NASA's Planetary Data System (PDS) Distribution - Delivery 1

(PDS-D D01)

CODATA 2002

Montreal, Canada

October 1, 2002

Outline

- What we did
- Why we did it
- How it works
- Why is it better this way
- Where do we go from here

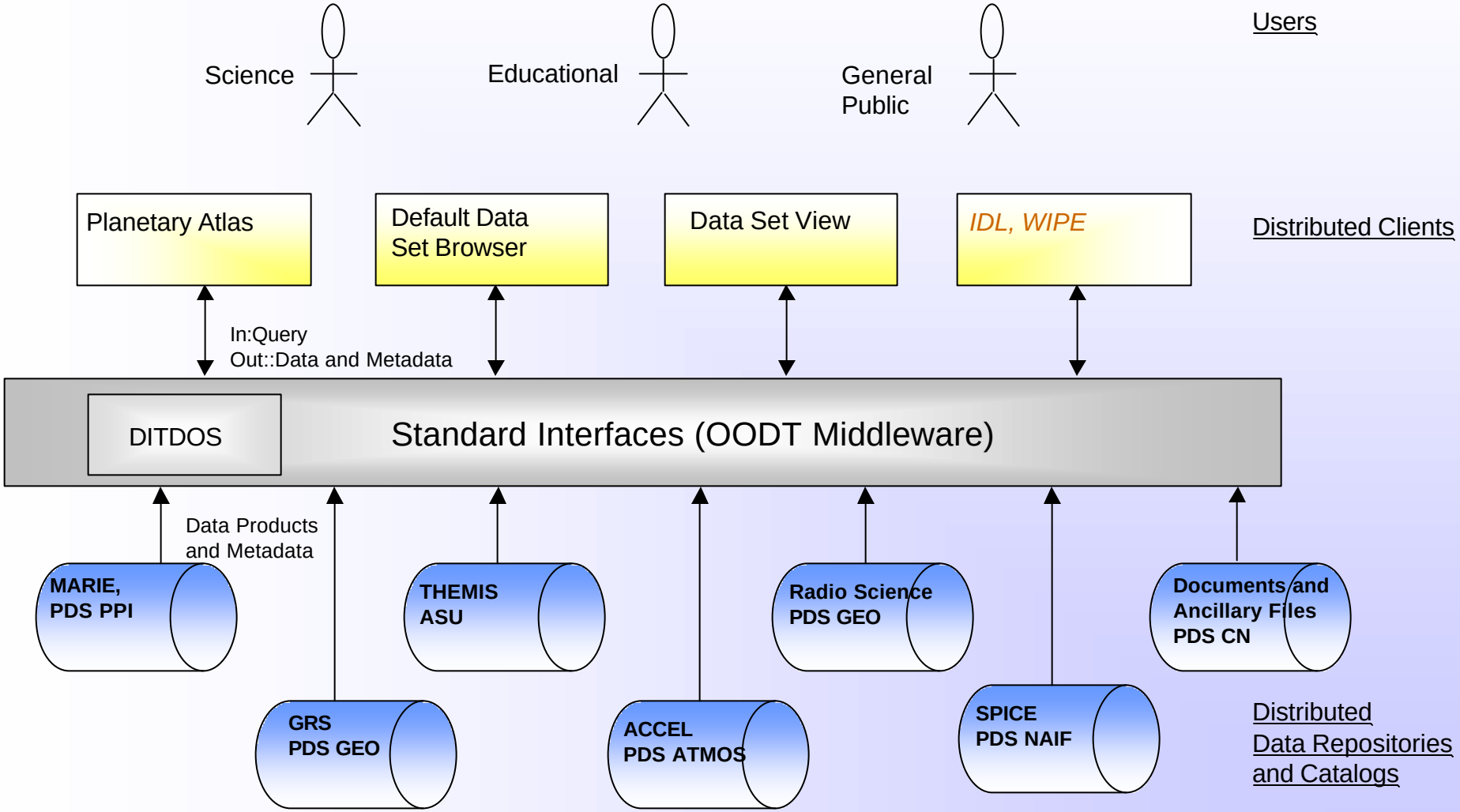
The Data

- ***Variety*** **and Volume**
  - 5TB of data from 30 years of exploration
  - ~700 Data Sets (hundreds of product types)
  - ~1700 Archive Volumes – CD/DVDs
  - Camera, Spectrometer, LECP, SAR, RS, …
  - Images, Time_Series, Spectra, Qubes, Tables, …
  - Binary and ASCII
  - Spacecraft and Earth Based
  - Many data representations
  - Geographically distributed
  - Multi-disciplinary
  - Maintain original bits and convert as needed

