

National Aeronautics and
Space Administration

EARTHDATA

The Signal Through the Noise

Making Sense of the Torrent of Metrics
Data

07/24/2025

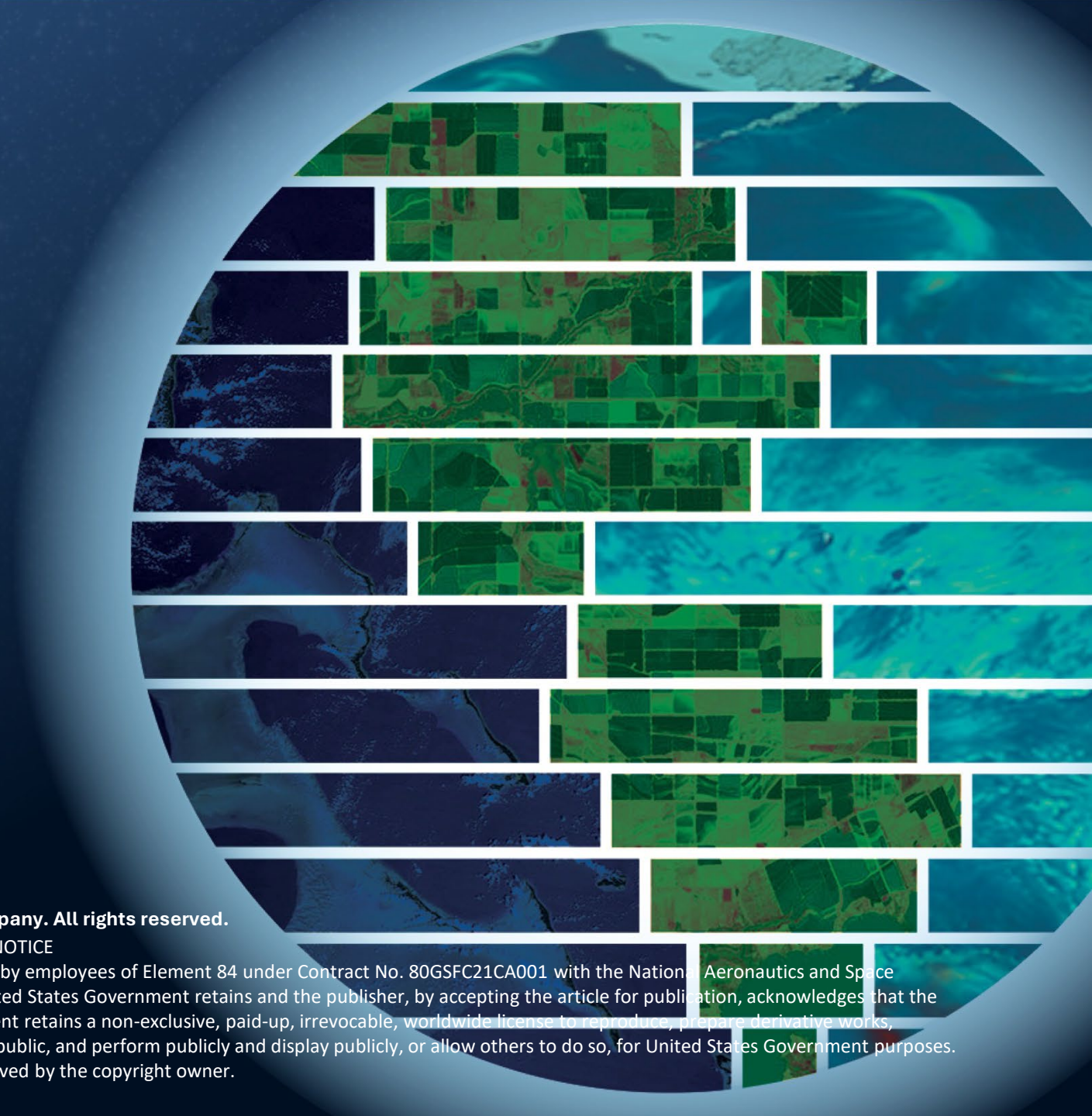
Jonathan Blake
Element 84

Product Owner, Global Metrics

©2025 Raytheon Company. All rights reserved.

GOVERNMENT RIGHTS NOTICE

This work was authored by employees of Element 84 under Contract No. 80GSFC21CA001 with the National Aeronautics and Space Administration. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, or allow others to do so, for United States Government purposes. All other rights are reserved by the copyright owner.



