Engaging with the Problem of Metadata Heterogeneity

Matthew S. Tiscareno,
Mia J. T. Mace, Mark R. Showalter, Mitchell K. Gordon, Robert S. French, Joseph N. Spitale
SETI Institute

PV2023 Meeting, Geneva, Switzerland
3 May 2023

An archivist's job can be complicated

- If data were all of the same type
  - Simply maximize usage of a one-size-fits-all container
  - Users come to the archive knowing what Qs to ask

What fun would that be?

Image from metrohighschool.files.wordpress.com

An archivist's job can be complicated

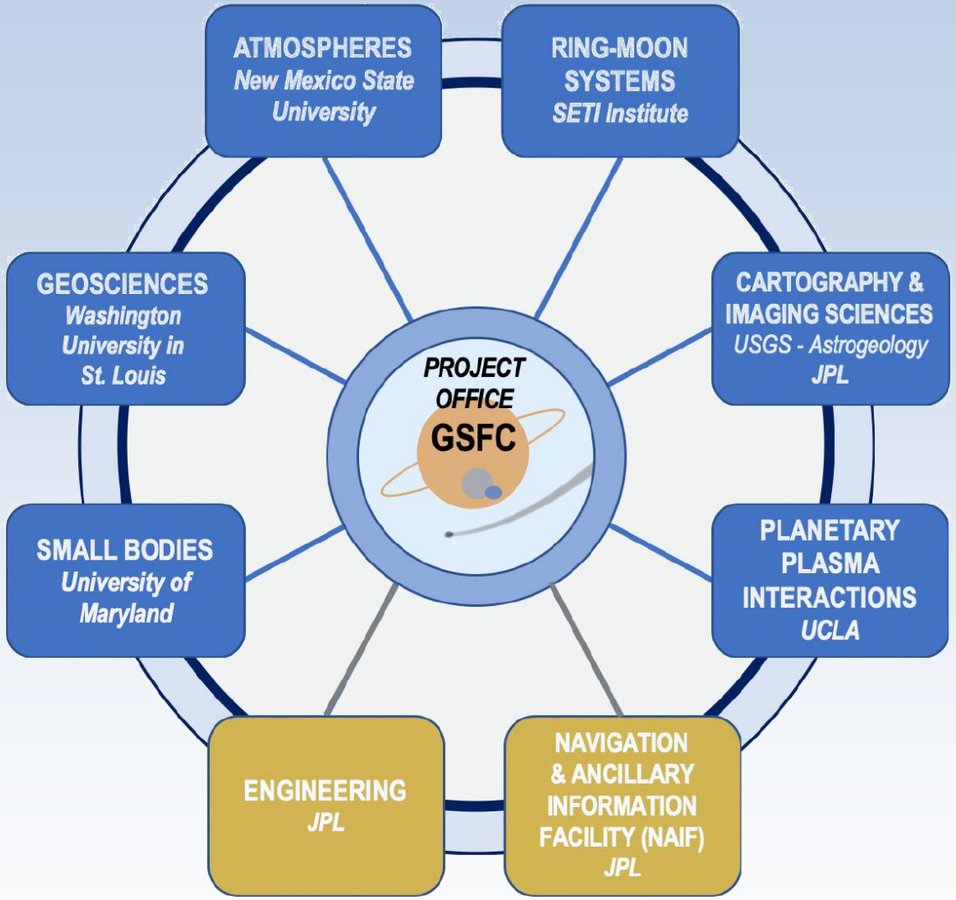
- Reality: Data is diverse
  - Many different kinds of things may be part of one archive
  - Some more than others
  - Containers must allow for all possibilities
  - Users come to the archive with diverse things in mind

This makes our job challenging (and interesting!)

Image from essexnewsdaily.com

The Planetary Data System and PDS4

- PDS is NASA's archive for Planetary Science
- The PDS4 Information Model provides framework
  - Rigorous structure and definition of metadata
  - Simplified file structures
  - Designed to remain functional for decades to come
- But is it enough?



