

ILLUMINATING THE DARKNESS: EXPLOITING UNTAPPED DATA AND INFORMATION RESOURCES IN EARTH SCIENCE

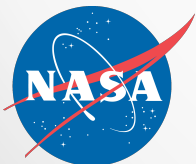
Manil Maskey* UAH, Rahul Ramachandran NASA/
MSFC, Peter Fox RPI, Steven Kempler NASA/GSFC

Project Team:

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

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

RPI: Stefan Zednik, Anirudh Prabhu



Earth Science Technology Forum, 2016



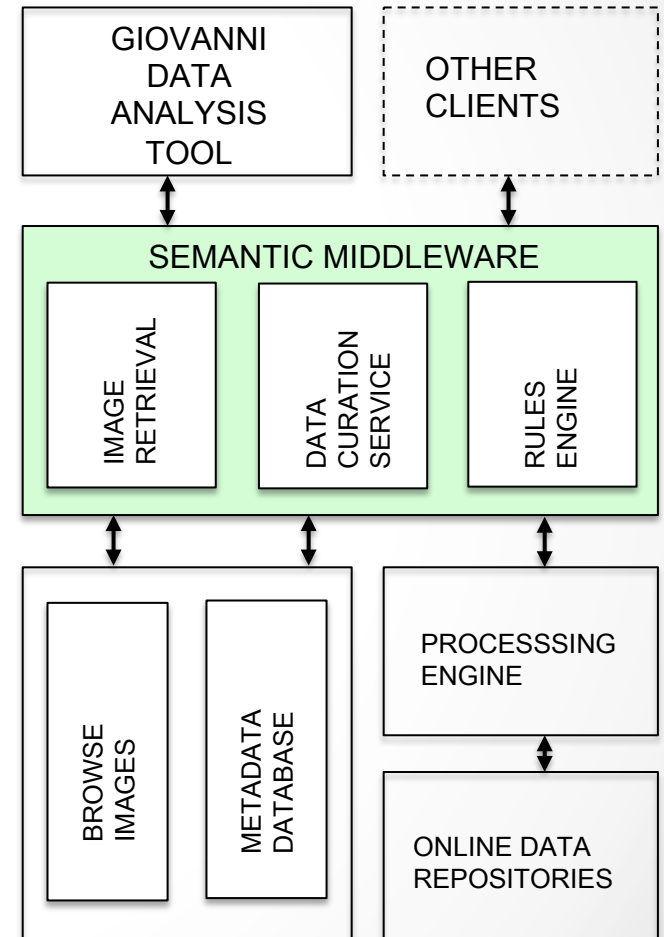
Earth Science Metadata: Dark Resources

- *Dark resources*: not used beyond **intended** purpose
 - Challenge: recognize, identify and effectively utilize for other **purposes**
- Metadata catalogs:
 - contain dark resources
 - structured information
 - free form descriptions: data and browse images
- NASA's Common Metadata Repository
 - > 6000 data collections
 - 270 million records for individual files
 - 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 Earth science and education communities.