Introductions

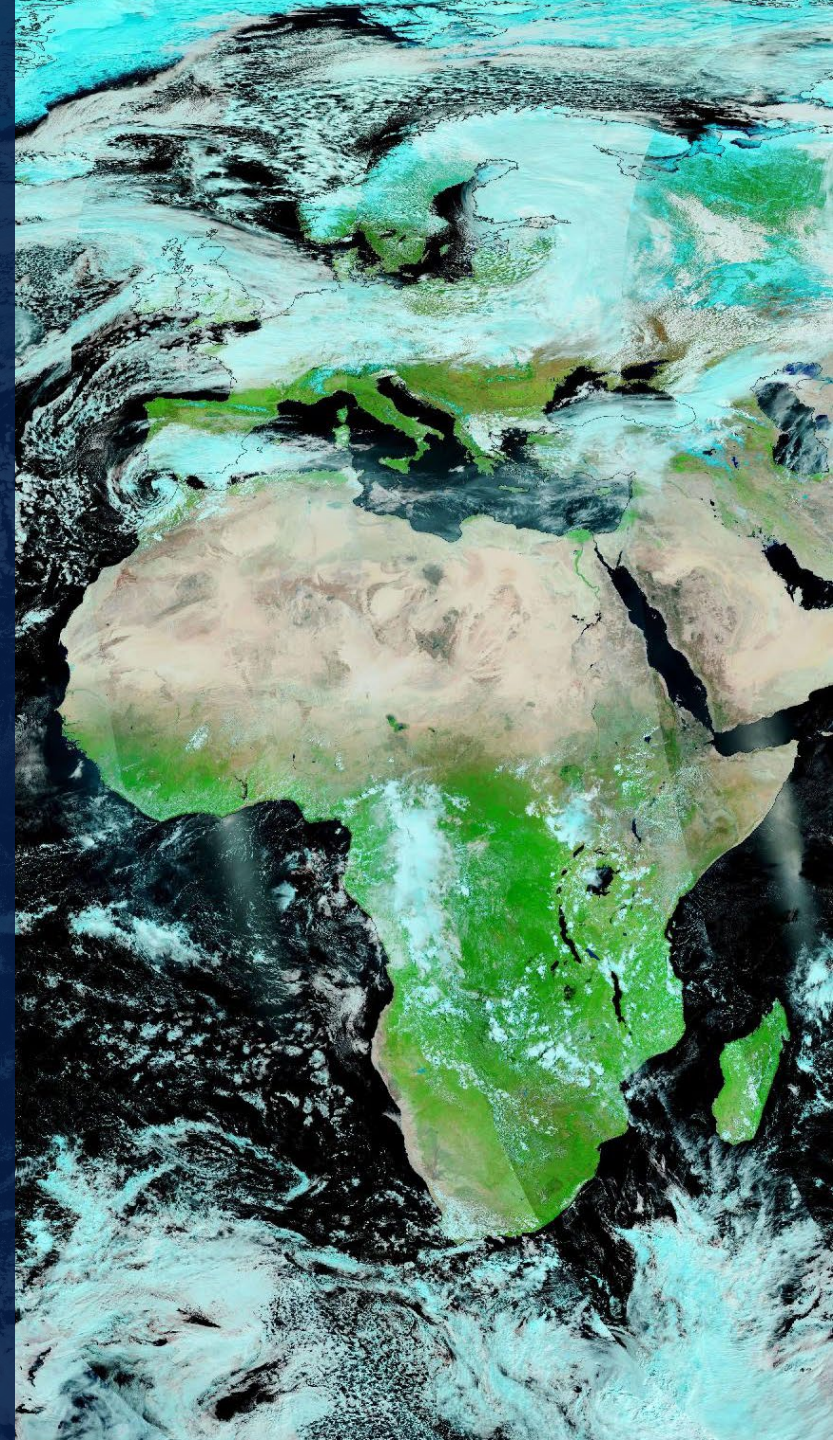


Jonathan Blake

NASA ESDIS

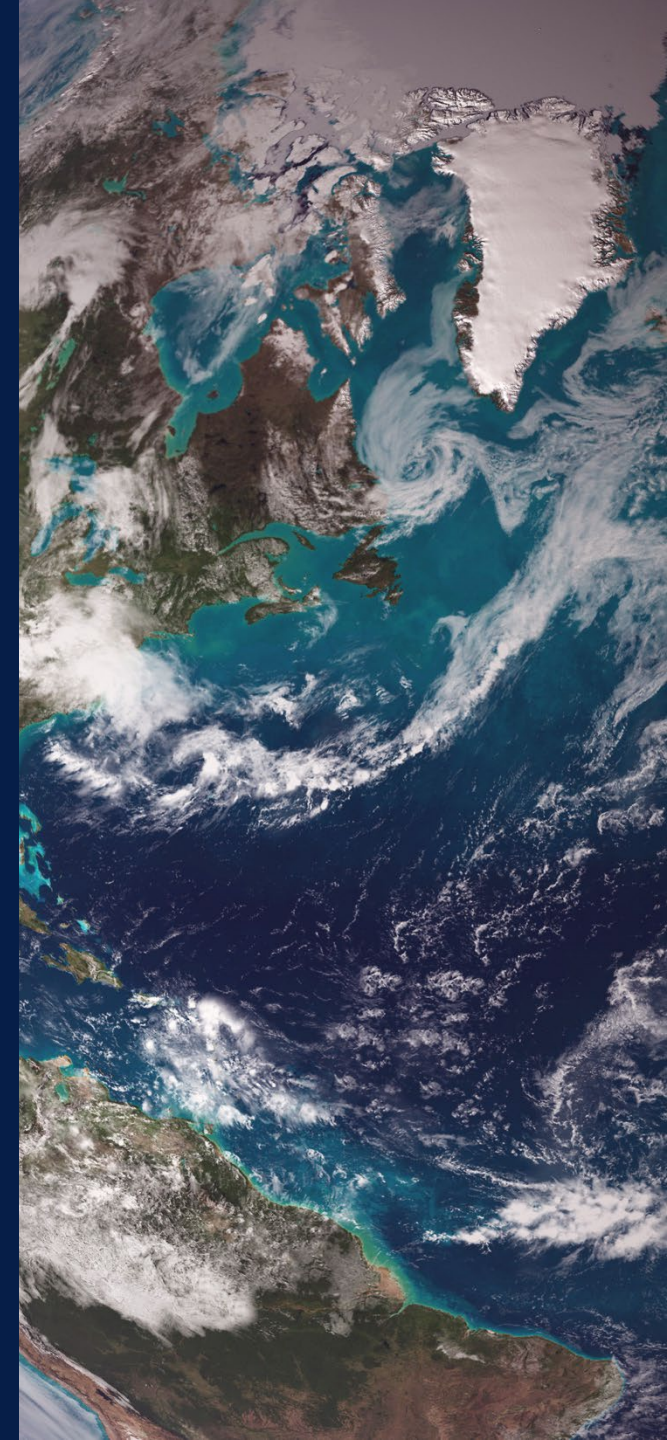
Global Metrics Product Owner

Jonathan.w.blake@nasa.gov

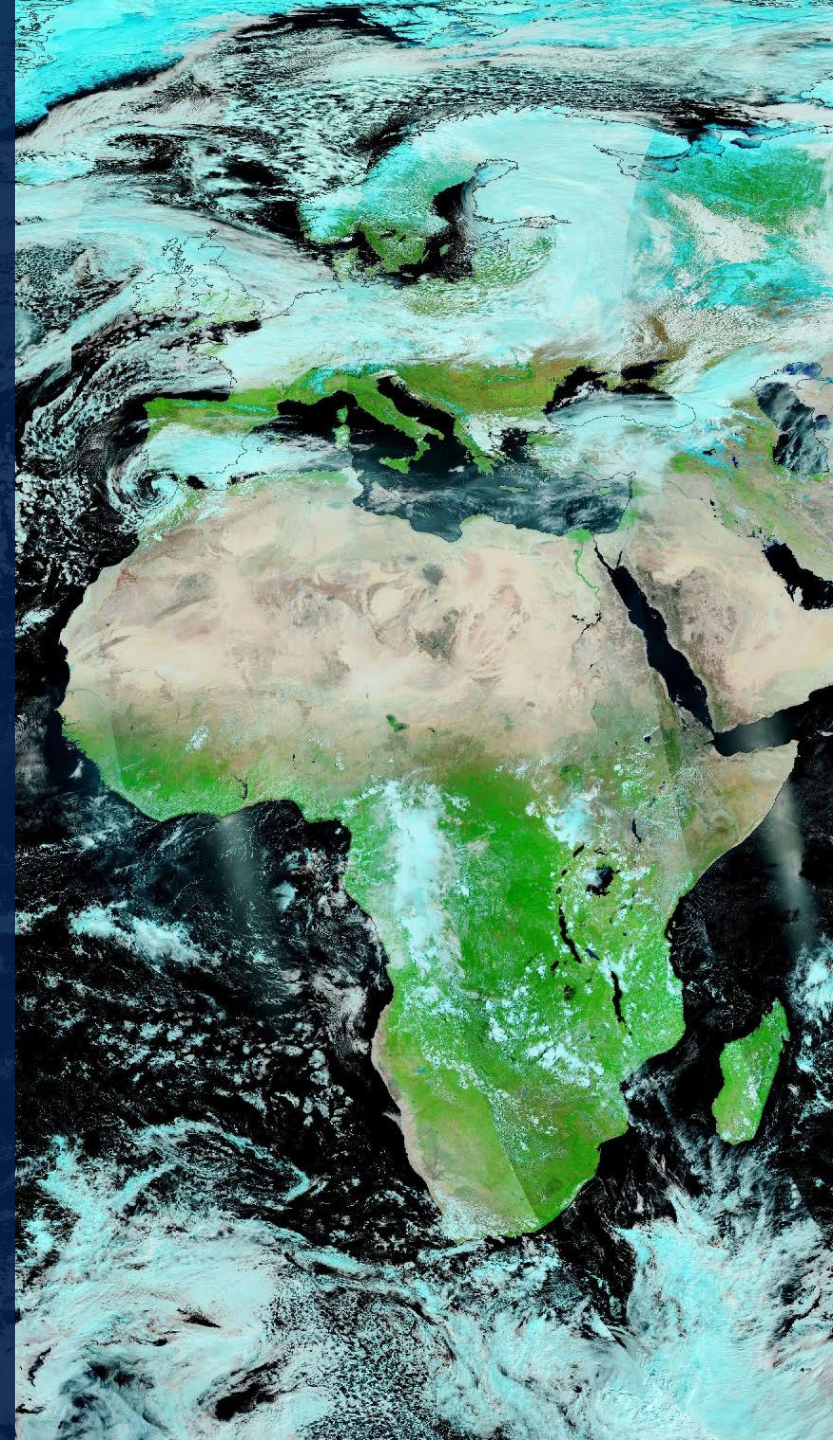


Agenda

- Business Intelligence as Storytelling
- Making Sense of the Data
- Using LLMs to Tell a Story



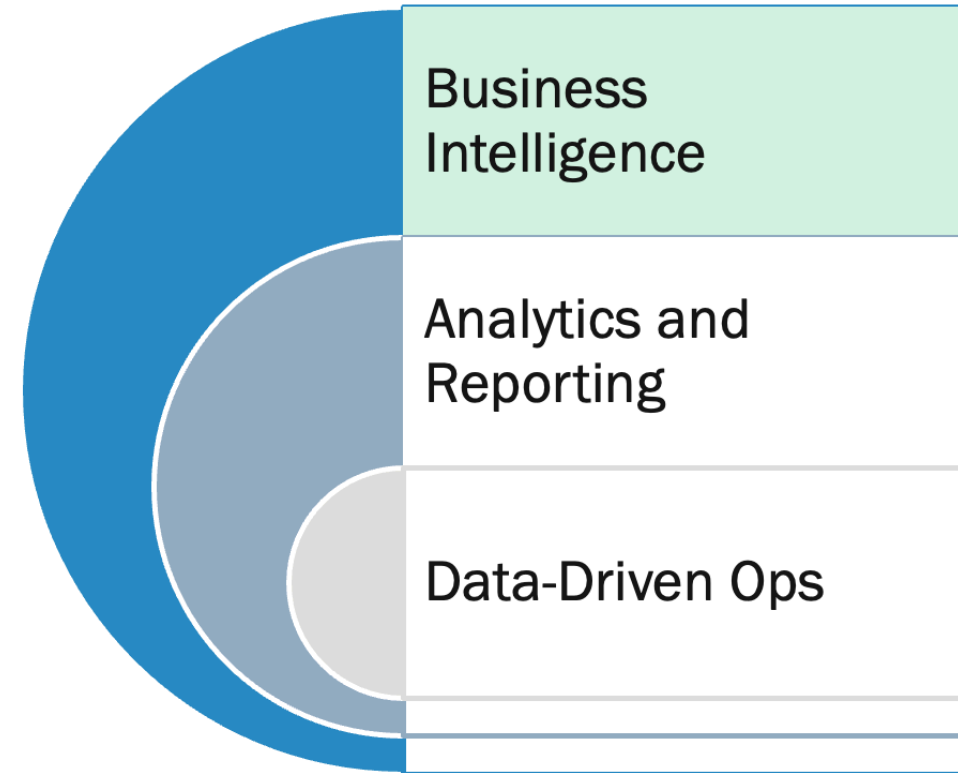
Business Intelligence as Storytelling



BI is not Your Typical Analytics

- Users at the highest levels of an organization have a **need for speed** when it comes to analytics
- BI tools should reduce the time and the complexity of telling these stories
 - Prioritize **efficiency** over **fidelity of information**
 - Prioritize **readability** over **thoroughness**