Heterogeneity in Data Type

- Many observing platforms
  - Astronomy
  - Spacecraft
  - Landed stations
  - Biology
  - Laboratories on Earth
  - and more…

To our knowledge, there is <u>no</u> single metadata attribute that applies to <u>every</u> data product in PDS

- Many types of instruments
  - Cameras
  - Spectrometers
  - Dust particle detectors
  - Charged particle detectors
  - Magnetic field detectors
  - Electric field detectors
  - Seismometers
  - and more…
- Many types of targets
  - Planets
  - Moons
  - Asteroids
  - Comets
  - Surface features
  - Magnetospheres
  - Weather stations
  - and more…
- Many targets within each type
  - Layered portable chemistry equipment
- Many coordinate systems
  - The naked eye
- and more…
  - and more…

Heterogeneity in Data Type

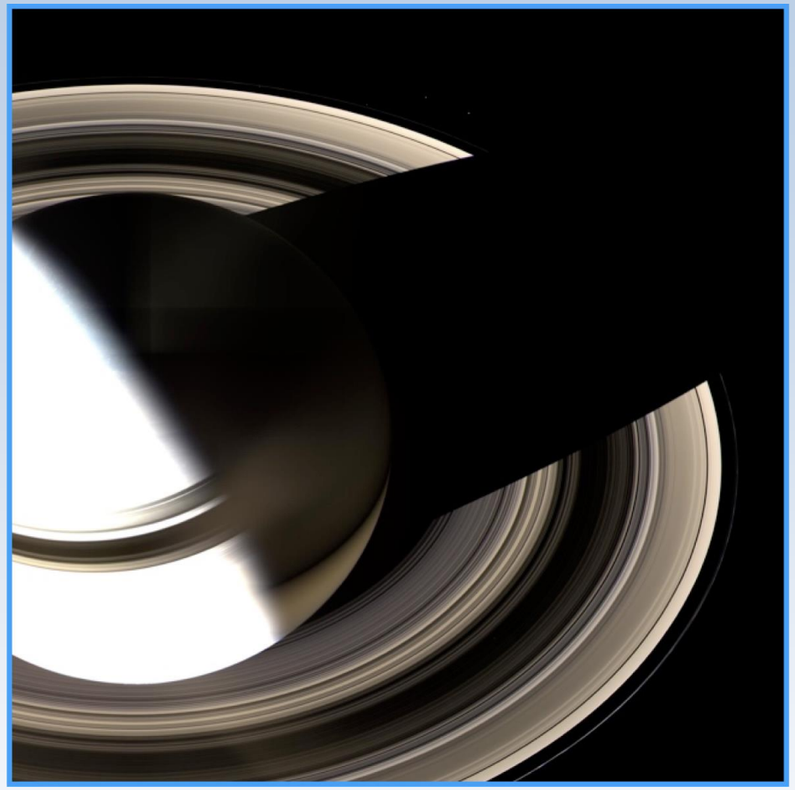
- Data diversity poses challenges for discoverability
- "One-stop shopping" is a great concept
  - Is it practical? Will it work?
  - The more diverse the data, the more merit in diverse discovery tools
  - Requires educating users

