

# Developing a Knowledge Base for NASA Earth Science and Hydrologic Applications

Amanda Weigel<sup>1,2</sup>, Patrick Gatlin<sup>1,3</sup>, Rahul Ramachandran<sup>1,3</sup>, JJ Miller<sup>1,2</sup>, Manil Maskey<sup>1,3</sup>, Jia Zhang<sup>1,3</sup>, Emily Berndt<sup>3</sup>

NASA/MSFC Data Science Informatics Group<sup>1</sup>

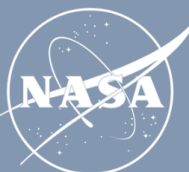
University of Alabama in Huntsville<sup>2</sup>

NASA Marshall Space Flight Center<sup>3</sup>

Carnegie Mellon University<sup>4</sup>



Carnegie  
Mellon  
University

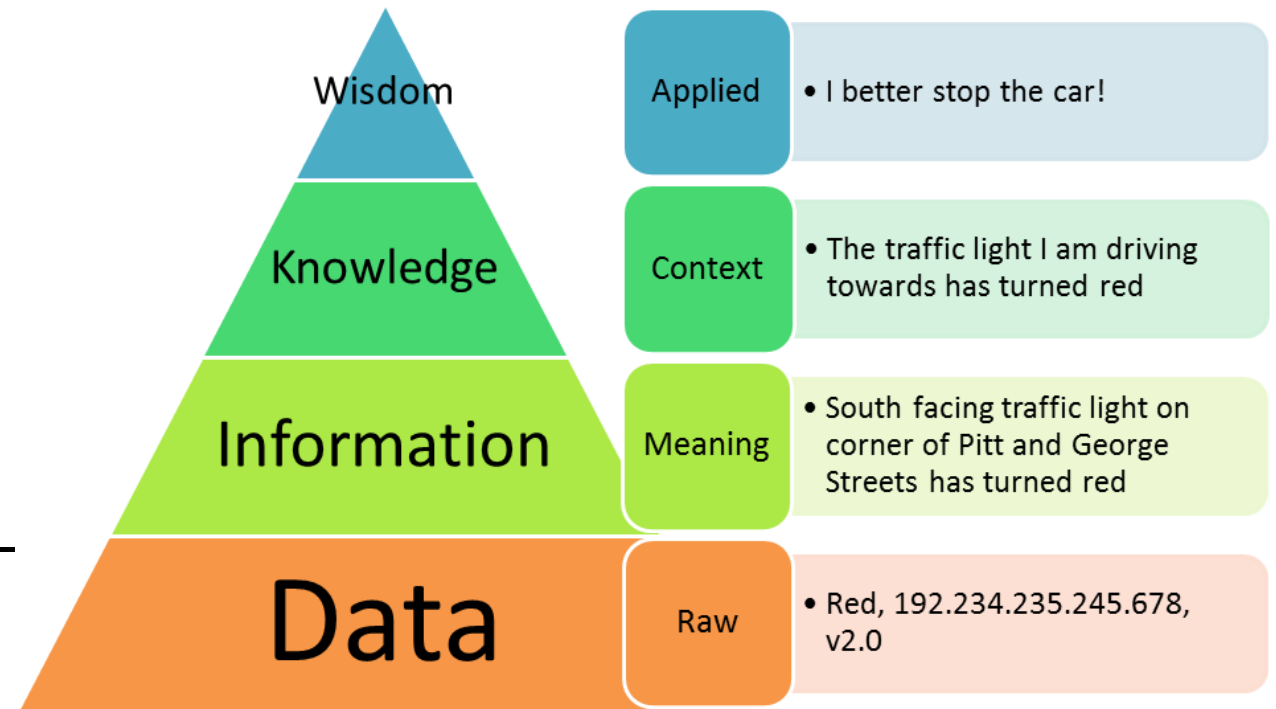




# Data Use Challenges

What common challenges do Earth science and hydrologic data users encounter?

- Data **discovery**
- Data **use**
- Identifying **key resources** about the data.
  - Accessing introductory material (for unfamiliar users).
- Determining what **methods** to use – data processing, quality control and analysis.



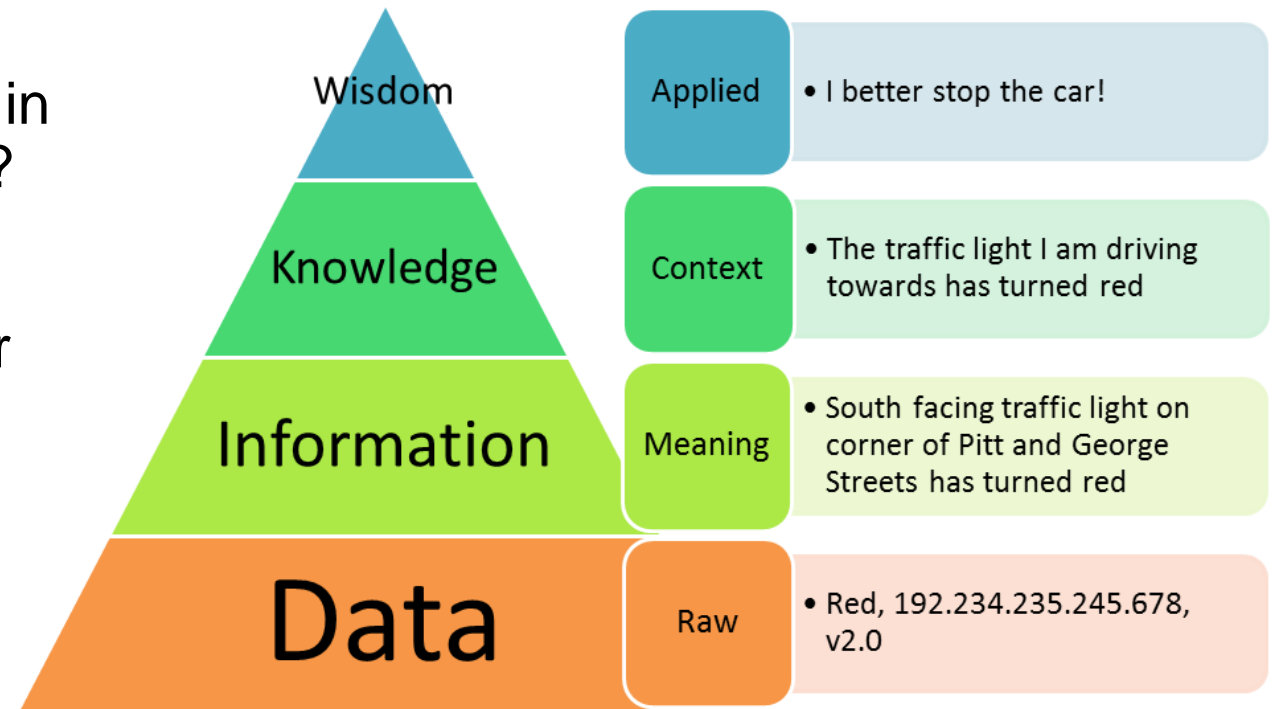
© 2011 Angus McDonald

# Data Use Challenges

**To address these challenges, what difficulties are presented?**

How can data and resources be linked in order to improve the data spin-up time?

How can we work to educate unfamiliar users?



