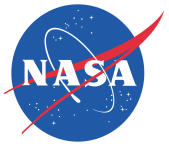


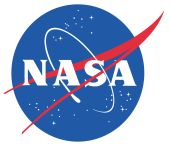
Exploring Earth Science Data with Giovanni: the Next Generation

Chris Lynnes

GES DISC

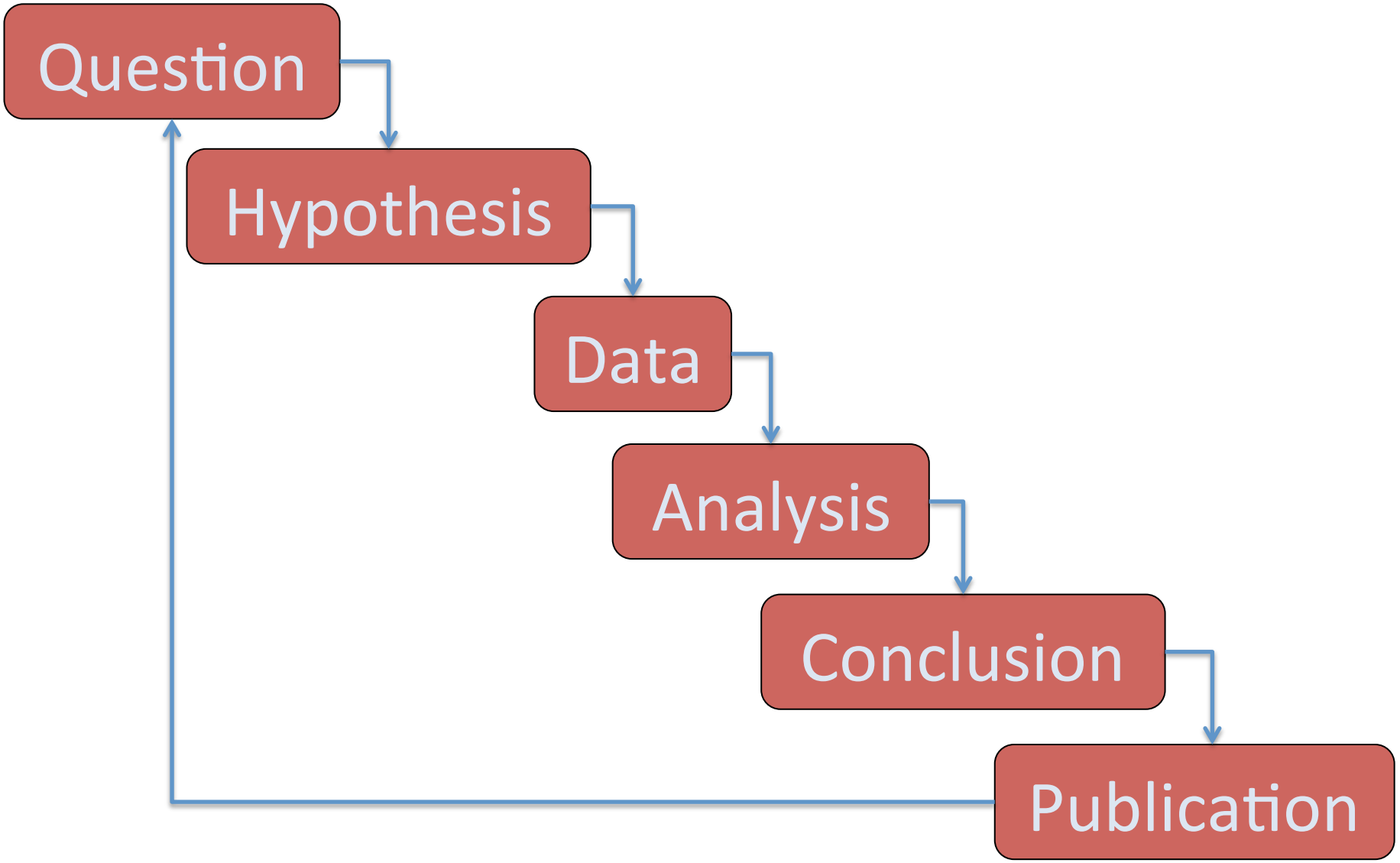
Past, Present, Future

- Part 1: Giovanni as Data Exploration Tool
- Part 2: Architecting for Data Exploration
- Part 3: The Future of Data Exploration in Giovanni

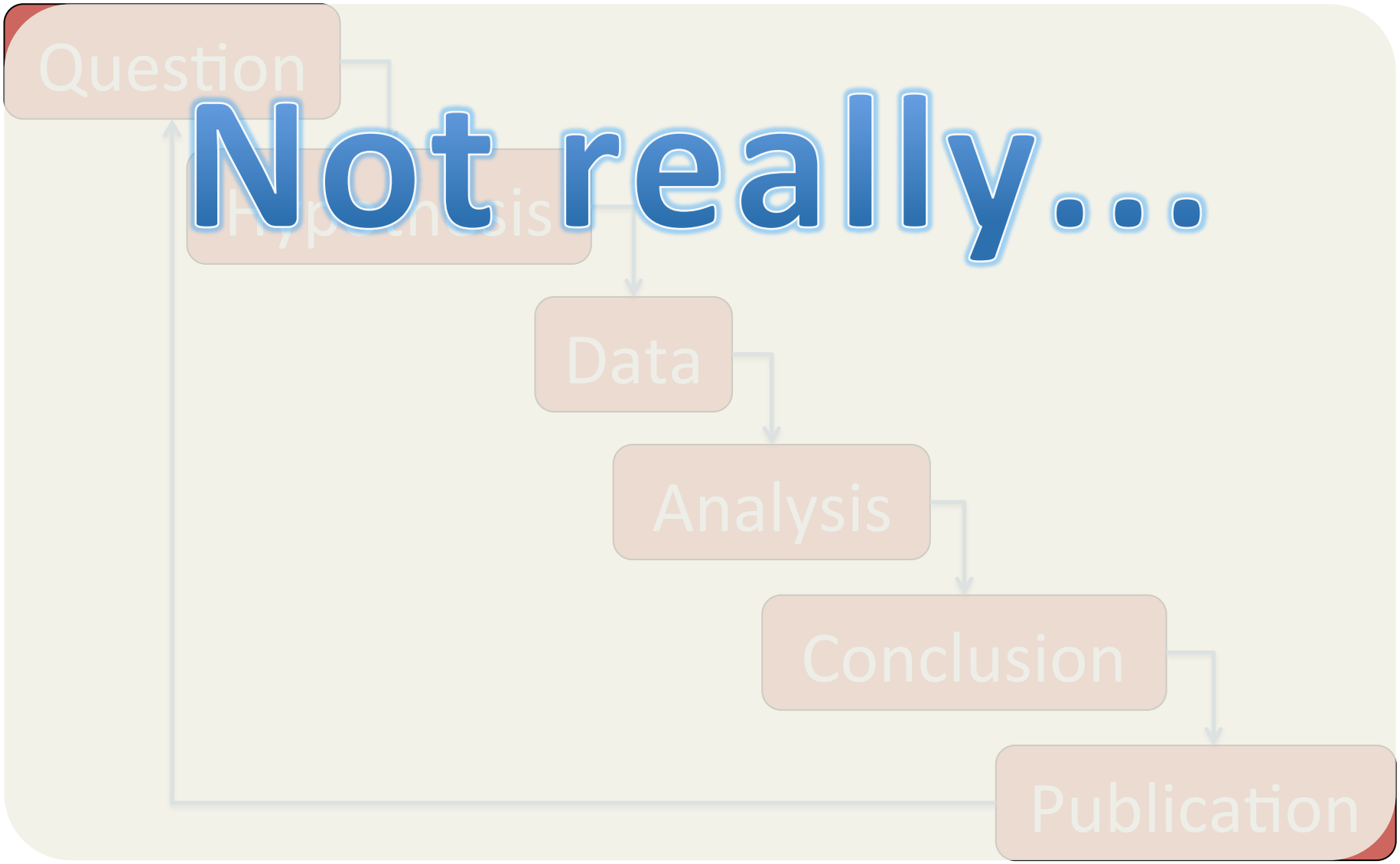


PART 1: GIOVANNI AS A DATA EXPLORATION TOOL

How Science Works

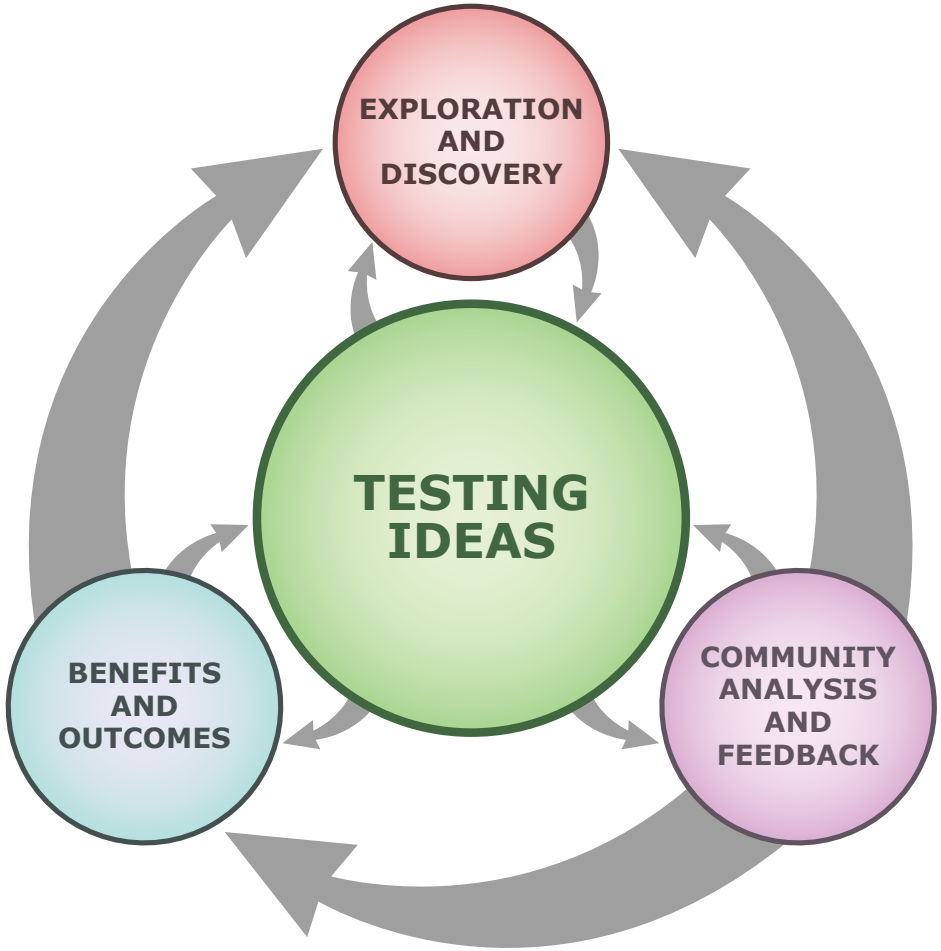


How Science Works



How Science Really Works

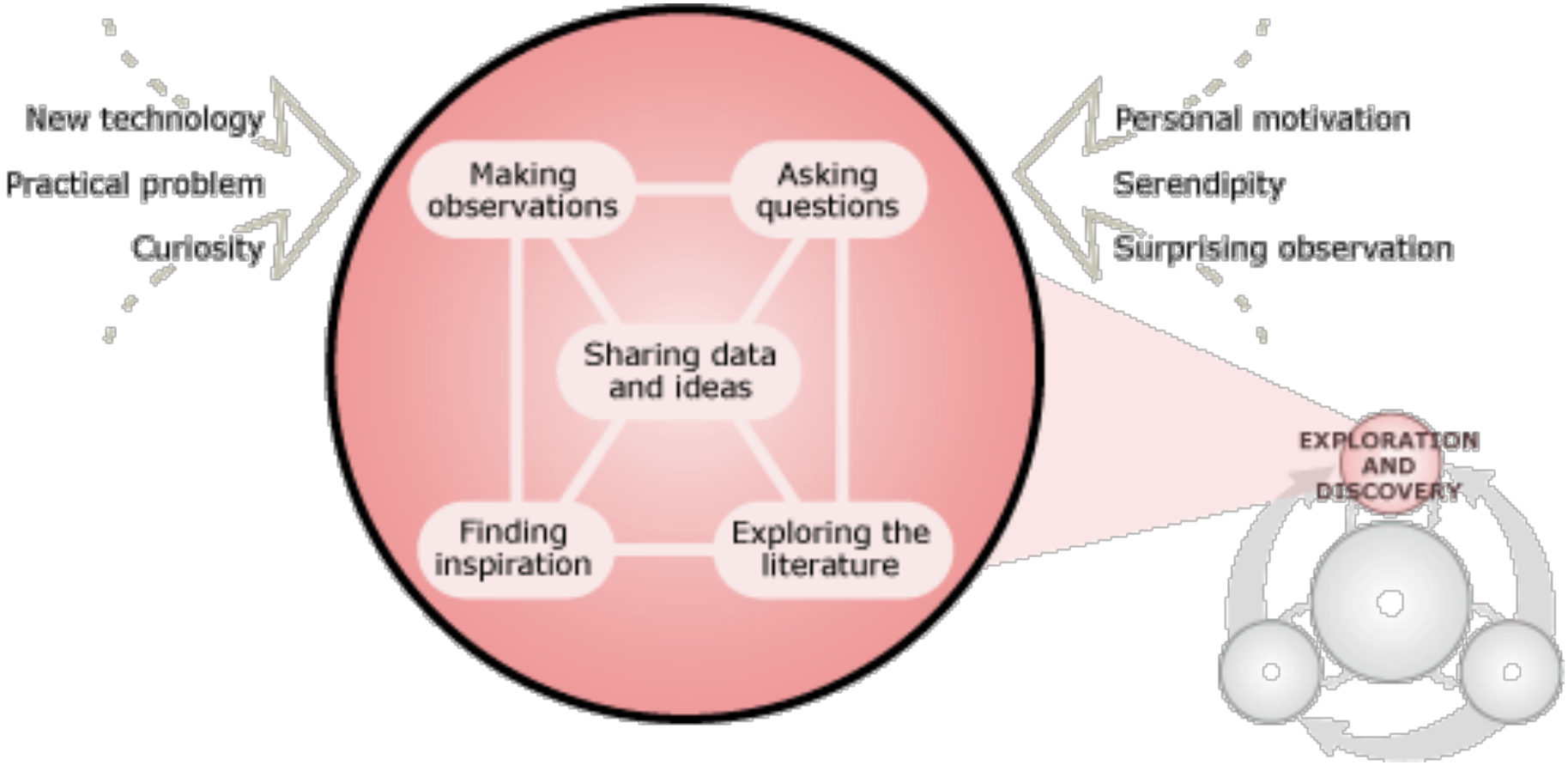
How science works



Understanding Science. 2013. University of California Museum of Paleontology. 17 Nov 2013 <http://www.understandingscience.org>