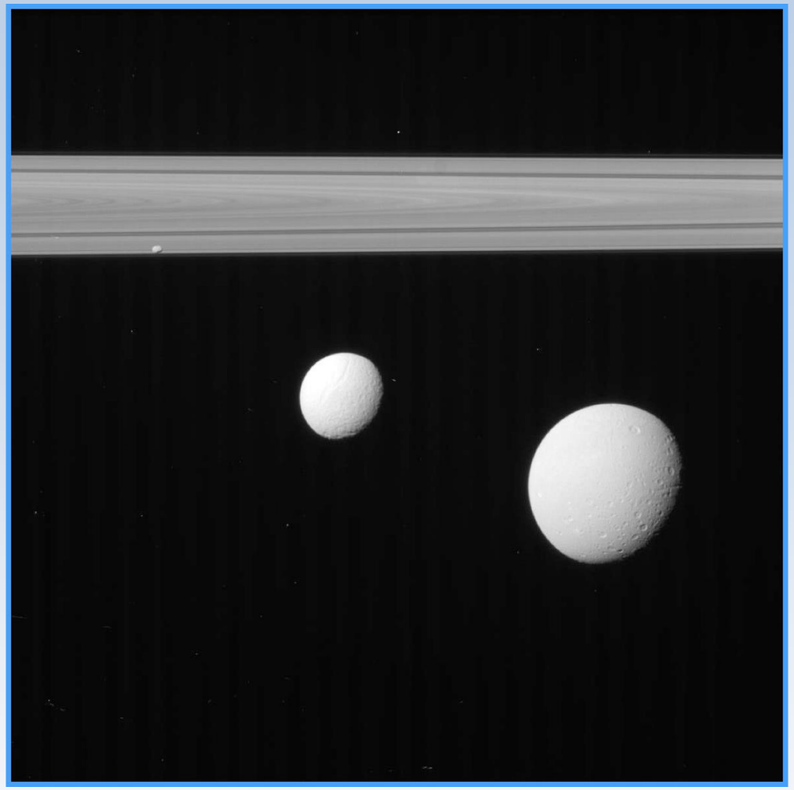
Image from one-stop-shop-online.com

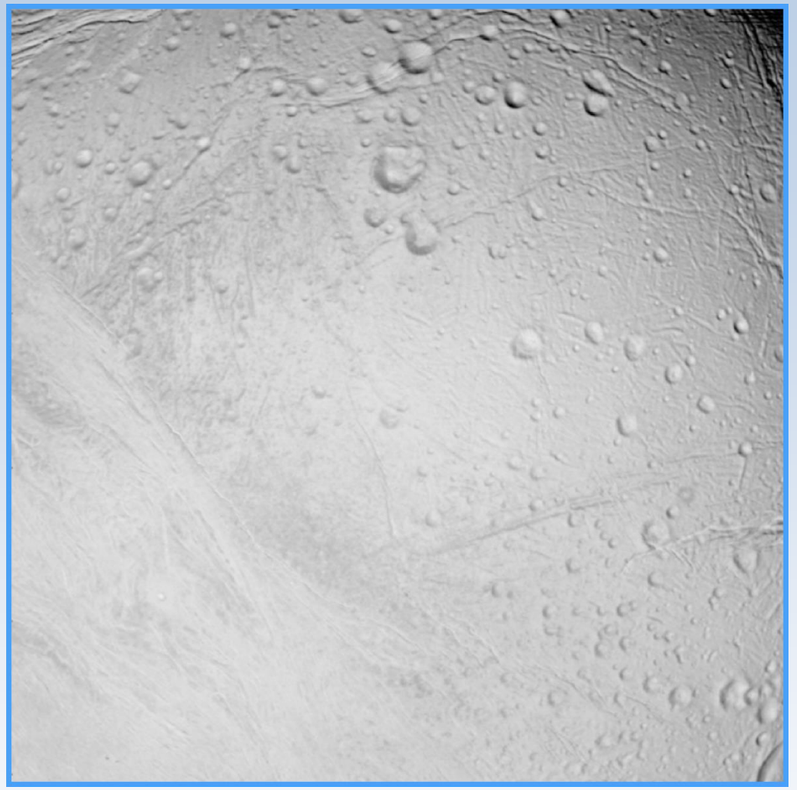
Heterogeneity in Detail

- A different kind of heterogeneity
  - Every data provider has their own context
  - Teams develop a characteristic way of doing things
  - Similar concepts and quantities expressed differently in archived metadata
- Example: Spectral Filters
  - Cassini ISS had a filter called RED, covering 0.57 μm to 0.72 μm
  - Hubble WCF3 has a filter called F689M, covering 0.65 μm to 0.72 μm
  - Archived metadata alone may not help you link these related observations
- Example: Coordinate Systems
  - One data provider gives planetocentric lat/lon, another planetographic
  - One data provider gives solar hour angle in hours, another longitude in °