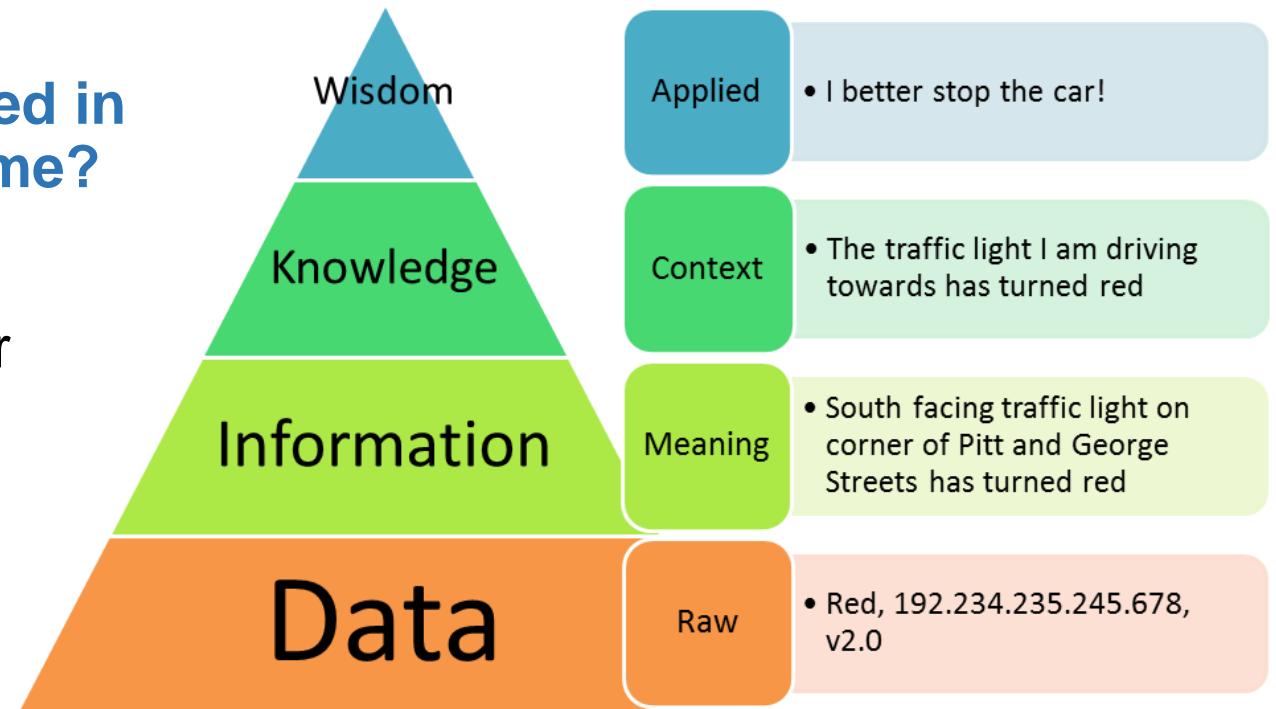
© 2011 Angus McDonald

# Data Use Challenges

To address these challenges, what difficulties are presented?

How can data and resources be linked in order to improve the data spin-up time?

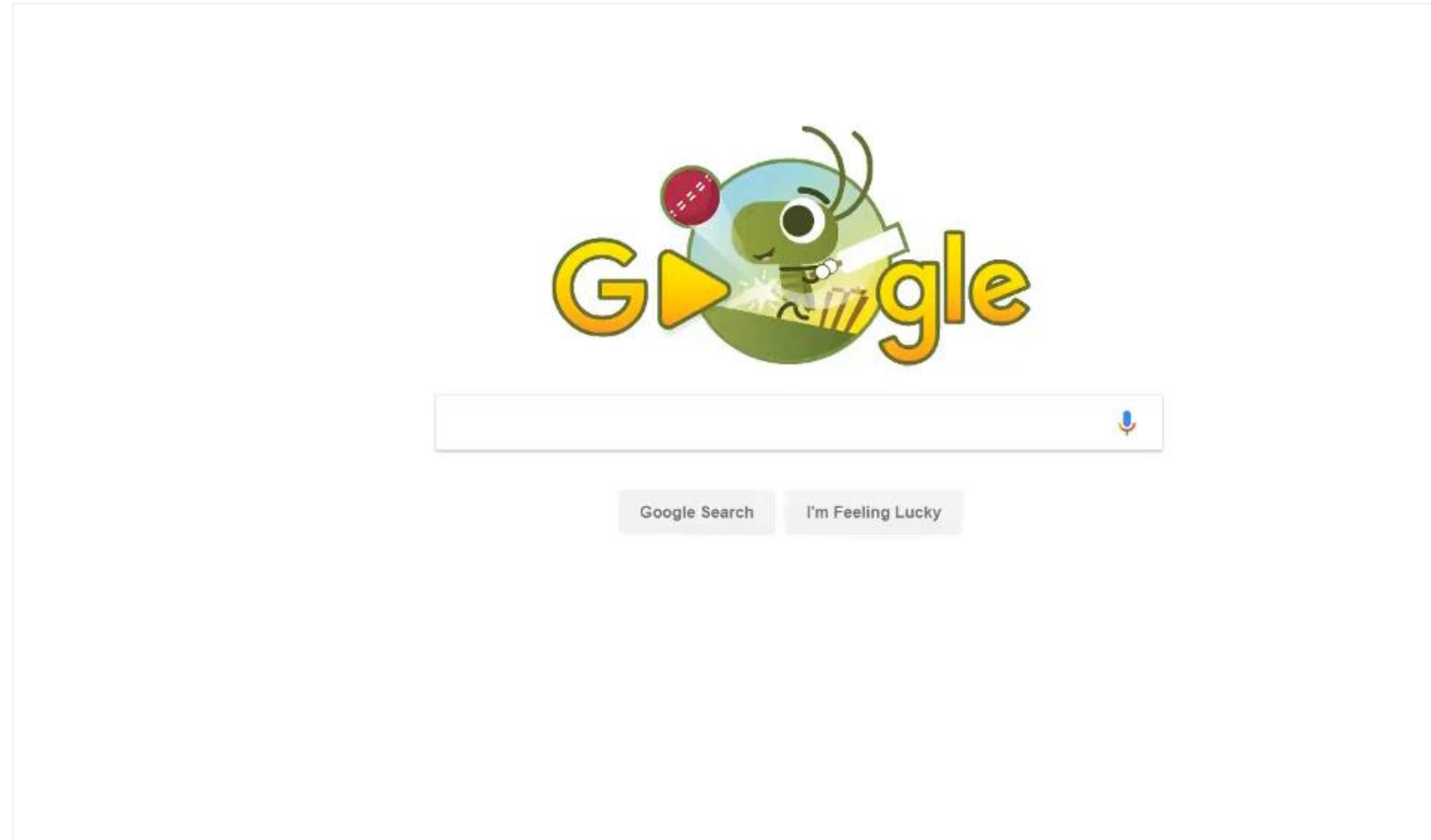
How can we work to educate unfamiliar users?