Project Goals

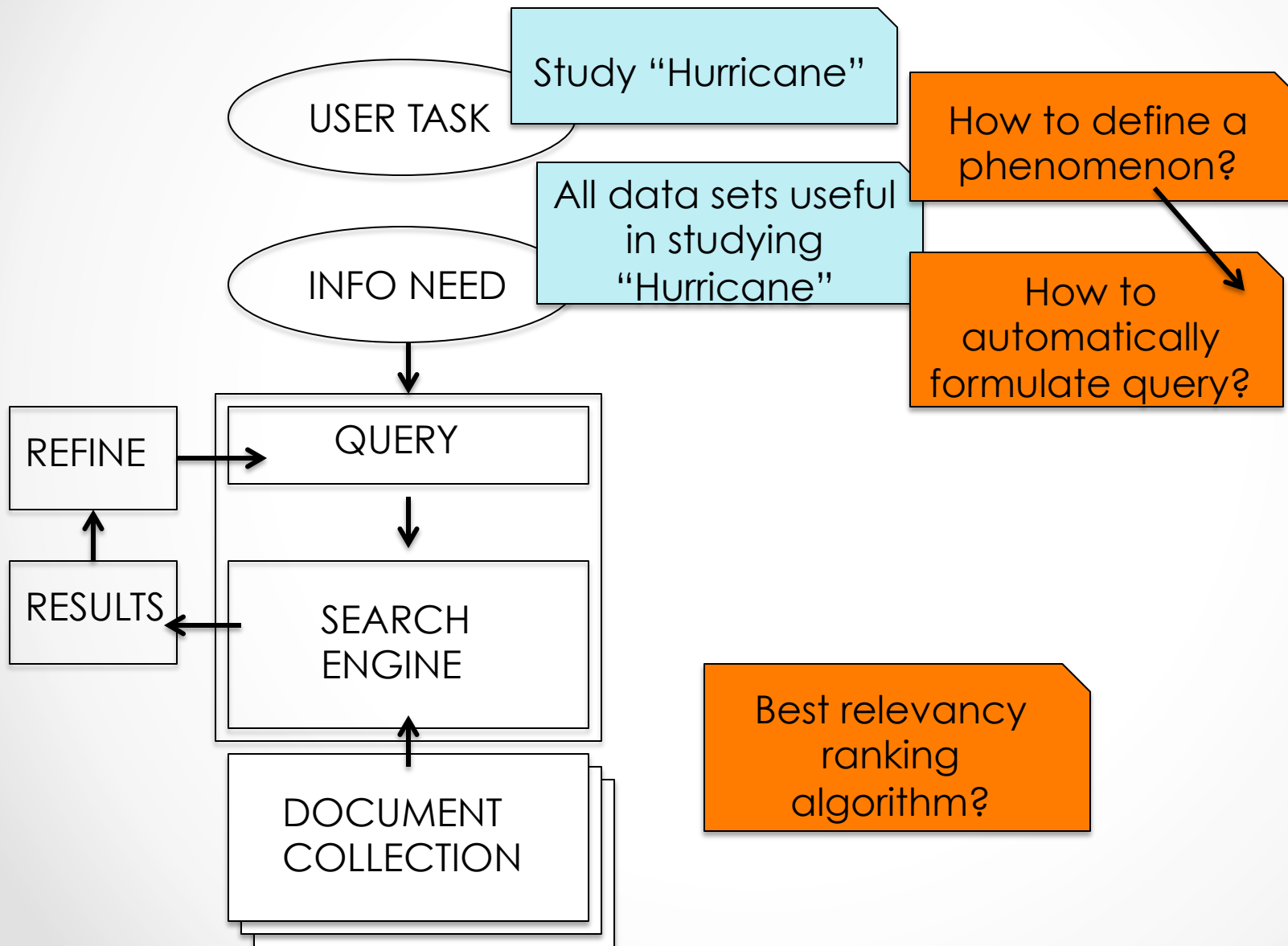
- Design a Semantic Middleware Layer (SML) to exploit metadata resources
 - provide novel **data discovery and exploration** capabilities that significantly reduce data preparation time.
 - utilize a varied set of semantic web, information retrieval and image mining technologies.
 - automate
- Design SML as a Service Oriented Architecture



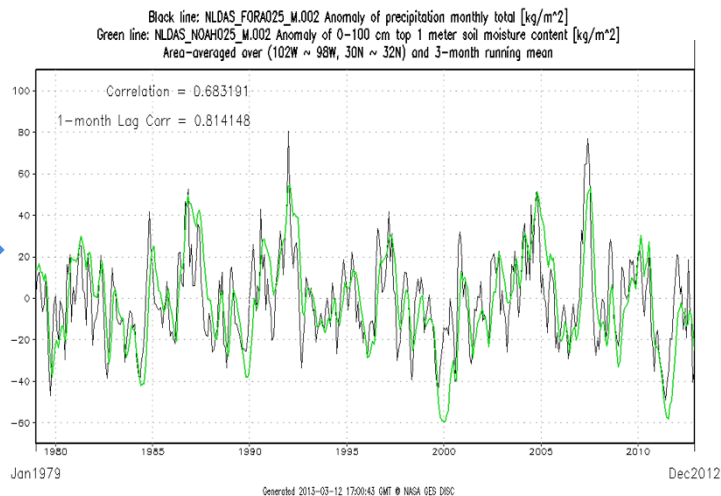
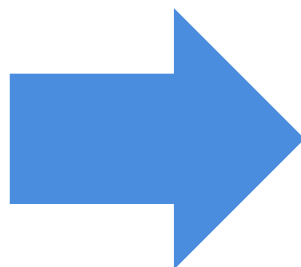
Data Curation Service

- Relevancy ranking algorithm for a set of **phenomena**
- Stand alone service
- Envisioned Use:
 - Given a phenomenon type (Ex: Hurricane), DCS returns a list of relevant data sets (variables)
 - <list of data sets (variables)> = DCS(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 Approach



Rules Engine: What settings should I use to visualize this event?