Heterogeneity in Detail: Geometric Metadata

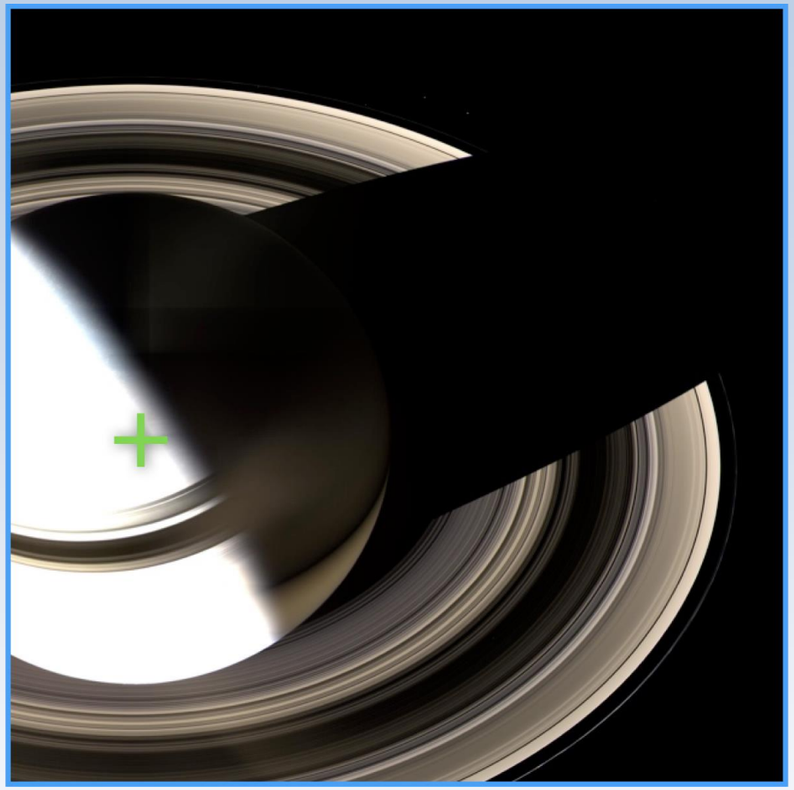


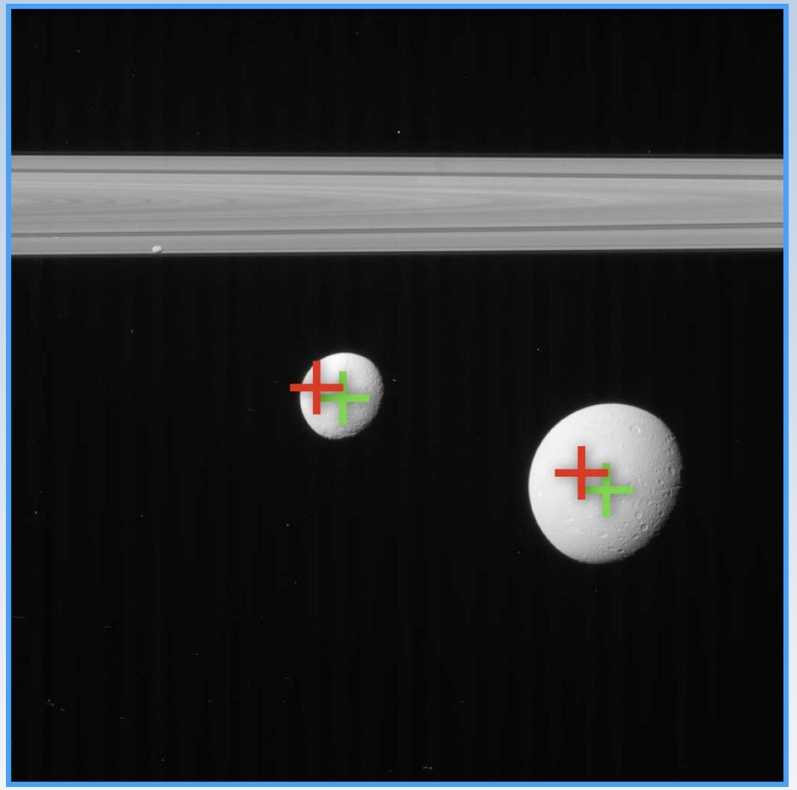


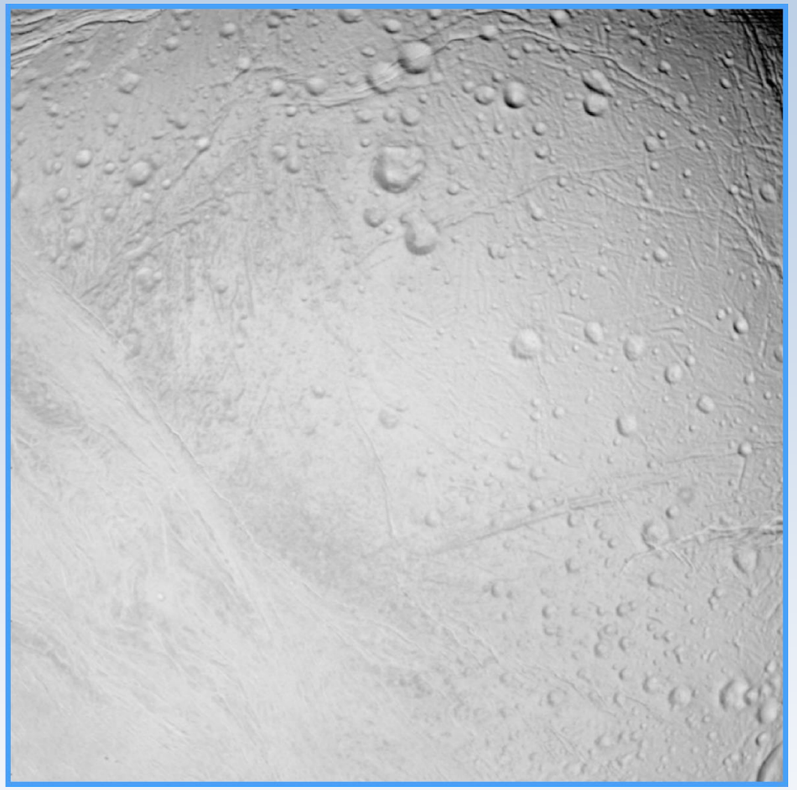


- Spacecraft teams generally provide metadata indicating the intended scientific target of each data product
  - But many products contain information relevant to more than one body
  - Moreover, serendipitous detections may not have metadata provided