© 2011 Angus McDonald

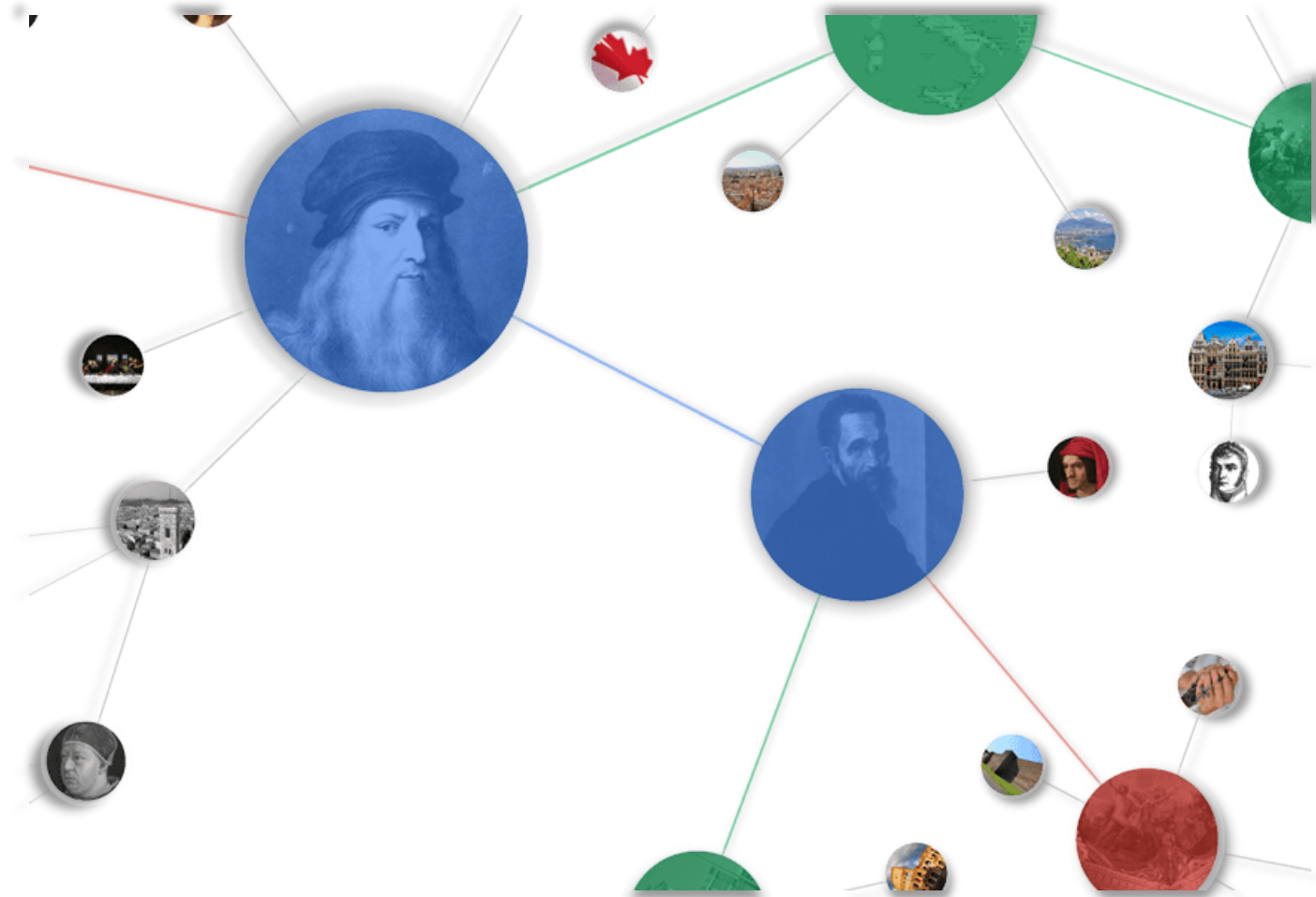
# What is a Knowledge Base?

- Think “**Google Search**”.
- Developed by Google in 2012 to enhance the results of its search engine by systematically linking information.
- Aggregates structured and detailed information about a defined topic.
- Enables users to resolve their query without having to navigate and assemble information manually.
- Why not apply it to Earth science and hydrologic data and information?



# Project Objectives

1. Identify key science information and develop an information model.
2. Extract key information from scientific literature (e.g. hypothesis, conclusions, methods, datasets, variables, etc.).
3. Link scientific knowledge to datasets, resources, services and scientists.
4. Develop a knowledge-based search capability for NASA Earth Science.



[Google, The Knowledge Graph](#)



# Technical Approach

## Terminology

### What is an **entity**?

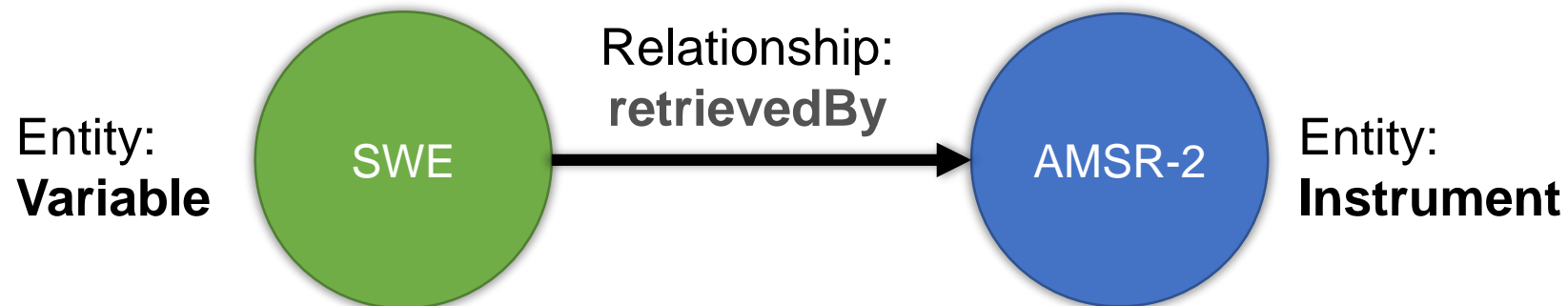
A thing with distinct and independent existence.

Examples: Variables, datasets, instruments, platforms etc.

### What is a **relationship**?

The connection between two entities.

Example: “Snow water equivalent (SWE) is retrieved by AMSR-2”.