Data Variable
?

Dataset
?
Visualization Type?

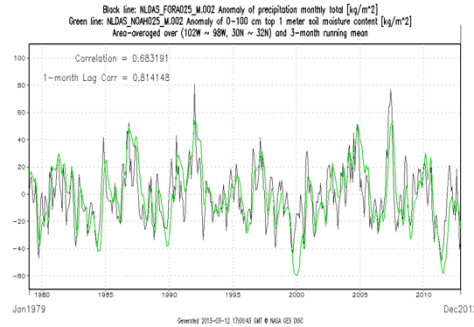
Goal: Automate data preprocessing and exploratory analysis and visualization tasks

Compute Compatibility

Use rules to make **assertions** about **compatibility** based on **multiple factors**



+



=



Phenomena:
Volcano - Ash
Plume

Service - Area
Averaged Time
Series

**STRONG
COMPATIBILITY
x2**

Temporal Evolution	Detection of Events
Strong	Strong

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

Service to generate and rank candidate workflow configurations

Images from , http://disc.sci.asfc.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

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):** 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, Longitude-Pressure; Cross Map, Time-Pressure; Vertical Profile.
- 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

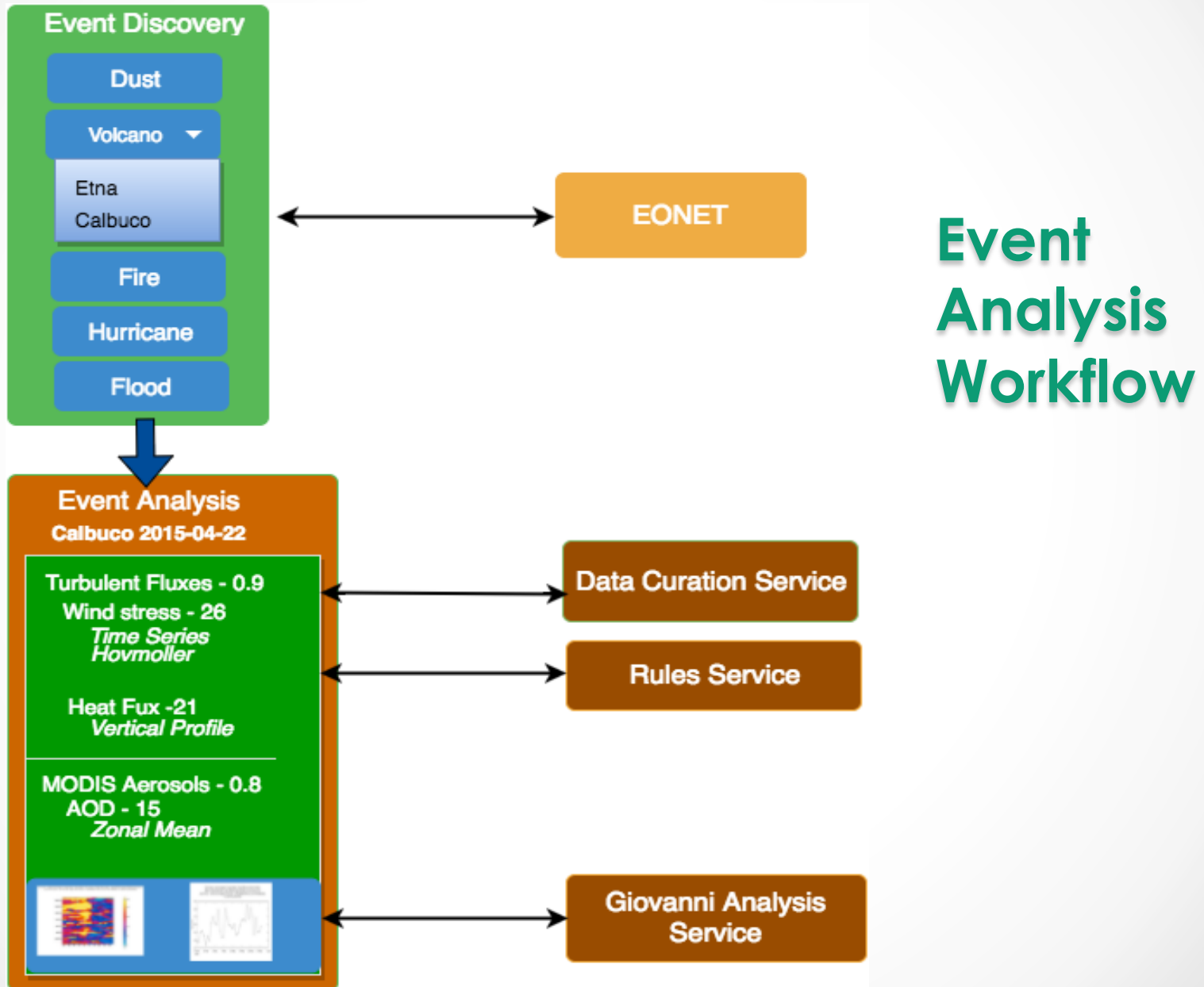
The screenshot shows the GIOVANNI web interface with the following elements:

- Select Plot:** Includes dropdowns for Maps, Comparisons, Vertical, Time Series (set to Area-Averaged), and Miscellaneous.
- Select Date Range (UTC):** Shows a date range from 2015-07-31 00:00 to 2015-11-24 23:59.
- Select Region:** Shows a bounding box for Manam Volcano with buttons for Show Map, Show Shapes, and Show Events.
- Select Variables:** A sidebar with expandable sections: Events (all), Events (by products), and Events (by variables). Under Events (by variables), 'Volcano (14)' is selected.
- Number of matching Variables:** 14 of 905. Total Variable(s) included in Plot: 1.
- Keyword:** A search input field.
- Event Selection Modal:** A window titled 'Event' showing a list of event types and a table of events. 'Manam Volcano' is selected.

Event types	Event
Landslides	<input type="radio"/> Calbuco Volcano, Chile
Manmade	<input type="radio"/> Cotopaxi Volcano, Ecuador
Sea and Lake Ice	<input checked="" type="radio"/> Manam Volcano
Severe Storms	<input type="radio"/> Masaya Volcano, Nicaragua
Snow	<input type="radio"/> Momotombo Volcano, Nicaragua
Temperature Extremes	<input type="radio"/> Mount Etna Volcano, Italy
Volcanoes	<input type="radio"/> Raung Volcano, Indonesia, July-Aug 2015

Curation
Service: event
type filters
relevant
variables

Giovanni - Dark Data Edition



Serendipitous Discovery

Data Curation Goal: map dataset keywords to granule variables

Application of Data Curation for Operational Use:

Data Curation Algorithm can be used to assess

- Metadata quality for both dataset and granules
 - Find incorrect/incomplete keyword annotations
- Automatically suggest science keywords

Operational Use: Prototype Variable Mapping

54.172.157.10:5000

News personal Mendeley GKeep NASA Demo H53 GHRC DarkData nspires RResp Unisys Weather - Ge

Data Parameter Mapping Tool

Datasets

AIRS/Aqua Level 2 Support retrieval (AIRS+AMSU) V005
GHRSST 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
MODIS/Terra Aerosol 5-Min L2 Swath 10km V005 NRT

Datasets

MODIS/Terra Aerosol 5-Min L2 Swath 10km V005 NRT

Science Keyword Map EDIT

ATMOSPHERE > AEROSOLS > PARTICULATE_MATTER 3
Deep_Blue_Aerosol_Optical_Depth_Land_STD : 1
Deep_Blue_Aerosol_Optical_Depth_550_Land : 1
Aerosol_Type_Land : 1
Aerosol_Cldmask_Byproducts_Ocean : 1
Deep_Blue_Aerosol_Optical_Depth_Land : 1
Aerosol_Cldmask_Byproducts_Land : 1
Deep_Blue_Aerosol_Optical_Depth_550_Land_STD : 1
Optical_Depth_Small_Average_Ocean : 0

Parameter Map EDIT

Optical_Depth_Small_Average_Ocean 3
ATMOSPHERE->AEROSOLS->AEROSOLS_OPTICAL_DEPTH/THICKNESS : 2
ATMOSPHERE->ATMOSPHERIC_RADIATION->OPTICAL_DEPTH/THICKNESS : 2
ATMOSPHERE->AEROSOLS->PARTICULATE_MATTER : 0
Asymmetry_Factor_Best_Ocean 0
Deep_Blue_Angstrom_Exponent_Land 0
Cloud_Fraction_Ocean 1
ATMOSPHERE->AEROSOLS->CLOUD_CONDENSATION_NUCLEI : 1