Three unique use cases for metrics data



Betting on LLMs

- Large Language Models are advanced storytelling tools!
- They can **efficiently** distil torrents of data to convey information in a **readable** way (plain English!)
- To leverage LLMs, we first need to produce data that the LLMs can understand

Amazon Quicksight with Q

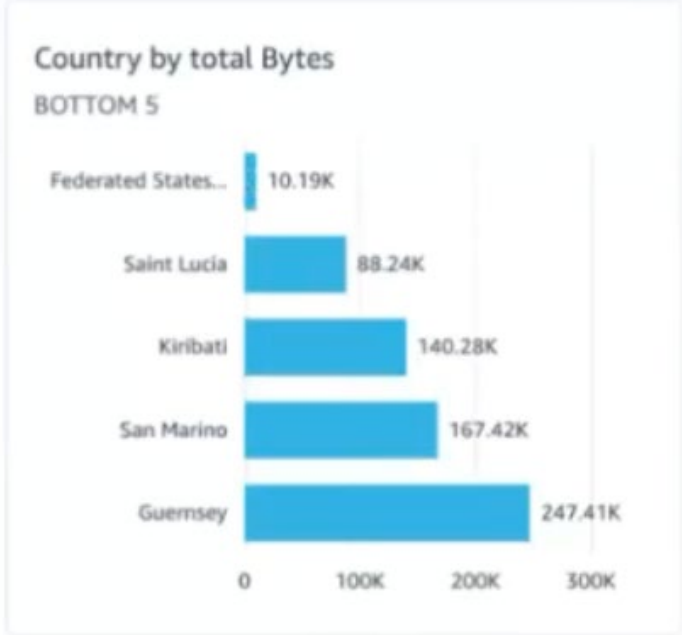
topic ▼ PINBOARD

What are the bottom 5 countries for bytes downloaded?

Interpreted as: **Bottom 5 Country by total Bytes.**

al of 225 unique total bytes for the tries is 653,543. th the lowest Federated States of 10,193 bytes. tom 5 countries Guernsey has and San Marino tes. There are 6 ave the same ss all countries.

