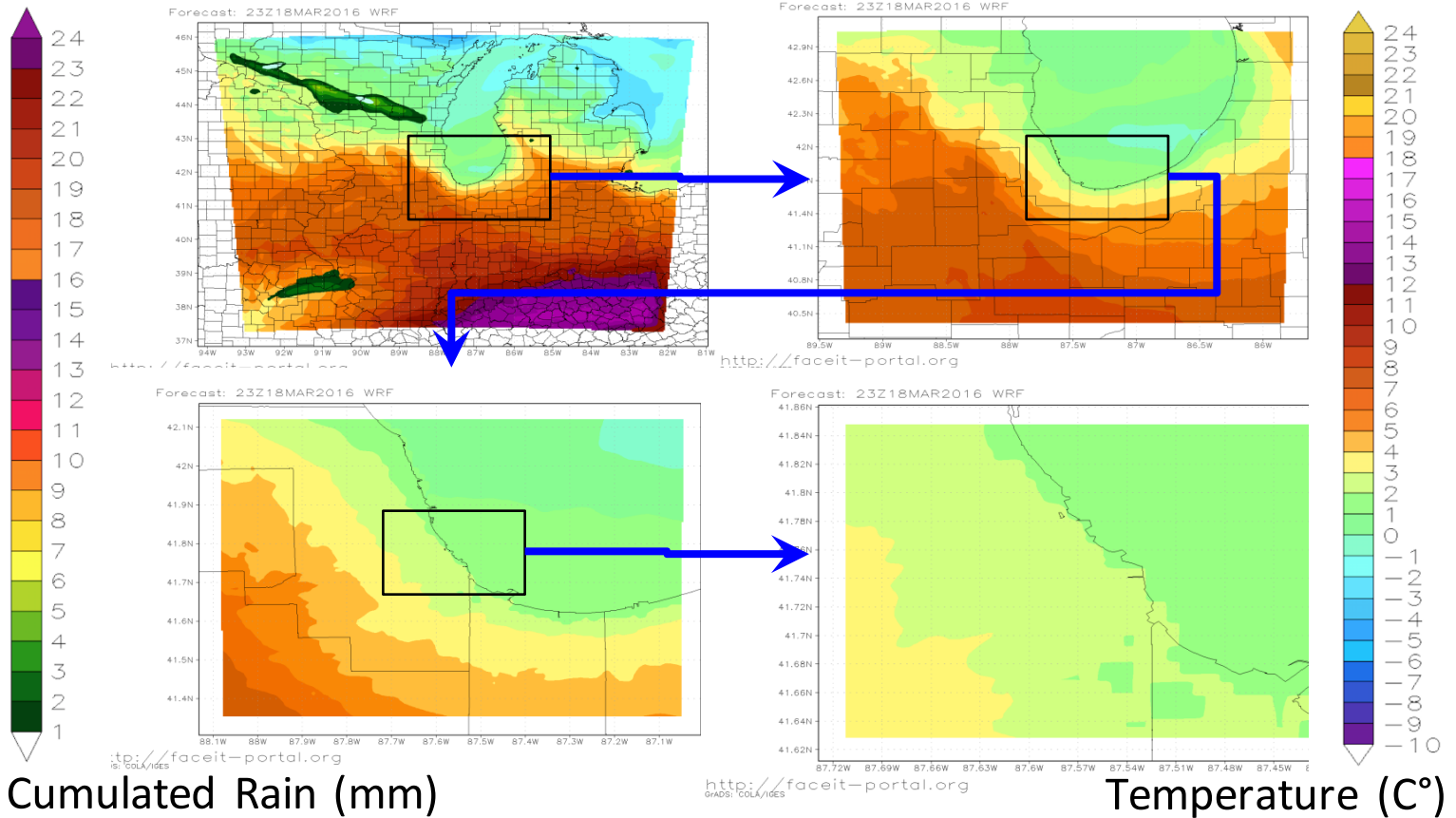




Weather Research and Forecast @AWS

Weather Forecast on Chicago area (333m)