MODIS/Terra Aerosol 5-Min L2 Swath 10km V005 NRT

- 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_RADIANCE : 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**
- ATMOSPHERE->AEROSOLS->SULFATE_PARTICLES : 2
- ATMOSPHERE->ATMOSPHERIC_RADIATION->RADIATIVE_FLUX : 2
- ATMOSPHERE->ATMOSPHERIC_RADIATION->REFLECTANCE : 2
- ✓ ATMOSPHERE->ATMOSPHERIC_RADIATION->OPTICAL_DEPTH/THICKNESS : 2

Remove

Remove

ATMOSPHERE->AEROSOLS->PARTICULATE_MATTER : 0 Remove

Edit/Save Mapping

Mapping Scores Generated by Algorithm

Opportunity to develop this prototype and infuse into operational use at DAACs to improve metadata quality

DEMO



SML Components

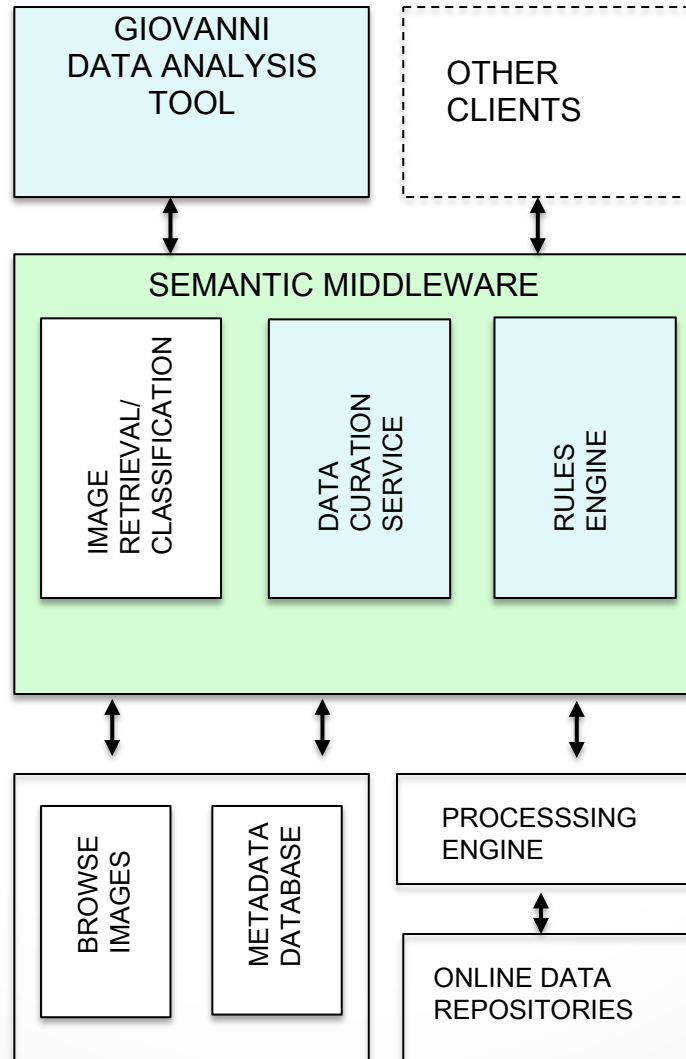
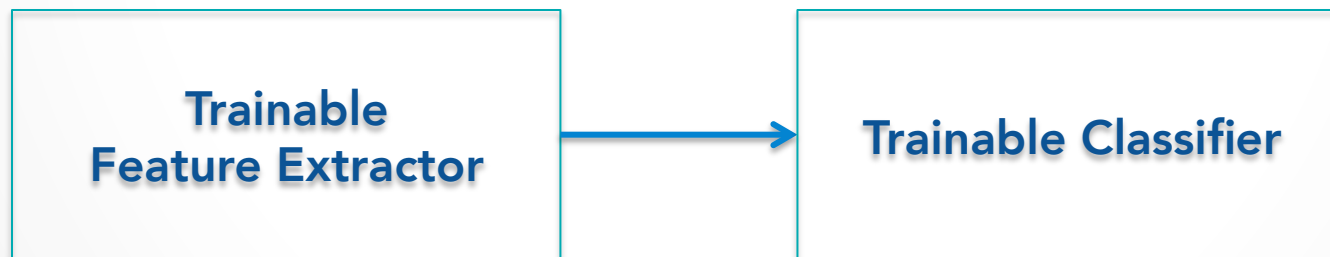


Image Classification

- Goal: label images in archives with a known Earth science phenomenon
- Challenge: “semantic gap”
 - low-level image pixels and high-level semantic concepts perceived by humans