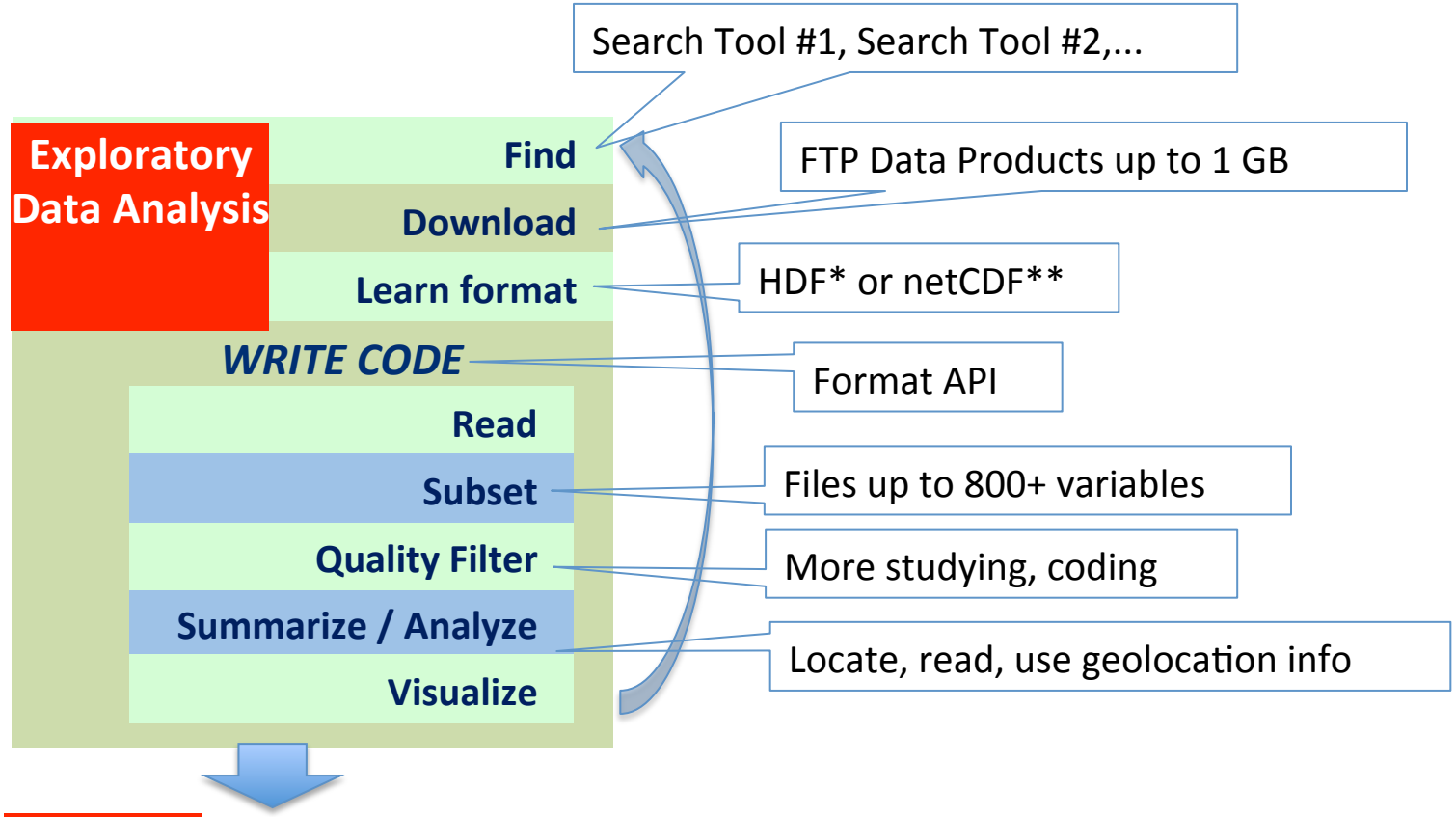
www.understandingscience.org
© 2008 The University of California Museum of Paleontology, Berkeley, and the Regents of the University of California

How Science Really Works



Understanding Science. 2013. University of California Museum of Paleontology. 17 Nov 2013 <http://www.understandingscience.org>

Remote Sensing Science



* Hierarchical Data Format ** network Common Data Form

http://giovanni.gsfc.nasa.gov/

Giovanni to the Rescue

Exploratory Data Analysis

- Find
- Download
- Learn format

WRITE CODE

- Read
- Subset
- Quality Filter
- Summarize / Analyze
- Visualize

Main Analysis Phase

- Select Data
- Analyze
- Derive Conclusions
- Publish

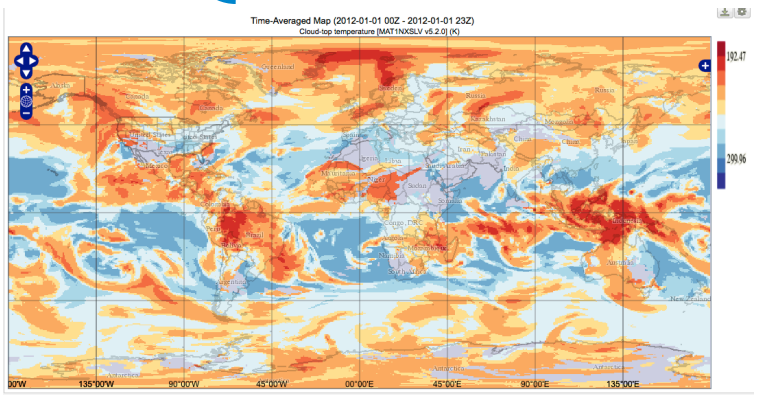
Giovanni provides *Quick-Start Exploratory Data Analysis:*

no coding necessary

Web-based Services

Giovanni

- Find
- Fetch
- Reformat
- Filter
- Regrid
- Summarize
- Visualize



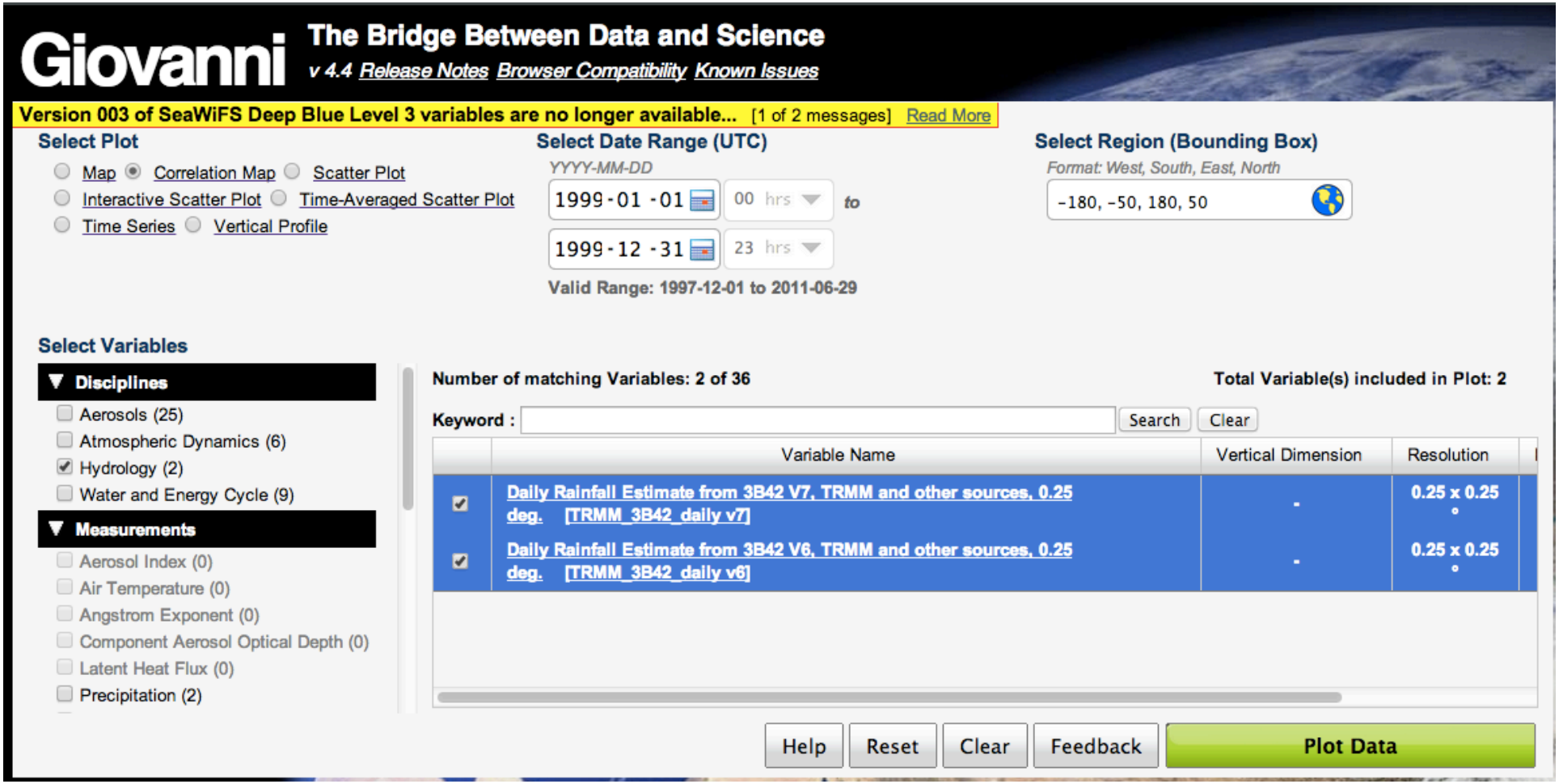
http://giovanni.gsfc.nasa.gov/

The Giovanni Legacy

Gregory Leptoukh
1953-2012

Yoram Kaufman
1948-2006

Giovanni User Interface



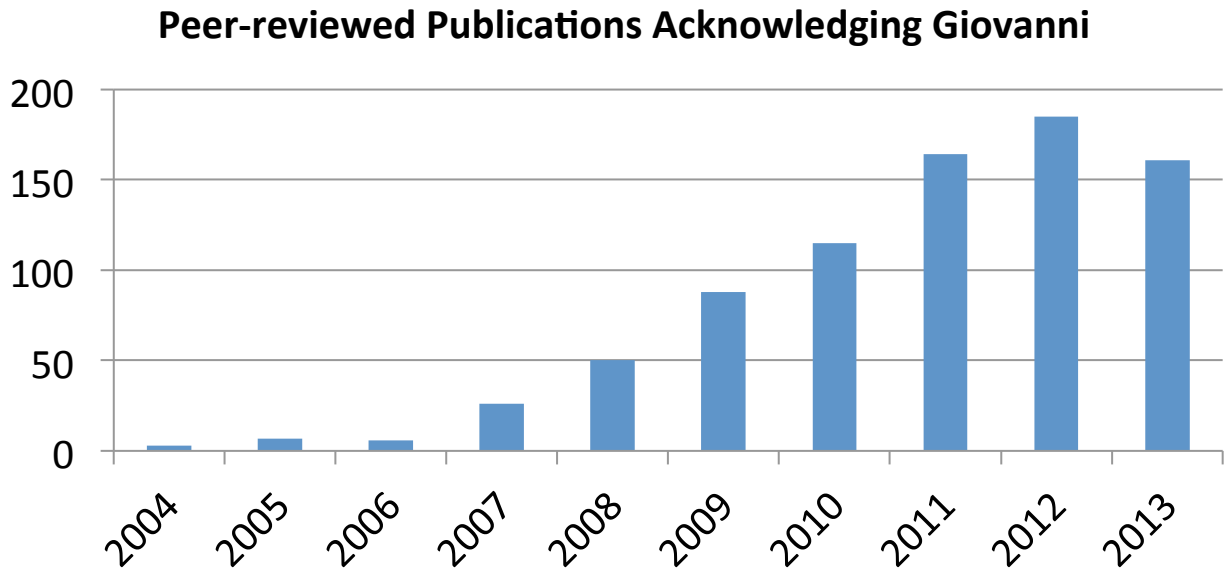
http://giovanni.gsfc.nasa.gov/

GES – DISC
Goddard Earth Sciences
Data Information Services Center

Giovanni Capabilities

- Visualizations (many)
- Data Comparison
- Basic Analysis
- Data Download
- Reformatting to netCDF, KMZ
- Lineage
- …

Giovanni by the Numbers



- ~4000 variables from 5 EOSDIS data centers, plus a few other data providers
- ~4000 unique hosts access per month

Reengineering Giovanni

1. Original architecture (Giovanni-3) was never fully implemented
   - Difficult to add major features
   - Expensive to maintain
2. New technologies enable new capabilities
3. Giovanni 4 is designed to be more agile
4. *Support for Data Exploration is emphasized*

Giovanni for Data Exploration

- Does the dataset show the phenomenon I am studying?
- How does dataset X compare with dataset Y?
  - Which one is better for my purpose?
  - How do they relate to each other?
- "That's funny..."*

* The most exciting phrase to hear in science, the one that heralds the most discoveries, is not 'Eureka!' (I found it!) but 'That's funny...'

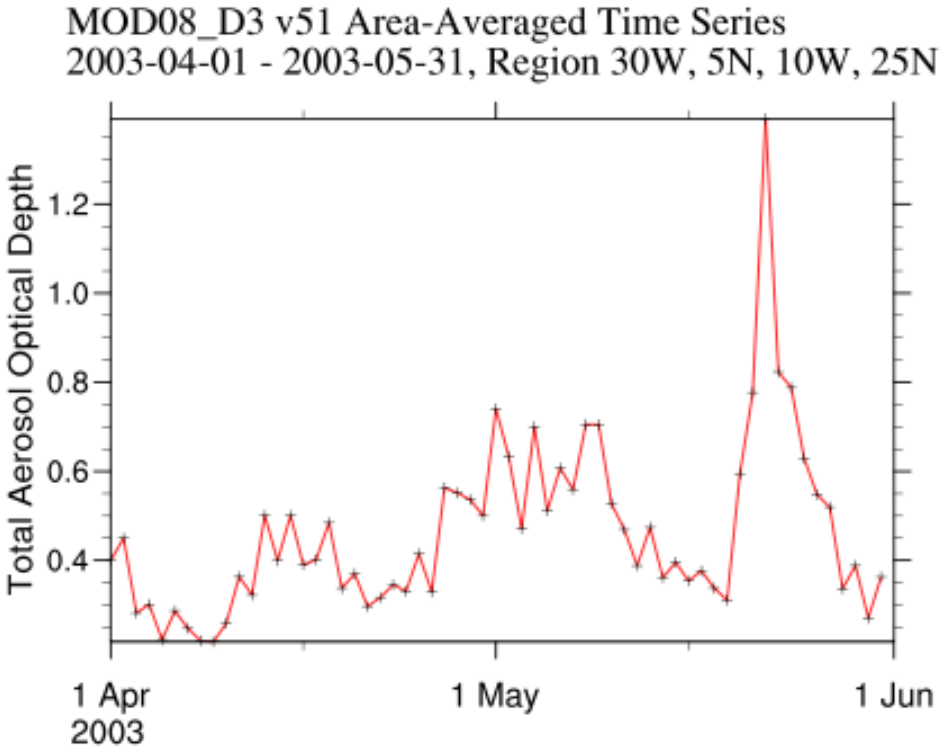
Isaac Asimov

What do we need for exploration?

- Visuals
- Access to the full data record
- Navigation and interrogation methods for vast quantities of data*
  - X, Y, Z, T
  - Data fields
- Analysis capabilities
- Interactivity & quick response
  - Not “What was I thinking?”