PDS-D (Distribution): Delivery D01

- **PDS-D integrates a collection of PDS resources and provides seamless access to distributed data repositories**
  - Provides unified web based search-and-retrieval user interface to novice and sophisticated users
  - Uses Internet as the primary method of data distribution
  - Supports real time (on demand) distribution of data to users
  - Provides ability for users to subscribe for notification of released data
  - Allows for mission(s) having more complex payloads and significantly larger data volumes

PDS-D D01 Architecture
Mars Odyssey



Development Approach

- **Implemented a multi-tiered information architecture**
  - Application Clients (Browsers/Interfaces)
  - Middleware (OODT)
  - Data and Metadata Servers (product server, profile server)
  - Data Repositories and Catalogs

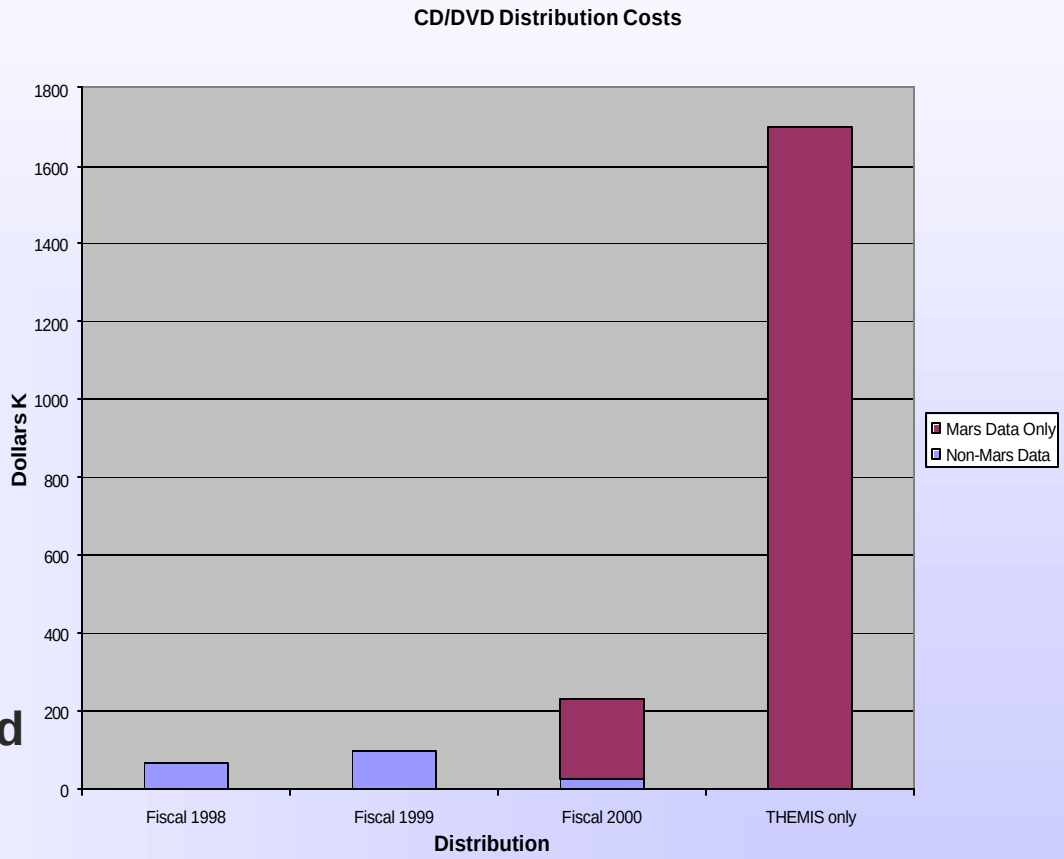
- **Simplified and *standardized system interfaces* through middleware**

- **Used existing PDS subsystems but *hid heterogeneity***
  - User Interfaces (Atlas, DITDOS)
  - Data repositories (disk farms)
  - Catalog databases (Sybase, Gatesware,…)
  - Remained geographically distributed and locally managed

- **Separated Data Architecture from Technology Architecture**
  - Used archive *metadata* to its full potential
  - Evolved technology architecture