Country by total Bytes
BOTTOM 5

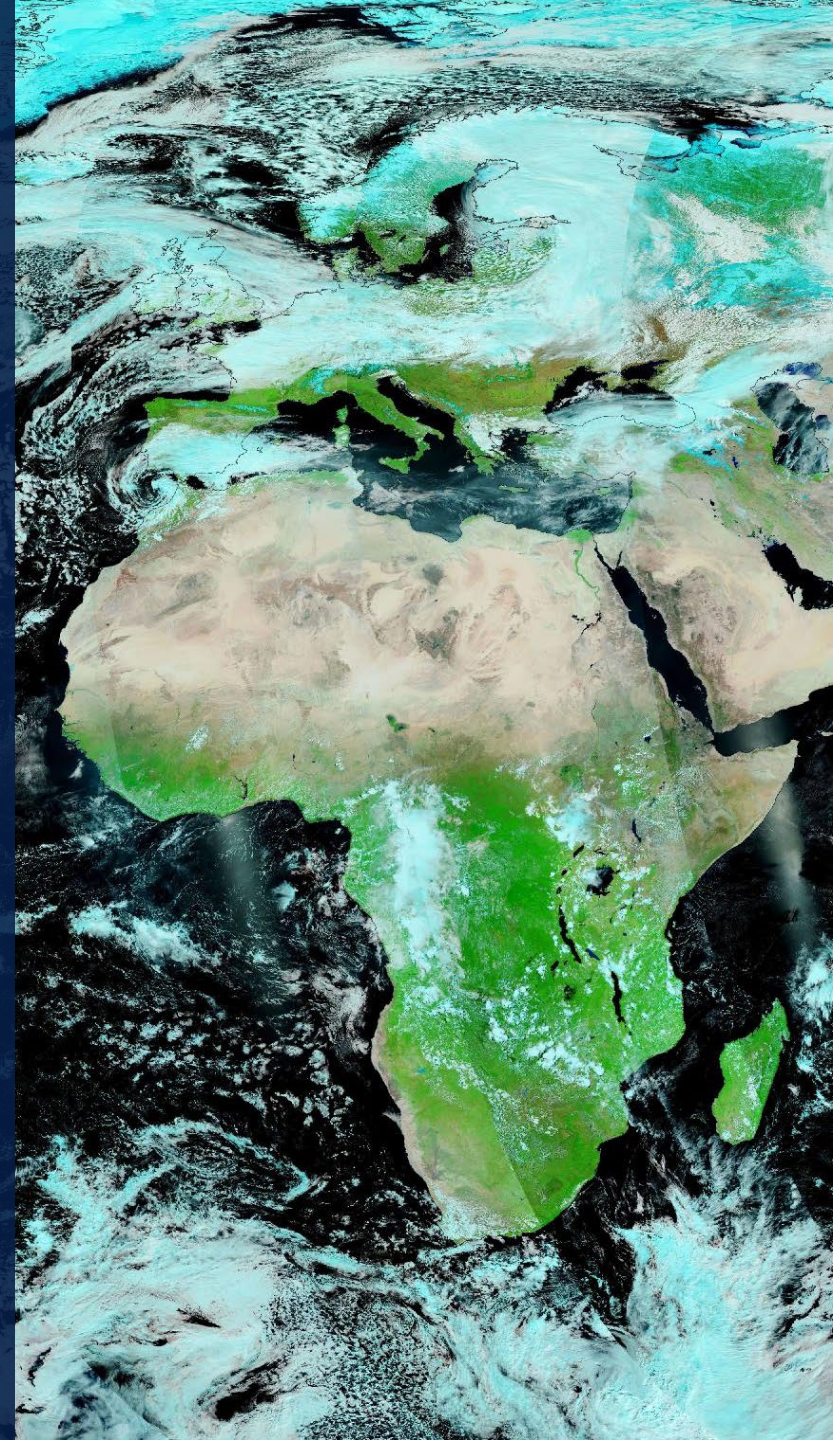


Country	Total Bytes
Federated States of Micronesia	10.19K
Saint Lucia	88.24K
Kiribati	140.28K
San Marino	167.42K
Guernsey	247.41K

Country and Bytes
BOTTOM 5 COUNTRYS BY BYTES

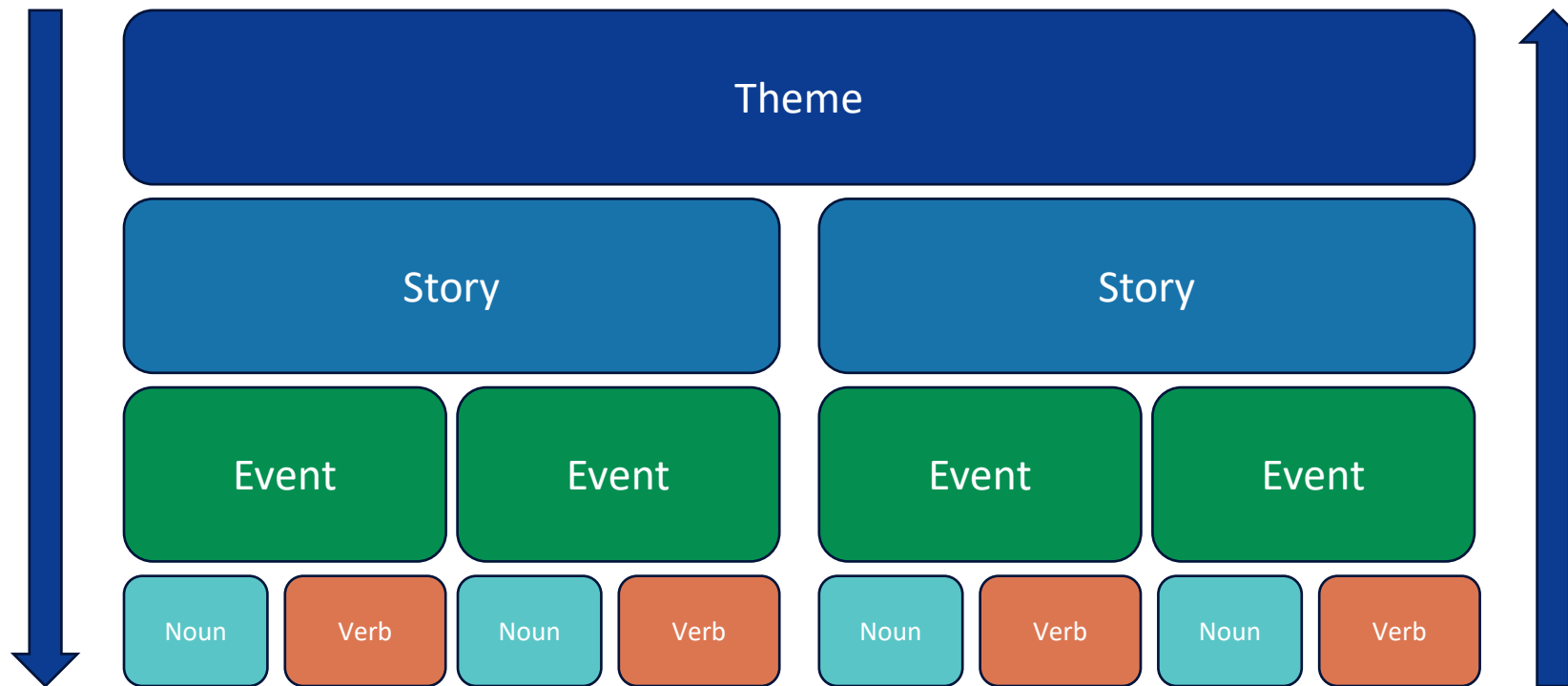
Country	Bytes
Guernsey	247.41K
San Marino	167.42K
Kiribati	140.28K
Saint Lucia	88.24K
Federated States of Micronesia	10.19K

Making Sense of the Data



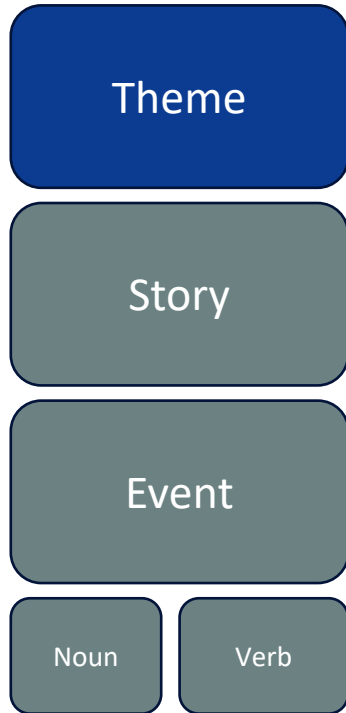
Storytelling 101

Start with a theme, find the stories and the events that comprise them (data)



Use the LLM to summarize the events into stories, and the stories into themes

Start with a Theme



Efficiency

Tell me stories about how efficient (or inefficient!) we are.

