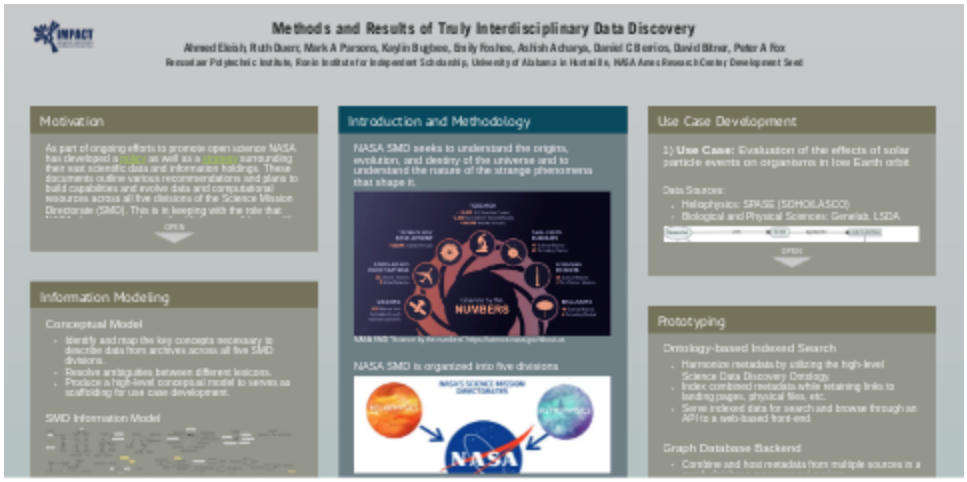
Methods and Results of Truly Interdisciplinary Data Discovery



This website uses cookies to ensure you get the best experience on our website. Learn more
Accept

Ahmed Eleish, Ruth Duerr, Mark A Parsons, Kaylin Bugbee, Emily Foshee, Ashish Acharya, Daniel C Berrios, David Bitner, Peter A Fox

Rensselaer Polytechnic Institute, Ronin Institute for Independent Scholarship, University of Alabama in Huntsville, NASA Ames Research Center, Development Seed

PRESENTED AT:

AGU FALL MEETING
New Orleans, LA & Online Everywhere
13–17 December 2021
Poster Gallery brought to you by WILEY



MOTIVATION

As part of ongoing efforts to promote open science NASA has developed a policy (https://science.nasa.gov/researchers/science-data/science-information-policy) as well as a strategy (https://science.nasa.gov/science-red/s3fs-public/atoms/files/SDMWG_Full%20Document_v3.pdf) surrounding their vast scientific data and information holdings. These documents outline various recommendations and plans to build capabilities and evolve data and computational resources across all five divisions of the Science Mission Directorate (SMD). This is in keeping with the role that NASA plays as a producer of publicly accessible scientific research products and in response to the anticipated growth in size of scientific data resources.

SMD *Catalog*

On this project our charge is to design and implement a cross-divisional data discovery solution or *catalog* that provides reliable access to SMD data, code, documentation and other research products.

Our interdisciplinary team is pursuing multiple system architectures that employ different technologies to explore the solution space for a good fit.