Heterogeneity in Detail: Geometric Metadata

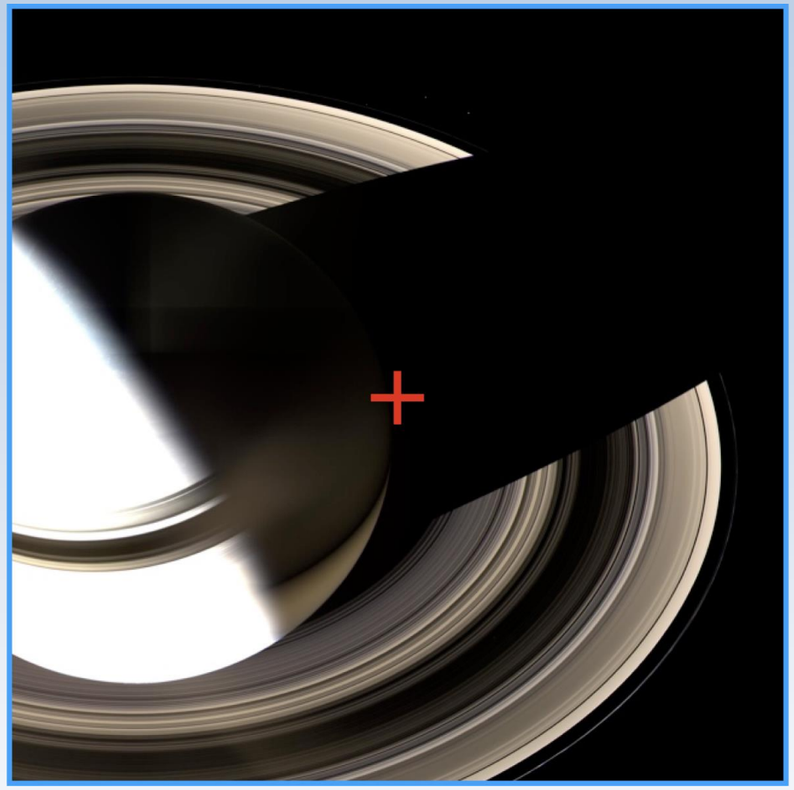


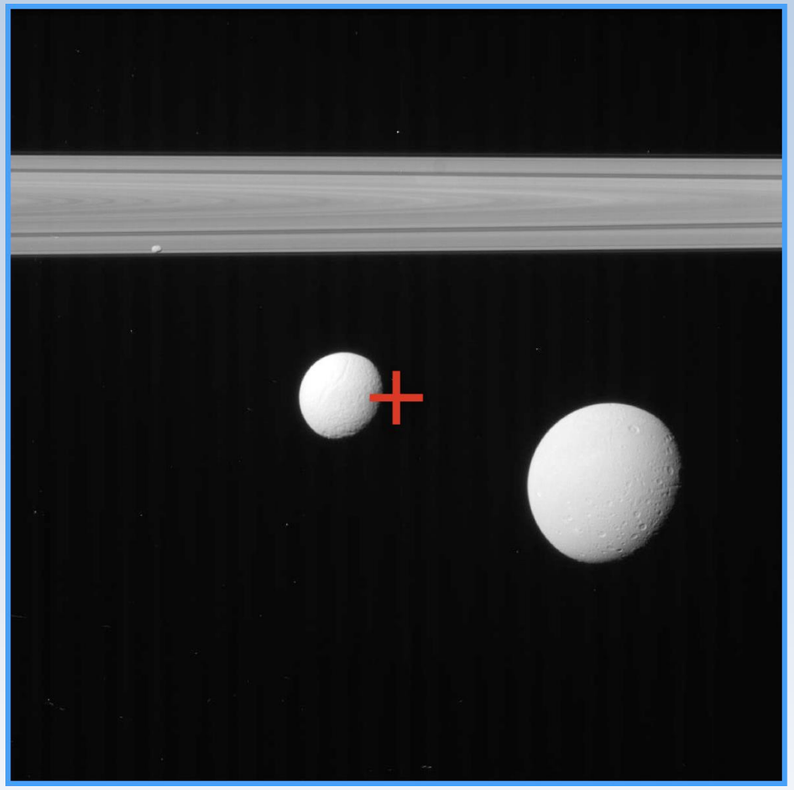


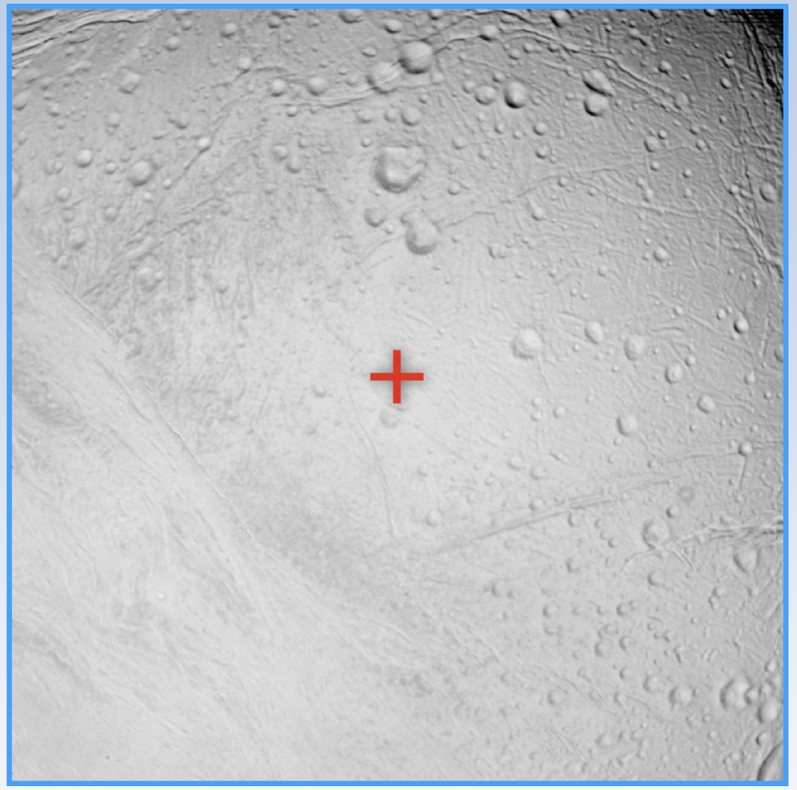


- Sometimes, team-provided metadata identifies the sub-spacecraft and sub-solar coordinates
  - However, this does not tell us whether the specified locations fall inside the field of view