“Deep” Architecture

- Features are key to recognition
- What about learning the features?
- Deep Learning
 - Hierarchical Learning
 - Mimics the human brain that is organized in a deep architecture
 - Multiple stages of representation



Convolutional Neural Network (CNN)

- ***Applicable to Images***
- ***Supervised***

Transfer Learning

- CNN requires large number of parameters
- Transfer learning
 - Use internal representation learned from one classification task to another
- Faster learning
- Better accuracy

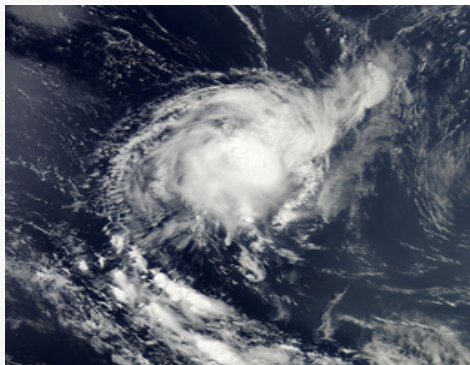
Applications: Searching for Events

- Detection of phenomena in Browse Imagery

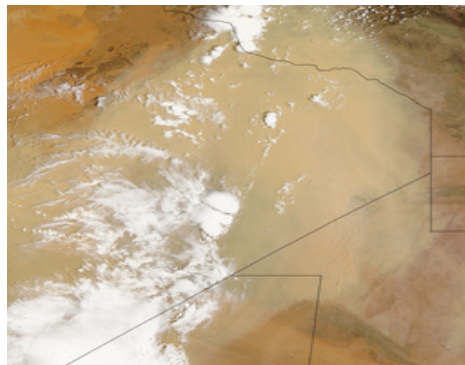
True/Pred	Dust	Hurricane	Smoke	Other
Dust	287	8	32	33
Hurricane	0	379	1	10
Smoke	12	12	443	9
Other	33	9	23	211