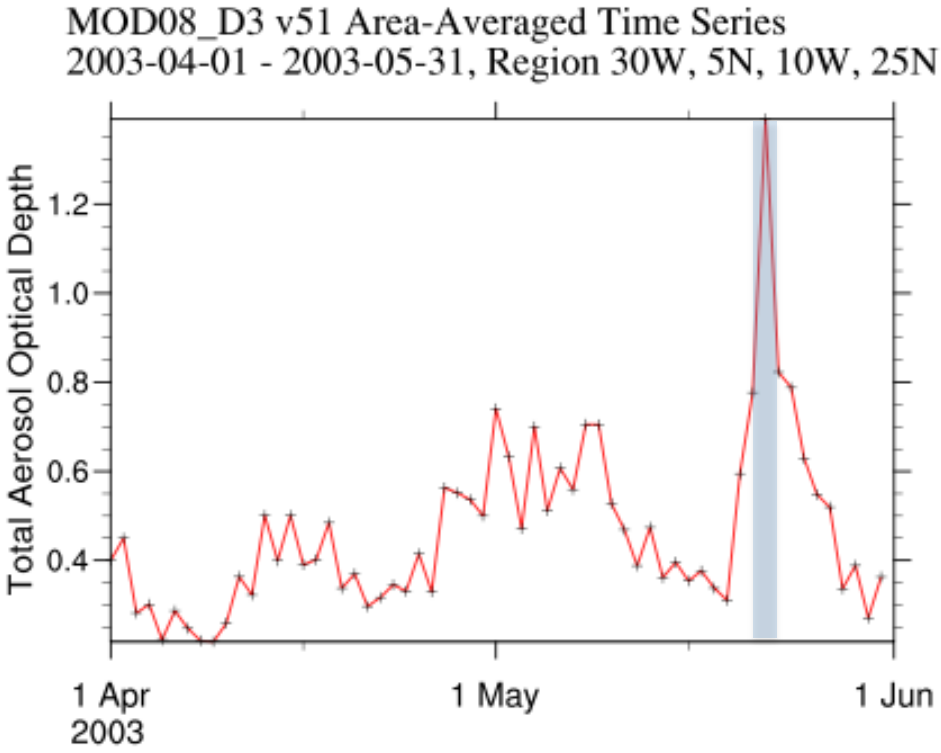
*aka “Big Data”

Example: Exploring in Time and Space

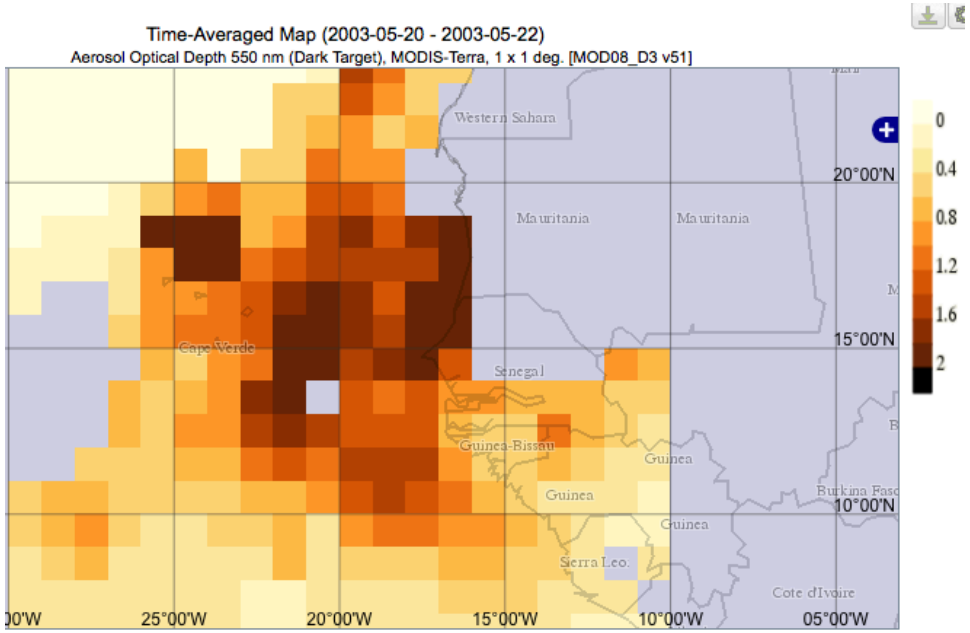


http://giovanni.gsfc.nasa.gov/

Example: Exploring in Time and Space



Time Averaged Map 20-22 May



http://giovanni.gsfc.nasa.gov/

Visually Summarizing Data

	Longitude	Latitude	Vertical	Time	Plot
2-D Color Slice	✔	✔	[i]	Σ	Time-averaged map
	✔	Σ	✔	Σ	Longitude Cross-Section
	Σ	✔	✔	Σ	Latitude Cross-Section
	✔	Σ	[i]	✔	Hovmoller (Longitude)
	Σ	✔	[i]	✔	Hovmoller (Latitude)
	Σ	Σ	✔	✔	Vertical-Time Cross-Section
1-D Line Plots	✔	Σ	[i]	Σ	Meridional Mean
	Σ	✔	[i]	Σ	Zonal Mean
	Σ	Σ	✔	Σ	Vertical Profile
	Σ	Σ	[i]	✔	Area-averaged Time Series

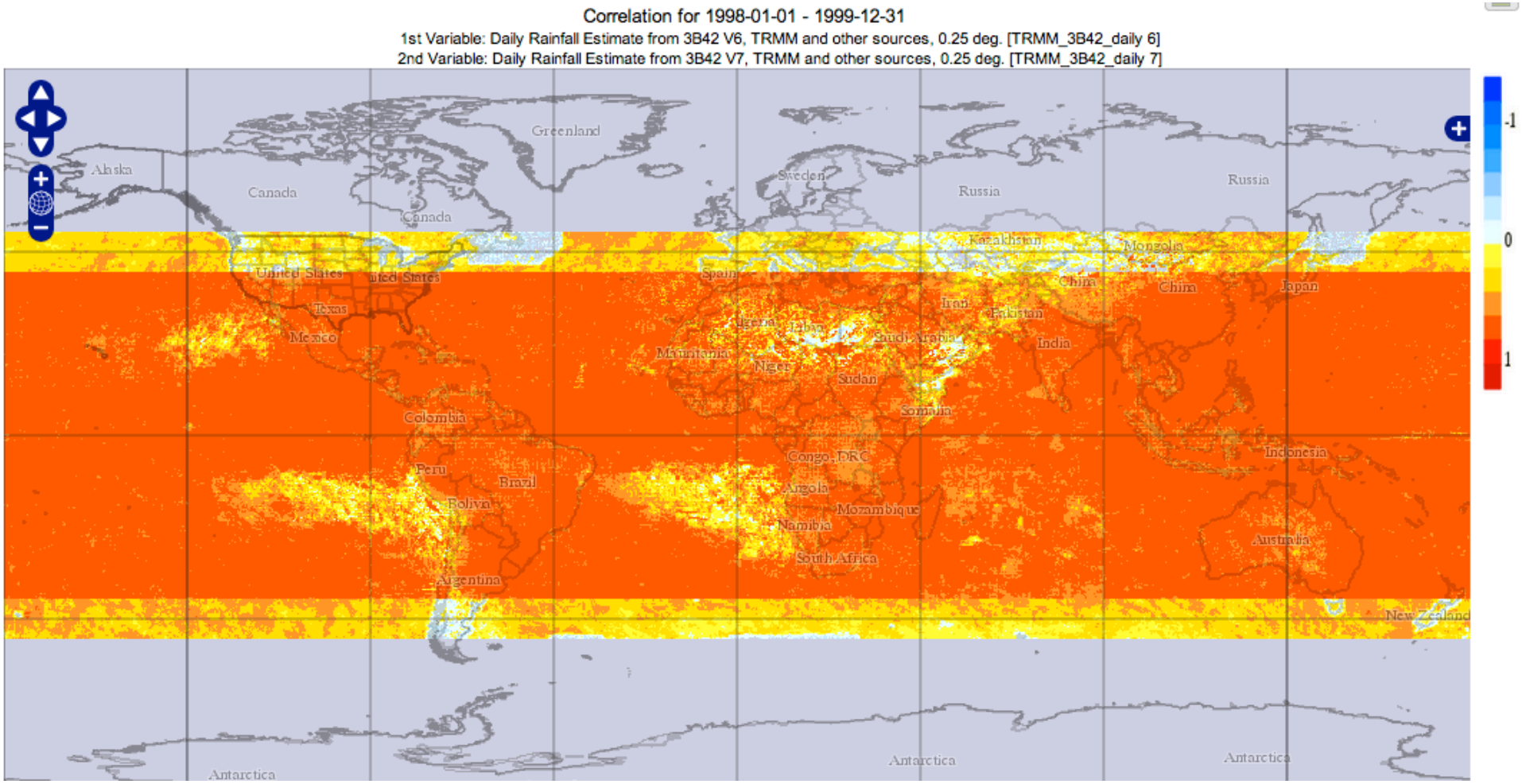
✔	Plot Axis
Σ	Aggregation (Summation, average, ...)
[i]	Select Layer or Level

Comparing Data

Latitude	Longitude	Z	Time	Plot Type
✔	✔	[i]	Σ	Correlation Map
✔	✔	[i]	Σ	Difference Map
Σ	Σ	[i]	✔	Difference Time Series
--	--	[i]	--	Scatterplot
✔	✔	[i]	--	Interactive Scatterplot+Map
✔	✔	[i]	Σ	Time-averaged Scatterplot+Map

Comparing Data with Giovanni

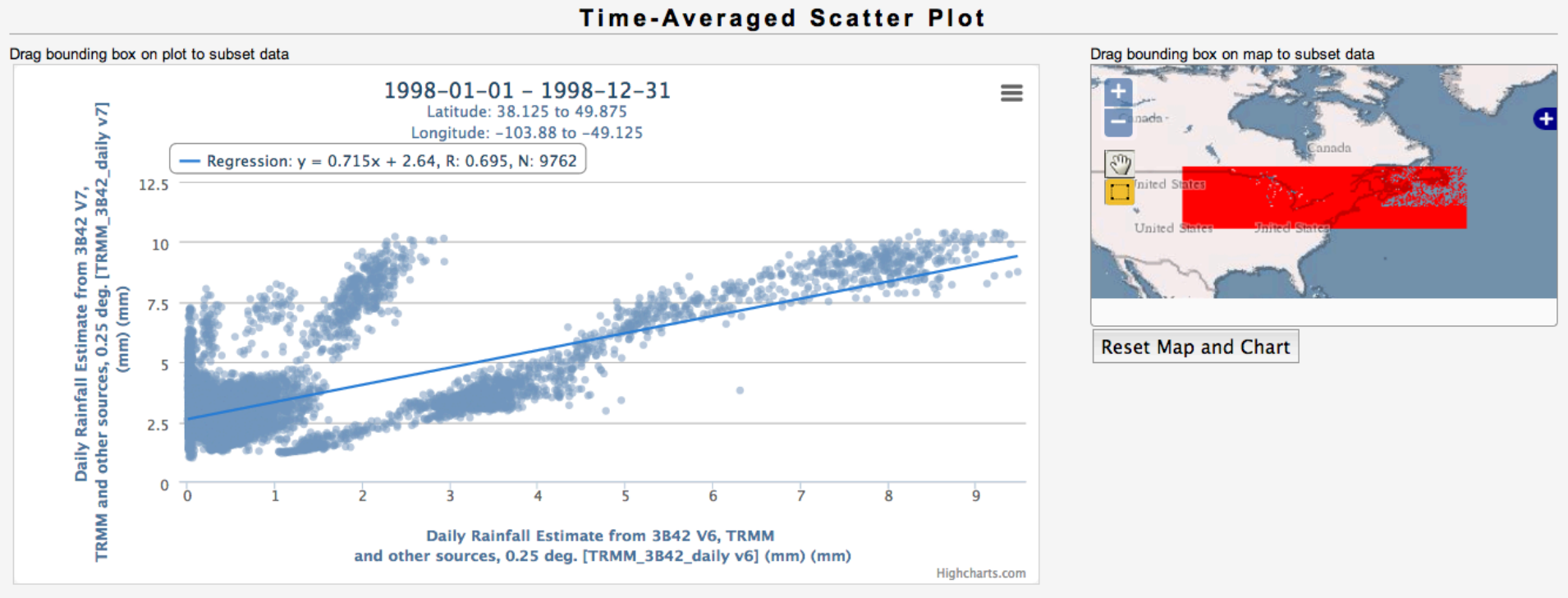
Correlation Map: TRMM 3B42 V6 vs. V7, 1998-1999



