

MINING DARK INFORMATION RESOURCES TO DEVELOP NEW INFORMATICS CAPABILITIES TO SUPPORT SCIENCE

Rahul Ramachandran NASA/MSFC, **Manil Maskey**
UAH, Kaylin Bugbee UAH

Project Team:

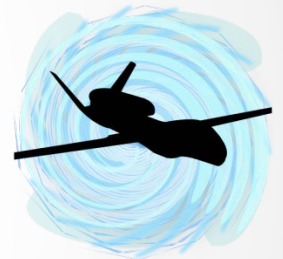
MSFC/UAH: Patrick Gatlin, Xiang Li, Amanda Weigel, JJ Miller,
Kaylin Bugbee, Ajinkya Kulkarni

GSFC: Steve Kempler, Suhung Shen, Chung-Lin Shie, Maksym
Petrenko

RPI: Peter Fox, Stefan Zednik, Anirudh Prabhu



European Geosciences Union General Assembly 2016
Vienna, Austria 17–22 April 2016



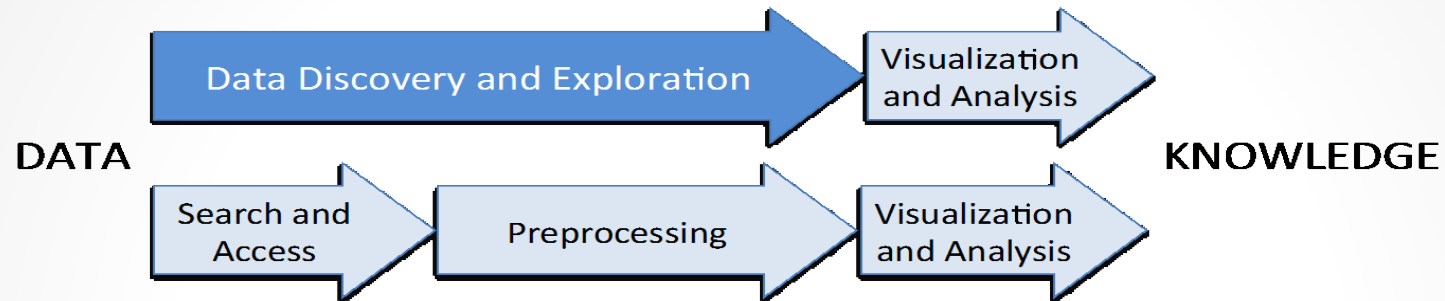
Outline

1. Project Overview
2. Data Curation Service
3. Rules Engine
4. Image Retrieval Service
5. Summary

Part 1: Project Overview

...

Motivation



- Data preparation steps are **cumbersome** and **time consuming**
 - Covers discovery, access and preprocessing
- Limitations of current Data/Information Systems
 - **Boolean search** on data based on instrument or geophysical or other **keywords**
 - Underlying **assumption** that users have sufficient knowledge of the **domain vocabulary**
 - **Lack support** for those **unfamiliar** with the domain vocabulary or the **breadth of relevant data** available

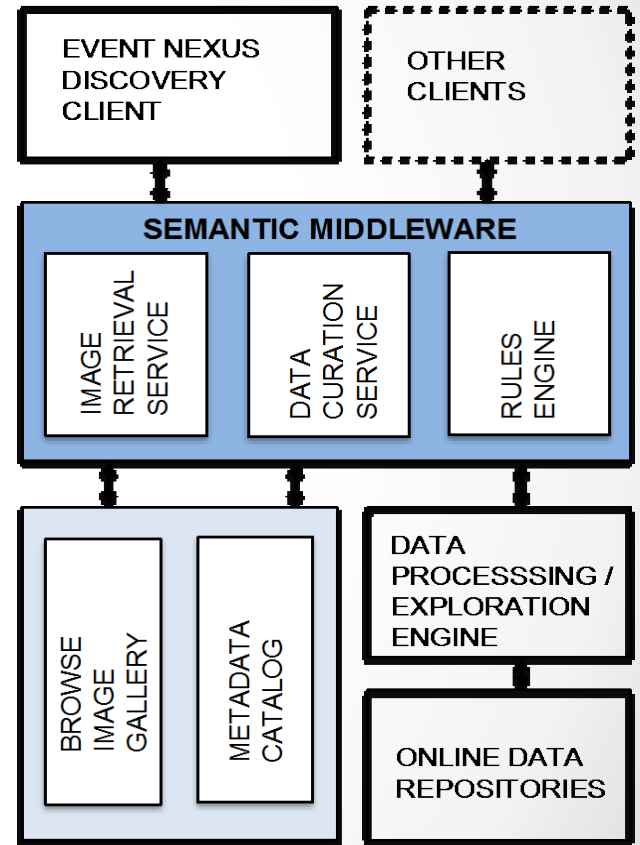
Earth Science Metadata: Dark Resources

- *Dark resources* - information resources that organizations collect, process, and store for regular business or operational activities but fail to utilize for **other** purposes
 - Challenge is to recognize, identify and effectively utilize these dark data stores
- Metadata catalogs contain dark resources consisting of structured information, free form descriptions of data and browse images.
 - NASA's Common Metadata Repository (ECHO) holds >6000 data collections, 270 million records for individual files and 67 million browse images.

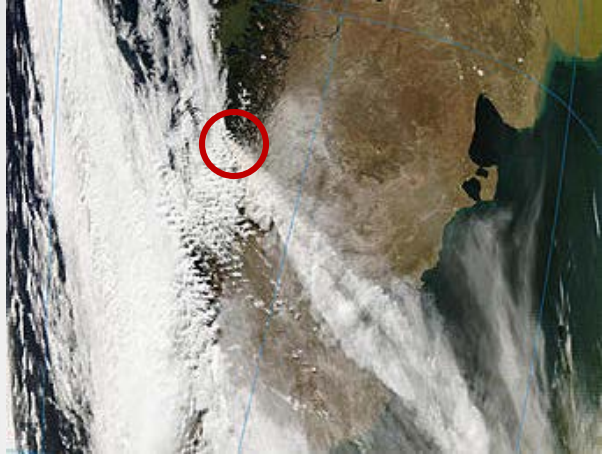
Premise: Metadata catalogs can be utilized *beyond their original design intent* to provide *new data discovery and exploration pathways* to support science and education communities.

Project Goals

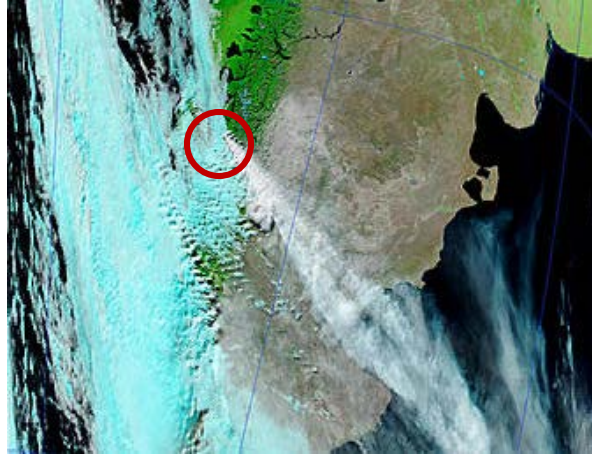
- Design a Semantic Middleware Layer (SML) to exploit these metadata resources
 - provide novel **data discovery and exploration** capabilities that significantly reduce data preparation time.
 - utilize a varied set of semantic web, information retrieval and image mining technologies.
- Design SML as a Service Oriented Architecture (SOA) to allow individual components to be used by existing systems



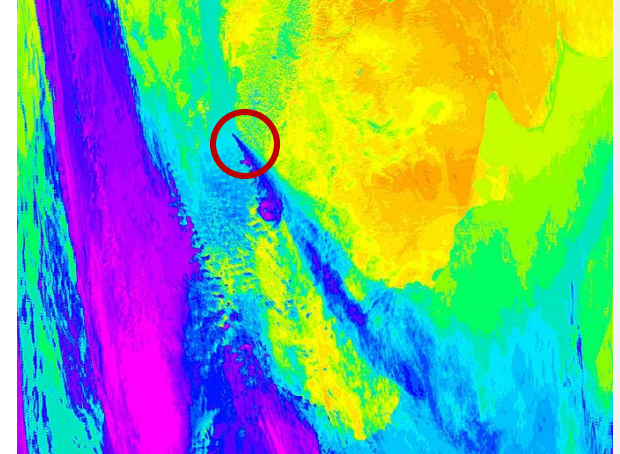
Use Case: *Find Interesting Events from Browse Images*