- WRF
- 4 nested domains
- 9km to 333m

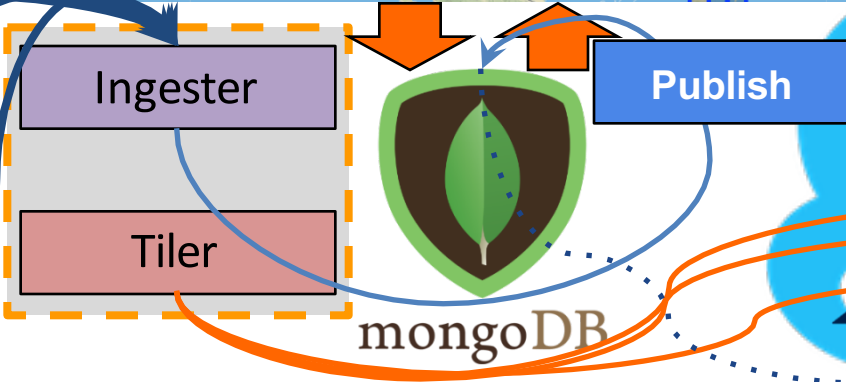
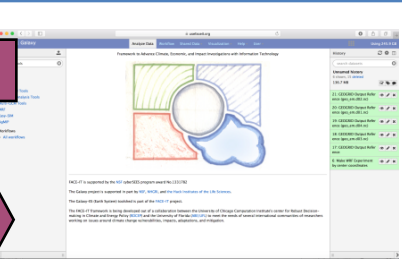
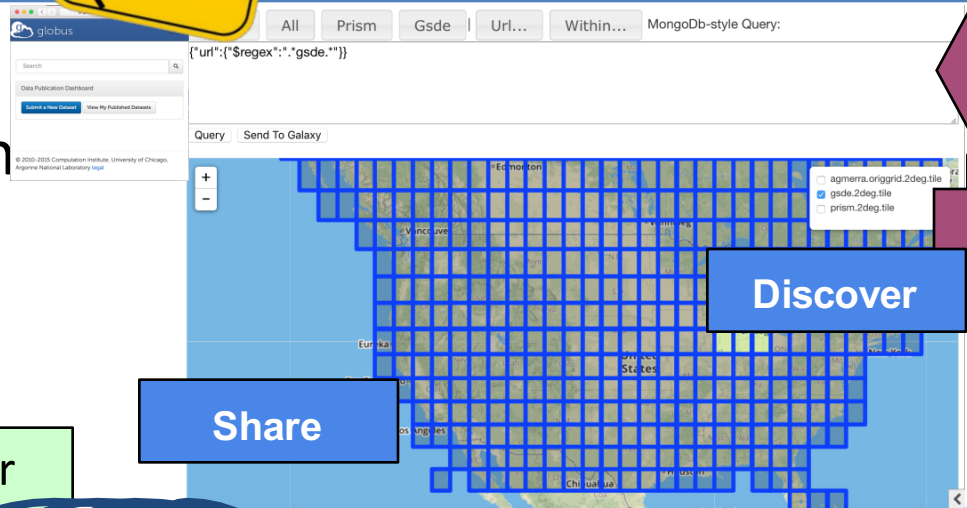
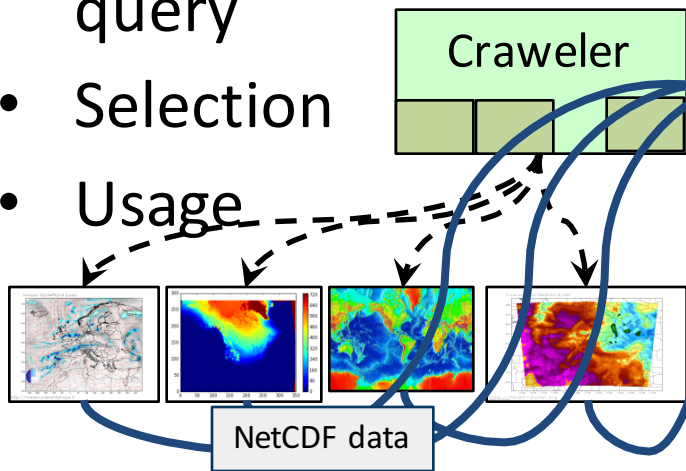


Breaking the fourth wall



NetCDF scavenging, discovery and provisioning

- NetCDF Scavenger
- Indexing / Ingestion
- Tiling
- Spatial / metadata query
- Selection
- Usage





Must be...
...continued!

Conclusions and [now] future works

- Face-IT Galaxy is a creative playground for the next generation of earth scientists powered by **Globus** for data movement and more.

<http://www.faceit-portal.org>

- **Propose** your application, write your code and **share it!**
- Spin-off projects: extreme weather simulations in the Bay of Napoli, IT (UniParthenope)