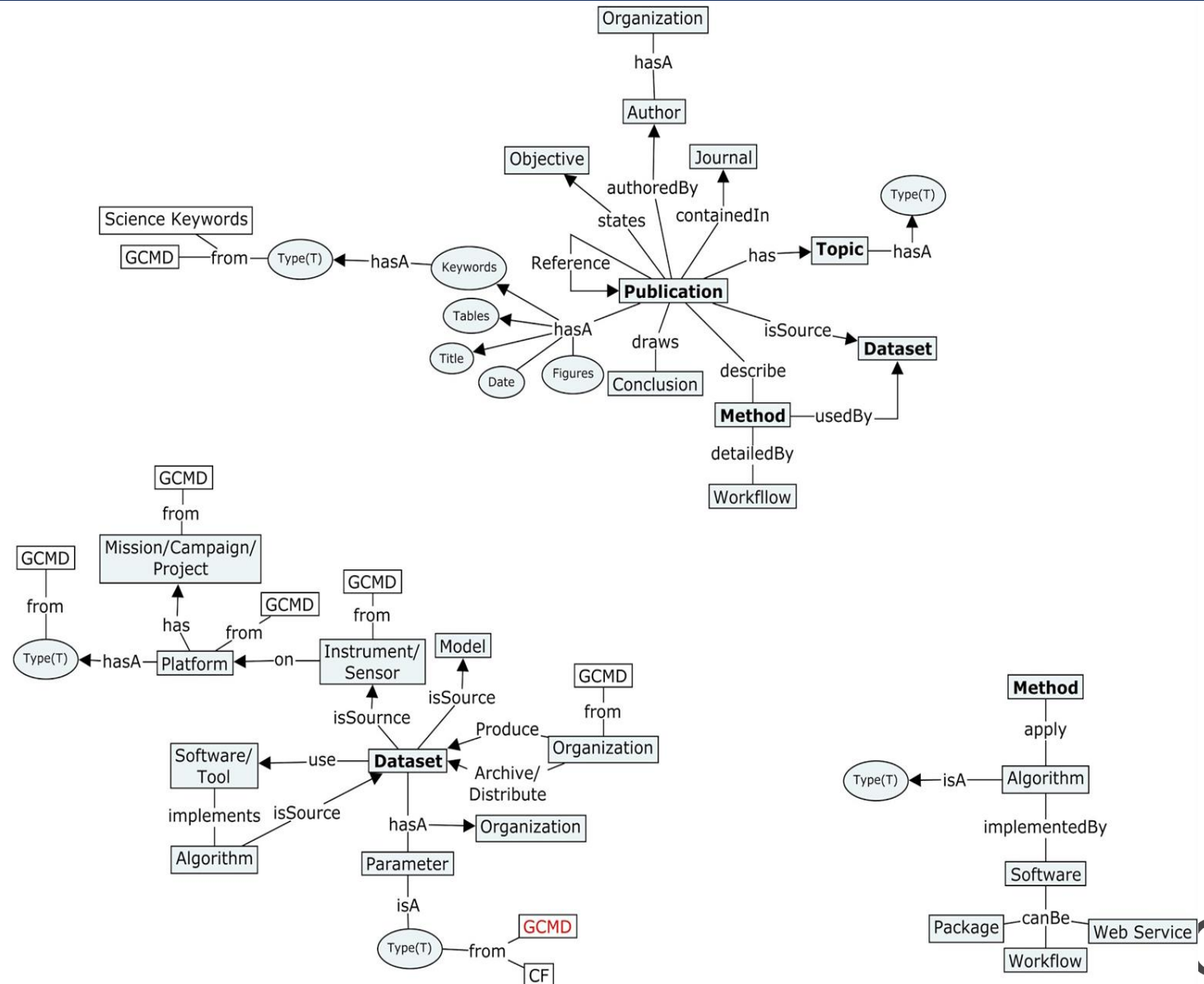


# Technical Approach

## Information Model

Development based on the Global Changing Information System (GCIS) information model

The information model defines entities and relationships pertinent to **NASA Earth science and hydrologic data, publications and resources.**





# Technical Approach

## Key Challenge

Knowledge base construction uses both structured and unstructured content (e.g., journal articles).

### Structured Content

- Metadata, tables, controlled vocabularies

```
File "AMSR_2_L3_DailySnow_P00_20170312.he5"
File type: Hierarchical Data Format, version 5
netcdf file:/C:/Users/AWeigel/Downloads/AMSR_2_L3_DailySnow_P00_20170312
{
  group: HDFEOS {
    group: ADDITIONAL {
      group: FILE_ATTRIBUTES {
      }
    }
    group: GRIDS {
      group: Northern_Hemisphere {
        dimensions:
          XDim = 721;
          YDim = 721;
        variables:
          short_HDFEOS_CRIS;
          :Projection = "HES_GCTP_LAMAZ";
          :UpperLeftPointMtrs = -9036843.073845, 9036843.073845; // do
          :LowerRightMtrs = 9036843.073845, -9036843.073845; // double
          :ProjParams = 6371228.0, 0.0, 0.0, 0.0, 0.0, 9.0E7, 0.0, 0.0
          :SphereCode = "19";
      }
    }
    group: Data_Fields {
      variables:
        byte_Flags_NorthernDaily(YDim=721, XDim=721);
        :_FillValue = -1UB; // byte
        :_Unsigned = "true";
    }
  }
}
```

### Unstructured Content

- Journal articles, ATB documents, user guides

**RESEARCH LETTER**  
10.1002/2015GL063325

**Impact of soil moisture on dust outbreaks in East Asia Using satellite and assimilation data**

Hyunglok Kim<sup>1</sup> and Minha Choi<sup>1</sup>

**Abstract** This study is the first assessment of the effects of soil moisture on satellite-derived aerosol optical depth (AOD) and global assimilation data on East Asia. The relationships among dust outbreaks, soil moisture, and wind speed over the Moderate Resolution Imaging Spectroradiometer and Global Land collected over 11 years (2003–2013). The mean AOD exponentially decreased under different wind speed conditions (average determination coefficient is 0.6). As conditions became stronger, the probability of a dust outbreak became greater. The threshold soil moisture for dust outbreaks increased with increasing wind speed. Our results have the capability to be used for dust-outbreak prediction and global-scale dust-emission studies.

**1. Introduction**  
Sand dust has direct implications for public health, transportation, and agricultural production. Dust is a major component of the atmospheric aerosol component, which plays an important role in climate systems and affects the atmospheric radiation budget by scattering and absorbing longwave and shortwave radiation (Sokolik and Tegen, 1996; Tegen et al., 1996; Kaufman et al., 2002; Myhre et al., 2003). Surface wind speed has generally been regarded as the main factor determining the mobilization of sand and dust in dust-source regions (Chomette et al., 1999). However, dust events are influenced not only by wind speed but also by surface conditions (Ravi et al., 2004; Natsopoulos et al., 2003; Wang et al., 2004; Ishizuka et al., 2005).

GEOPHYSICAL RESEARCH LETTERS, VOL. 36, L01807, doi:10.1029/2008GL035846, 2009

**Saharan dust particles nucleate droplets in eastern Atlantic clouds**

Cynthia H. Twohy,<sup>1</sup> Sonia M. Kreidenweis,<sup>2</sup> Trude Eidhammer,<sup>2</sup> Edward V. Browell,<sup>3</sup> Andrew J. Heymsfield,<sup>4</sup> Aaron R. Banerjee,<sup>5</sup> Bruce E. Anderson,<sup>3</sup> Gao Chen,<sup>3</sup> Syed Ismail,<sup>3</sup> Paul J. DeMott,<sup>2</sup> and Susan C. Van Den Heever<sup>2</sup>

Received 28 August 2008; revised 3 November 2008; accepted 25 November 2008; published 13 January 2009.

[1] Many soil-derived particles dominated by insoluble mineral, including Saharan dusts, are known to act as ice nuclei. If, however, dust particles can compete with other atmospheric particle types to form liquid cloud droplets, they have a greater potential to change climate through indirect effects on cloud radiative properties and to affect the hydrological cycle through precipitation changes. By directly collecting and analyzing the residual nuclei of small cloud droplets, we demonstrate that Saharan dust particles do commonly act as cloud condensation nuclei (CCN) in the eastern North Atlantic. Droplet activation calculations support the measurements by showing that due to its slightly hygroscopic nature, even submicron dust can be important as CCN. Given the dual nature of Saharan dust particles as CCN and ice nuclei, this infusion of dust is expected to impact not only droplet size and albedo in small clouds, but ice formation in deep convective clouds. Citation: Twohy, C. H., et al. (2009), Saharan dust particles nucleate droplets in eastern Atlantic clouds, *Geophys. Res. Lett.*, 36, L01807, doi:10.1029/2008GL035846.

**2. Measurements**  
[4] On 5 Sept 2006, a mission was flown from the Cape Verde Islands (near the west coast of Africa) and over the Sahara Desert in Mauritania. This area is a major dust source region (Goudic and Middleton, 2001). Figure 1 shows vertical profiles of aerosol scattering ratio (1a and 1b) and water vapor mixing ratio (1c) from a differential absorption lidar [Browell et al., 2005], looking down through the dust layer which extended to about 5.5 to

# Technical Approach

## How do we extract entities from unstructured content?

- Natural language processing (**NLP**) – Computers analyzing, understanding and deriving meaning from human language.
- Semantic Entity Recognition (**SER**) – NLP technique used to identify entities in text.
- Use NLP (SER) techniques to identify entities within the unstructured text.
- Apply to journal publication text to extract and identify data, models, methods, people, and institutions (i.e., entities).
- Generate a truth set – Dictionary of known models, science keywords, CMR NASA Earth science data catalogue.