Band 1-4-3 (true color)



Band 7-2-1



LST

Example: MODIS-Aqua 2008-05-03 18:45 UTC

Chaitén Volcano Eruption

Eruption Time period: May 2 – Nov 2008

Location: Andes region, Chile (-42.832778, -72.645833)



Image Retrieval Service can be used to find volcanic ash events in browse imagery

Suggest Relevant Data

Total SO₂ mass:

e.g. **Chaitén** is 10 (kt) =(kilotons) , (1kt= 1000 metric tons)

ftp://measures.gsfc.nasa.gov/data/s4pa/SO2/MSVOLSO2L4.1/MSVOLSO2L4_v01-00-2014m1002.txt

Daily SO₂:

OMI/Aura Sulphur Dioxide (SO₂) Total Column Daily L2 Global 0.125 deg

http://disc.sci.gsfc.nasa.gov/datacollection/OMSO2G_V003.html

Calibrated Radiances:

MODIS/Aqua Calibrated Radiances 5-Min L1B Swath 1km

<http://dx.doi.org/10.5067/modis/myd021km.006>

Aerosol Optical Thickness:

MODIS/Aqua Aerosol 5-Min L2 Swath 10km

<http://modis-atmos.gsfc.nasa.gov/MOD08>

SeaWiFS Deep Blue Aerosol Optical Depth Data 13.5km

<http://disc.gsfc.nasa.gov/datacollection>

Data Curation Service
recommends relevant
datasets to support event
analysis

IR Brightness Temperature:

NCEP/CPC 4-km Global (60 deg N - 60 deg S) Merged IR Brightness Temperature Dataset

Generate Giovanni SO2 Plots

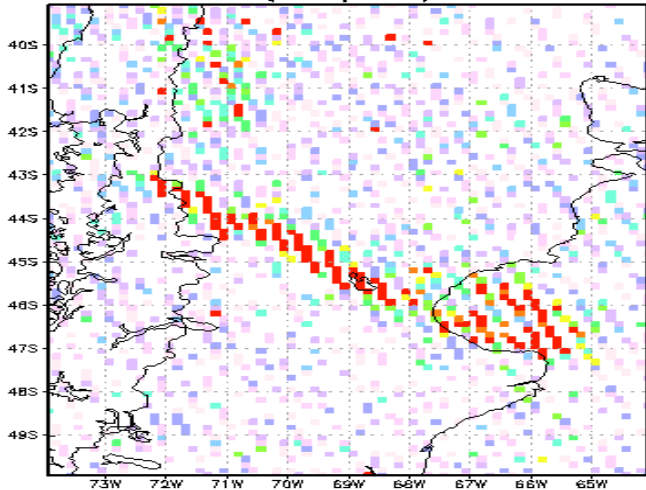
MODIS-Aqua 2008-05-03 18:45 UTC



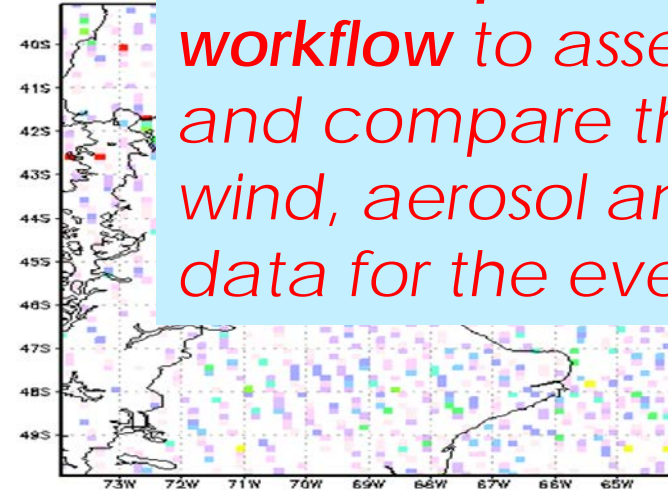
MODIS-Aqua 2008-05-05 18:30 UTC



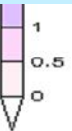
2G.003 SO2 Column Amount (Planetary Boundary Layer) (03May2008)



2G.003 SO2



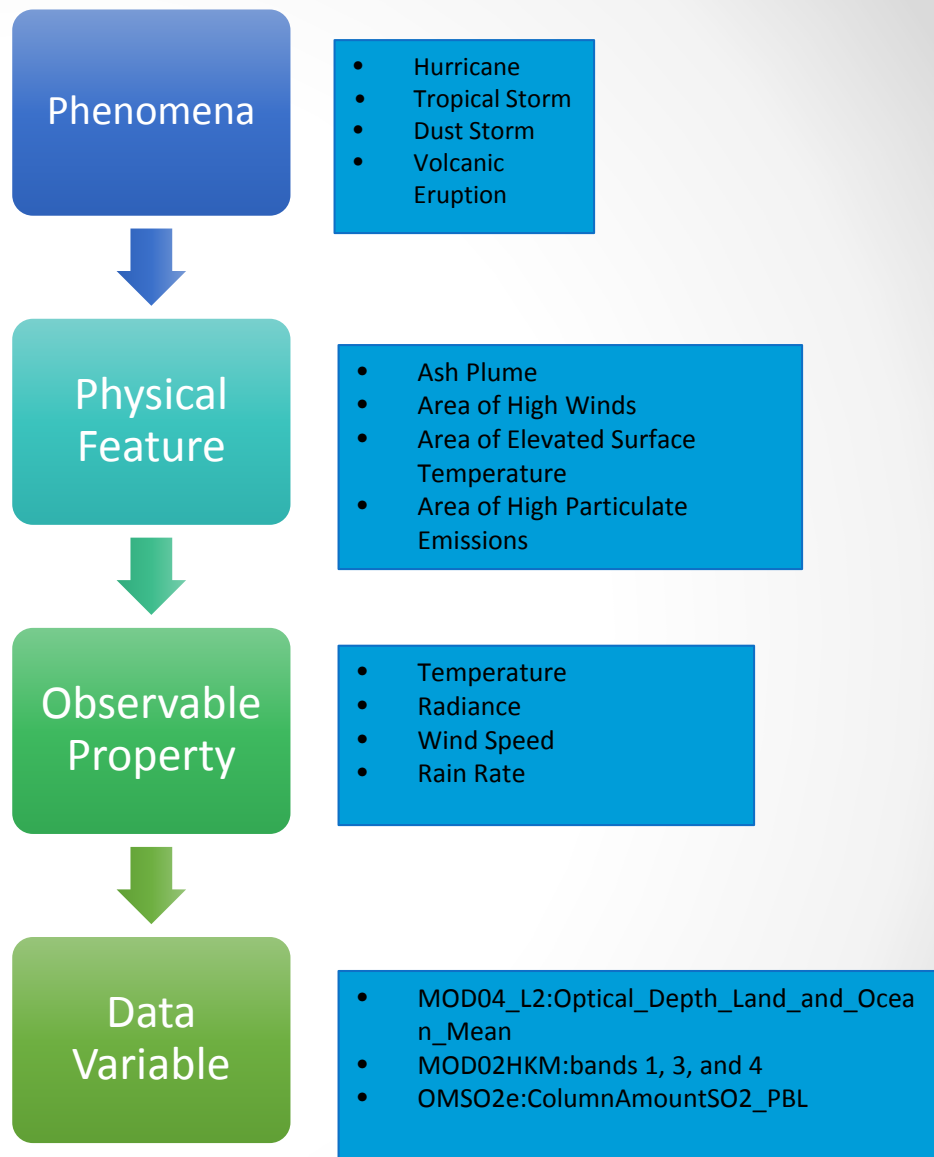
Rules Engine invokes a Giovanni processing workflow to assemble and compare the wind, aerosol and SO2 data for the event



http://gdata2.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=omil2g

Conceptual Model

- **Phenomena**
 - Event type
- **Physical Feature**
 - Manifestation / Driver of phenomena
 - Has space/time extent
 - Can precede or linger after what is generally thought of as the phenomena event
- **Observable Property**
 - Characteristic/property of physical feature
- **Data Variable**
 - Measurement/estimation of observable feature