Find Stories to Convey the Theme



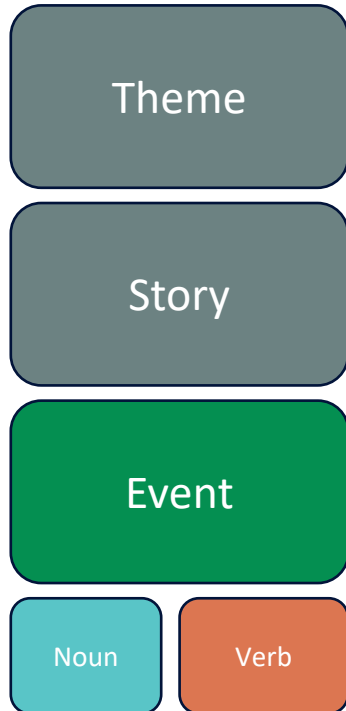
What do we do?

Look for the verbs!

- **Process, archive, and distribute** Earth science data
- **Provide tools** to facilitate the above

“Cloud-hosted DAACs
distributed science data
efficiently.”

Find Events to Tell the Story

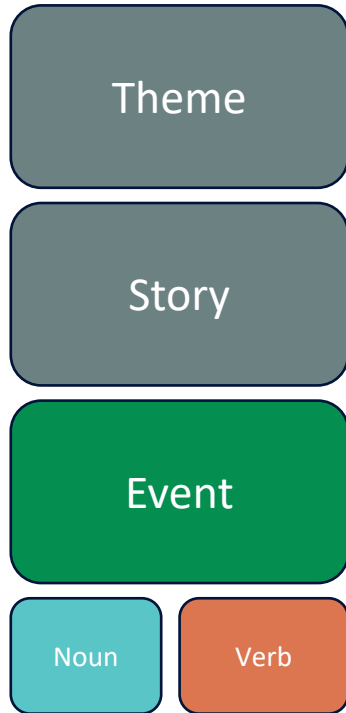


How do we distribute efficiently?

You might see:

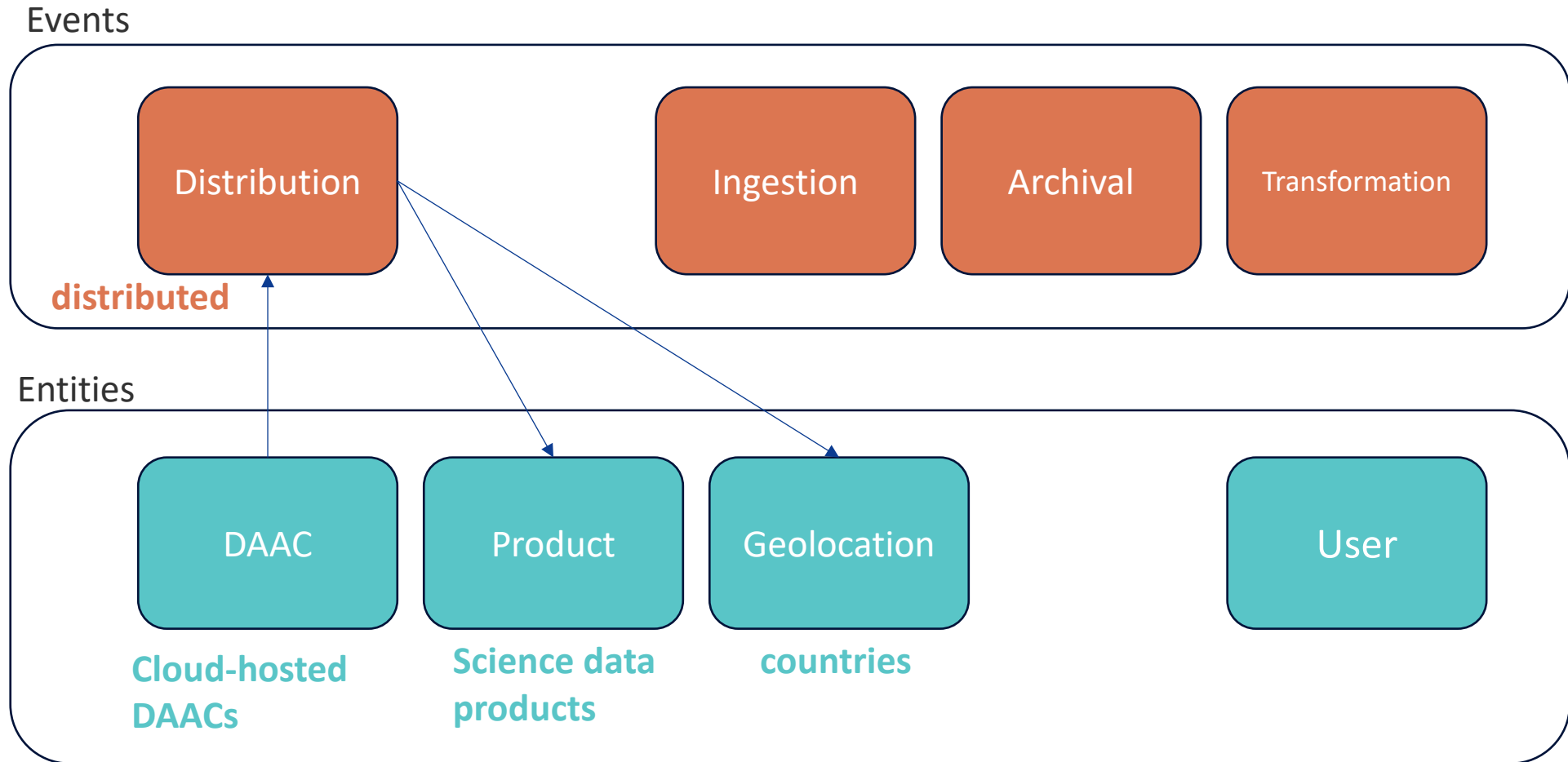
- Large **volumes** of distribution, in terms of
 - Product count
 - Terabytes
- Distribution **spread** to a large variety of
 - Countries
 - Users
 - Companies / organizations

Break the Events Down



“Cloud-hosted DAACs distributed science data products to _____ countries in 2024.”