Atmospheric Research 166 (2015) 165–181

Contents lists available at ScienceDirect

**Atmospheric Research**

journal homepage: [www.elsevier.com/locate/atmos](http://www.elsevier.com/locate/atmos)

**Satellite tools to monitor and predict Hurricane Sandy (2012): Current and emerging products**

Michael J. Folmer<sup>a,\*</sup>, Mark DeMaria<sup>b</sup>, Ralph Ferraro<sup>c</sup>, John Beven<sup>b</sup>, Michael Brennan<sup>b</sup>, Jaime Daniels<sup>c</sup>, Robert Kuligowski<sup>c</sup>, Huan Meng<sup>c</sup>, Scott Rudlosky<sup>c</sup>, Limin Zhao<sup>d</sup>, John Knaff<sup>e</sup>, Sheldon Kusselson<sup>d</sup>, Steven D. Miller<sup>f</sup>, Timothy J. Schmit<sup>g</sup>, Chris Velden<sup>h</sup>, Brad Zavodsky<sup>i</sup>

<sup>a</sup> University of Maryland/ESSIC/CICS, College Park, MD, United States  
<sup>b</sup> NOAA/NWS/NCEP/NHC, Miami, FL, United States  
<sup>c</sup> NOAA/NESDIS/STAR, College Park, MD, United States  
<sup>d</sup> NOAA/NESDIS/OSPO, College Park, MD, United States  
<sup>e</sup> NOAA/NESDIS/STAR, Fort Collins, CO, United States  
<sup>f</sup> Colorado State University/CIRA, Fort Collins, CO, United States  
<sup>g</sup> NOAA/NESDIS/STAR, Madison, WI, United States  
<sup>h</sup> University of Wisconsin-Madison/CIMSS, Madison, WI, United States  
<sup>i</sup> NASA/SFRT, Huntsville, AL, United States

**ARTICLE INFO**

Article history:  
Received in February 2015  
Received in revised form 2 June 2015  
Accepted 3 June 2015  
Available online 11 June 2015

**Keywords:**  
Remote sensing  
Passive microwave  
Hurricane  
GOES-R  
JPSS  
Sandy

**ABSTRACT**

Hurricane Sandy – a tropical cyclone that transitioned into an extratropical cyclone near the time of landfall along the east coast of the United States – caused historic damage in many regions which rarely receive such a direct hit from a storm of this magnitude, including many of the large metropolitan areas along the U.S. eastern seaboard. Specifically, Sandy generated record low-pressure, a large wind field with corresponding storm surge and copious amounts of precipitation in some areas, including record snowfall in mountainous regions. Sandy presented several forecast challenges to the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS). Satellites played an integral role in the analysis and forecast of Sandy's track and intensity. The NOAA National Hurricane Center, Ocean Prediction Center, and Weather Prediction Center all relied on information from satellites to make critical warning decisions using various satellite products that assist with diagnosing tropical cyclone intensity, surface winds over the ocean, and heavy precipitation. All of the skillful global forecast models used satellite data for initiation to better forecast the track and intensity of Sandy. As part of the **Geostationary Operational Environmental Satellite – R-series (GOES-R)** and **Joint Polar Satellite System (JPSS)** Proving Ground activities, new satellite products were available to forecasters at these national centers in experimental form to assist with observing this unique, high impact event. This paper will demonstrate how the current satellite products assisted NOAA forecasters during Sandy and introduce some new satellite products that could be used to analyze and predict future high impact weather systems.

© 2015 Elsevier B.V. All rights reserved.

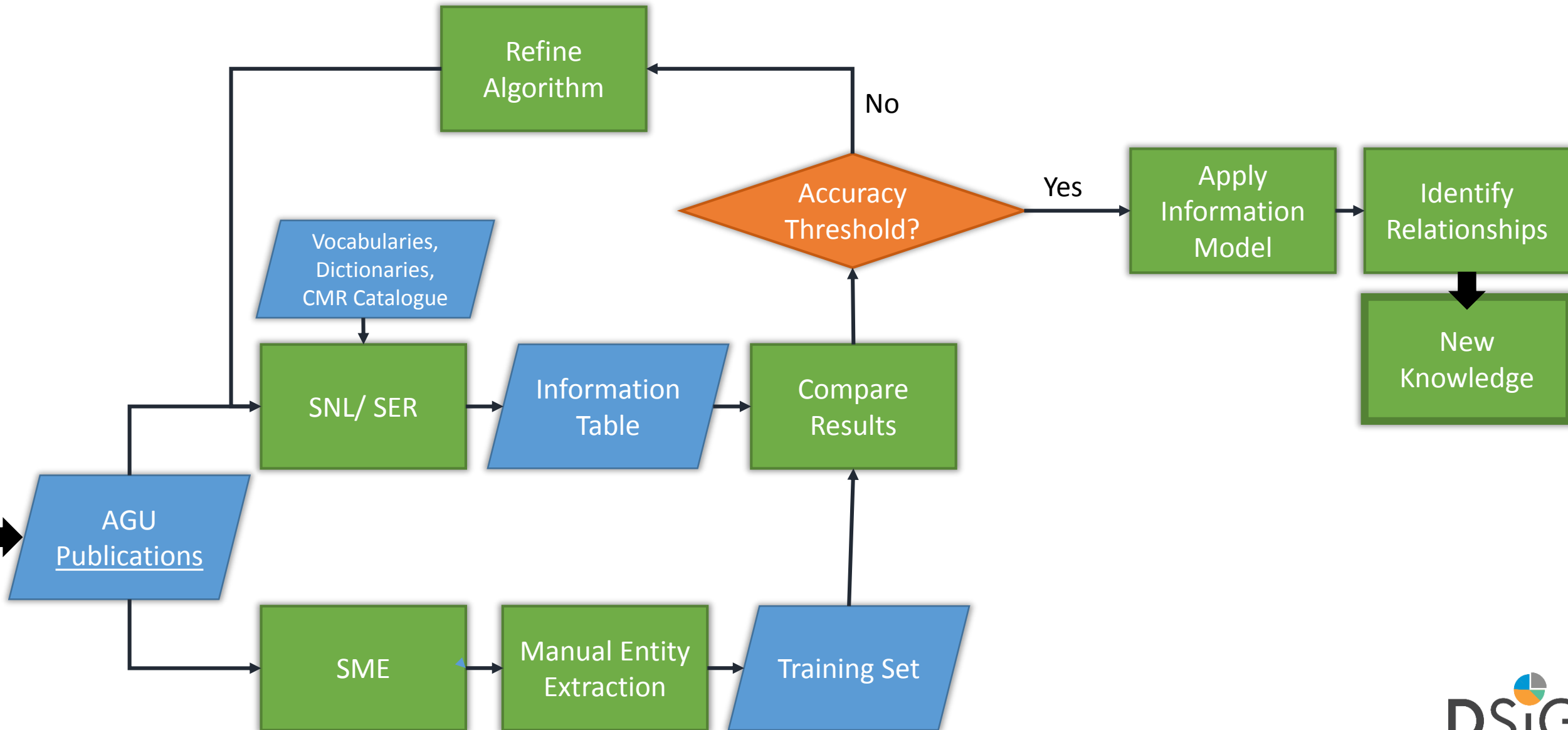
**1. Introduction**

From a meteorological perspective, Hurricane Sandy (henceforth simply referred to as Sandy) was a "perfect storm". The convergence of several synoptic features phased together along the U.S. Mid-Atlantic coastline to create record low-pressure, a huge wind field with corresponding storm surge, and copious amounts of precipitation in some areas, including record snowfall. Sandy caused over 250 deaths and upwards of \$70 billion in damage and economic loss during its trek from the Caribbean northward to the mid-Atlantic and Northeastern United States (Blake et al., 2013).