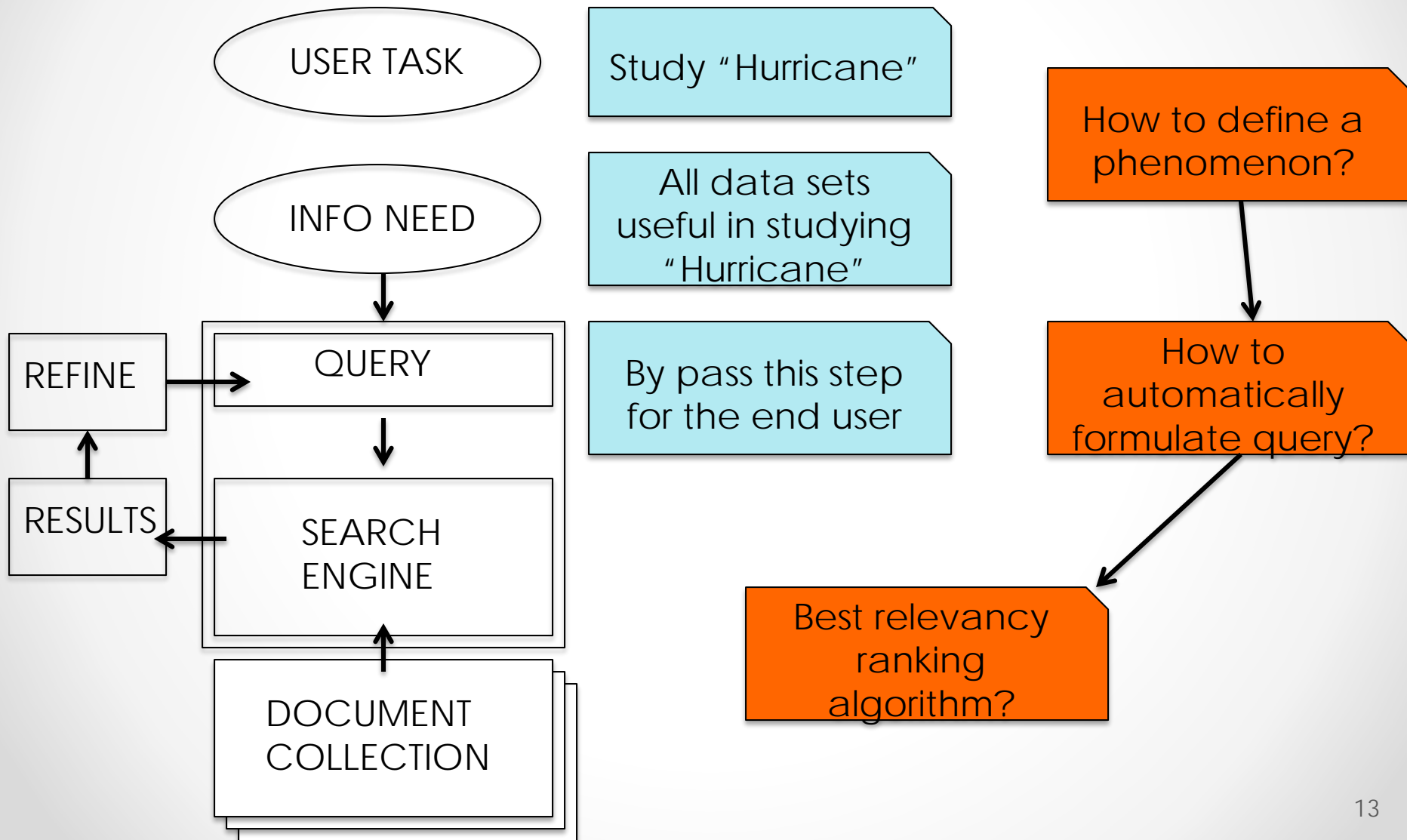
Part 2: Data Curation Algorithm for Phenomena

...

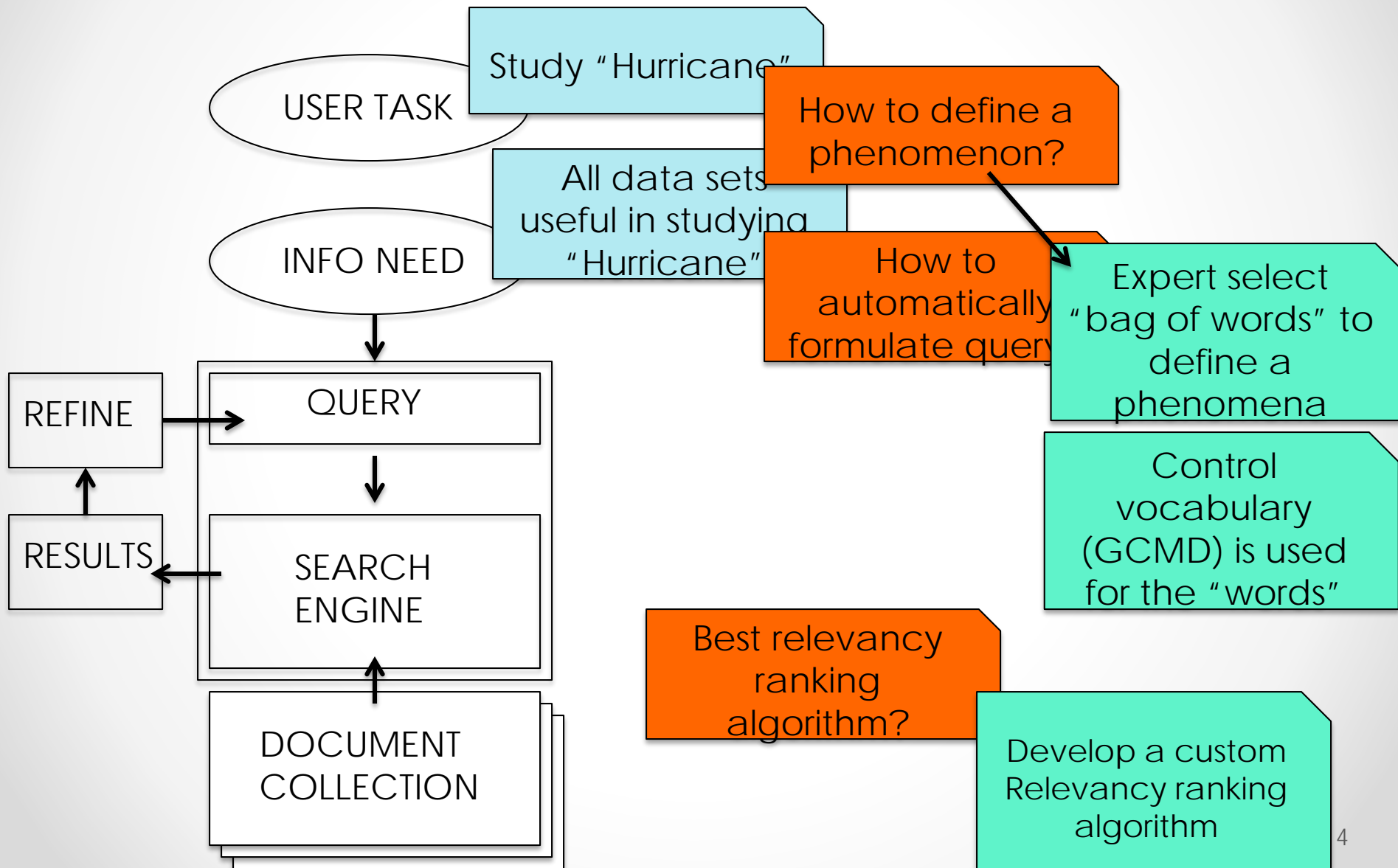
Objectives

- Design a data curation (relevancy ranking) algorithm for a set of **phenomena**
- Provide the data curation algorithm as a stand alone service
- Envisioned Use:
 - Given a phenomenon type (Ex: Hurricane), DCS returns a list of relevant data sets (variables)
 - $\langle \text{list of data sets (variables)} \rangle = \text{DCS}(\text{Phenomenon Type})$
 - For a specific phenomenon instance (event: Hurricane Katrina), these curated datasets can be filtered based on space/time to get actual granules