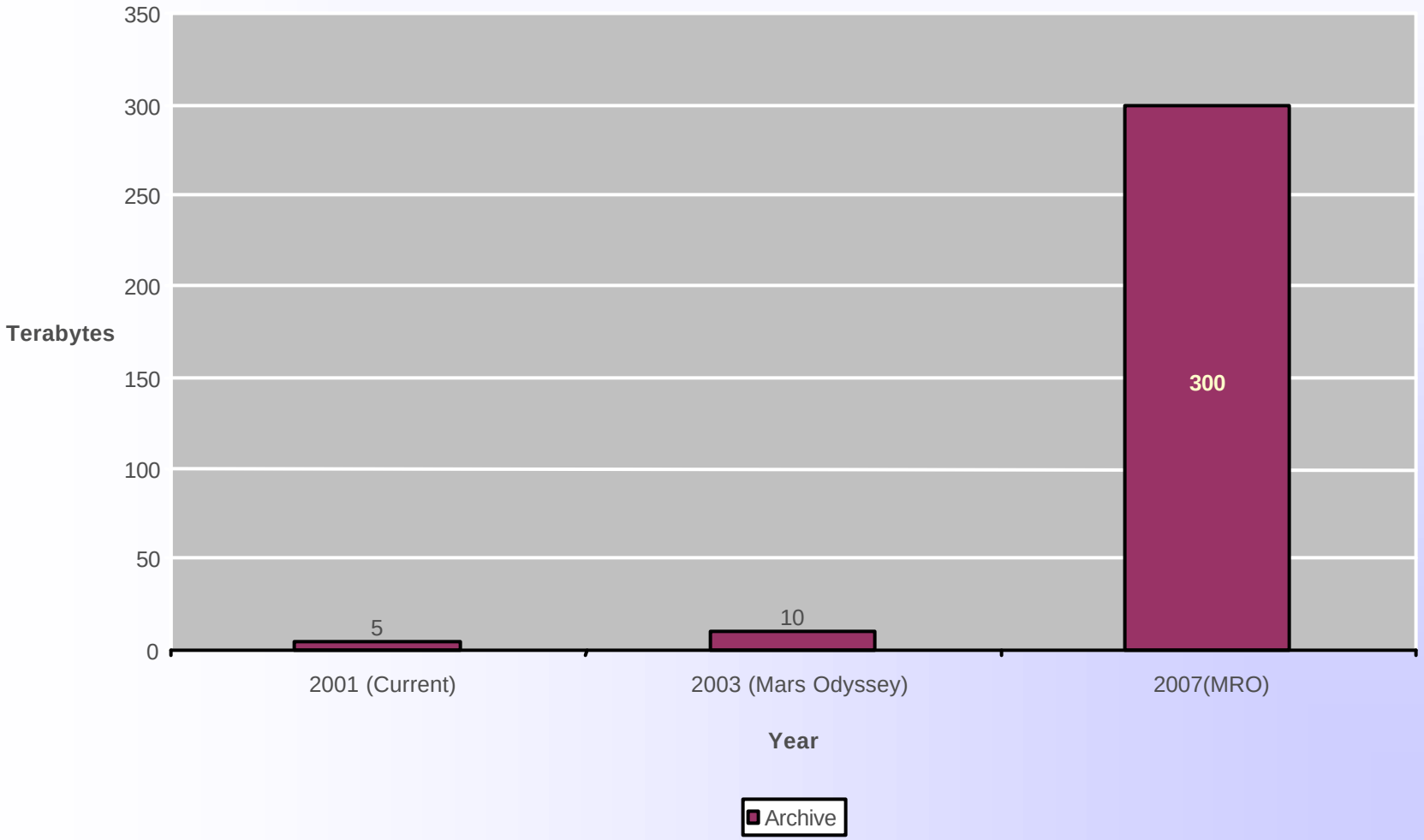
Why a new development?