Build the Semantic Model



Turn Conceptual to Actual

Cloud-hosted
DAACs

DISTRIBUTION	
REQUEST_ID	TEXT
TIMESTAMP	TIMESTAMP
CLIENT_IP	TEXT
USER_ID	TEXT
DAAC	TEXT
OPERATION	TEXT
KEY	TEXT
BYTES	INTEGER
GEO_IP	OBJECT
REQUEST_TIME_MS	INT
...	...

distributed

GEOIP	
IP_ADDRESS	TEXT
LOCATION	TEXT
NETWORK	TEXT
POSTAL_CODE	TEXT
LATITUDE	INTEGER
LONGITUDE	INTEGER
ORGANIZATION	TEXT
DOMAIN	TEXT
...	...

countries

PRODUCT	
PRODUCT	VARCHAR
FILENAME	VARCHAR
METADATALONGNAME	VARCHAR
PRODUCT_LEVEL	VARCHAR
DAAC	VARCHAR
PRODUCT_TYPE	CHAR
DISCIPLINE	VARCHAR
INSERT_TIME	TIMESTAMP
UPDATE_TIME	TIMESTAMP
...	...

Science data
products

ETL

Extract / Translate / Load







Where can we find this info?



ETL

Extract / Translate / Load

- NLP and LLMs require data to be input in a **language** that can be interpreted (English!)
- Raw data inputs (L0 data) will have to be made readable (L1) through various data translations

 geoip.ip	34.73.128.2
 geoip.latitude	32.861
 geoip.location	POINT (-79.9746 32.8608)
 geoip.longitude	-79.975
 geoip.postal_code	29415
 geoip.region_code	SC
 geoip.region_name	South Carolina
 geoip.timezone	America/New_York

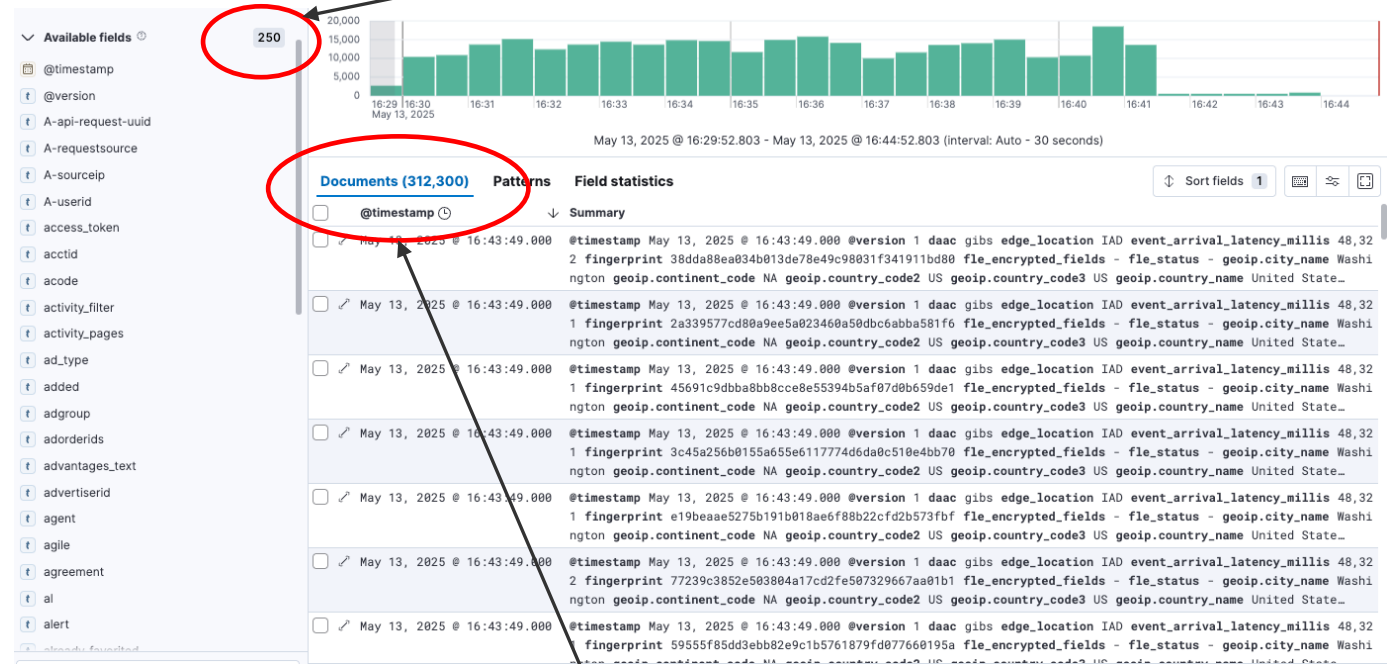


ETL

Extract / Translate / Load

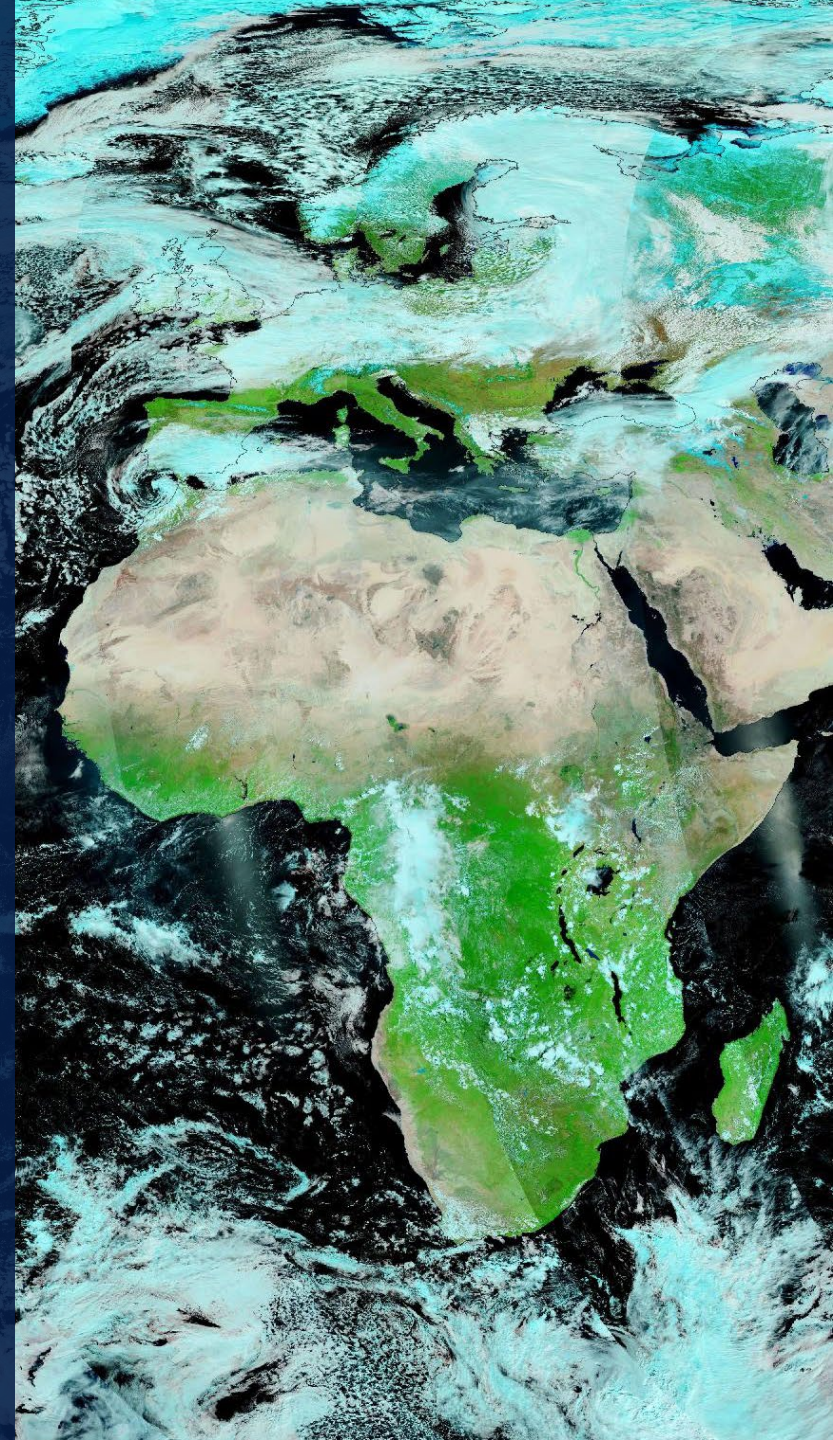
- Raw data will be noisy!
- For best LLM performance, we want to reduce the data we load without compromising any info vital to our story.
- Think **lossless compression**