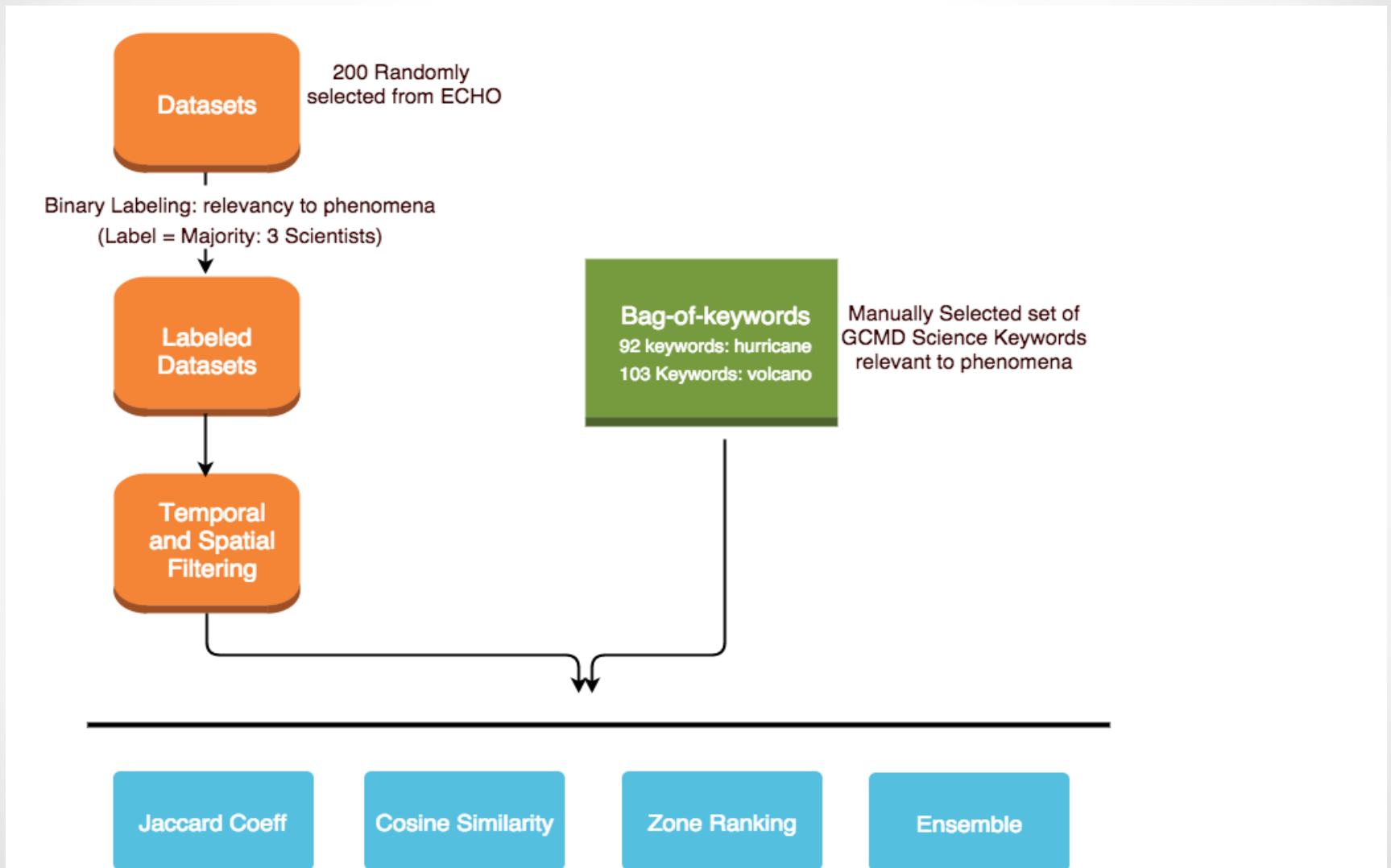
Data Curation is a Specialized Search Problem



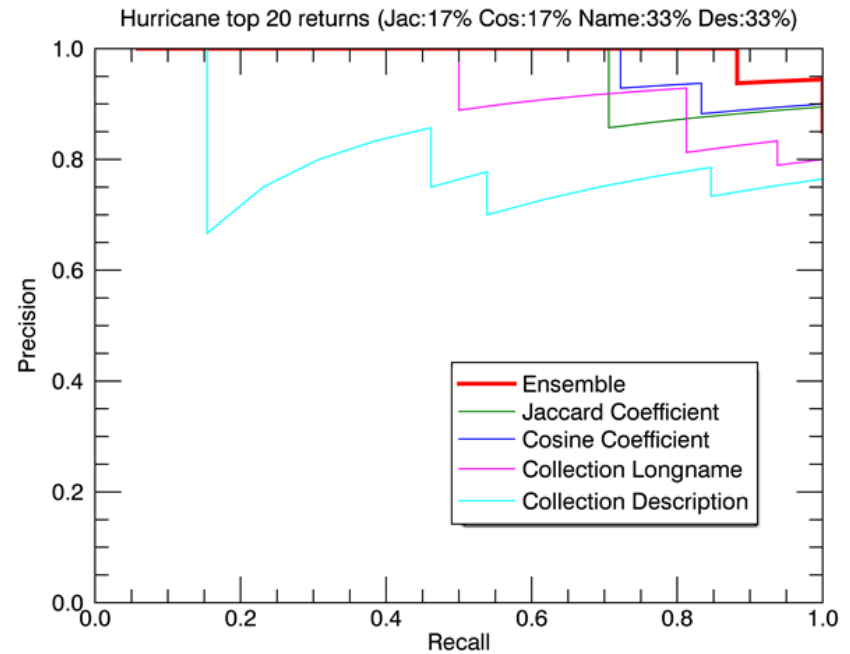
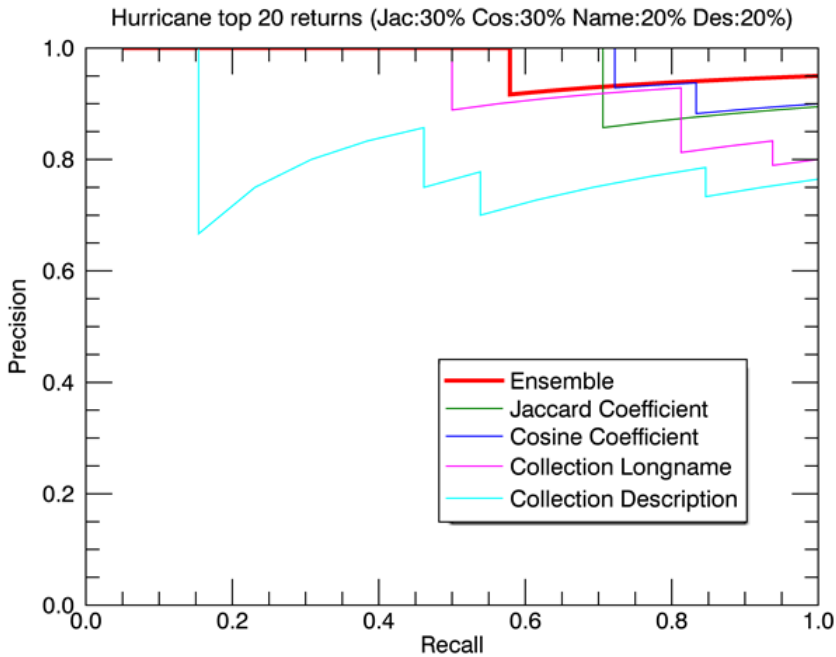
Our Approach



Experiment Setup



Top 20 returns (Hurricane)



Next: Find relevant data fields

- Need actual data variables
 - Example: Giovanni uses these fields for visualization
- What we know
 - Data set (Collection) level science keywords (GCMD) – Experts
 - Granule data fields and metadata – Auto extract*
- How do we map?
 - Start with GCMD to CF Standard name
 - Most don't follow CF Standard names

Approach

Dataset

Extract Science
Keywords

Text processing

Normalization

Bag-of-words

Granules

Extract Variables
and Descriptions

Text processing

Look up Table

Normalization

Bag-of-words

OPeNDAP, netCDF Libs, ...

Remove special characters, Tokenize,
...