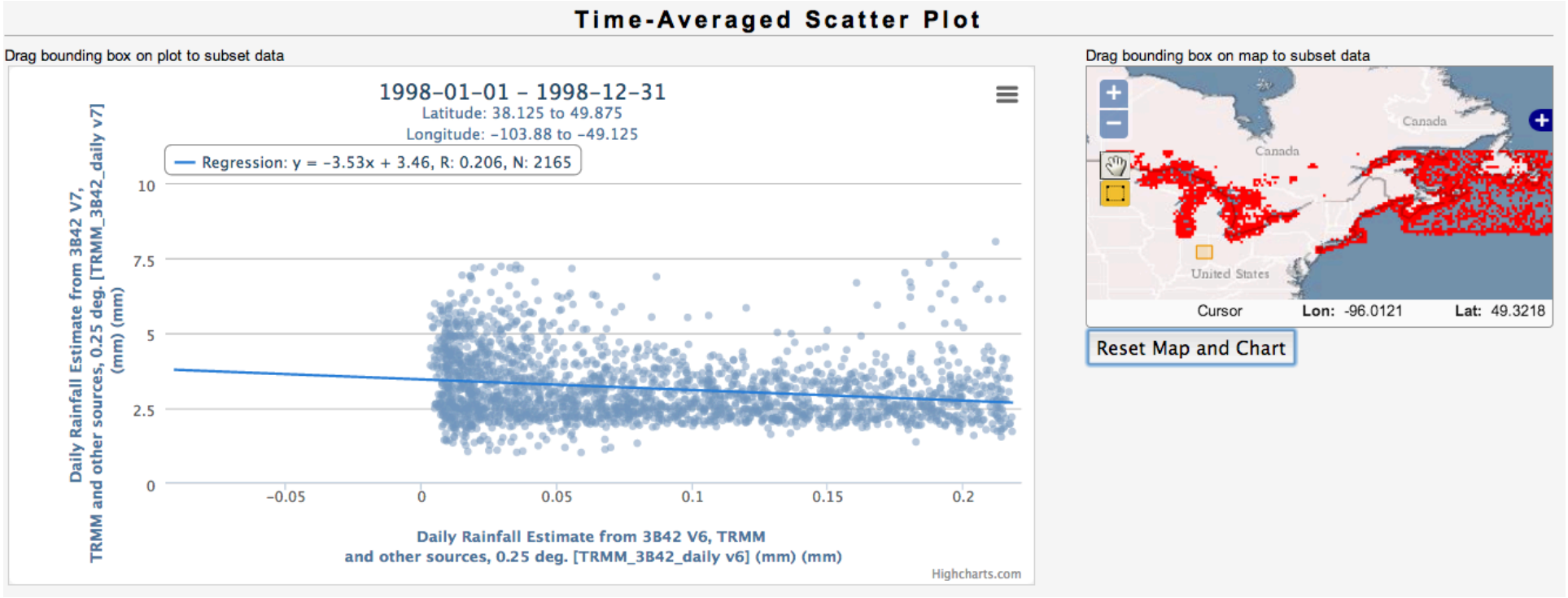
http://giovanni.gsfc.nasa.gov/

GES – DISC
Goddard Earth Sciences
Data Information Services Center

Interactivity via Client-side Viz



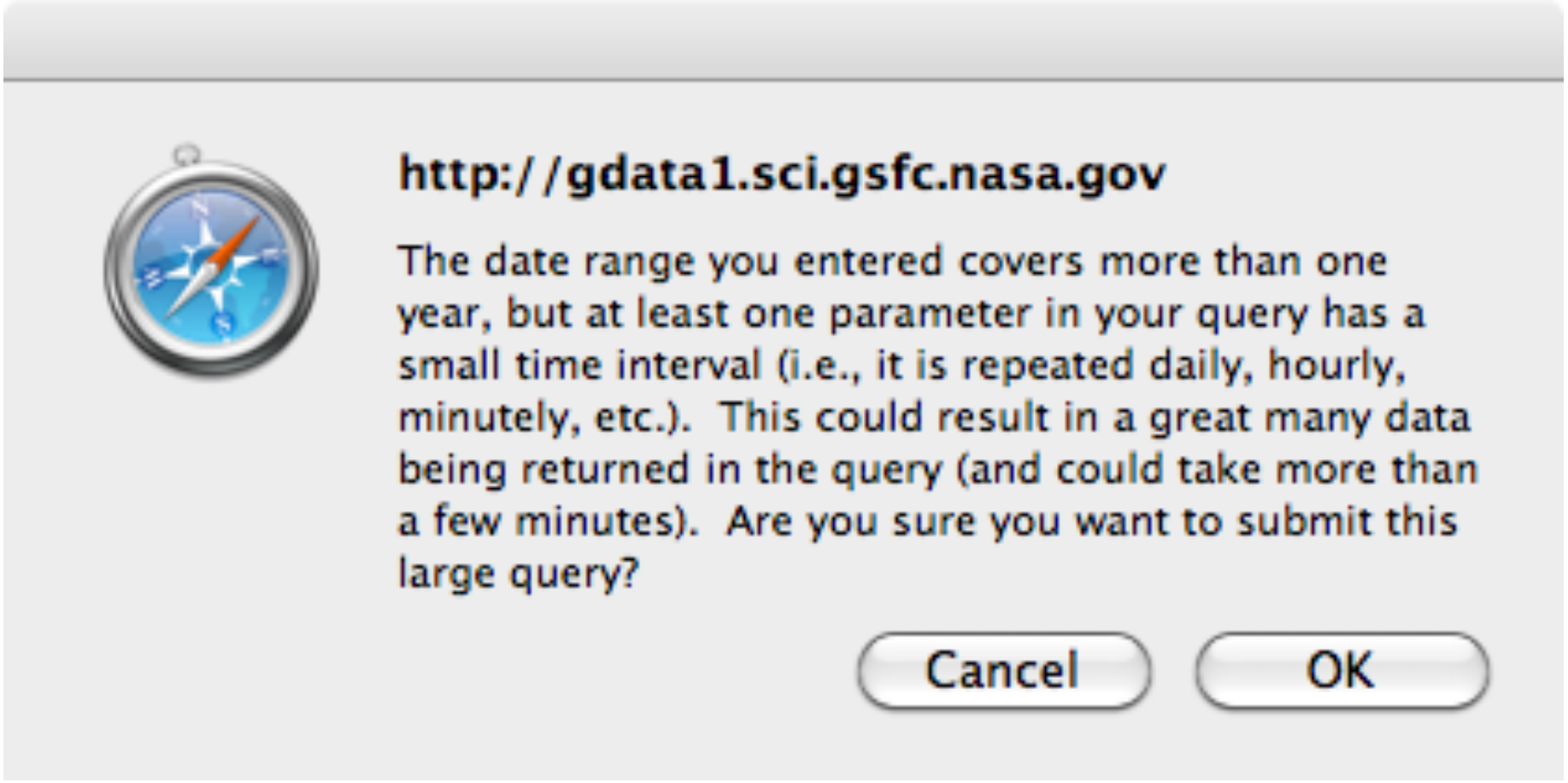
Interactivity via Client-side Viz



http://giovanni.gsfc.nasa.gov/

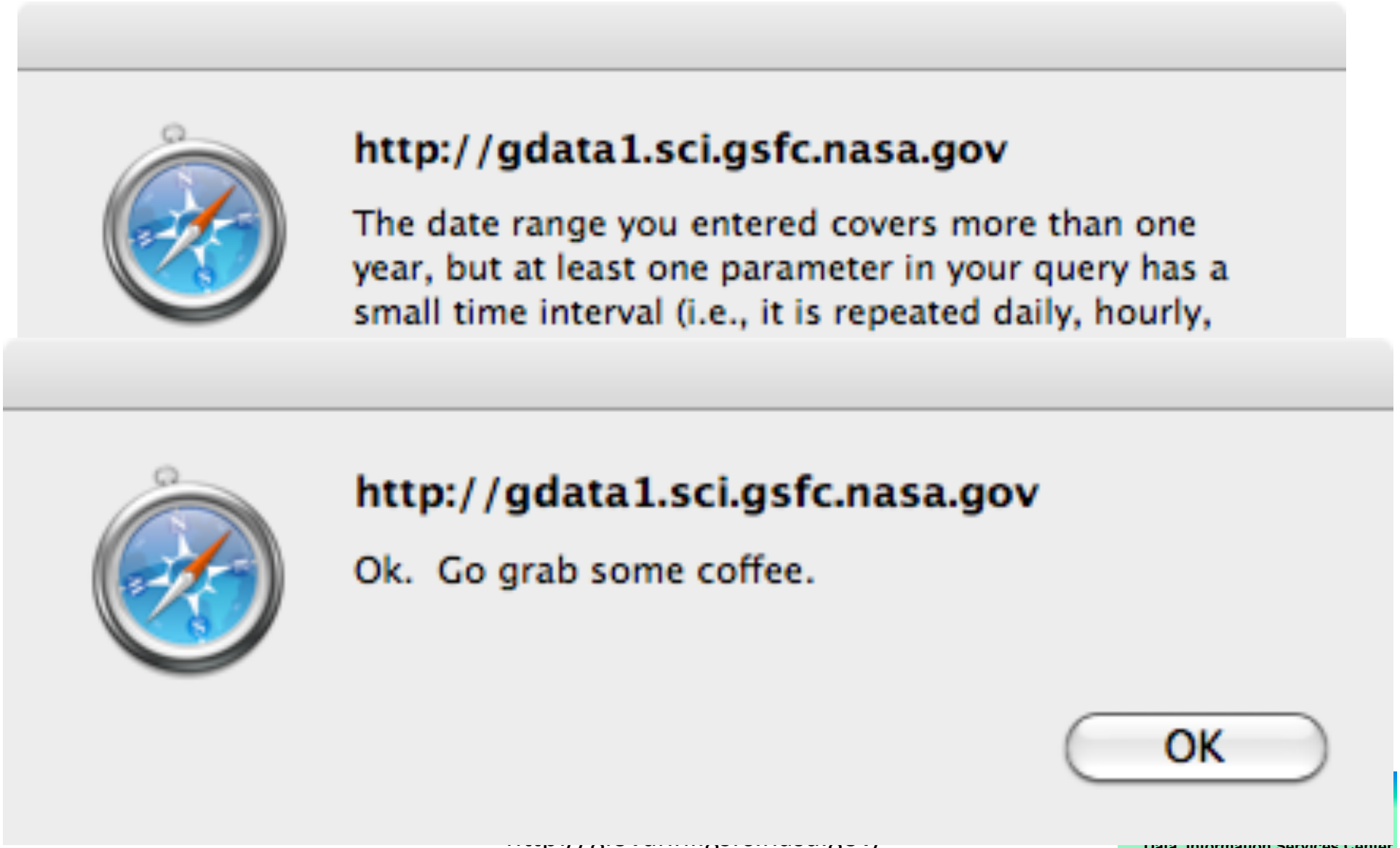
Speed: Breaking the Coffee Barrier

1 Year of MODIS L3 Daily Aerosol data in today's Giovanni-3:

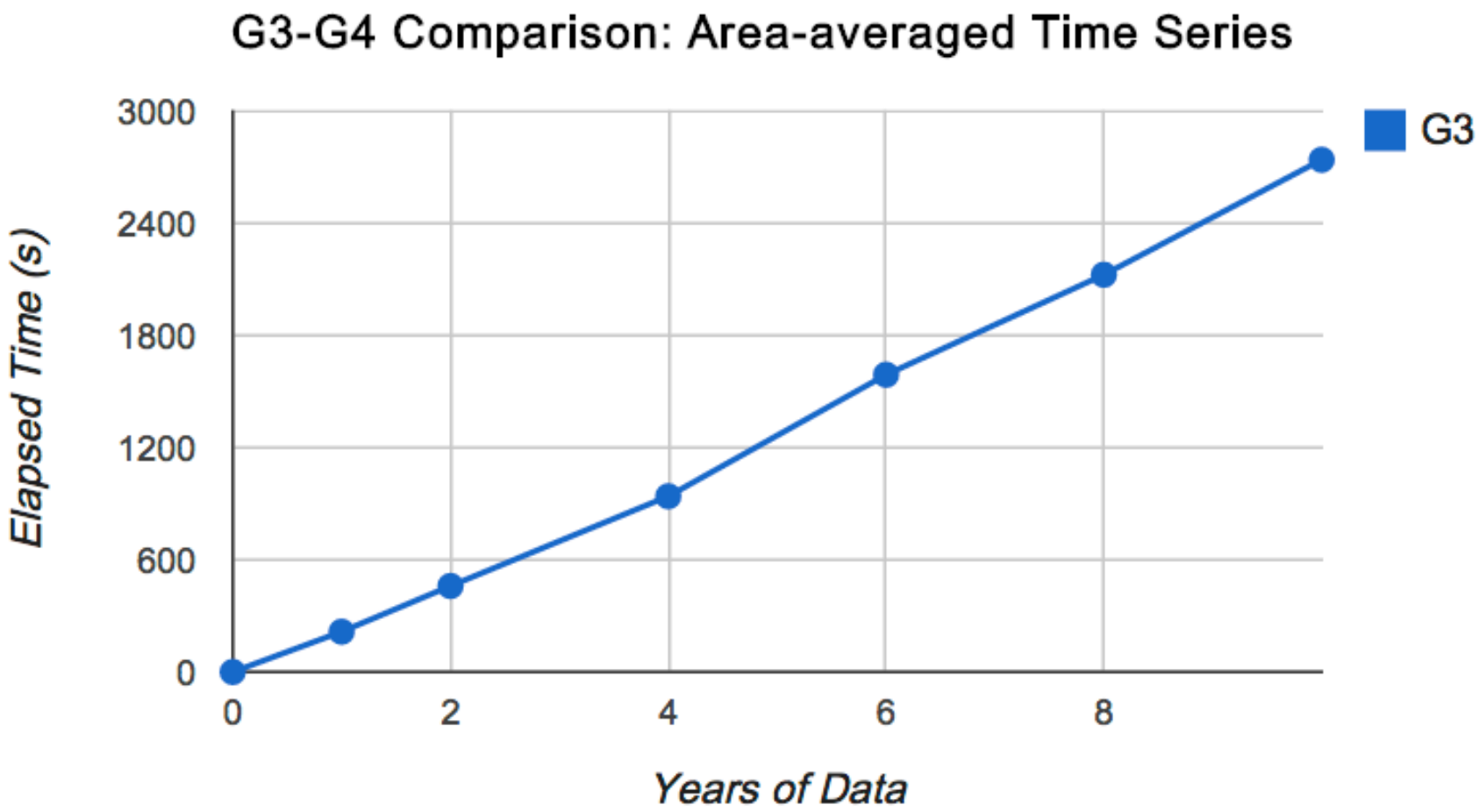


Speed: Breaking the Coffee Barrier

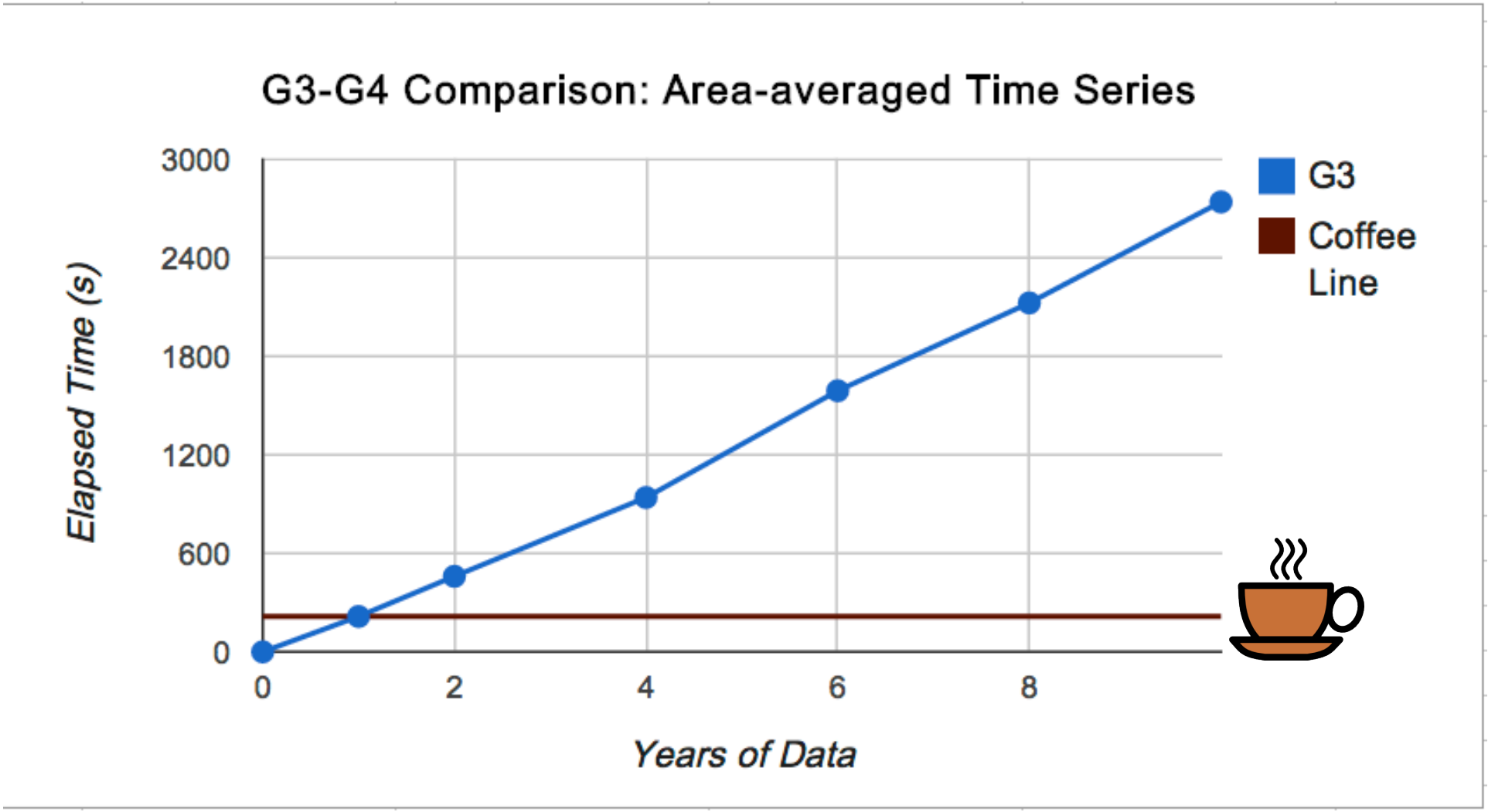
1 Year of MODIS L3 Daily Aerosol data in today's Giovanni-3:



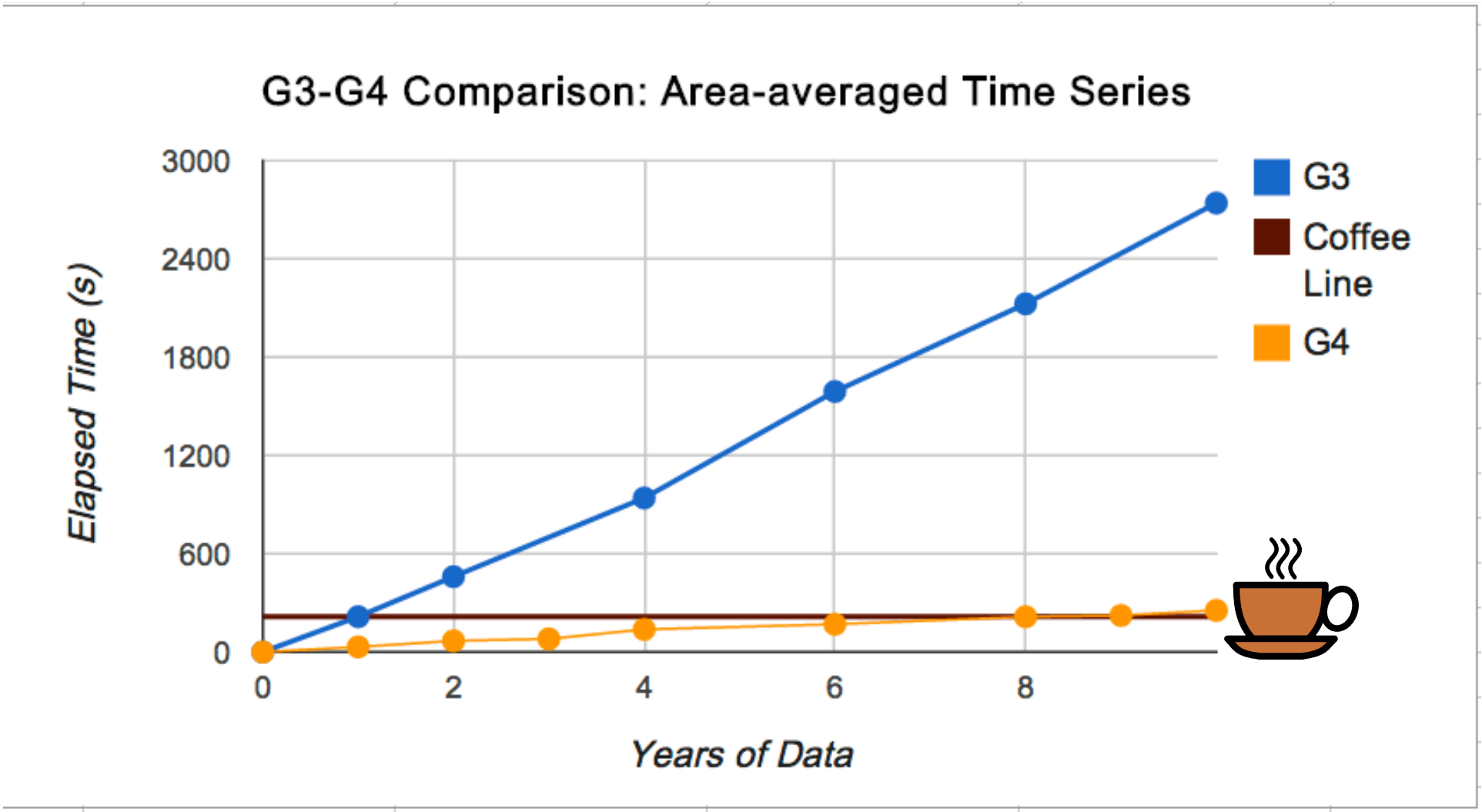
Time to Plot Time Series of MODIS Daily Aerosols

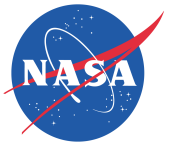


Time to Plot Time Series of MODIS Daily Aerosols



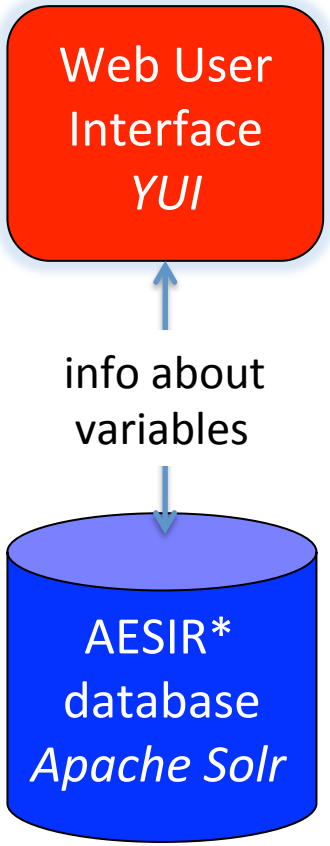
Time to Plot Time Series of MODIS Daily Aerosols





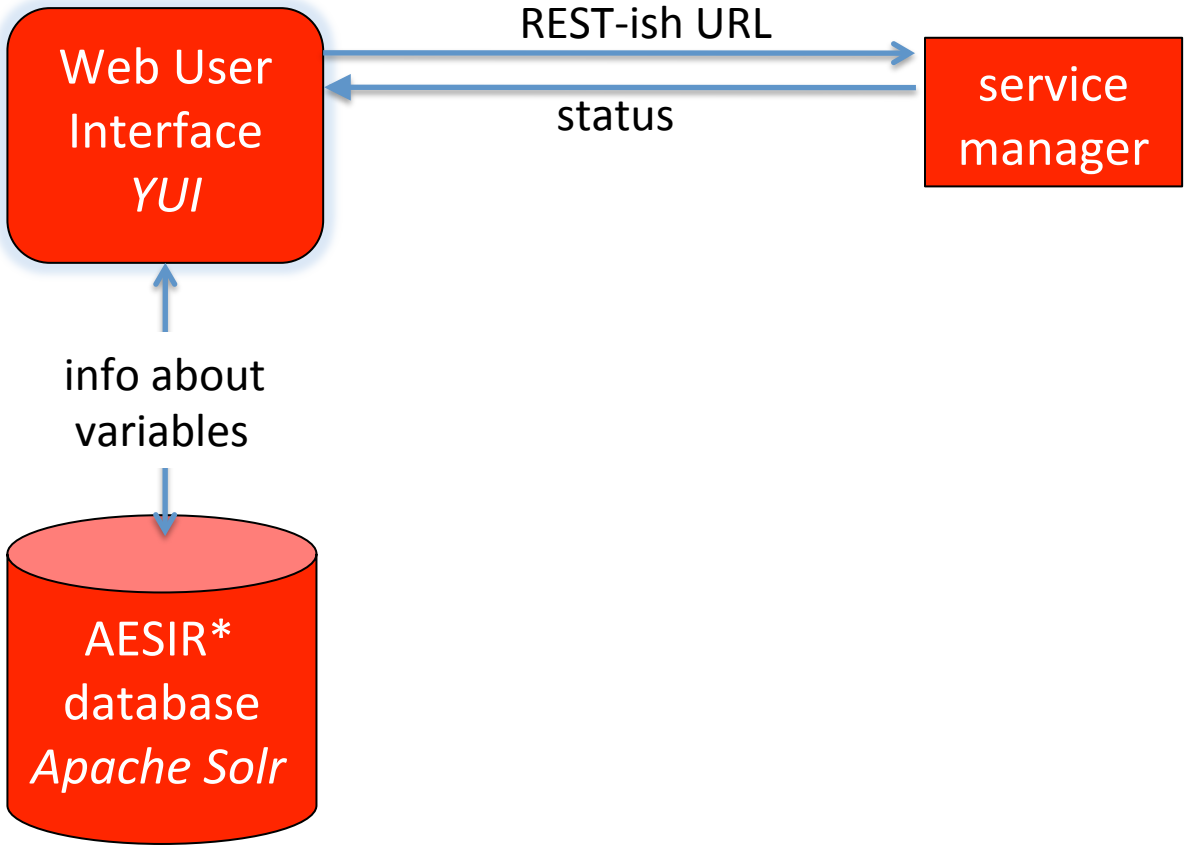
PART II – ARCHITECTING FOR EXPLORATION

Giovanni-4 Architecture

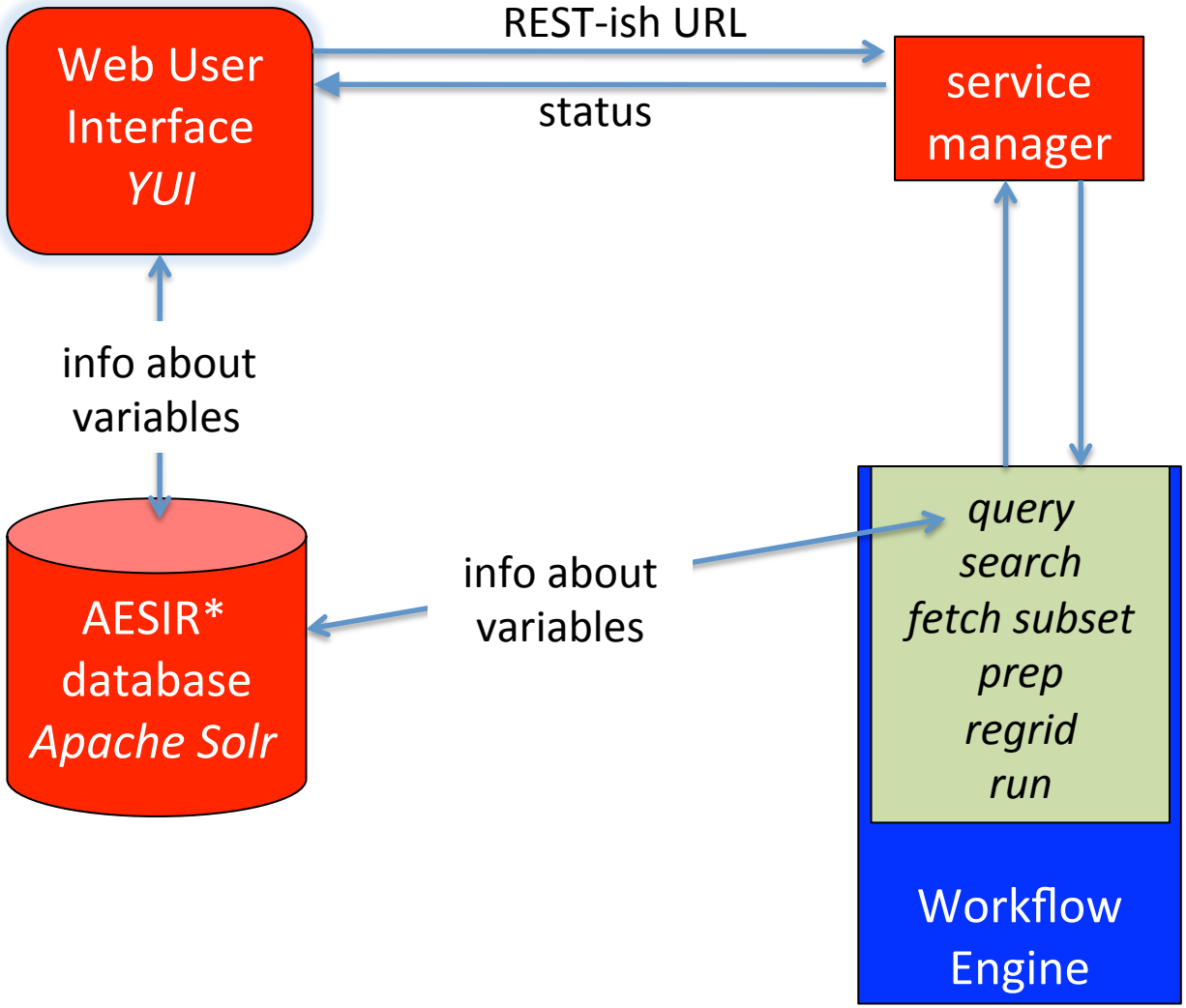


*Application-friendly Earth Science Information Retriever

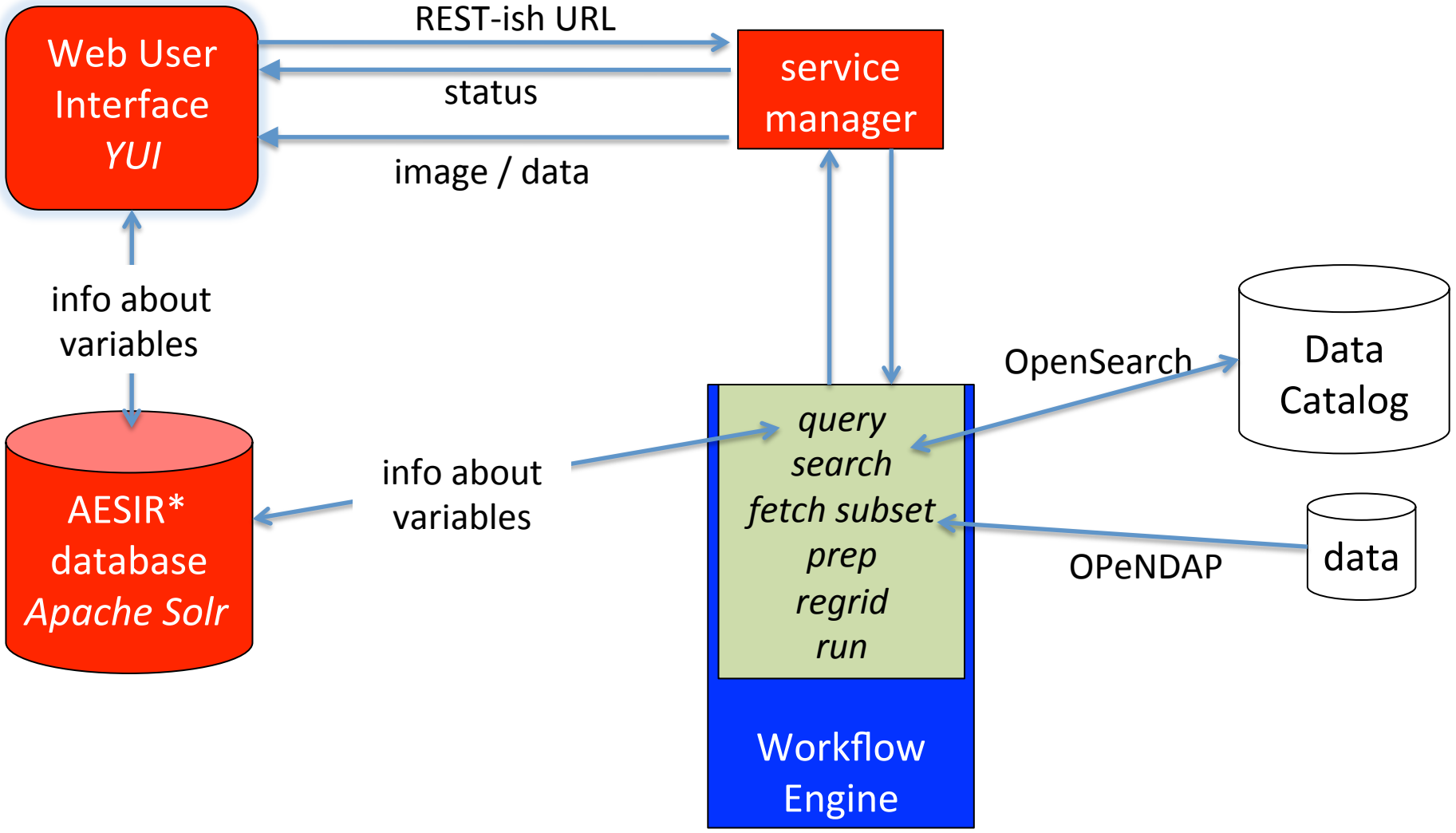
Giovanni-4 Architecture



Giovanni-4 Architecture



Giovanni-4 Architecture



Performance Approach

- Cache: Preprocess data and save in cache
- Fast code: Use nco for processing (Run)
- Lazy subsetting: Defer spatial subsetting until processing
- Results Reuse: Reuse intermediate results from previous runs
- Reuse Corollary: Rerun as few steps as necessary based on user selection

Reused Components for Exploration

Reused Component	Usage	Benefit
YUI Web widgets, Solr Database	Faceted browse	Navigating many variables
nco	Processing netCDF data	*Fast*
HighCharts	Browser-side plotting	Interactive plots
MapServer, OpenLayers	Map visualization	Zoom / Pan
Kepler	Workflow engine	...