L0 logs contain **fields with no relevance to our story**



L0 logs capture **every data access event**. This amount of information is generally not relevant to our story.

Using LLMs to Tell our Stories



Metrics Lakehouse

BI Tool



Amazon QuickSight



Amazon Q

Compute Engine



Amazon Athena

Data Catalog



AWS Glue

Storage Format

ICEBERG



Apache Iceberg

Storage



S3



S3

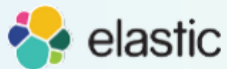


S3

ETL



Distribution Logs



Product Metadata



Geographic Data

Quicksight & Q for BI

Accelerate decision-making and enhance business productivity with Generative BI capabilities powered by Amazon Q in QuickSight.

- Build, uncover, and share valuable insights using **natural language prompts**.
- Simplify data understanding for business users through a **context-aware Q&A experience**.

Source: <https://aws.amazon.com/quicksight/q/>

NLP Analytics with Quicksight & Q

Theme

Story

Event

Noun

Verb

Include	Friendly name	Synonyms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Average Ingest latency	Ingest Latency X Latency X +
<input type="checkbox"/>	<input checked="" type="checkbox"/> Data center	DAACS X DAAC X +
<input type="checkbox"/>	<input checked="" type="checkbox"/> Discipline	Science Discipline X Subject X Keyword X +

- With our data ETL'd, there's one more step!
- Q requires **synonyms** for the data fields that we've provided, so that it can better understand the questions we ask

NLP Analytics with Quicksight & Q

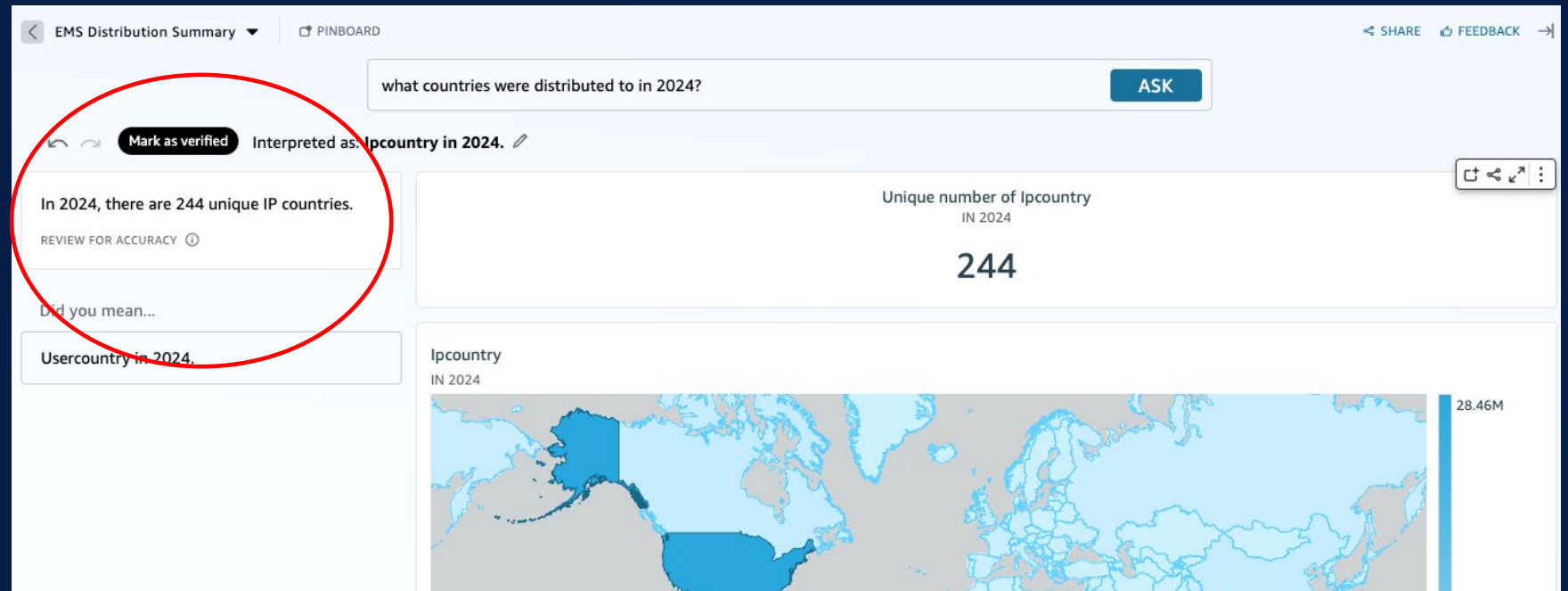
Theme

Story

Event

Noun

Verb



“Cloud-hosted DAACs distributed science data products to **244** countries in 2024.”