Acronym/Abbreviation expansion, CF

Remove stopwords/Stem/Lemmatize

Intersection

NLP
Learn Patterns

Suggest Keywords
Assess Metadata

Example: GLAS/ICESat L2 Global Thin Cloud/Aerosol Optical Depths Data (HDF5) V033 – Dataset Metadata

EARTHDATA β Temporal Spatial Feedback

[Back to Granules](#)

GLAS/ICESat L2 Global Thin Cloud/Aerosol Optical Depths Data (HDF5) V033 VERSION 33

John.P.Dimarzio.1@nasa.gov
ICESat Science Investigator-led Processing System (I-SIPS)
757-864-1238 (phone)
David.W.Hancock@nasa.gov

NASA DAAC at the National Snow and Ice Data Center
303-492-6199 (phone)
303-492-2468 (fax)
nsidc@nsidc.org

Science Keywords:
Earth Science Atmosphere Clouds
Earth Science Atmosphere Aerosols

MONTH
+ **GLAS/ICESat L2 Global Thin Cloud/Aerosol Optical Depths Data (HDF5) V033**
- Aug Sep Oct Nov Dec Jan 2015 Feb Mar Apr May Jun

Example: GLAS/ICESat L2 Global Thin Cloud/Aerosol Optical Depths Data (HDF5) V033

Sample file:
GLAH11.033/2006.10.25/GLAH11_633_2117_001_1275_0_01_0001.H5

Data Variables

The screenshot shows the HDFView 2.10.1 interface. The 'Recent Files' list contains the file 'arkDataAll-ECHO-hurricane-datasets\hurricane_top_30_datagranules\extra_copy_GLAS_ICESat_granule\GLAH11_633_2117_001_1275_0_01_0001.H5'. The file tree on the left is expanded to show the following structure:

- GLAH11_633_2117_001_1275_0_01_0001.H5
 - ANCILLARY_DATA
 - BROWSE
 - Data_1HZ
 - Angle
 - DS_Cloud_Layer_10
 - DS_UTCTime_1
 - Flags
 - Geolocation
 - Geophysical
 - r_Surface_pres
 - r_Surface_relh
 - r_Surface_temp
 - r_Surface_wdir
 - r_Surface_wind
 - r_cld1_grd_det
 - OD1064CloudLayers
 - OD532CloudLayer
 - i_cld1_qf
 - i_cld1_uf
 - r_MRg_cldbot_pres
 - r_MRg_cldbot_relh
 - r_MRg_cldbot_temp
 - r_MRg_cldtop_pres
 - r_MRg_cldtop_relh
 - r_MRg_cldtop_temp
 - r_cld1_bot
 - r_cld1_msf
 - r_cld1_od
 - r_cld1_top

Example: GLASICESat L2 Global Thin Cloud Aerosol Optical Depths Data (HDF5) V033

Science keyword to variable mapping

- r_Surface_relh | Surface Relative Humidity
 - No match
- r_Surface_temp | Surface Temperature
 - No match
- r_Surface_wind | Surface Wind Speed
 - No match
- r_cld1_od | Cloud Optical Depth at 532 nm
 - Score=3 keyword: ATMOSPHERE->CLOUDS->CLOUD OPTICAL DEPTH/THICKNESS
 - Score=2 keyword: ATMOSPHERE->AEROSOLS->AEROSOL OPTICAL DEPTH/THICKNESS

Variable to keyword mapping

- ATMOSPHERE->CLOUDS->CLOUD OPTICAL DEPTH/THICKNESS
 - Score=3 name: r_cld_ir_OD | Cloud Optical Depth at 1064 nm
 - score=3 name: i_cld1_qf | Cloud optical depth flag for 532 nm
 - Score=3 name: i_cld1_uf | Cloud optical depth flag for 532 nm
 - Score=3 name: r_cld1_od | Cloud Optical Depth at 532 nm

- *Found incorrect/incomplete keyword annotation*
- *Can be used **assess metadata quality and suggest keyword annotation!!***

Parameter Mapping Tool

The screenshot shows the 'Data Parameter Mapping Tool' interface. At the top, there's a browser address bar with '54.172.157.10:5000'. Below it, a navigation bar contains various folders like 'personal', 'Mendeley', 'GKeep', 'NASA', 'Demo', 'H53', 'GHRC', 'DarkData', 'nspires', 'RResp', and 'Unisys Weather - G...'. The main header is 'Data Parameter Mapping Tool'. Underneath, there's a 'Datasets' section with a list of datasets including 'AIRS/Aqua Level 2 Support retrieval (AIRS+AMSU) V005', 'GHRSSST Level 2P USA NASA MODIS Aqua SST:1', 'MODIS/Terra Temperature and Water Vapor Profiles 5-Min L2 Swath 5km V005', 'LIS/OTD 2.5 DEGREE LOW RESOLUTION DIURNAL CLIMATOLOGY (LRDC) V2.3.2013', and 'MODIS/Terra Aerosol 5-Min L2 Swath 10km V005 NRT'. A green label 'Datasets' points to this list. Below the datasets, a specific dataset 'MODIS/Terra Aerosol 5-Min L2 Swath 10km V005 NRT' is selected. The interface is split into two main panels: 'Science Keyword Map' and 'Parameter Map', both with 'EDIT' buttons. The 'Science Keyword Map' shows a tree view with 'ATMOSPHERE' expanded to 'AEROSOLS' and 'PARTICULATE_MATTER' selected. A green label 'Science Keyword' points to this tree. The 'Parameter Map' shows a list of parameters with their scores. A green label 'Parameter' points to 'Optical_Depth_Small_Average_Ocean' with a score of 3. A green label 'Mapped Science Keywords' points to 'Optical_Depth_Small_Average_Ocean' with a score of 0. A green label 'Mapped Parameters' points to 'Aerosol_Type_Land' with a score of 1.

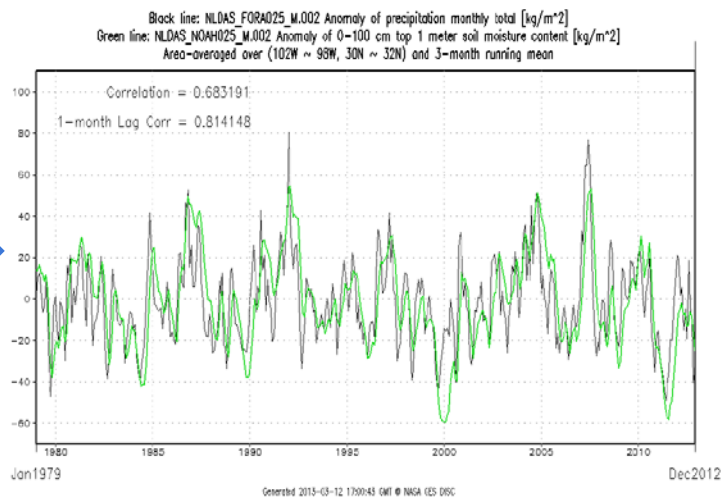
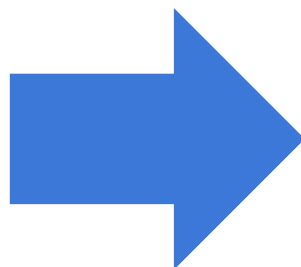
This close-up shows a dropdown menu with a list of atmospheric parameters. The parameters include: 'ATMOSPHERE->AEROSOLS->AEROSOL_PARTICLE_PROPERTIES : 2', 'ATMOSPHERE->AEROSOLS->CLOUD_CONDENSATION_NUCLEI : 2', 'ATMOSPHERE->AEROSOLS->AEROSOL_EXTINCTION : 2', 'ATMOSPHERE->AEROSOLS->AEROSOLS_OPTICAL_DEPTH/THICKNESS : 2', 'ATMOSPHERE->AEROSOLS->AEROSOL_RADIANC : 2', 'ATMOSPHERE->AEROSOLS->CARBONACEOUS_AEROSOLS : 2', 'ATMOSPHERE->AEROSOLS->DUST/ASH/SMOKE : 2', 'ATMOSPHERE->AEROSOLS->NITRATE_PARTICLES : 2', 'ATMOSPHERE->AEROSOLS->ORGANIC_PARTICLES : 2', 'ATMOSPHERE->AEROSOLS->PARTICULATE_MATTER : 2' (highlighted in blue), 'ATMOSPHERE->AEROSOLS->SULFATE_PARTICLES : 2', 'ATMOSPHERE->ATMOSPHERIC_RADIATION->RADIATIVE_FLUX : 2', 'ATMOSPHERE->ATMOSPHERIC_RADIATION->REFLECTANCE : 2', and 'ATMOSPHERE->ATMOSPHERIC_RADIATION->OPTICAL_DEPTH/THICKNESS : 2' (checked with a green checkmark). Each parameter has a 'Remove' button to its right.

Edit/Save Mapping

Mapping Scores Generated by Algorithm

Part 3: Rules Engine

What settings should I use to visualize this event?



Data
Variable
?

Dataset
?
Visualization
Type?



Goal: Automate data preprocessing and exploratory analysis and visualization tasks

Strategy

- Service to generate and rank candidate workflow configurations
- Use rules to make **assertions** about **compatibility based on multiple factors**
 - does this data variable make sense for this feature?
 - does this visualization type make sense for this feature?
 - does the temporal / spatial resolution of this dataset make sense for this feature?
- Each compatibility assertion type is assigned weights.
 - ex: Strong = 5, Some = 3, Slight = 1, Indifferent = 0, Negative = -1.
- Based on the aggregated compatibility assertions, we calculate the score for each visualization candidate.