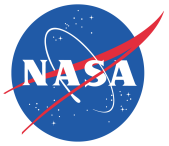
Reused Components for Exploration

Reused Component	Usage	Benefit
YUI Web widgets, Solr Database	Faceted browse	Navigating many variables
nco	Processing netCDF data	*Fast*
HighCharts	Browser-side plotting	Interactive plots
MapServer, OpenLayers	Map visualization	Zoom / Pan
~~Kepler~~	Workflow engine	...

Lesson: Performance is the Achilles Heel of Reuse

Giovanni Workflow Engine Needs

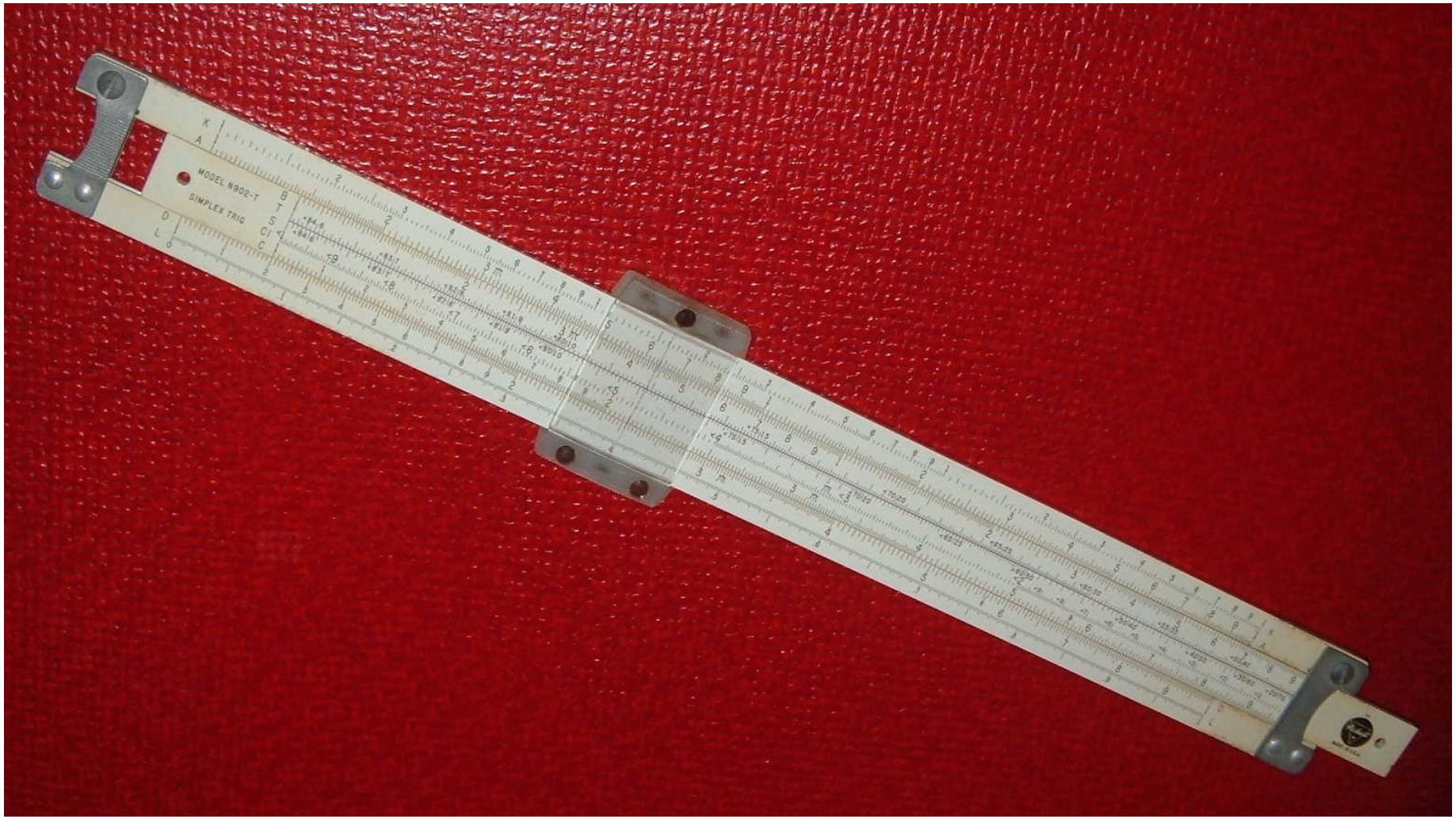
- Run reliably
- Run fast
- Run minimally (backward chaining)
  - Go only as far back as absolutely needed in the workflow to restart a subsequent request
  - Driven by the iterative nature of the data exploration Use Case



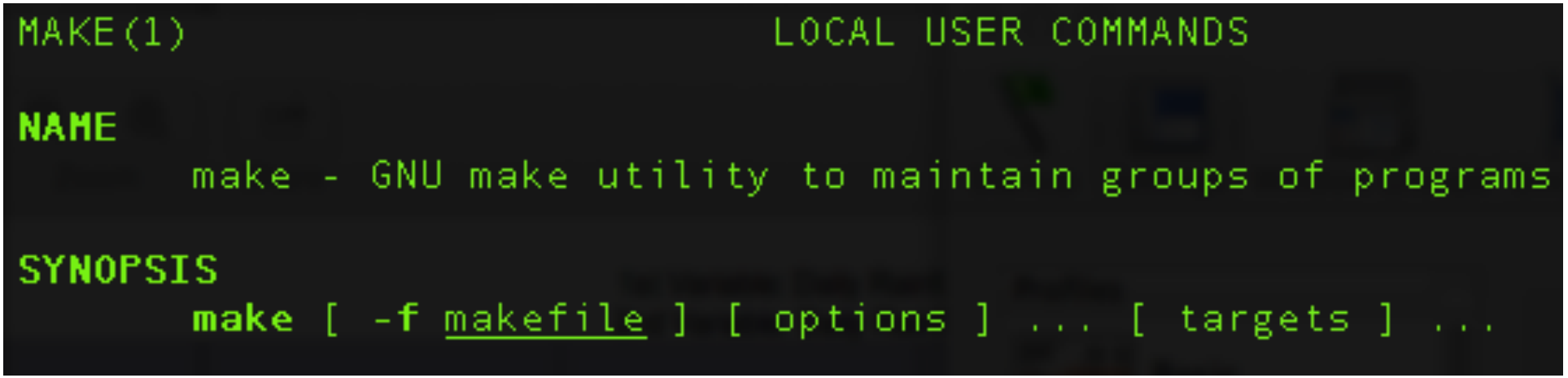
Tapping an Ancient Computing Technology...

Nope, too far back...

Right millenium...



Here we go...



- Runs fast, reliably
- Simple parallelization
- Simple load throttling
- The Last Word in Backward Chaining

PART III – THE FUTURE OF DATA EXPLORATION WITH GIOVANNI

Potential directions

- Interactivity++
- Speed++
- Domain-Specific
  - Seasonal analysis
  - Level 2 Swath data
  - Shapefile masking
- Participatory Design (myGiovanni)
- Collaboration with gSocial
- Feature Detection
- Federated Giovanni

Interactivity++

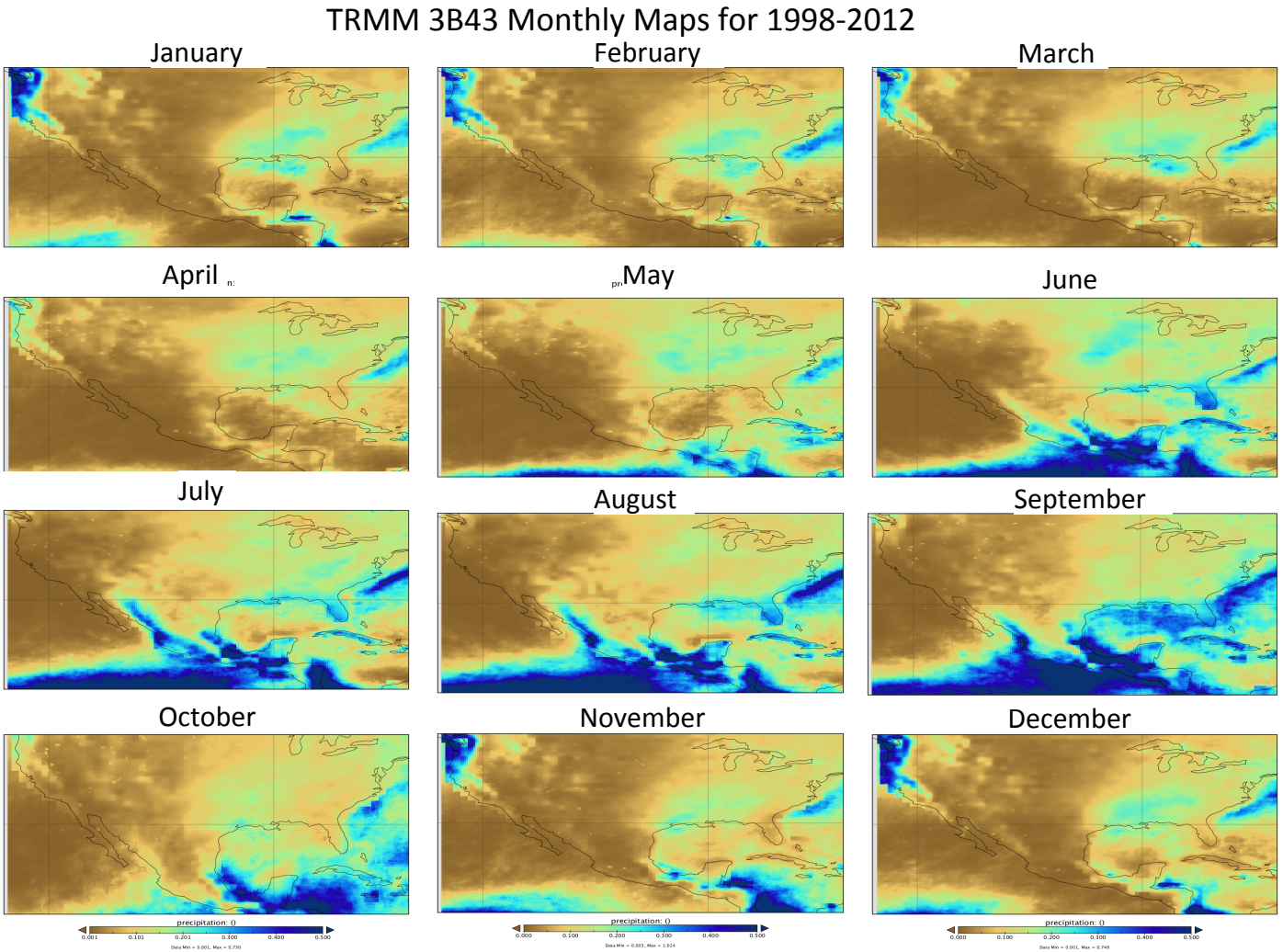
- More client-side viz.?
- Slider-comparison maps?
- Use Giovanni results as input to subsequent workflows?

Speed++

- Coarse-grained parallelism with make
- Cloud deployment of Giovanni
- MapReduce processing?