Satellite data are utilized in three primary ways for tropical cyclone (TC) analysis and prediction. First, they are used for situational awareness. For example, water vapor imagery loops from geostationary satellites are often used subjectively to assess the flow in the cyclone environment. Second, quantitative satellite products are used for storm analysis, such as satellite-based cyclone position and intensity estimates and sea surface temperature analyses. Third, satellite data are assimilated into numerical forecast models. All three of these applications were heavily utilized in Sandy. For example, numerical weather forecasts were generally very good, with models such as the National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Prediction's (NCEP) Global Forecast System (GFS) and

\* Corresponding author at: NOAA Center for Weather and Climate Prediction (NCWCP), 5830 University Research Court, Rm. 4000, College Park, MD 20740.  
E-mail addresses: [Michael.Folmer@noaa.gov](mailto:Michael.Folmer@noaa.gov) (M.J. Folmer), [Mark.DeMaria@noaa.gov](mailto:Mark.DeMaria@noaa.gov) (M. DeMaria), [Ralph.Ferraro@noaa.gov](mailto:Ralph.Ferraro@noaa.gov) (R. Ferraro).


# Technical Approach





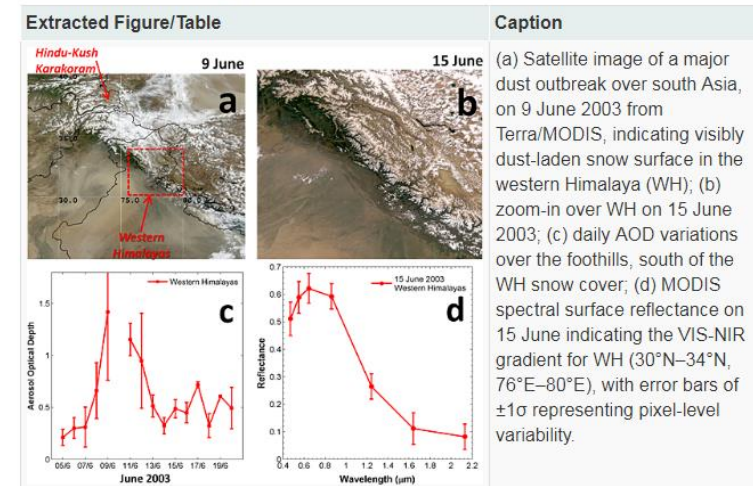
# Preliminary Results

Example of extracted and populated information from unstructured sources

 <span style="float: right;"> <a href="#">Recommendation</a> <a href="#">Resources</a> <a href="#">Publication</a> <a href="#">Analytics</a> <a href="#">Collaboration</a> <a href="#">Scientist</a> </span>	
<h2>Satellite observations of desert dust-induced Himalayan snow darkening</h2> <p> <a href="#">MetaData</a> <a href="#">Figures and Captions</a> <a href="#">Analytics</a> <a href="#">Sections</a> <a href="#">Micro Article</a> <a href="#">References</a> <a href="#">Evaluation</a> <a href="#">Proofread</a> </p>	
<b>Paper Title</b>	Satellite observations of desert dust-induced Himalayan snow darkening
<b>Authors</b>	Ritesh Gautam , Teppei J. Yasunari , Ritesh Gautam , N. Christina Hsu , William K.-M. Lau , Teppei J. Yasunari
<b>Missing Acronym</b>	BC ( black carbon ); HTP ( Himalaya-Tibetan Plateau ); WH ( western Himalaya ); WH ( western Himalaya ); SCF ( snow cover fraction ); IGP ( Indo-Gangetic Plains ); WH ( western Himalaya ); AOD ( Aerosol Optical Depth ); (NIR) ( n ); TOA ( top-of-atmosphere ); $\mu\text{m}$ ( ); RTM ( radiative transfer model ); SSA ( single scattering albedo );
<b>Id</b>	55
<b>Title</b>	Satellite observations of desert dust-induced Himalayan snow darkening
<b>Year</b>	2013
<b>Date</b>	16-Mar 2013
<b>Url</b>	<a href="http://onlinelibrary.wiley.com/doi/10.1002/grl.50226/full">http://onlinelibrary.wiley.com/doi/10.1002/grl.50226/full</a>
<b>Document Id</b>	55
<b>Flag</b>	0
<b>Subject Category</b>	Geology
<b>Channel Name</b>	GEOPHYSICAL RESEARCH LETTERS
<b>Key Words</b>	HYDROLOGICAL CYCLE; SPECTRAL ALBEDO; TIBETAN PLATEAU; MONSOON; AEROSOLS; PRODUCTS; GLACIERS; IMPACTS; CLIMATE; COVER
<b>Author Keyword</b>	Dust, Snow; Remote Sensing

<b>Topic</b>	climate(87.45%); dust(12.54%); aerosols(0.01%);
<b>Topic Keywords</b>	HYDROLOGICAL CYCLE; SPECTRAL ALBEDO; TIBETAN PLATEAU; MONSOON; AEROSOLS; PRODUCTS; GLACIERS; IMPACTS; CLIMATE; COVER
<b>Instruments</b>	MODIS(Moderate-Resolution Imaging Spectroradiometer)
<b>Datasets</b>	Transit (C1214558371-NOAA_NCEI) Terra 1 MODIS Imagery (C1214598068-SCIOPS) Aqua 1 MODIS Imagery (C1214598104-SCIOPS)
<b>Variables</b>	Toa cloud radiative effect, Solar zenith angle, Snow grain size, Normalized difference vegetation index, Surface snow area fraction, Tropopause instantaneous radiative forcing
<b>Organizations</b>	TOA, SCF, SSA, IGP, HTP
<b>Sponsored Organizations/Projects</b>	NASA