Ruleset Development

Survey asked users to rate characteristics of phenomena features

Feature characteristics for analysis *

What characteristics are of interest when analyzing the feature?

	negative value	indifferent	slight value	some value	strong value
east-west movement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
north-south movement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
temporal evolution	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
spatial extent of event	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
year-to-year variability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
may impact seasonal variation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
variation with atmospheric height	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
global phenomena	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
detection of events	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Survey results used to formulate rules

[rule1:

```
(?feature rdf:type  
dd:AshPlume)
```

```
->
```

```
(?feature  
dd:strongCompatibilityFor  
dd:temporal_evolution),
```

```
(?feature  
dd:indifferentCompatibilityFor  
dd:east-west-movement),
```

```
...
```

```
]
```

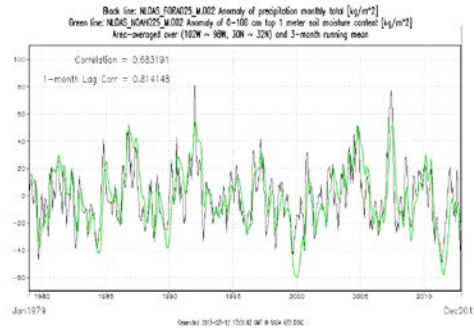
Phenomena Feature Characteristic Mappings

Phenomena	East-West Movement	North-South Movement	Temporal Evolution	Spatial Extent of Event	Year-to-Year Variability	May Impact Seasonal Variation	Variation with Atmospheric Height	Global Phenomena	Detection of Events
Volcano - Ash Plume	Indifferent	Indifferent	Strong	Slight	Strong	Strong	Strong	Strong	Strong
Flood	Some	Some	Strong	Some	Some	Strong	Some	Slight	Some
Dust Storm	Strong	Strong	Strong	Strong	Indifferent	Indifferent	Strong	Indifferent	Some

Service to Characteristic Mappings

Service	Visualization	East-West Movement	North-South Movement	Temporal Evolution	Spatial Extent of Event	Year-to-Year Variability	Seasonal Variation	Variation with Atmospheric Height	Global Phenomena	Detection of Events
Time-averaged Map	Color-Slice Map				✓					
Area-averaged Time Series	Time Series			✓						✓
User-defined Climatology	Color-Slice Map						✓			
Vertical Profile	Line Plot							✓		
Seasonal Time Series	Time Series					✓				
Zonal Means	Line Plot								✓	
Hovmoller (Longitude)	Color-Slice Grid	✓								
Hovmoller (Latitude)	Color-Slice Grid		✓							

Compute Compatibility



Phenomena:
 Volcano - Ash
 Plume

Service - Area
 Averaged Time
 Series

**STRONG
 COMPATIBILITY
 x2**

Temporal Evolution	Detection of Events
Strong	Strong

Area Averaged Time Series : bestFor →	Temporal evolution; Detection of events
---------------------------------------	---

Images from , http://disc.sci.gsfc.nasa.gov/datareleases/images/nldas_monthly_climatology_figure_9.gif, <http://www.clipartbest.com/cliparts/biy/bAX/biybAXGIL.png>

volcanic ash image - By Boaworm (Own work) [CC BY 3.0 (<http://creativecommons.org/licenses/by/3.0/>)], via Wikimedia Commons

Integrating Services in Giovanni

- **Tool:** Giovanni is a popular on-line environment that lets users discover, plot, and download a number of geophysical parameters (data variables)
- **Goal:** Leverage Dark Data services and technologies to assist Giovanni users in discovering and exploring data

'Success will be realized when Giovanni requests can be automatically invoked with the appropriate spatial and temporal extents, variables and workflow / visualization type for a particular event'

Giovanni – Standard Edition

The screenshot shows the GIOVANNI web interface with the following elements:

- Navigation:** EARTHDATA, Data Discovery, DAACs, Community, Science Disciplines.
- Header:** GIOVANNI The Bridge Between Data and Science v 4.17.2 [Release Notes](#) [Browser Compatibility](#) [Known Issues](#)
- Select Plot:** Maps: Time Averaged Map (selected), Comparisons: Select..., Time Series: Select..., Vertical: Select..., Miscellaneous: Select...
- Select Date Range (UTC):** YYYY-MM-DD HH:mm to YYYY-MM-DD HH:mm. Valid Range: 1979-01-01 to 2016-02-04. Note: Please specify a start date.
- Select Region (Bounding Box or Center):** Format: West, South, East, North. Example: -180, -90, 180, 90.
- Select Variables:**
 - Disciplines:** Aerosols (122), Atmospheric Chemistry (37), Atmospheric Dynamics (144), Cryosphere (5), Hydrology (369), Ocean Biology (11), Oceanography (8), Water and Energy Cycle (391).
 - Measurements:** Aerosol Index (3), Air Pressure (24), Air Temperature (39), Albedo (11), Altitude (4), Angstrom Exponent (16), Atmospheric Moisture (42), Buoyancy (1), CH4 (8), CO (8), CO2 (2), Canopy Water Storage (3), Chlorophyll (2).
- Number of matching Variables:** 721 of 975. Total Variable(s) included: 0.
- Keyword:** [Empty field]
- Table of Variables:**

Variable Name	Source	Resolution	Frequency	Latitude	Start Date	End Date
<input type="checkbox"/> Aerosol Angstrom Exponent 550/865 nm (Dark Target, Ocean-only) (MOD08_D3 v051)	MODIS-Terra	1°	Daily	1°	2000-03-01	2016-02-01
<input type="checkbox"/> Aerosol Angstrom Exponent 470/660 nm (Dark Target, Land-only) (MOD08_D3 v051)	MODIS-Terra	1°	Daily	1°	2000-03-01	2016-02-01
<input type="checkbox"/> Aerosol Optical Depth 550 nm (Dark Target) (MOD08_D3 v051)	MODIS-Terra	1°	Daily	1°	2000-03-01	2016-02-01
<input type="checkbox"/> Pixel Count of Aerosol Optical Depth 550 nm (Dark Target) (MOD08_D3 v051)	MODIS-Terra	1°	Daily	1°	2000-03-01	2016-02-01
<input type="checkbox"/> Aerosol Optical Depth 550 nm (Deep Blue, Land-only) (MOD08_D3 v051)	MODIS-Terra	1°	Daily	1°	2000-03-01	2007-12-31
<input type="checkbox"/> Aerosol Angstrom Exponent 550/865 nm (Dark Target, Ocean-only) (MYD08_D3 v051)	MODIS-Aqua	1°	Daily	1°	2002-07-04	2015-09-30
<input type="checkbox"/> Aerosol Angstrom Exponent 470/660 nm (Dark Target, Land-only) (MYD08_D3 v051)	MODIS-Aqua	1°	Daily	1°	2002-07-04	2015-09-30
<input type="checkbox"/> Aerosol Optical Depth 550 nm (Dark Target) (MYD08_D3 v051)	MODIS-Aqua	1°	Daily	1°	2002-07-04	2015-09-30
<input type="checkbox"/> Pixel Count of Aerosol Optical Depth 550 nm (Dark Target) (MYD08_D3 v051)	MODIS-Aqua	1°	Daily	1°	2002-07-04	2015-09-30
<input type="checkbox"/> Aerosol Optical Depth 550 nm (Deep Blue, Land-only) (MYD08_D3 v051)	MODIS-Aqua	1°	Daily	1°	2002-07-04	2015-09-30
- Vertical Choices:**
 - Cross Map, Latitude-Pressure
Cross Map, Latitude-Pressure
[Details...](#)
 - Cross Map, Longitude-Pressure
Cross Map, Longitude-Pressure
[Details...](#)
 - Cross Map, Time-Pressure
Cross Map, Time-Pressure
[Details...](#)
 - Vertical Profile
Vertical Profile
[Details...](#)
- Buttons:** Help, Reset, Feedback, Plot Data.