Domain-Specific Features

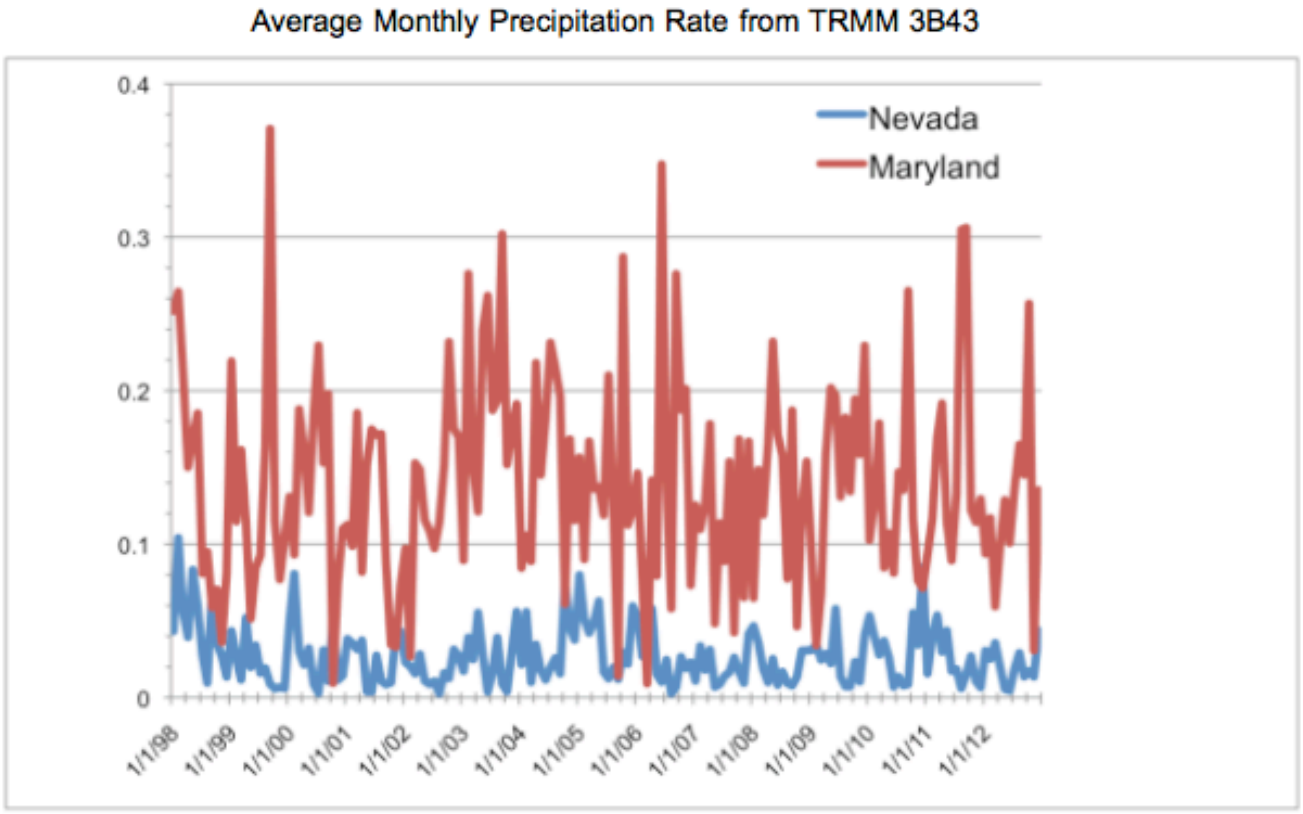
- Seasonal / Recurring-monthly analysis



http://giovanni.gsfc.nasa.gov/

Domain-Specific Features

- Seasonal / Recurring-monthly analysis
- Shapefile masking
  - Area-average calculations

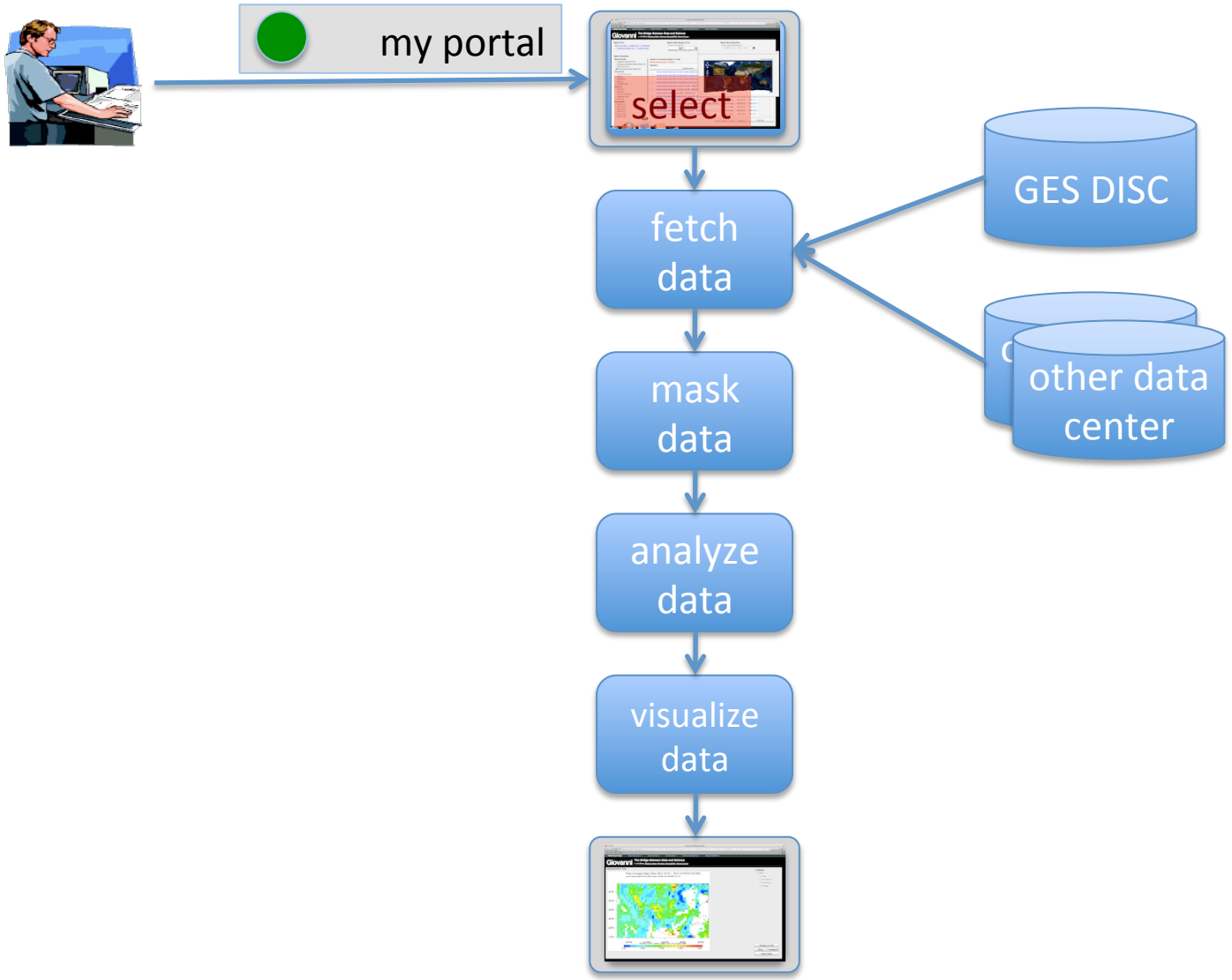


http://giovanni.gsfc.nasa.gov/

Domain-Specific Features

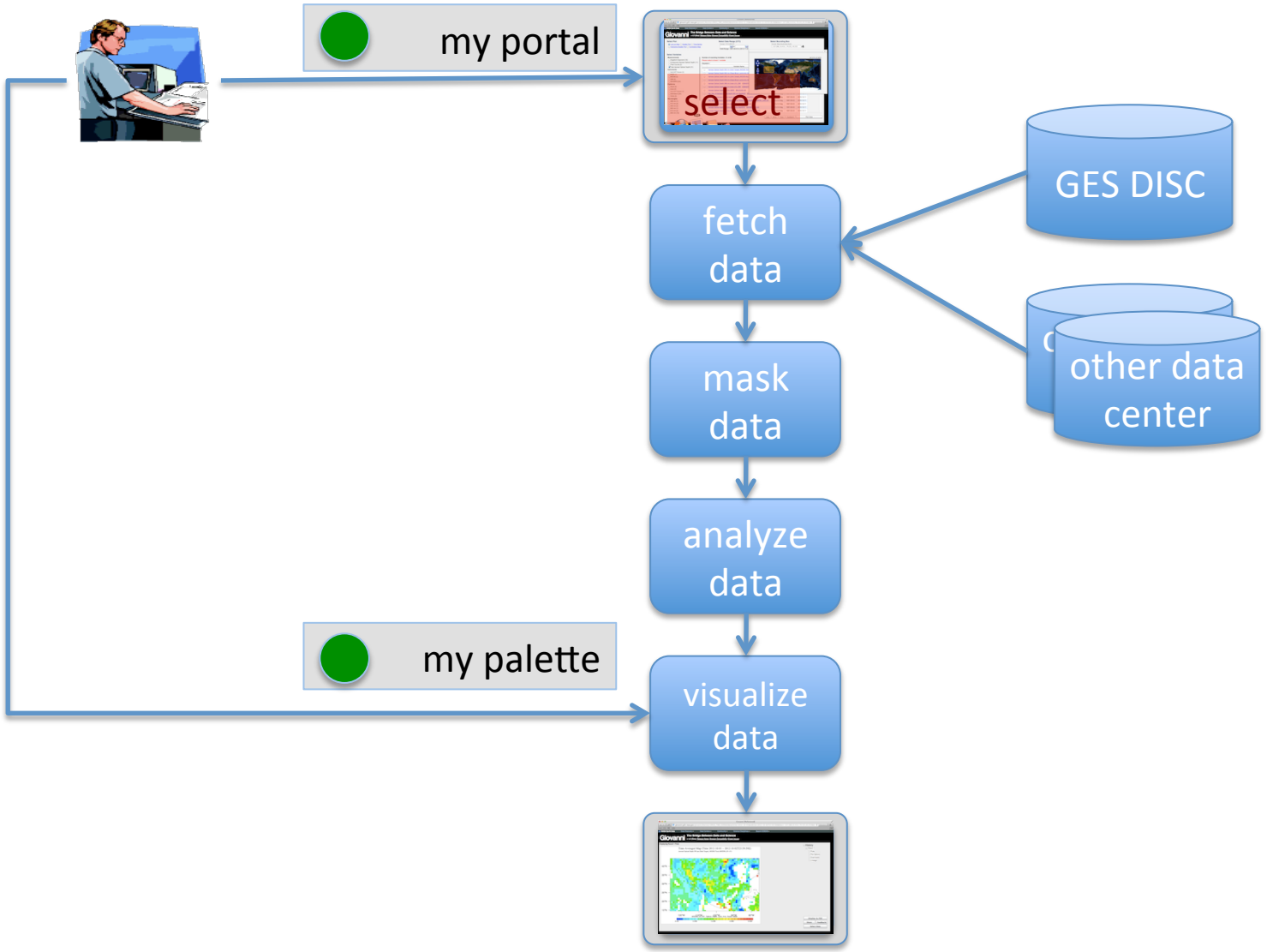
- Seasonal / Recurring-monthly analysis
- Shapefile masking
  - Area-average calculations
- Drill-down to Level 2 orbit/swath data?
- Contextual annotations
  - "The difference between V6 and V7 of TRMM 3B42 for 40-50 N and S during 1998-Feb. 2000 is that V6 results are based on the occasional microwave overpass, while V7 has both microwave and IR."