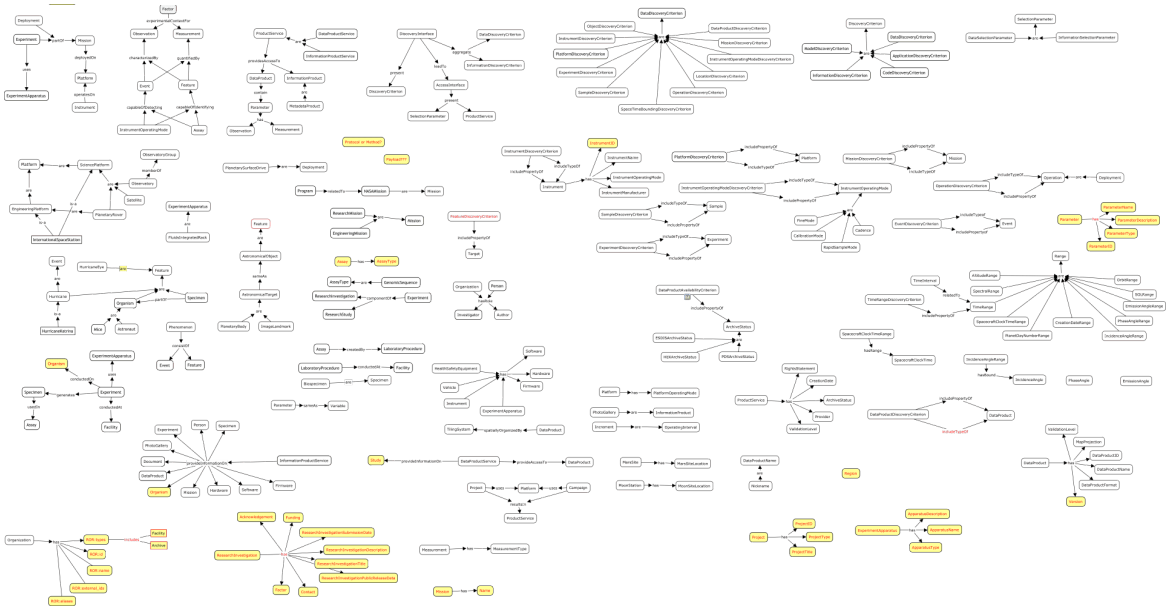
This is in conjunction with related efforts to develop a cross-disciplinary information model as well as identify the relevant standards to drive the work and enact the new policy. See IN45H-0520 Developing the Cross-Disciplinary Information Model for NASA's Science Mission Directorate and IN51B-01 Holistic and pragmatic standards processes enable interdisciplinary science.

INFORMATION MODELING

Conceptual Model

- Identify and map the key concepts necessary to describe data from archives across all five SMD divisions.
- Resolve ambiguities between different lexicons.
- Produce a high-level conceptual model to serves as scaffolding for use case development.