- **Increased data volumes too much for PDS CD distribution system**
  - MGS CD costs exceeded $500K for 600GB
  - THEMIS alone will produce over 4TB of data
  - CD/DVD distribution too costly and cumbersome
- **Data will be physically distributed among various sites and sources**
- **Data will be more complex and heterogeneous**
- **Scientists want online access to all available data**



Mars Exploration Program: Very Large Data Volumes

Growth in Size of PDS Archive



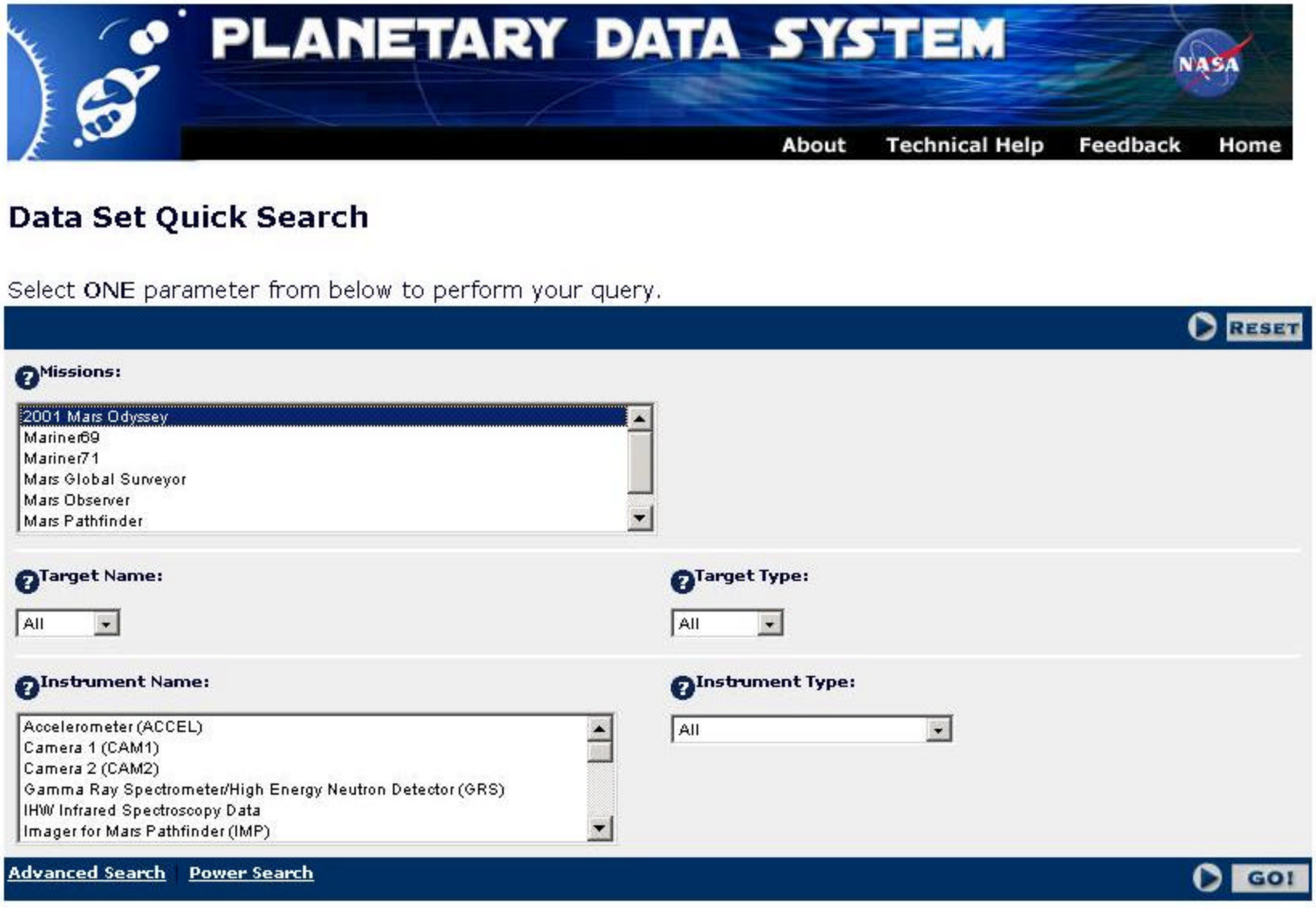
Object Oriented Data Technology (OODT)

- Framework for sharing data among heterogeneous and distributed data repositories
- Provides standardized system interfaces through middleware
- Differentiates between data architecture and technology architecture
- Used in medical science, space science, and other domains
- See poster