### 3 Extracted Figure/Table

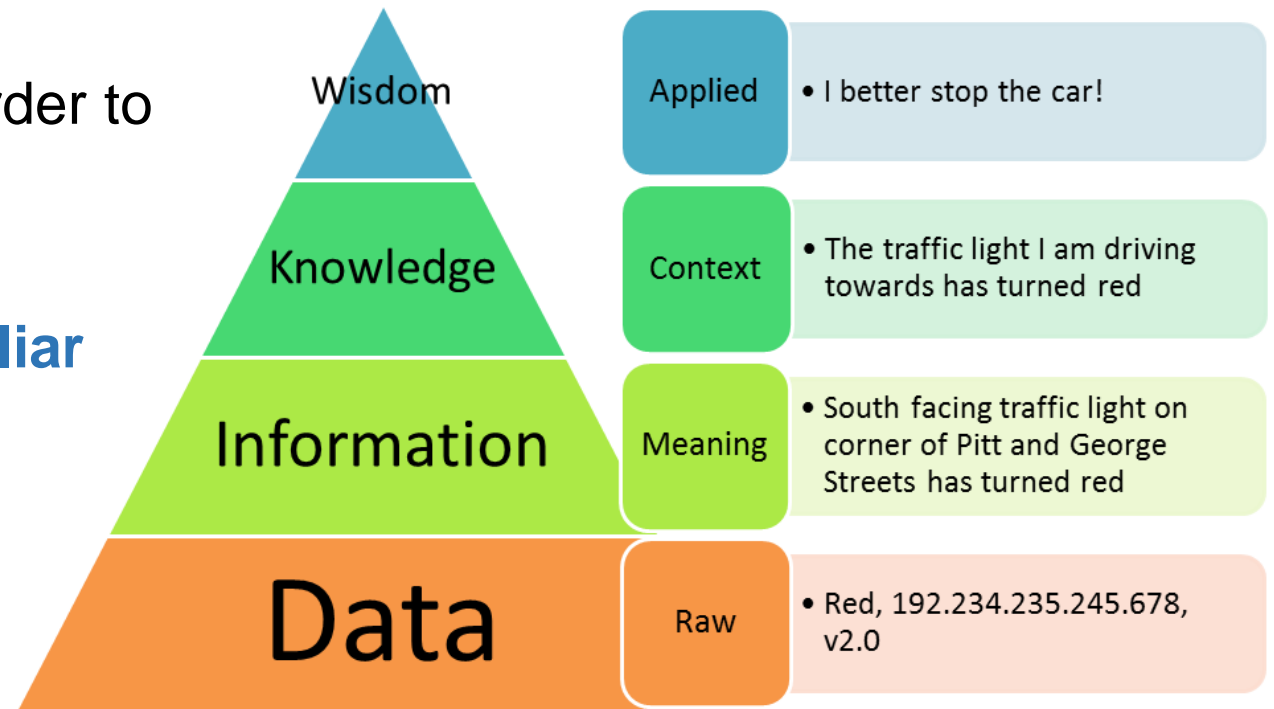


# Data Use Challenges

To address these challenges, what difficulties are presented?

How can data resources be linked in order to improve the data spin-up time?

How can we work to educate unfamiliar users?



© 2011 Angus McDonald

# Other Resources

Challenge - Publications and technical documents often prove difficult for new and unfamiliar users to digest.

NASA Global Hydrology Resource Center (**GHRC**) Data Active Archive Center (**DAAC**)

NASA Short-term Prediction Research and Transition Center (**SPoRT**)

**What resources are available to introduce data, methods and concepts?**

- GHRC DAAC Data Recipes
- GHRC DAAC Micro Articles
- NASA SPoRT Quick Guides





# NASA GHRC DAAC Data Recipes

## What is a Data Recipe?

Tutorials or step-by-step instructions to help users learn how to discover, visualize and use data, information, software and techniques.

## Types of Data Recipes

- Using netCDF data in ArcGIS
- GHRC tool tutorials
- Python notebooks and scripts
- Data format conversions and georeferencing

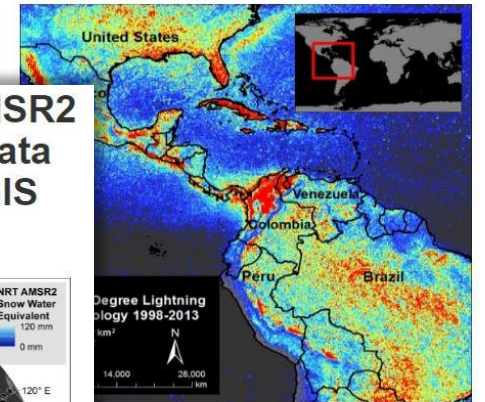
[Discover GHRC Data Recipes Here](#)

## Using ArcGIS to Convert LIS Very High Resolution Gridded Lightning Climatology NetCDF Data to GeoTIFF Format

[Description](#) | [How to Use](#) | [Dataset Information](#) | [Key Parameters](#)

### Description

The Lightning Imaging Sensor (LIS) aboard the Tropical Rainfall



### Supporting Software Information

ACCESS



Restricted, License Required

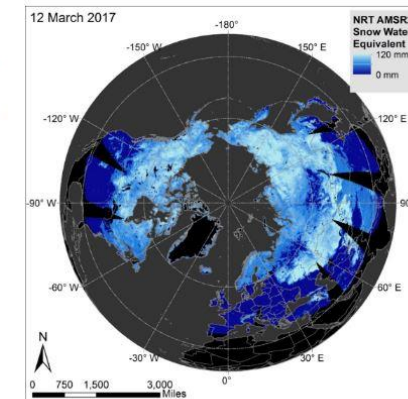
## How to Georeference and Convert NRT AMSR2 Snow Water Equivalent Polar EASE-Grid Data to GeoTIFF Format using Python and ArcGIS

[Description](#) | [How to Use](#) | [Dataset Information](#) | [Key Parameters](#)

### Description

The near real-time (NRT) Land Atmosphere Near real-time Capability for EOS (LANCE) AMSR2 Daily Global Snow Water Equivalent (SWE) EASE-Grids dataset contains SWE and quality assurance flag information for the Northern and Southern Hemispheres. These data are available in HDF-EOS5 format requiring georeferencing in order to display according to the NSIDC EASE-Grid format. This data recipe employs Python to georeference and create four GeoTIFFs of the SWE and quality assessment flag layers within the HDF-EOS5 files. The data are brought into ArcMap where the projection is defined to enable further data processing, analysis and map making. This data recipe requires a pre-installed version of ArcMap, Python and the necessary Python packages.

*Image created using the NRT AMSR2 Snow Water Equivalent Polar EASE-Grid dataset in ArcMap 10.2*



### Data Recipe Type



Data Format Conversion

### Supporting Software Information

#### TYPE



Python Script



ArcMap 10.2+

#### ACCESS



Open Source



Restricted, license required

# NASA GHRC DAAC Micro Articles

## What is a Micro Article?

A short, interesting document that brings together data and key science concepts

Creates a knowledge base for users by curating around GHRC's data and science thematic areas

## Types of Micro Articles

- Instruments
- Phenomena
- Events or Case Studies
- Publications

[Discover Micro Articles Here](#)

Home >> GHRC Micro Articles >> Phenomenon >> Lightning

## LIGHTNING



Atmospheric Phenomenon

### WHAT IS LIGHTNING?

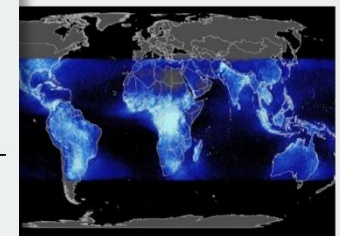
Lightning is the electrical discharge between positively and negatively charged regions within clouds. The electrical discharge serves as an equalization process between the charged regions, and can travel from cloud-to-cloud, cloud-to-ground, or cloud-to-air. Visually, lightning appears as a bright flash of light, or a stroke,



[https://commons.wikimedia.org/wiki/File:Pink\\_Lightning.jpg](https://commons.wikimedia.org/wiki/File:Pink_Lightning.jpg)

### How does lightning occur?

The particles within a cloud interact with each other, causing the particles to fracture and break apart. It is currently believed that smaller ice particles tend to acquire a positive charge, while the larger particles acquire a more negative charge. Under the influences of thunderstorm updrafts, these particles separate until the upper portion of the cloud acquires a net positive charge, and the lower portion of the cloud becomes negatively charged. This separation of charge creates an electrical potential both within the cloud and between the cloud and ground. Eventually, the electrical resistance in the air between the charged regions breaks down and a flash begins. The resulting lightning strokes are an electrical discharge between positive and negative regions of a thunderstorm.



<http://dx.doi.org/10.5067/LIS/LIS/DATA301>



Volcanic eruptions



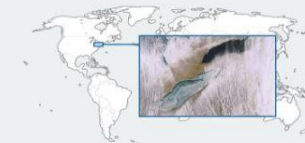
Wildfires

## Lake Effect Snow Event during GCPEX Field Campaign

### Event

#### What happened and why it happened

A lake effect snow event occurred during the GPM Cold-season Precipitation Experiment (GCPEX) field campaign in Ontario, Canada during the 2011-2012 winter season. Cold, northwest winds moved across the Georgian Bay in eastern Lake Huron and picked up moisture from the lake that fed the development of clouds and snow that varied considerably across a small region south of the bay. The snow clouds developed into persistent narrow bands that resulted in 2 inches of snowfall accumulation at one of the ground sites whereas only 12 miles away they produced 16 inches of snow.