Creating Analyses with Quicksight

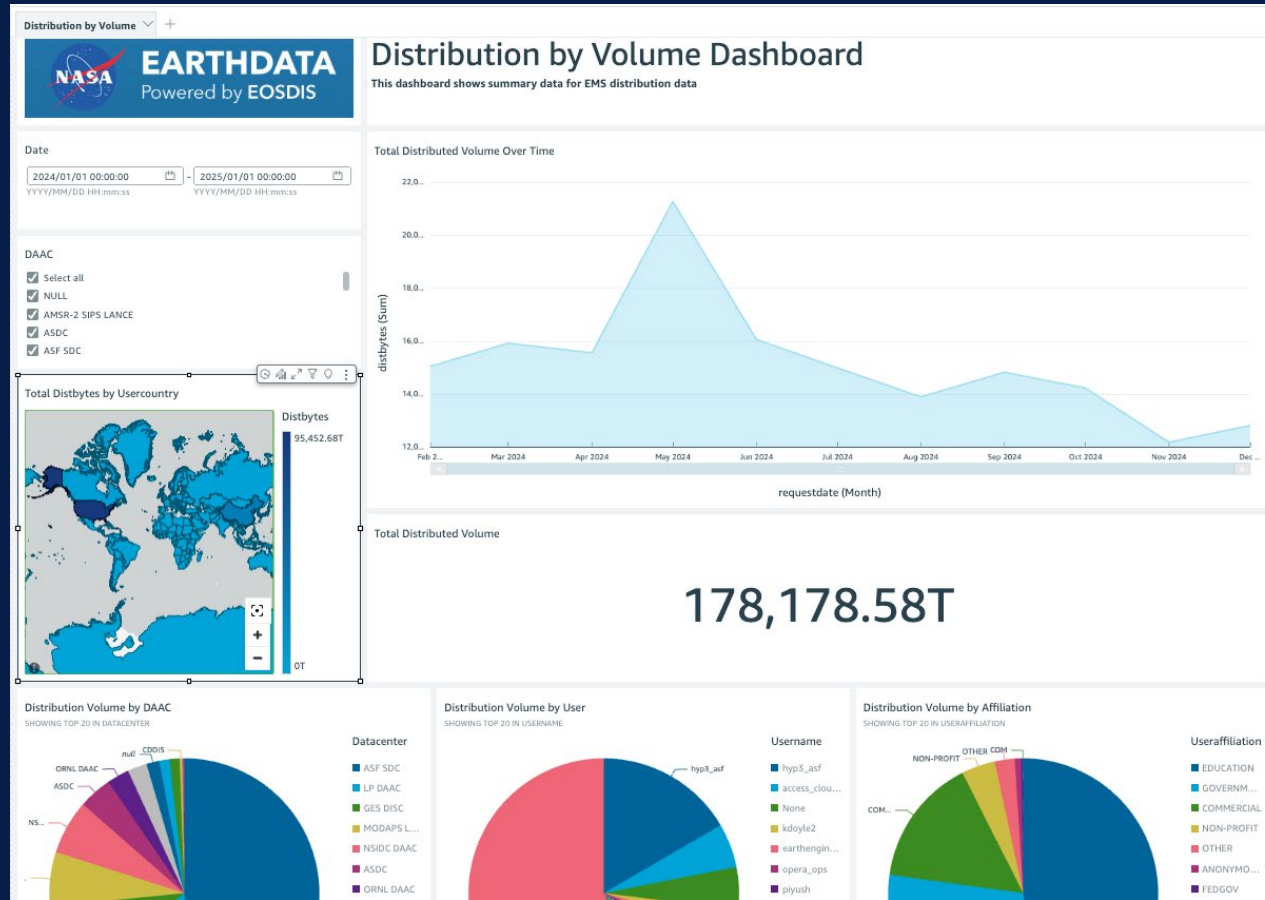
Theme

Story

Event

Noun

Verb



“Cloud-hosted DAACs distributed science data efficiently.”

Using natural language, we can build analyses to tell our story.

Creating Reports with Quicksight

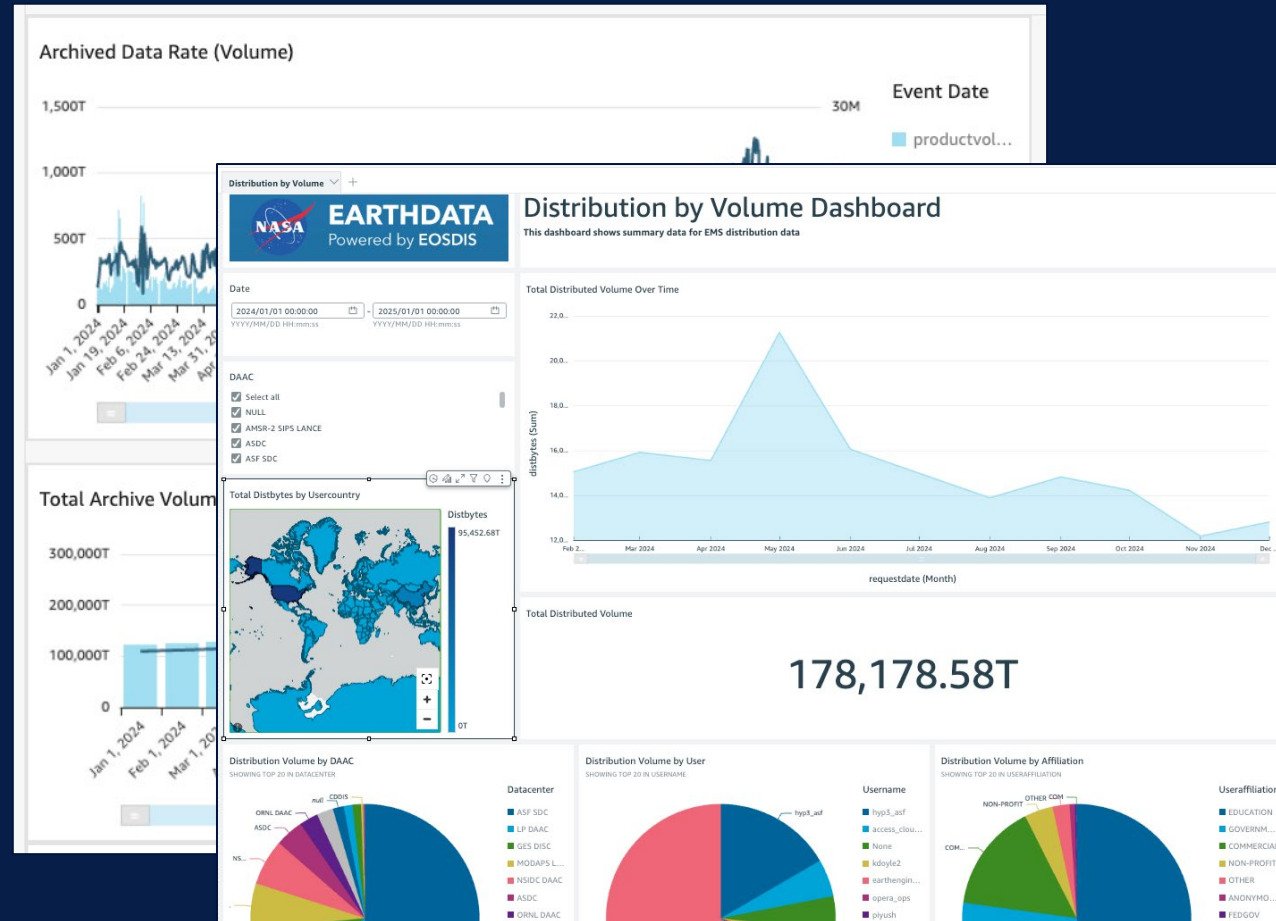
Theme

Story

Event

Noun

Verb



“Cloud hosted DAACs distributed, processed, and archived data efficiently.”

Analyses can be aggregated into dashboards and reports to convey the theme of efficiency.

EARTHDATA

earthdata.nasa.gov

Thank you!

This work was supported by NASA/GSFC under Raytheon Company
contract number 80GSFC21CA001