Heterogeneity in Detail: Geometric Metadata

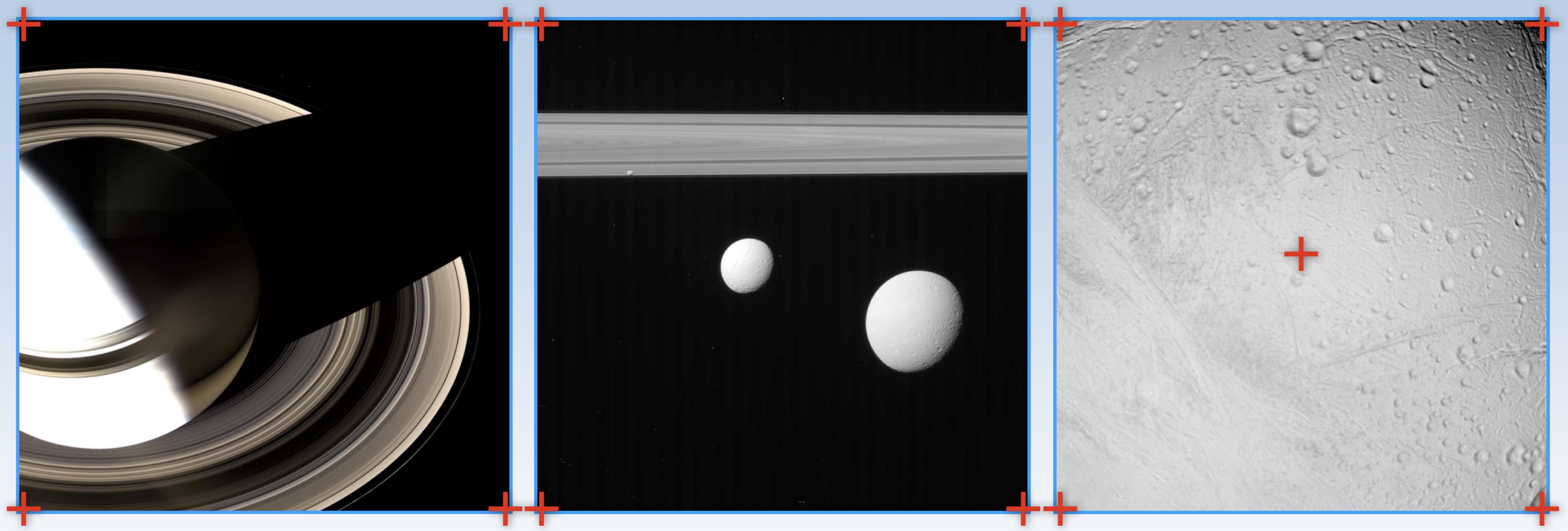






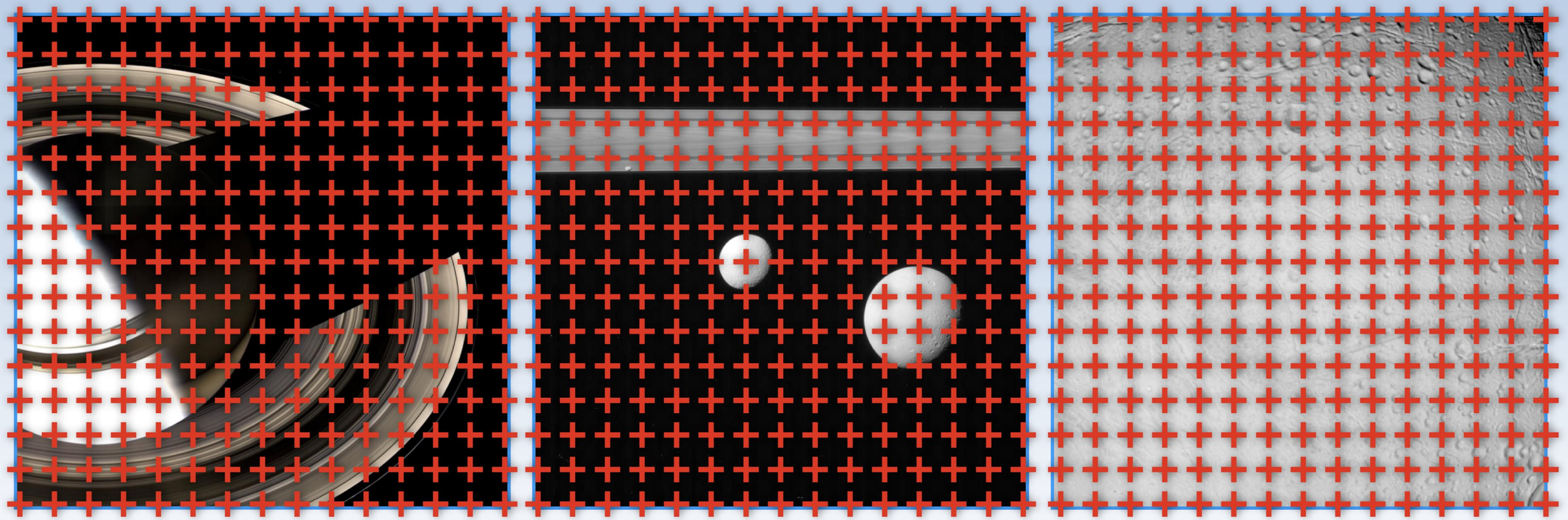
- Sometimes, team-provided metadata provides information about the center of the field of view
  - However, we cannot assume that these values are in any way representative of the product's actual content