Confusion Matrix

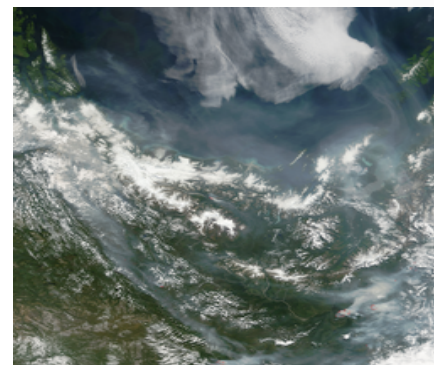
Overall Accuracy = **87.88%**



Hurricane – True Positive



Dust – True Positive



Smoke– True Positive

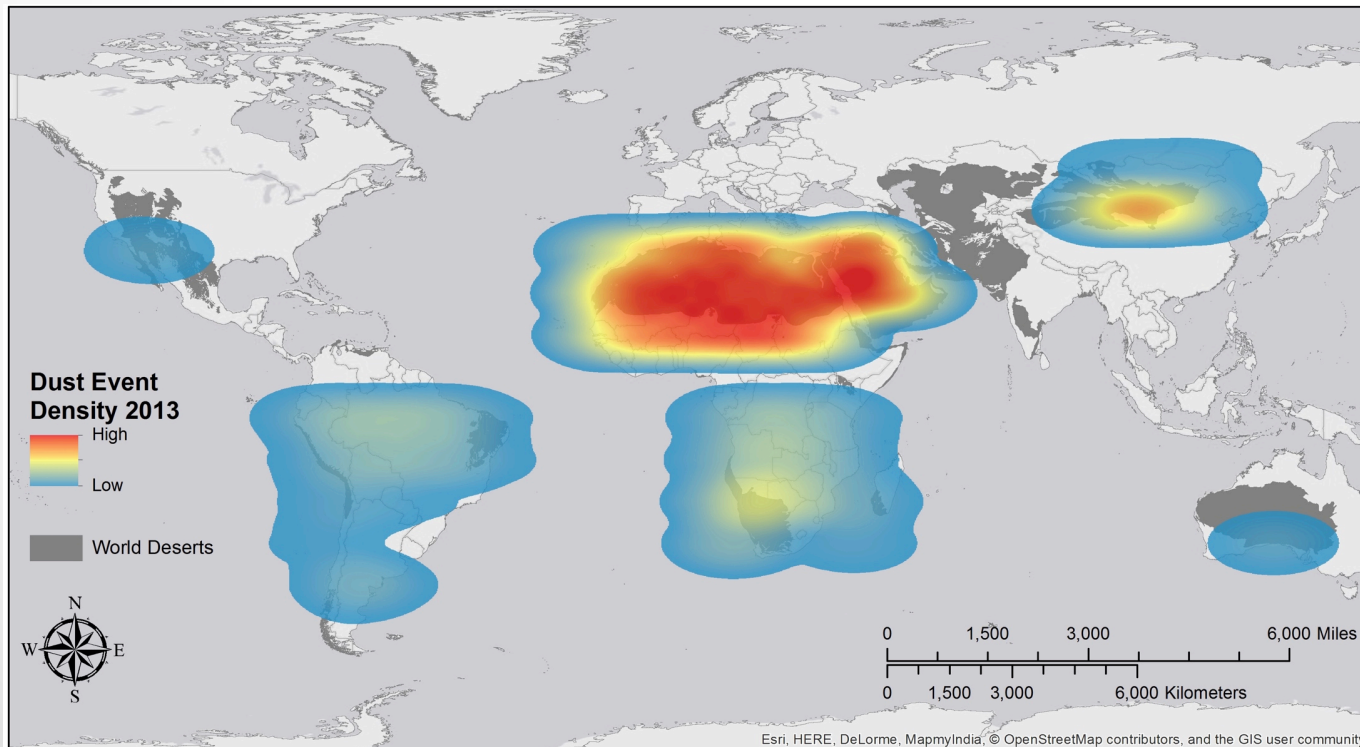
Applications: Enabling new science

- Dust climatology – Collaboration with Sundar Christopher, UAH Atmospheric Science Professor

True\Predicted	Dust	Other	Total
Dust	1379	379	1758
Other	260	4932	5192
	1639	5311	6950

Validation
Accuracy = **91%**

Confusion Matrix



Based on GIBS

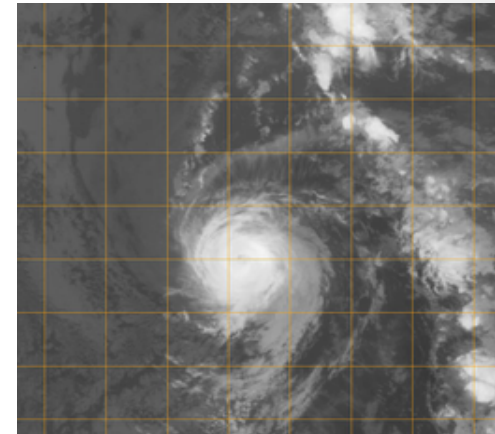
Applications: Improving forecast operations

- Hurricane intensity estimation - Collaboration with Dan Cecil, NASA/MSFC Atmospheric Scientist

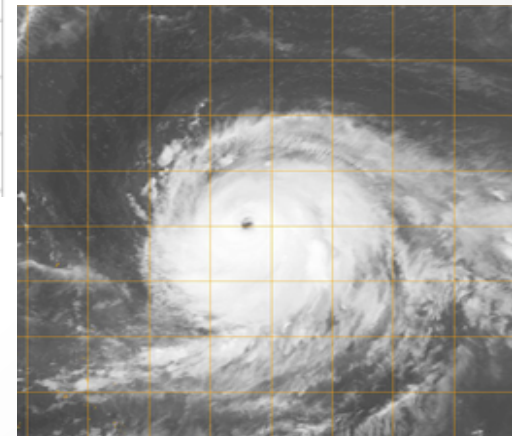
True\Predicted	td	ts	h1	h2	h3	h4	h5	no_cat	total
td	3168	335	0	1	0	0	0	6	3510
ts	489	4823	159	5	11	3	6	0	5496
h1	9	484	1158	92	20	6	1	0	1770
h2	3	76	214	513	145	4	0	5	960
h3	6	40	33	155	689	55	0	0	978
h4	1	18	17	12	142	810	32	0	1032
h5	2	2	0	0	27	59	216	0	306
no_cat	22	0	0	0	0	0	0	32	54
	3700	5778	1581	778	1034	937	255	43	14106

Overall Accuracy : 81 %(Top 2 Probabilities 95.73%)