Dan Crichton – PI

Steve Hughes – Co-I

How PDS-D D01 works
Data Set View



http://starbrite.jpl.nasa.gov/pds

Data Set View – Results

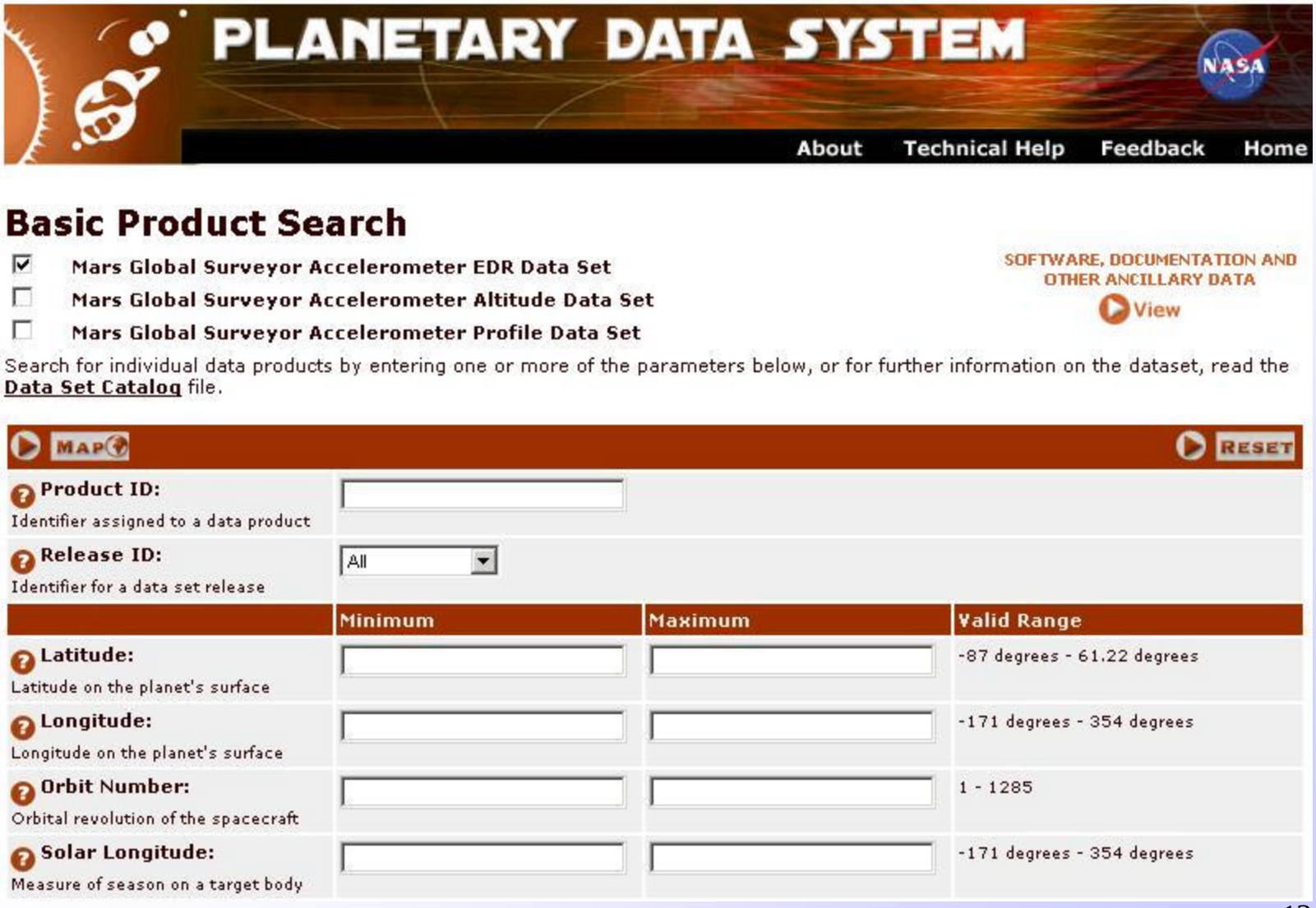
Search Results (5 data sets found)

Data Set	Instrument Host	Information About the Data Set	Data Products & Related Files	Other Resources
1. 2001 Mars Odyssey Gamma Ray Spectrometer, Neutron Spectrometer, and High Energy Neutron Detector Experiment Data Records	ODY	View Information	Product Search	• 2001 Mars Odyssey Data Archives
2. 2001 Mars Odyssey Thermal Emission Imaging System Infrared Experiment Data Records	ODY	View Information	Product Search	• Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas
3. 2001 Mars Odyssey Thermal Emission Imaging System Visible Experiment Data Records	ODY	View Information	Product Search	• Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas
4. 2001 Mars Odyssey Thermal Emission Imaging System Infrared Reduced Data Records	ODY	View Information	Product Search	• Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas
5. 2001 Mars Odyssey Thermal Emission Imaging System Visible Reduced Data Records	ODY	View Information	Product Search	• Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas

Page 1 | New Search

Custom Data Set Browser – THEMIS/GRS

Default Data Set Browser – Radio Science



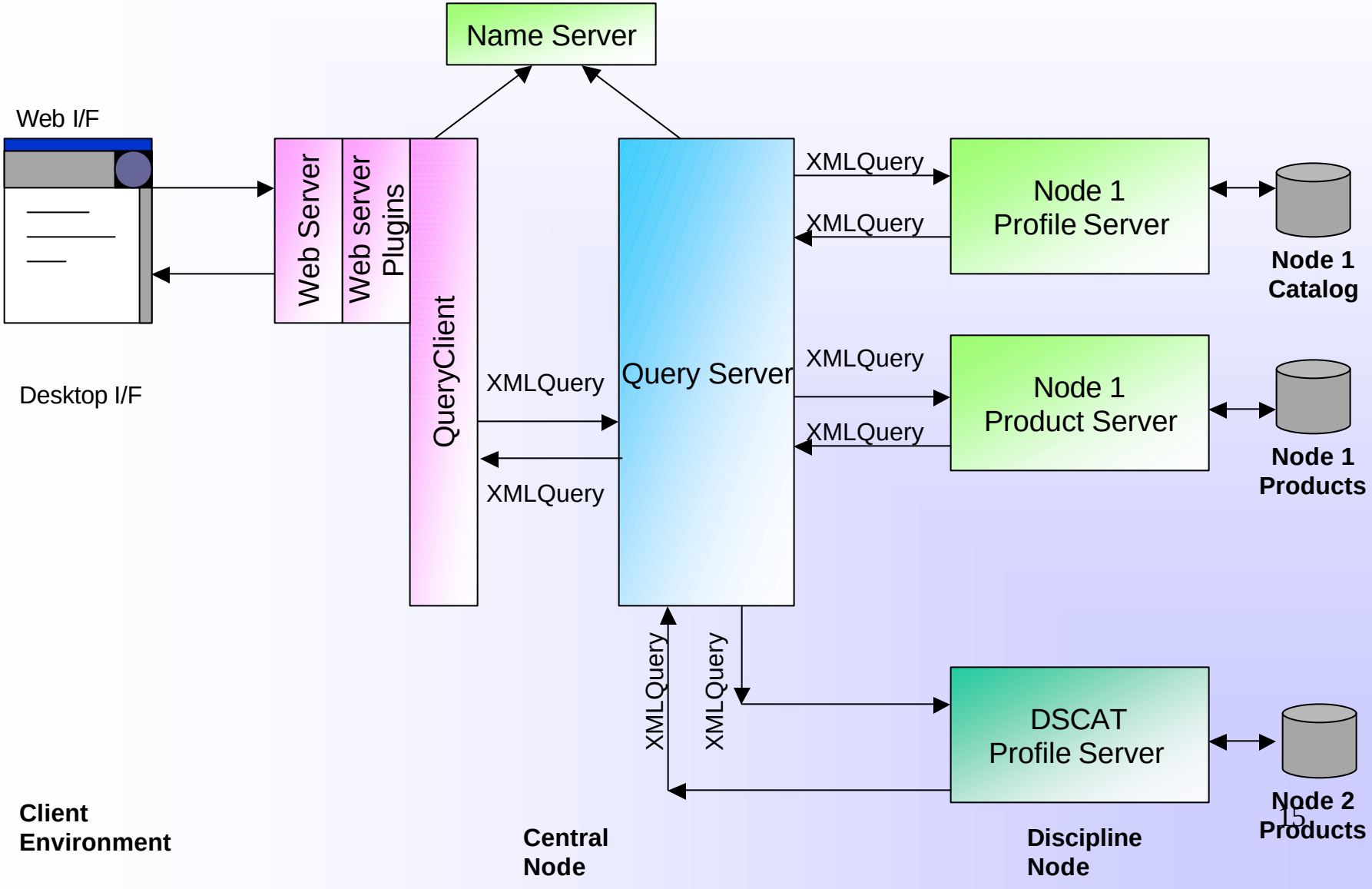
Default Data Set Browser – Radio Science