Heterogeneity in Detail: Geometric Metadata



- Sometimes, team-provided metadata provides information about what falls at the four corners of the image
  - However, we cannot assume that these values are in any way representative of the product's actual content

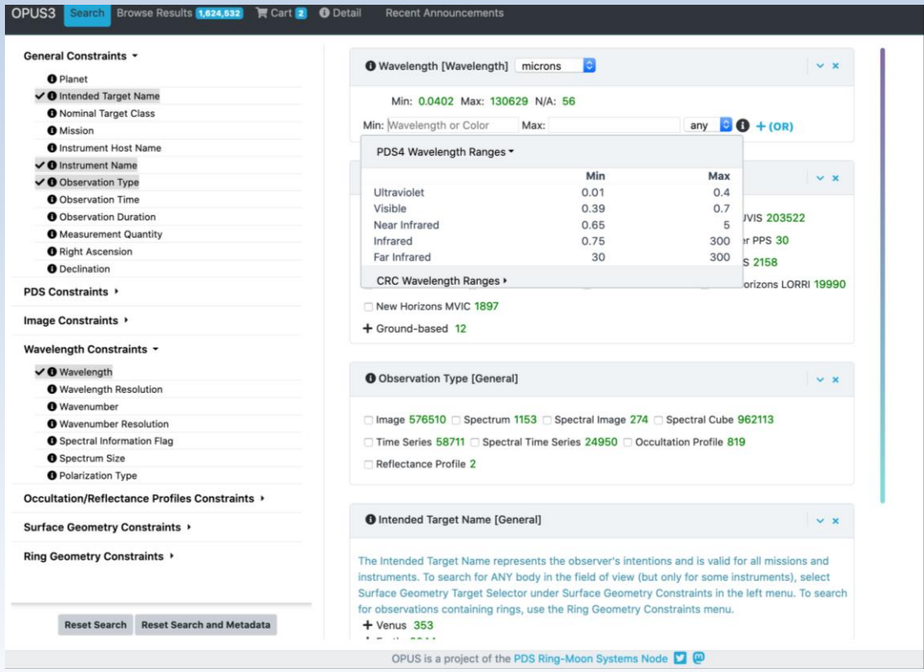
Heterogeneity in Detail: Geometric Metadata



- OPUS metadata identifies every body/ring inside the field of view
- OPUS metadata is derived by densely sampling the field of view, and describing images by a range rather than a single value
- This approach provides a much more accurate description of the field of view

Mitigating Metadata Heterogeneity

- Maximizing the usefulness of cross-mission search requires
  - Curation,
  - Interpretation, and
  - Standardization of metadata
- At the PDS Ring-Moon Systems Node (RMS), we prioritize
  - Generating our own supplemental metadata
  - Using standardized attributes
  - Carrying out detailed geometric re-calculations



- This approach drives improved cross-mission (even cross-discipline) search on our OPUS search tool (opus.pds-rings.seti.org)

Mitigating Metadata Heterogeneity

- However, our approach is not well-suited to every data set
- Within PDS, other Discipline Node teams operate other data discovery tools that serve the needs of their user communities
- There is room for improvement
  - Better user education
  - Better common entry point
- We believe the best common entry point
  - Is not a single unified tool
  - But a platform that quickly guides users to the existing focused tool that best meets their needs

How Data Discovery Tools Proliferate