Data: NRL Images, HURDAT

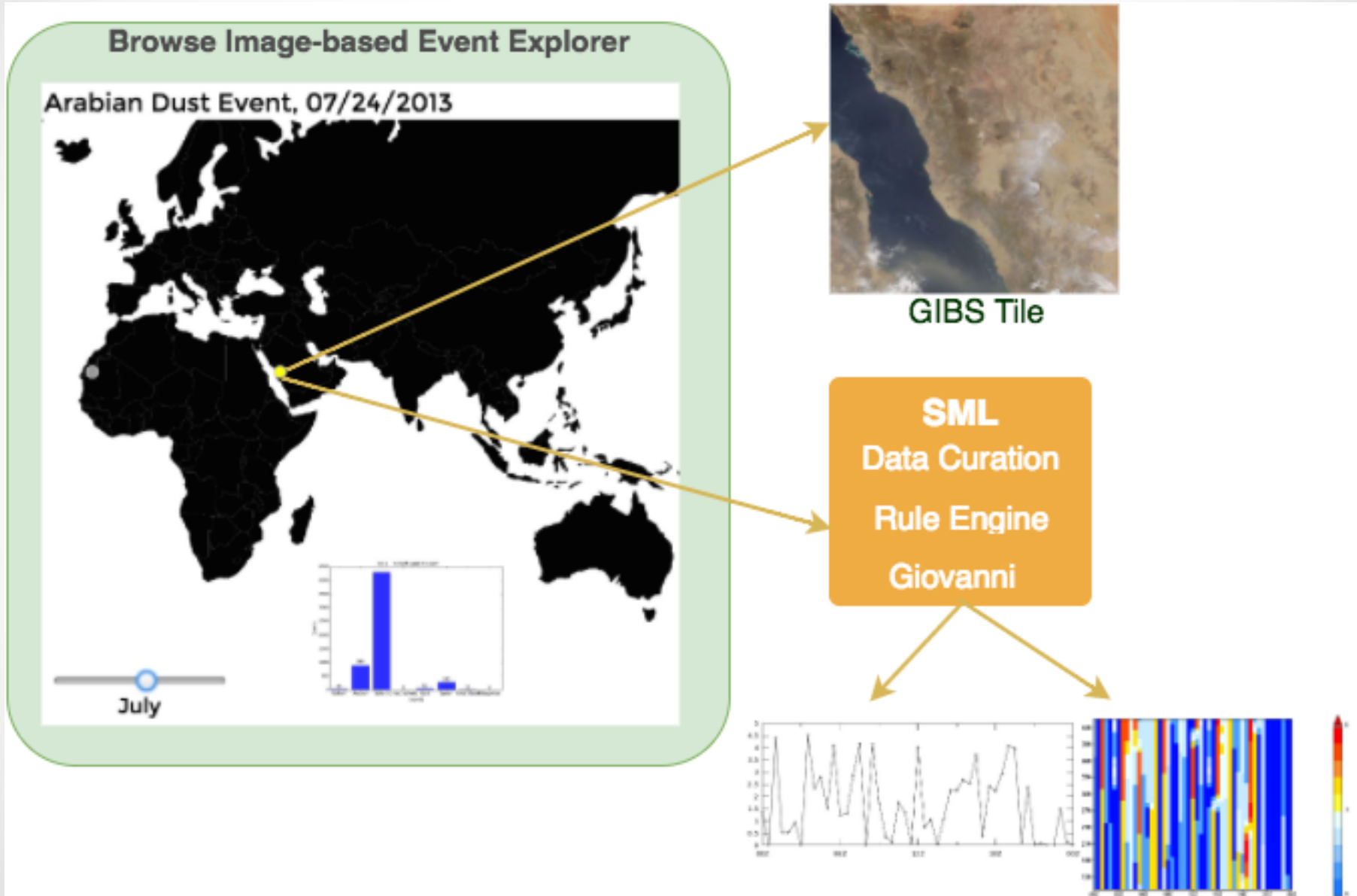


Cat 2 Hurricane



Cat 4 Hurricane

Ongoing Work



Journal Publications

- Submitted:
 - *Deep Learning for Phenomena-based Classification of Earth Science Images* - IEEE Geoscience and Remote Sensing Letters
- In Progress:
 - *Relevancy Algorithm to Curate Earth Science Data for Different Phenomena* – to be submitted to Computers and Geoscience
 - *Detecting Transverse Cirrus Bands using Deep Learning* - IEEE Geoscience and Remote Sensing Letters (collaboration with U.S. Nair)
- Planned:
 - *Dust climatology* (collaboration with Sundar Christopher)
 - *Hurricane intensity estimation* (collaboration Dan Cecil)

**Thanks to NASA ESTO for their
support**

Contact

Manil Maskey

manil.maskey@nasa.gov

Rahul Ramachandran

rahul.ramachandran@nasa.gov