This search found 3170 products
Products 1 - 10 are displayed

PREVIOUS 10 NEXT

Product	Information	Download	Select for Download (All / None)
1. **MGS_AERODYNAMIC_PROPERTIES** Product description goes here.	View Label	Get Product	☐
2. **PERIAPSIS_ORBITAL_ELEMENTS_P0040** Product description goes here.	View Label	Get Product	☐
3. **PERIAPSIS_ORBITAL_ELEMENTS_P0041** Product description goes here.	View Label	Get Product	☐
4. **PERIAPSIS_ORBITAL_ELEMENTS_P0042** Product description goes here.	View Label	Get Product	☐
5. **PERIAPSIS_ORBITAL_ELEMENTS_P0043** Product description goes here.	View Label	Get Product	☐
6. **PERIAPSIS_ORBITAL_ELEMENTS_P0044** Product description goes here.	View Label	Get Product	☐
7. **PERIAPSIS_ORBITAL_ELEMENTS_P0045** Product description goes here.	View Label	Get Product	☐
8. **PERIAPSIS_ORBITAL_ELEMENTS_P0046** Product description goes here.	View Label	Get Product	☐
9. **PERIAPSIS_ORBITAL_ELEMENTS_P0047** Product description goes here.	View Label	Get Product	☐

Conceptual PDS Implementation



Why is it better this way

- **Seamless search and retrieval of data products**
  - Users can access all resources without knowing their location (location transparency)
  - Integration of PDS Nodes (one PDS instead of 7!)
  - Primary method of data distribution
  - Supports heterogeneous data repositories
  - Geographically distributed data repositories
- **Standard interfaces for software developers**
- **Plug-ins for favorite tools**
- **PDS can evolve as basic information technology changes**
- **Better tracking of data set releases**
- **Mission more involved with the PDS sooner**

Scalability

- **Number of system component interconnections increases linearly**
  - Nodes added as needed
  - One-to-one connections from each component to middleware
  - Exponential number of inter-operational connections made dynamically via message passing

- **Since distribution system is built as a light layer on top of the archive system, it will scale as long as the archive system scales**

- **Archive can remain distributed as needed to support larger data repositories (e.g. MRO)**
  - **Parallel load balancing**

- **Smaller frequently used data repositories can be mirrored for better performance**

Where we go from here

- **Collect requirements from missions and users to analyze out future releases of PDS-D**
- **Plan releases of PDS-A**
  - Automated archive product creation work flow
    - Product generation, labeling, validation, and ingestion
    - Derived product processing and versioning
- **Overhaul PDS Data Model**
  - XML modeling and interfaces
  - Ground-based data sets
  - Wavelength regimes
  - Targets with multiple identifiers and types

Backup

Correlative Search the Simple Way

- **All data resources in the system are profiled**
- **Submit a query that describes what you want**
  - *Not how* to get what you want
- **System returns all matching data profiles**
  - Provides identification and description information
    - Provides location information
    - Provides all PDS metadata to support correlative science
    - Information is machine and human readable
- **Submit query to retrieve data**