User needs to decide:

- Variable(s)
- Time
- Space
- Plot type

<http://giovanni.sci.gsfc.nasa.gov/giovanni/>

Giovanni – Dark Data Edition

Selected event & its time Event Client

Rules Service:
highlights
suitable plots
based on
selected event
& variables

Curation
Service: event
type filters
relevant
variables

The screenshot displays the GIOVANNI web interface with the following elements:

- Header:** EARTHDATA, Data Discovery, DAACs, Community, Science Disciplines, GIOVANNI The Bridge Between Data and Science v 4.18, Release Notes, Browser Compatibility, Known Issues.
- Select Plot:** Maps: Select..., Comparisons: Select..., Vertical: Select..., Time Series: Area-Averaged (highlighted), Miscellaneous: Select...
- Select Date Range (UTC):** YYYY-MM-DD HH:mm to YYYY-MM-DD HH:mm. Range: 2015-07-31 00:00 to 2015-11-24 23:59. Valid Range: 2000-03-01 to 2016-01-19.
- Select Region (Bounding Box or Shapefile or Event):** Format: West, South, East, North. Volcanoes: Manam Volcano. Buttons: Show Map, Show Shapes, Show Events.
- Select Variables:**
 - Events (all): Hurricane (14), Volcano (14)
 - Events (by products): Hurricane (2), Volcano (2)
 - Events (by variables): Hurricane (14), Volcano (14) (checked)
- Number of matching Variables:** 14 of 905. Total Variable(s) included in Plot: 1.
- Keyword:** [input field]
- Variable List:** Aerosol Opt Target (M...), Aerosol Opt Target (M...), Precipitable V..., Total Column Weighted Me..., Cirrus Reflec Mean (MY...), Ice Cloud Op Mean (MY...), Liquid Water Mean of Dail..., Cloud Top Pr..., Cloud Top Pressure: Mea...
- Event types:** Landslides, Manmade, Sea and Lake Ice, Severe Storms, Snow, Temperature Extremes, Volcanoes (selected), Water Color.
- Event List:** Calbuco Volcano, Chile; Cotopaxi Volcano, Ecuador; Manam Volcano (selected); Masaya Volcano, Nicaragua; Momotombo Volcano, Nicaragua; Mount Etna Volcano, Italy; Raung Volcano, Indonesia, July-A 2015.
- Source:** EOnet
- Buttons:** Done, Clear Event Selection, Help, Reset, Feedback, Plot Data, Go to Results.

Part 4: Image Retrieval

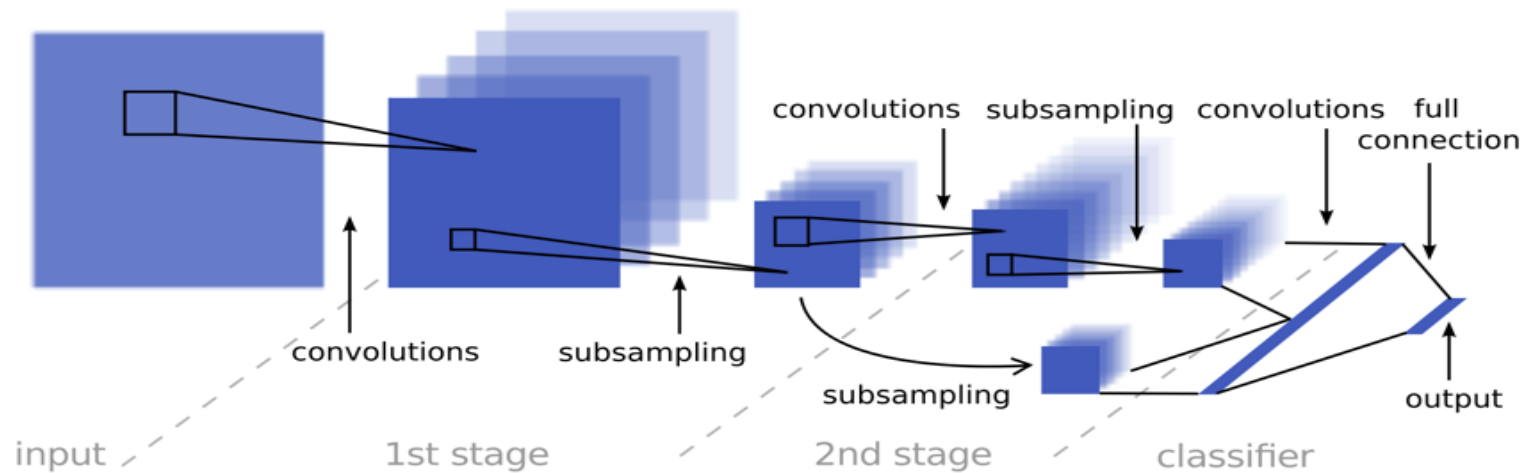
...

Image Retrieval

- Goal: given an image of Earth science phenomenon retrieve similar images
- Challenge: “semantic gap”
 - low-level image pixels and high-level semantic concepts perceived by humans

Deep Learning

- Mimics the human brain that is organized in a deep architecture
 - Processes information through multiple stages of transformation and representation
- Learns complex functions that directly map pixels to the output, without relying on *human-crafted features*



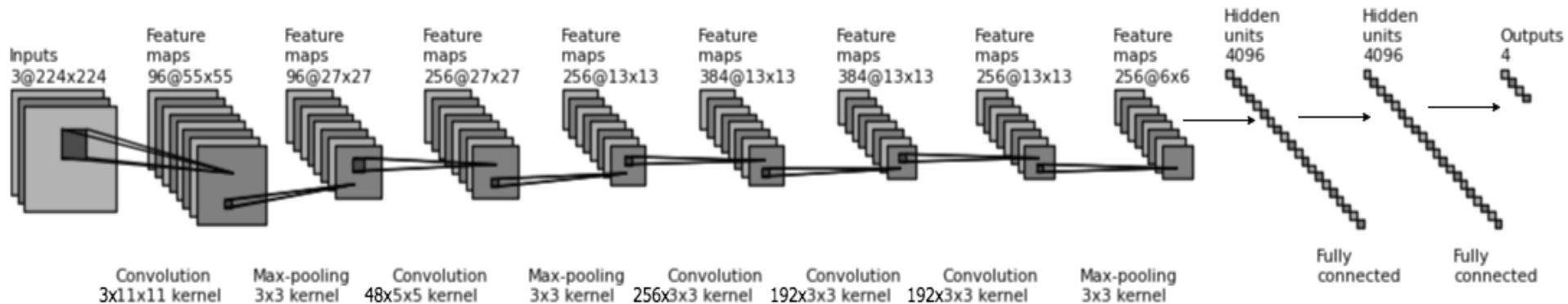
Convolution neural network

Transfer Learning

- CNN requires large number of parameters
- Learning parameters from *a few thousand training samples* is unrealistic
- Transfer learning
 - Use internal representation learned from one classification task to another
 - AlexNet architecture - Krizhevsky et. al.
 - Weights learned from ImageNet 1.3 million high-resolution images
 - State-of-the-art classification accuracy