SMD Information Model



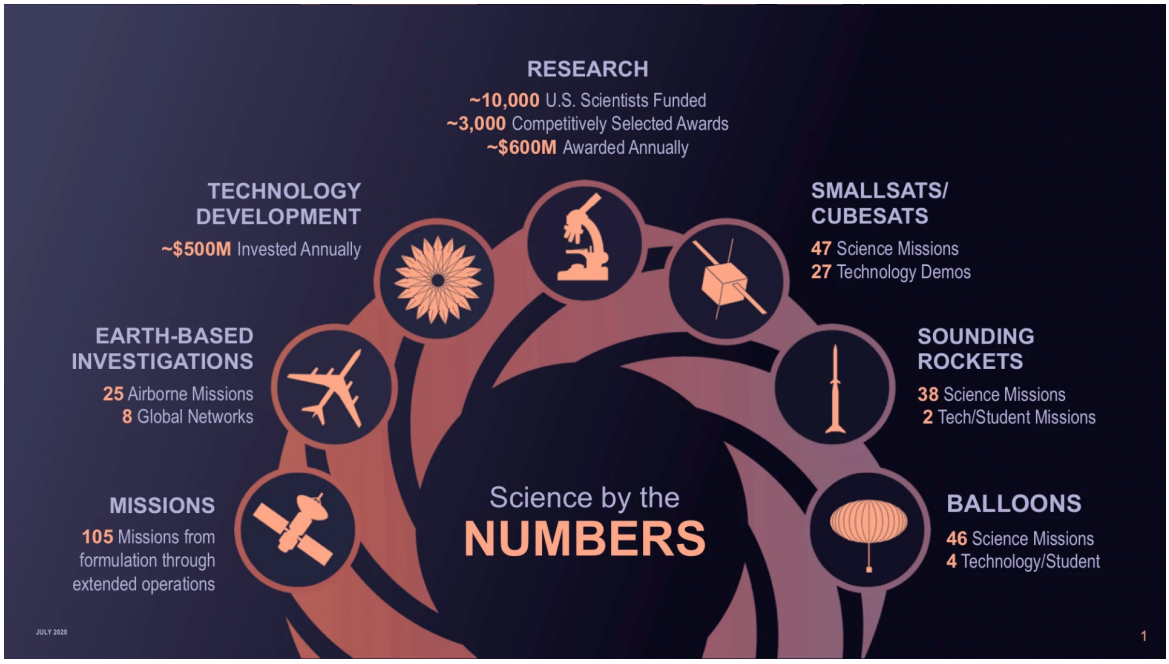
Logical Model

- Go from concepts to classes and properties.
- Re-use existing ontologies and complying with standards where possible.
- Intergrate feedback and lessons from use case development.

Science Data Discovery Ontology (https://github.com/NASA-IMPACT/sddo)

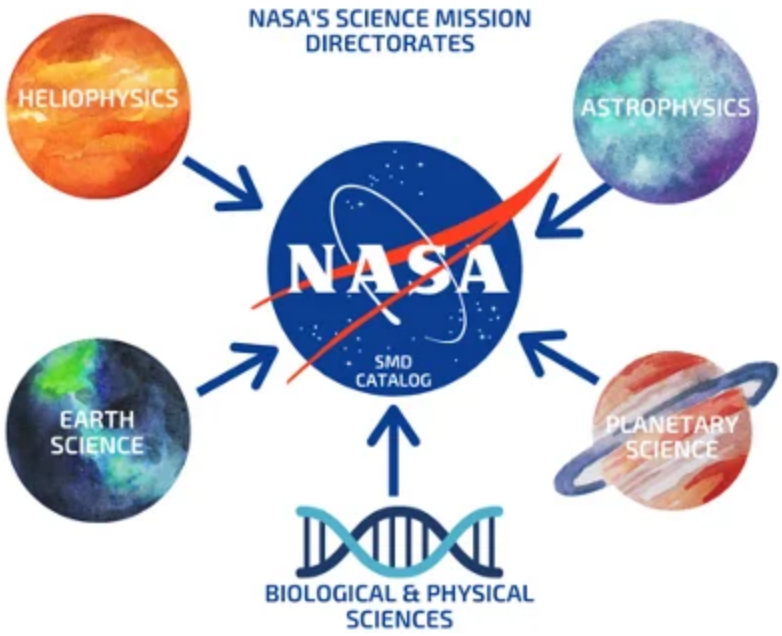
INTRODUCTION AND METHODOLOGY

NASA SMD seeks to understand the origins, evolution, and destiny of the universe and to understand the nature of the strange phenomena that shape it.



NASA SMD "Science by the numbers" https://science.nasa.gov/about-us

NASA SMD is organized into five divisions



Data and computing systems and resources under the five divisions are further organized into archives and nodes with each division uniquely structured according to the needs of its community.



In order to tackle this immense task we employ the following methodology:

1) Information Modeling

We explore the SMD ecosystem in depth to learn its knowledge structure.

2) Use Case Development

We organize and conduct workshops to engage with data producers and consumers for their unique perspectives .