### Science Question

Lake effect snow is generated when cold air moves over warm lake waters such that narrow bands of snow clouds develop. The warmer lake waters heat the lower portions of air causing it to become less dense and begin to rise. As this moisture-laden, warmer air rises it begins to cool leading to condensation and the formation of clouds that can become rather tall enabling the growth of very large snowflakes. Lake effect snow bands can produce snowfall rates exceeding 5 inches an hour, especially if the wind is directed along the largest width of the lake so that a great deal of moisture is continually supplied to the clouds.



#### SPATIAL COVERAGE

[N: 47, W: -80.2, E: -67.7, S: 43.5] degrees



#### TIME RANGE

February 10-12, 2012



#### EVENT TYPE

Lake Effect Snow

### Get Data

The GPM Cold Season Precipitation Experiment (GCPEX) was a field campaign that occurred in Ontario, Canada during the 2012 winter season. The objective of the GCPEX campaign was to study snowfall's physical and radiative properties from the ground through the atmosphere. These measurements are used to help scientists understand the minimum snow rate that can be detected from space and also how well space sensors can discriminate between snow, rain and clear air. Measurements were taken from five ground sites and three research aircraft to provide as complete a sampling as possible.



# NASA SPOrT Quick Guides

## What is a Quick Guide?

Short, easy to use resources that highlight key aspects of a data product or tool.

Intended to assist forecasters in quickly recalling information during times of operation.

## Available Forms

- Download/print
- Interactive web browser
- Interactive through personal display system

[Discover SPOrT Quick Guides here](#)

**Daytime Convection RGB Quick Guide**

**Why is the Convection RGB imagery important?**

The Daytime Convective Storms (Convection) RGB was designed for identification of convection with strong updrafts and small ice particles indicative of severe storms. This RGB helps increase nowcasting capabilities of severe storms by identifying the early stage of strong convection. Knowing the microphysical characteristics of convective clouds helps determine storm strength and stage to improve nowcasts and short-term forecasts. Bright yellow in the RGB indicates strong updrafts prior to the mature storm stage.

**Convection RGB Recipe**

Color	Band / Band Diff. (μm)	Physically Relates to...	Small contribution to pixel indicates...	Large Contribution to pixel indicates...
Red	6.2 – 7.3	Cloud height	Low clouds	High clouds
Green	3.9 – 10.4	Particle size	Large ice or water particles, weak updrafts	Small ice or water particles, strong updrafts
Blue	1.6 – 0.64	Cloud phase	Ice clouds	Water clouds

**Impact on Operations**

**Primary Application**  
**Convection and Severe Weather:** identify intense updrafts that indicate strong convection.  
 Strong convection is bright yellow...

**Limitations**

**Daytime only application:** the RGB relies on solar reflectance from visible, near-IR, and shortwave IR channels.  
 Direct solar impact by sun (satellite viewing...)

**Limitations**

...true color reflectance channels, and ... at night.

...distinguish snow and clouds...



# Other Resources

## **How do these benefit Earth science and hydrologic applications?**

- These resources provide introductory information that is easy to read and understand without overwhelming users.
- Each point to additional documents for more detailed information.
- Each contain information on commonly used data, models, and software.
- They link directly to data, helping users understand a dataset and how to apply it towards research or applications.
- Populating this information within a knowledge graph allows users to search and discover information on data and methods for a broad user community.

# Next Steps

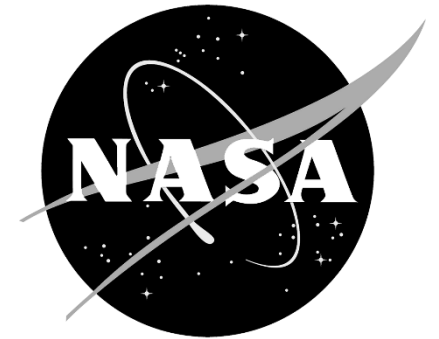
- Investigate generating easy to understand resources in a structured format to allow more seamless integration within the knowledge graph.
- Continue refining SER for Earth Science
- Continue building and evaluate a training set for SER (working with graduate students and SMEs)
- Scale efforts to to all Earth Science related journal titles in the Wiley Online Library
- Begin mining graphs to obtain new information
  - Prediction of relationship between entities (i.e., Network Link Prediction)
  - Automatic generation of new content (e.g., MicroArticles)

# Benefits to NASA Earth Science and Hydrology

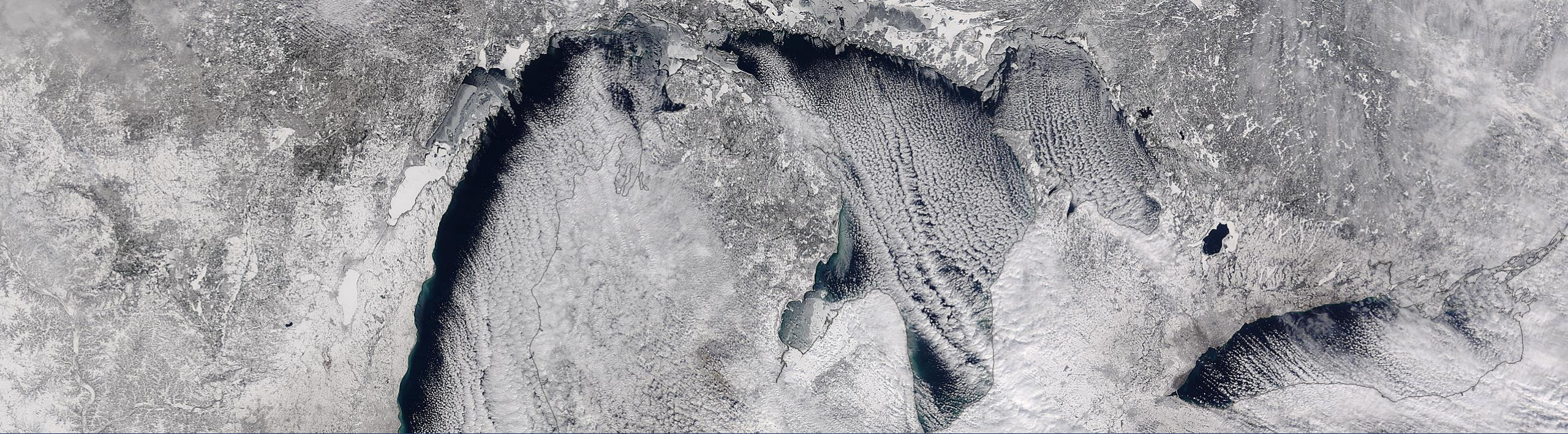
- Addresses the challenge in navigating the increasing volume of data and information.
- To provide an operational knowledge base to enhance NASA's Earth science research.

## Beneficial Applications

- Hypothesis formulation and testing:
  - Automate the search for and compilation of background information.
  - Given a topic, what hypotheses have been tested?
  - What data/tools are being used to test a hypothesis?
  - Common paths to knowledge discovery.
- Mission development/review:
  - What kinds of instruments/parameters are needed to specify science objectives?
  - Impact of a mission by linking it with publications and dataset distribution.







# Thank you, questions?

[amanda.m.weigel@nasa.gov](mailto:amanda.m.weigel@nasa.gov)

NASA/MSFC Data Science Informatics Group



Carnegie  
Mellon  
University

**UAH**  
THE UNIVERSITY OF  
ALABAMA IN HUNTSVILLE