Participatory Design in myGiovanni



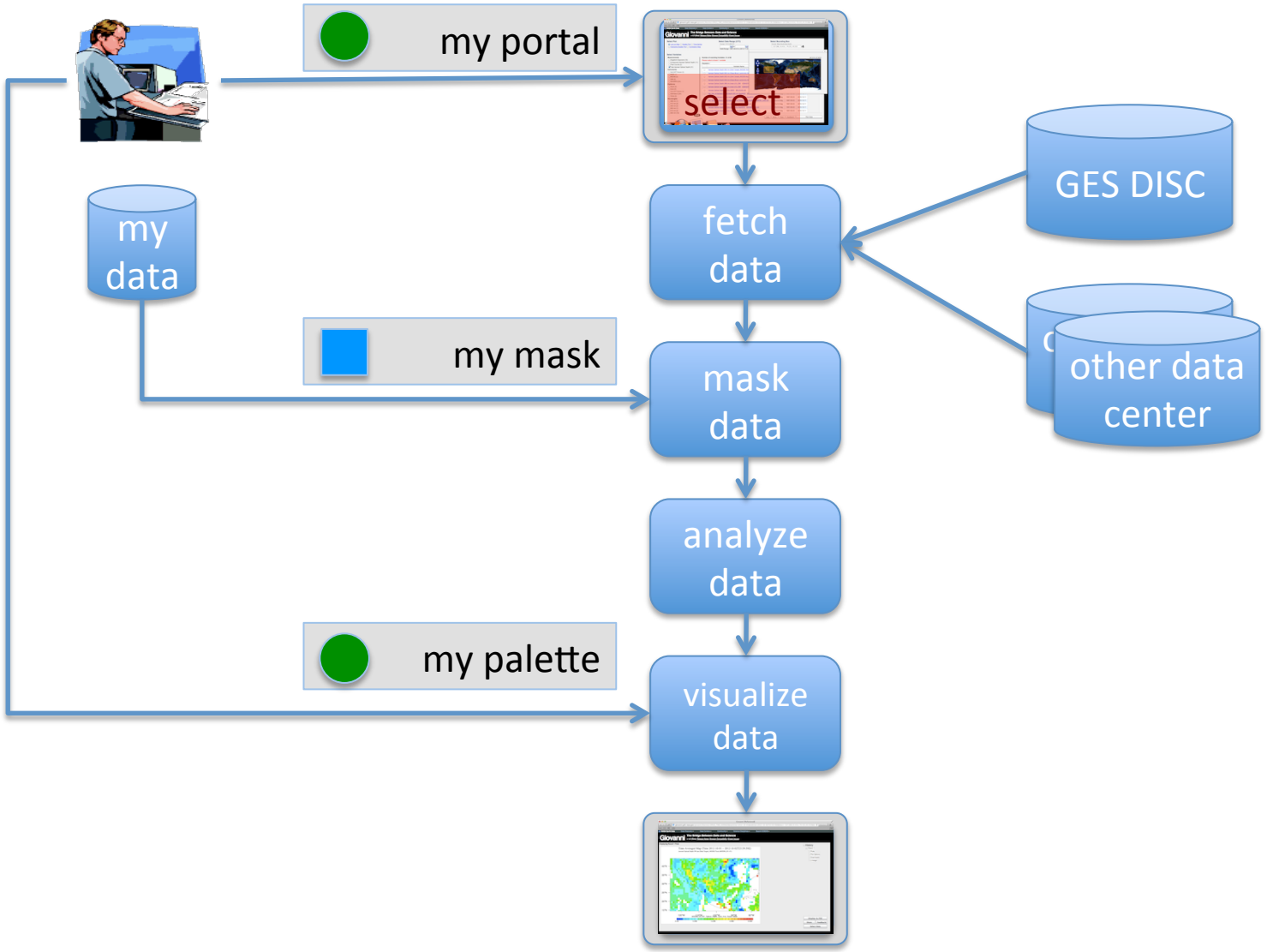
http://giovanni.gsfc.nasa.gov/

Participatory Design in myGiovanni



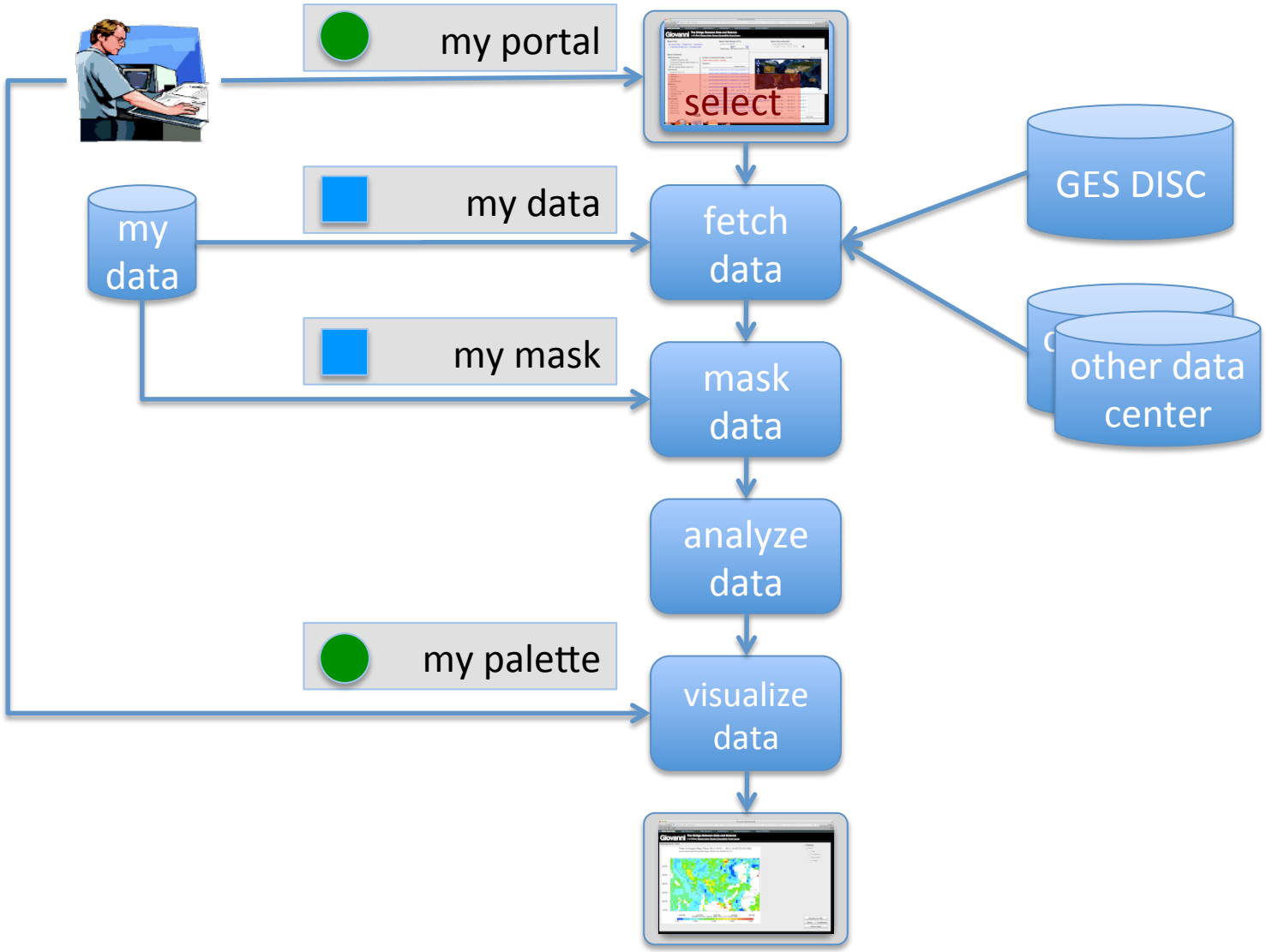
http://giovanni.gsfc.nasa.gov/

Participatory Design in myGiovanni



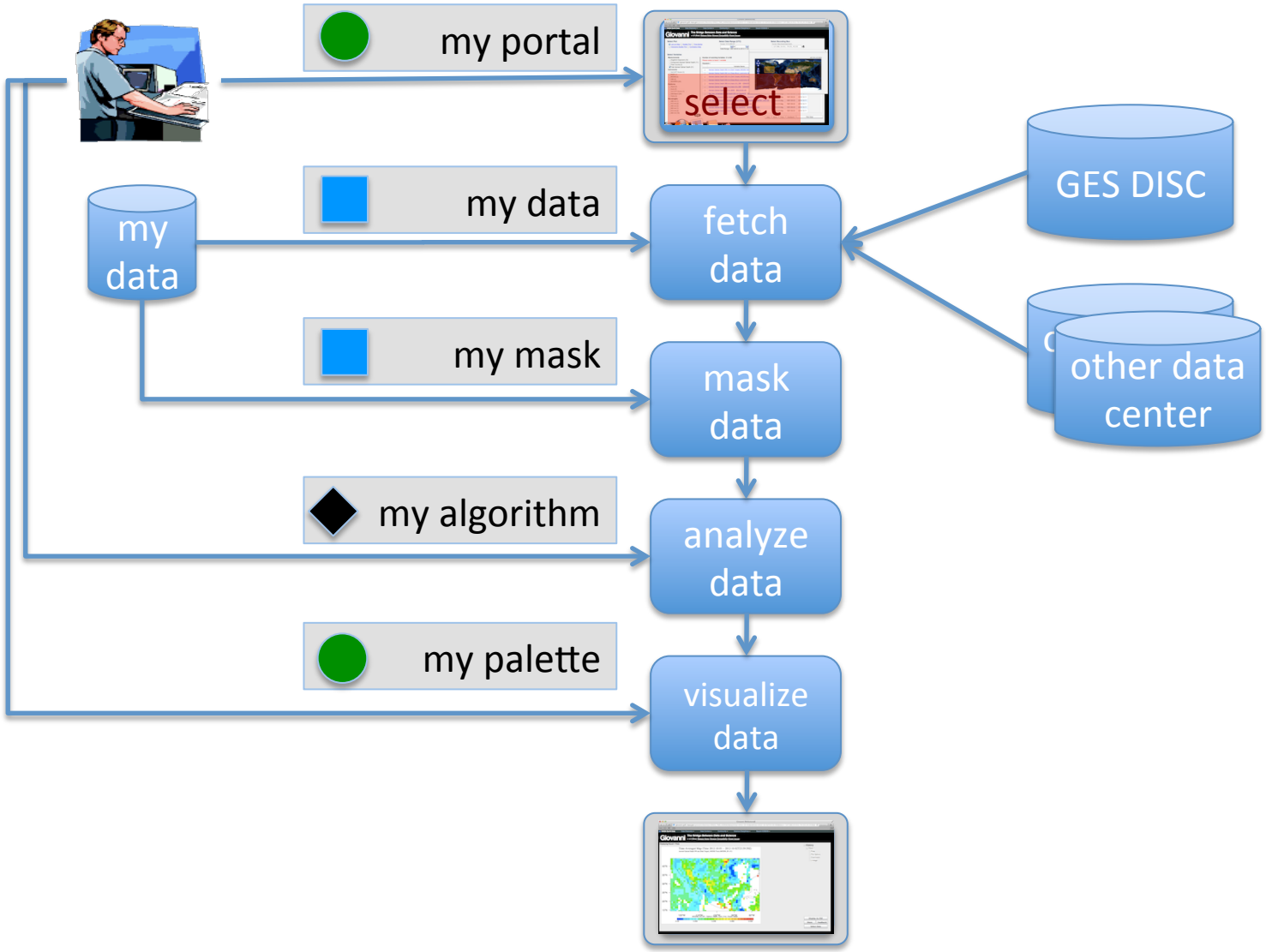
http://giovanni.gsfc.nasa.gov/

Participatory Design in myGiovanni



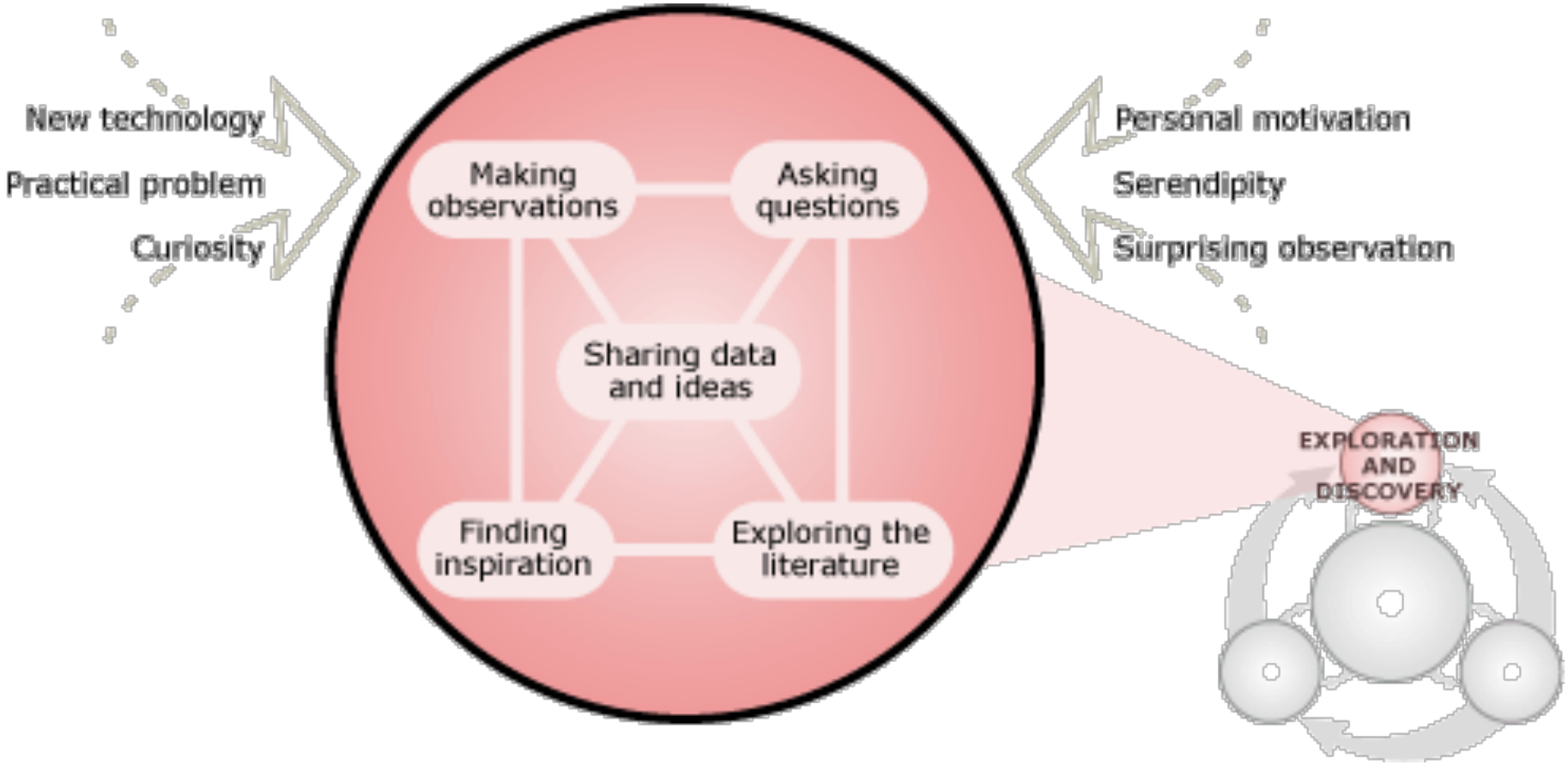
http://giovanni.gsfc.nasa.gov/

Participatory Design in myGiovanni

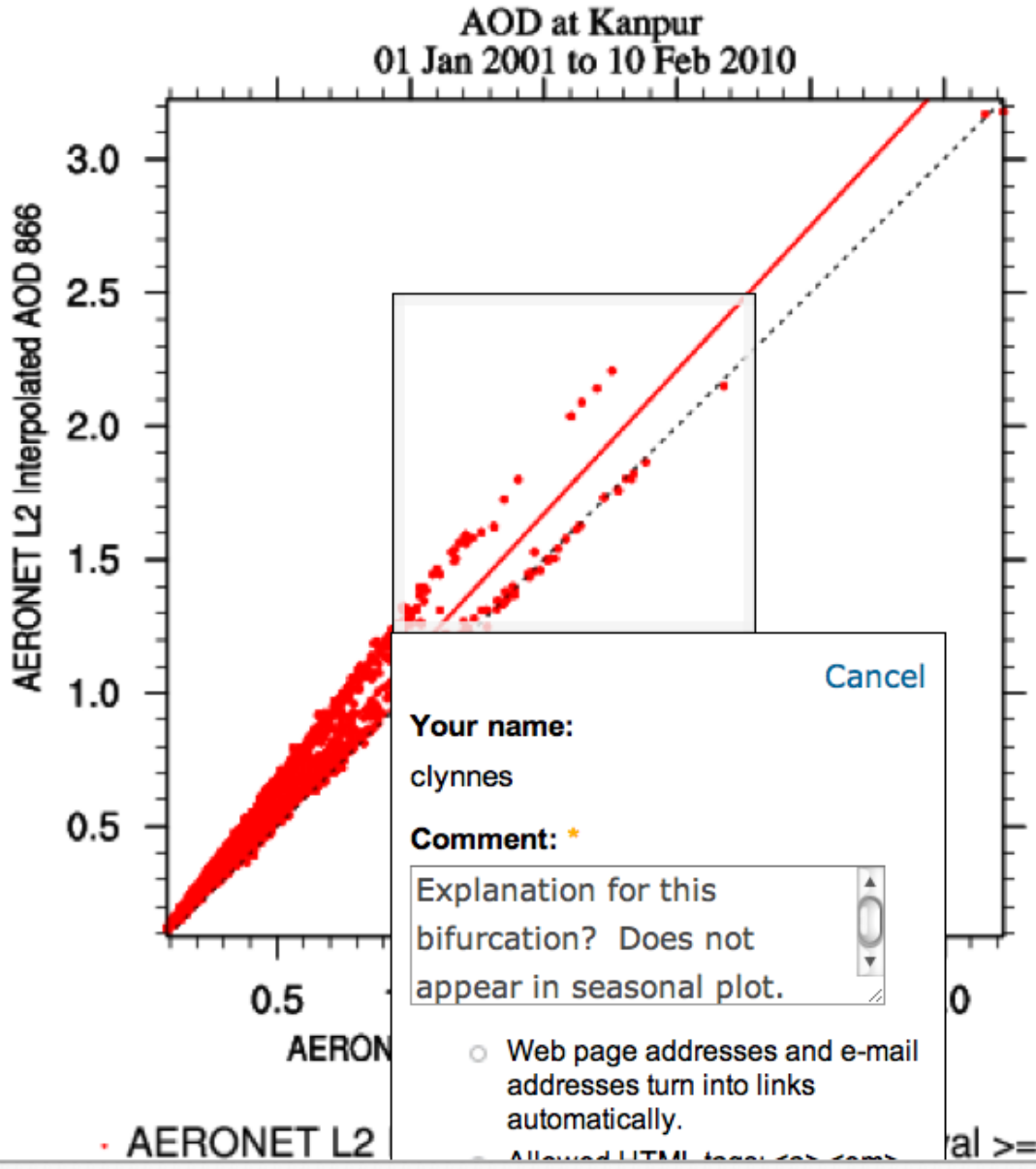


http://giovanni.gsfc.nasa.gov/

Collaboration in Exploration

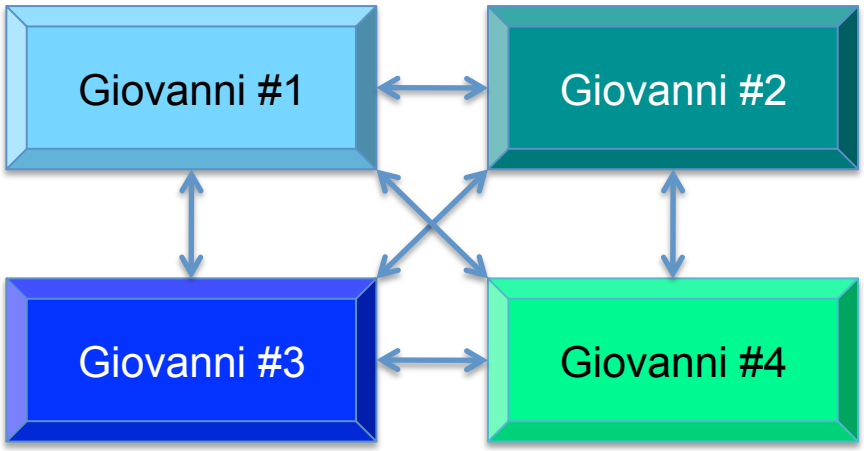


Collaboration with gSocial



Feature Detection

- Data exceeding a given threshold?
  - E.g., rain rate
  - Giovanni configuration value
  - User specified value
- Data patterns detected?
  - e.g., Cyclonic
  - User-contributed algorithms

Federated Giovanni

- Enlist other data and algorithm providers
- Add / share data
- Add / share algorithms
- Add / share Giovanni enhancements