PDS-D D01 Architecture

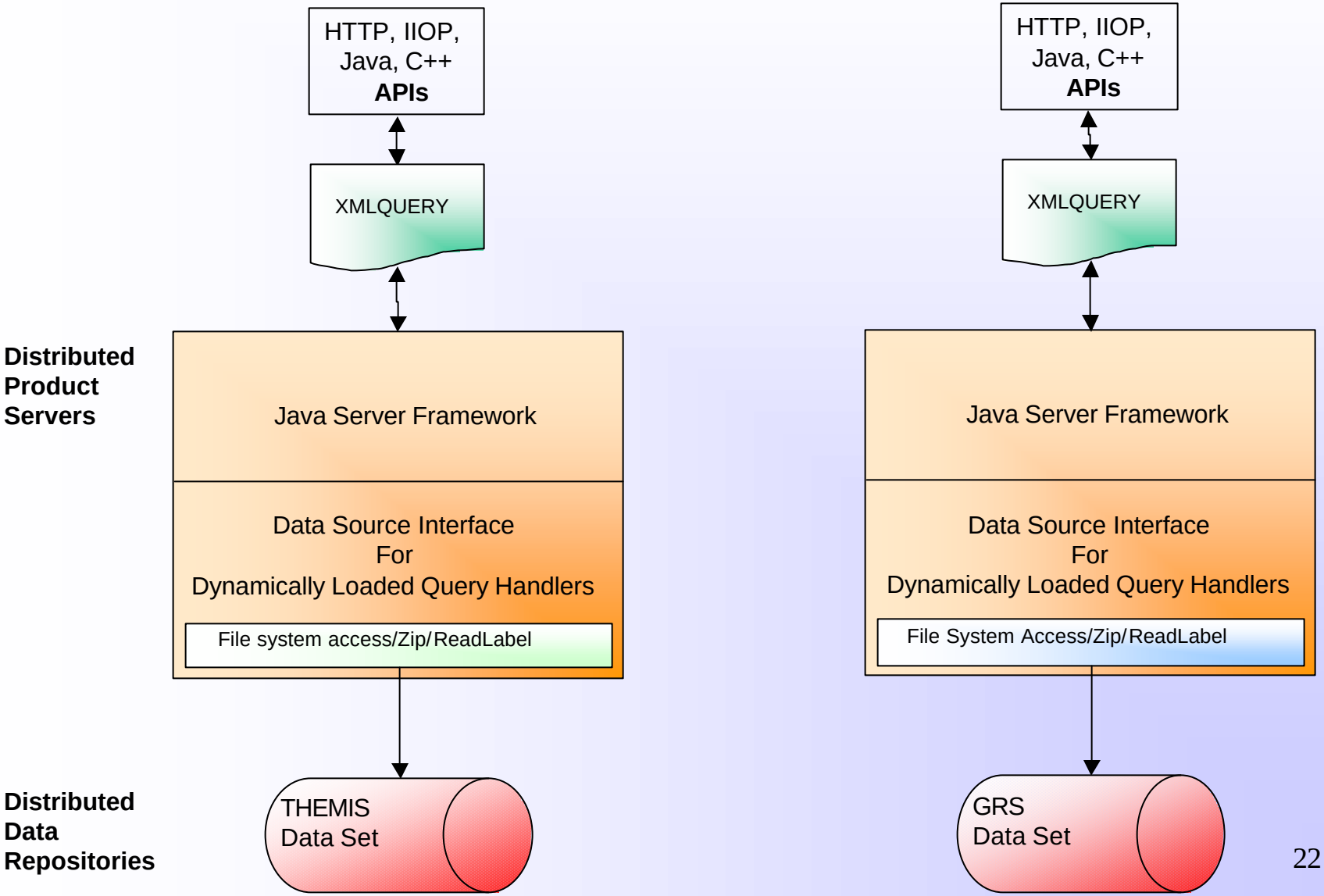
- **The architecture provides *seamless access* to distributed data repositories and catalogs**
  - Location independence (e.g. the data appears local)
  - Information hiding (e.g standard interfaces to differing catalogs and repositories)
  - Data driven architecture (i.e. use PDS data architecture)
  - Scalable and extensible (e.g. Odyssey through MRO)
  - Client APIs for search and retrieval of data and metadata

- **Maintains *geographically distributed* data archives**
  - Conform to CODMAC recommendation for discipline data systems
  - Keeps data in the hands of the scientific experts
  - Promote closer ties with mission instrument teams

- **The architecture has *minimum impact* on existing PDS resources**
  - Maintain original bits and convert as necessary
  - Leverage development using existing system resources

- **Separates data architecture from the technology architecture**

Product Server Architecture



Product Server Performance

Product	Size	Zip	Xfer	Route (source, dest)
Themis	6.9MB	6sec->2.3MB	11sec @ 100 KB/s	ASU/CN/CN
Themis	"	"	8 min @ 2.7 KB/s	ASU/CN/Home1
Themis	"	"	27sec @ 36 KB/s	ASU/CN/Home2
Themis	"	"	24sec @ 50 KB/s	ASU/CN/Flag
MOLA	37MB	8sec->17MB	60sec @ 50 KB/s	GEO/CN/Flag
GAMMA	144MB	65sec->9.7MB	8sec @ 180 KB/s	CN/CN/CN
NAIF SPK	150MB	120s->56.5MB	135sec @350 KB/s	CN/CN/CN
RS RSR	351MB	300s->328MB	300sec @1 MB/s	CN/CN/CN
		GRS/CN - ~ 1 second per MB		Home1 – 34Kbps Home2- DSL

Distribution Prior to Odyssey

- **Data product search and retrieval (e.g. images)**
  - Discipline *expertise is local*, at the nodes
  - Data placed online in *local repositories*
  - Product catalogs developed for each product type
    - Voyager images (~50 attributes)
    - Galileo images (>100 attributes)
    - MGS MOC images (~40 attributes)
  - Custom user interfaces developed for catalogs
  - Primary customers are *discipline oriented*

- **Data set search (e.g. all Viking images)**
  - Catalog search but not data access