3) Agile Software Development

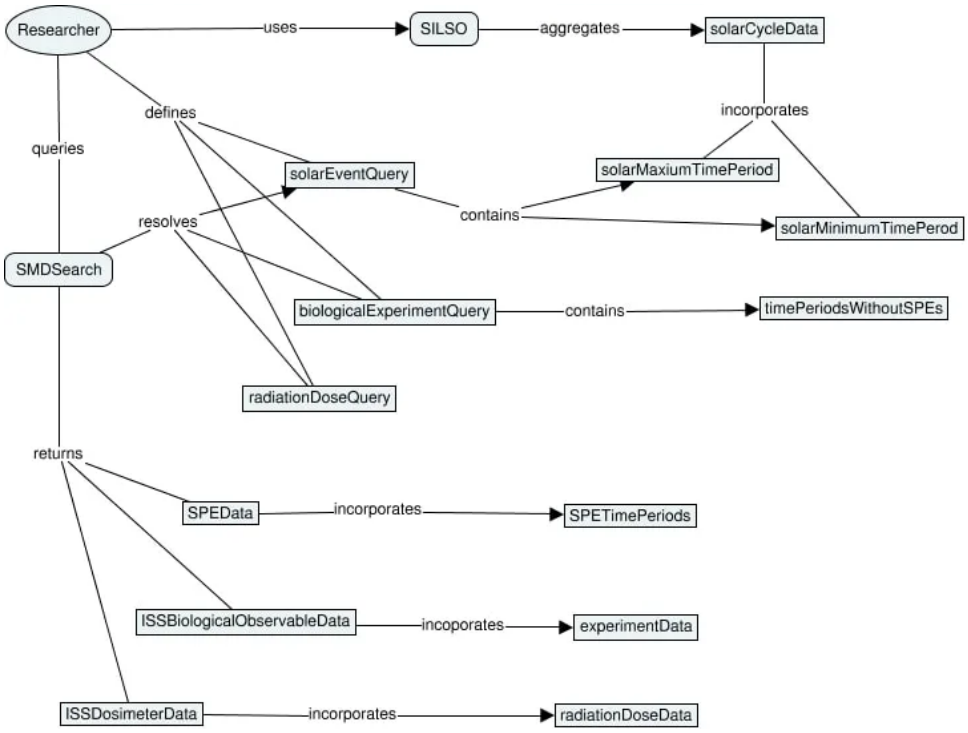
We follow an agile process and commit to rapid prototyping.

USE CASE DEVELOPMENT

1) **Use Case:** Evaluation of the effects of solar particle events on organisms in low Earth orbit

Data Sources:

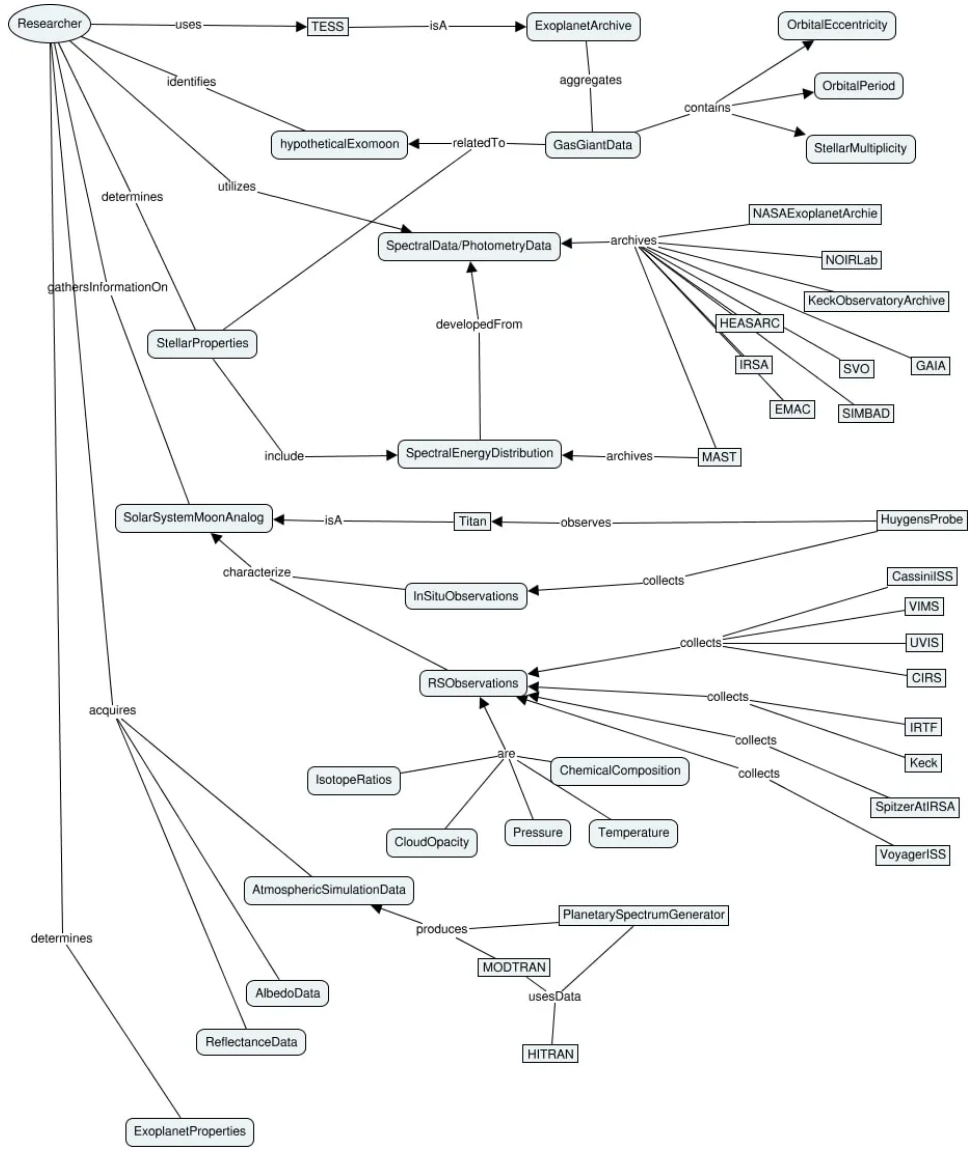
- Heliophysics: SPASE (SOHO/LASCO)
- Biological and Physical Sciences: Genelab, LSDA