Experiment: CNN Configuration

Text

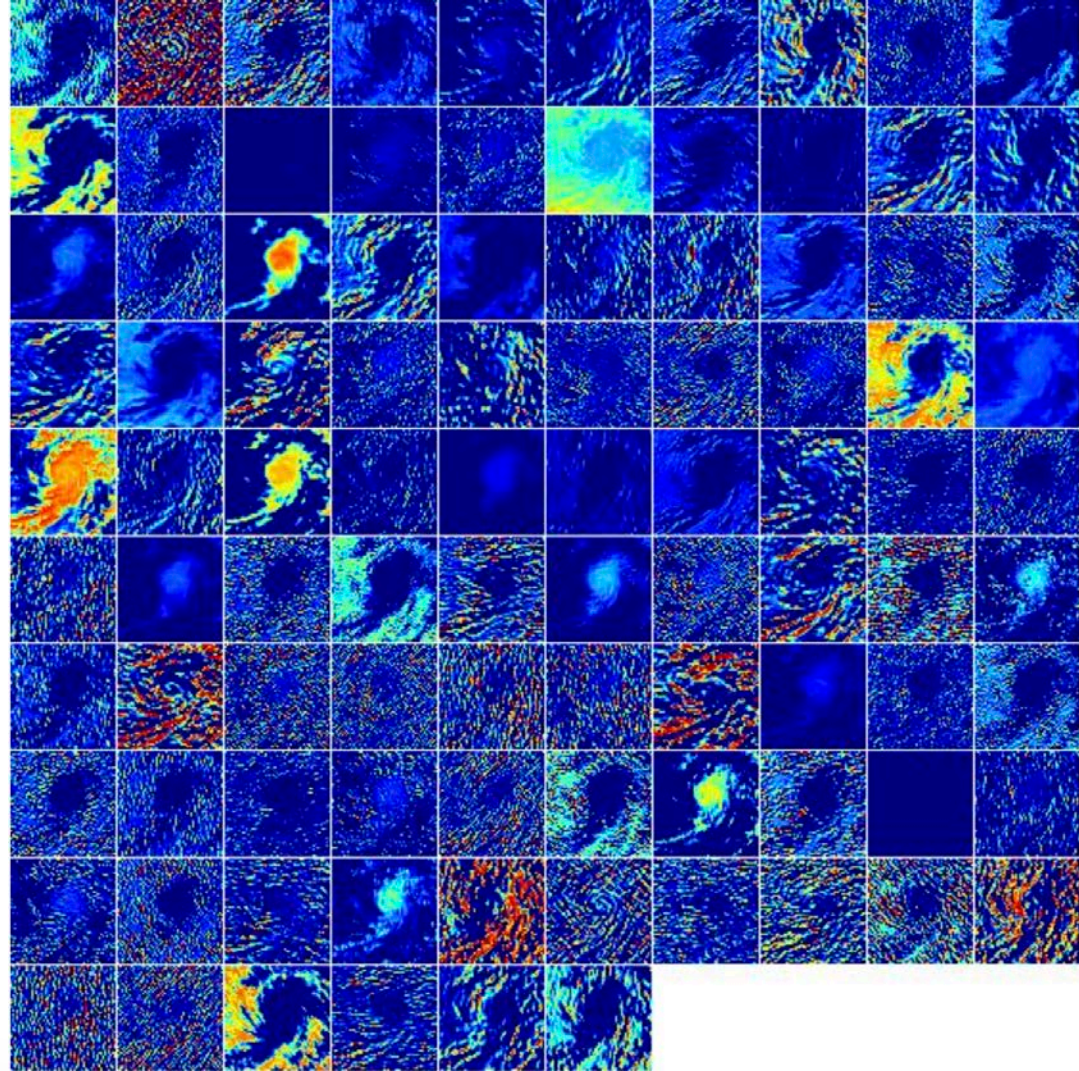


- AlexNet architecture
 - Initialized weights with ImageNet trained model
 - Adaptive learning rate
 - GPU implementation

Experiment CNN – Visualization

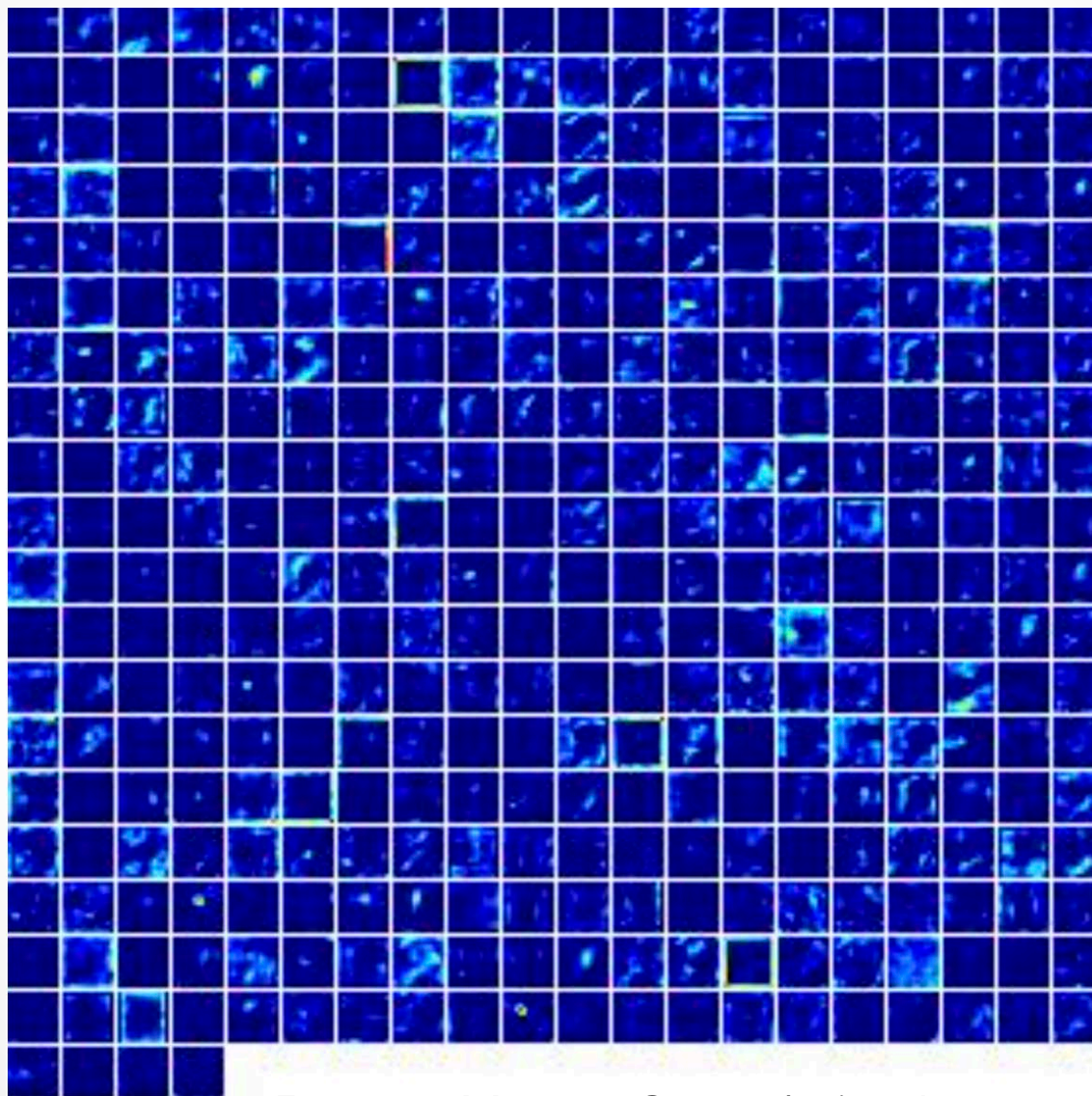


Input Image



Feature Maps – Convolution Layer 1

Experiment CNN – Visualization



Feature Maps – Convolution Layer 3

Results: Confusion Matrix

MODIS Rapid Response Test Images (Images are New to Trained CNN)

True/Pred	Dust	Hurricane	Smoke	Other
Dust	287	8	32	33
Hurricane	0	379	1	10
Smoke	12	12	443	9
Other	33	9	23	211

Overall Accuracy = **87.88%**

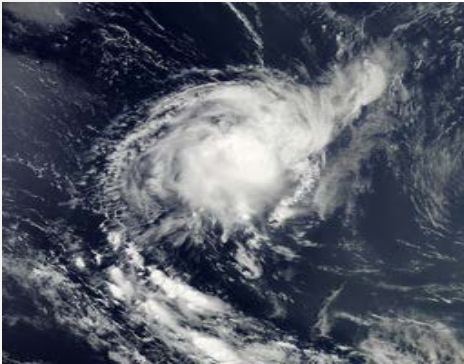
Producer's Accuracy

Dust 86.45%
Hurricane 92.89%
Smoke 88.78%
Other 80.23

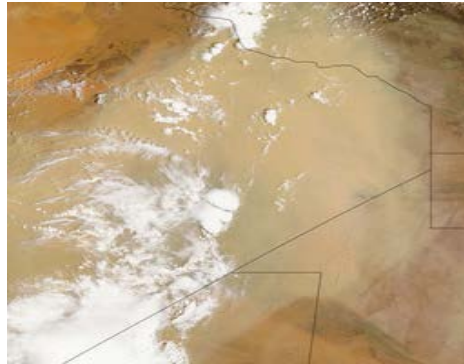
User's Accuracy

Dust 79.72%
Hurricane 97.18%
Smoke 93.07%
Other 76.45%

Results (MODIS Rapid Response)



Hurricane – True Positive



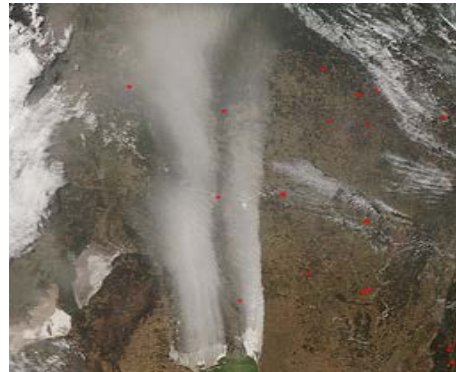
Dust – True Positive



Smoke – True Positive



Hurricane – False Negative



Dust – False Positive



Smoke – False Positive

Summary

- Building three specific semantic middleware components
 - *Image retrieval service*
 - *Data curation service*
 - *Semantic rules engine*

Infuse the entire middleware or the components into existing NASA data and information system

Dr. Rahul Ramachandran

Deputy Editor, Earth Science Informatics Journal

Manager, Global Hydrology Resource Center

NASA's Distributed Active Archive Center

Earth Science Office (ZP11)

NASA / Marshall Space Flight Center

Huntsville, Alabama 35812, USA

(w) 256.961.7620

(c) 256.226.6854

OrcId: orcid.org/0000-0002-0647-1941

www.linkedin.com/in/ramachandran05/