2) **Use Case:** Where Are The Organic Exomoons?

Data Sources:

- Astrophysics: TESS
- Planetary Sience: Cassini, Juno
- Earth Science: MERRA, Aqua


Researcher
uses
TESS
isA
ExoplanetArchive
OrbitalEccentricity
identifies
aggregates
OrbitalPeriod
contains
hypotheticalExomoon
relatedTo
GasGiantData
StellarMultiplicity
determines
utilizes
NASAExoplanetArchie
SpectralData/PhotometryData
archives
NOIRLab
gathersInformationOn
KeckObservatoryArchive
developedFrom
HEASARC
StellarProperties
IRSA
SVO
GAIA
EMAC
SIMBAD
include
SpectralEnergyDistribution
archives
MAST
SolarSystemMoonAnalog
isA
Titan
observes
HuygensProbe
characterize
collects
InSituObservations
CassiniISS
VIMS
collects
UVIS
CIRS
RSObservations
collects
IRTF
acquires
collects
Keck
are
IsotopeRatios
ChemicalComposition
collects
SpitzerAtIRSA
CloudOpacity
Pressure
Temperature
VoyagerISS
AtmosphericSimulationData
PlanetarySpectrumGenerator
determines
produces
MODTRAN
AlbedoData
usesData
ReflectanceData
HITRAN
ExoplanetProperties

PROTOTYPING

Ontology-based Indexed Search

- Harmonize metadata by utilizing the high-level Science Data Discovery Ontology.
- Index combined metadata while retaining links to landing pages, physical files, etc.
- Serve indexed data for search and browse through an API to a web-based front-end.

Graph Database Backend

- Combine and host metadata from multiple sources in a graph database management system.
- Make use of graph functions to handle search and discovery queries.
- Take advantage of graph properties to reveal insights into the relationships among the data.

Intelligent Search (Sinequa)

- Utilize an enterprise software platform to provide search capabilities and gain insights from research content.
- This includes a multitude of components and features to enable search as well as analytics across all data sources.
- An extensible front-end framework enables customization of user intergace elements offering high flexibility if presentation.