SCT

SECRETARÍA DE
COMUNICACIONES
Y TRANSPORTES

AEM

AGENCIA ESPACIAL
MEXICANA

Coordinación de Formación de Capital Humano en el Campo Espacial

Dirección de Divulgación de la Ciencia y Tecnología Espacial

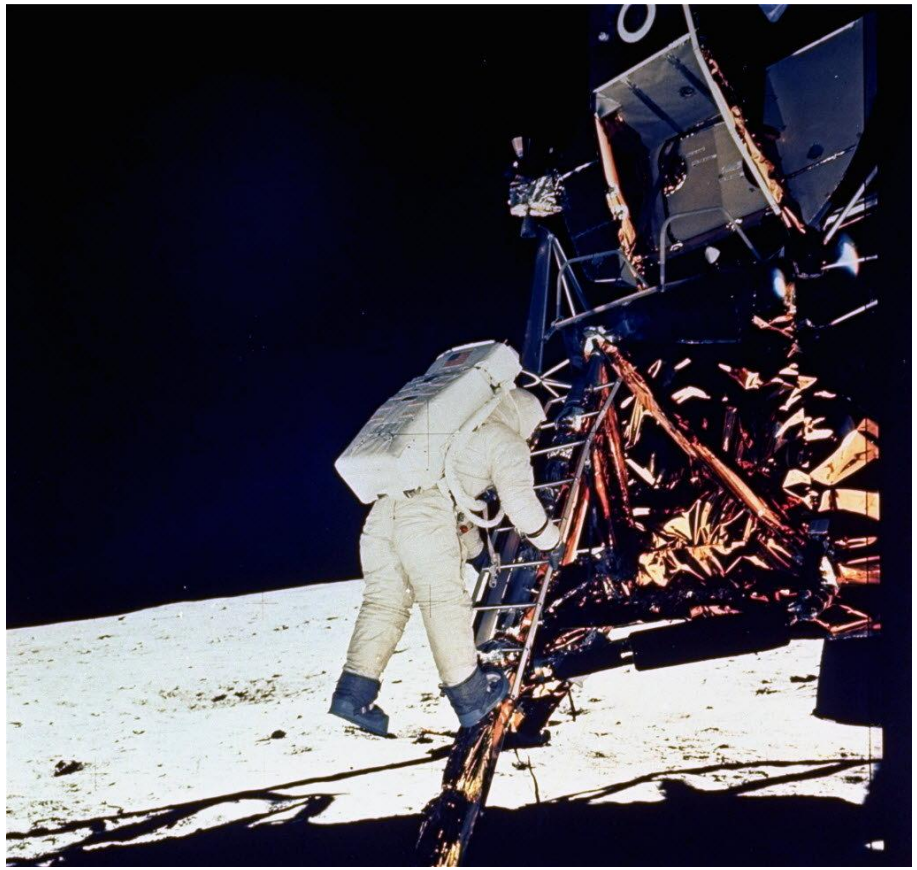
Big Data

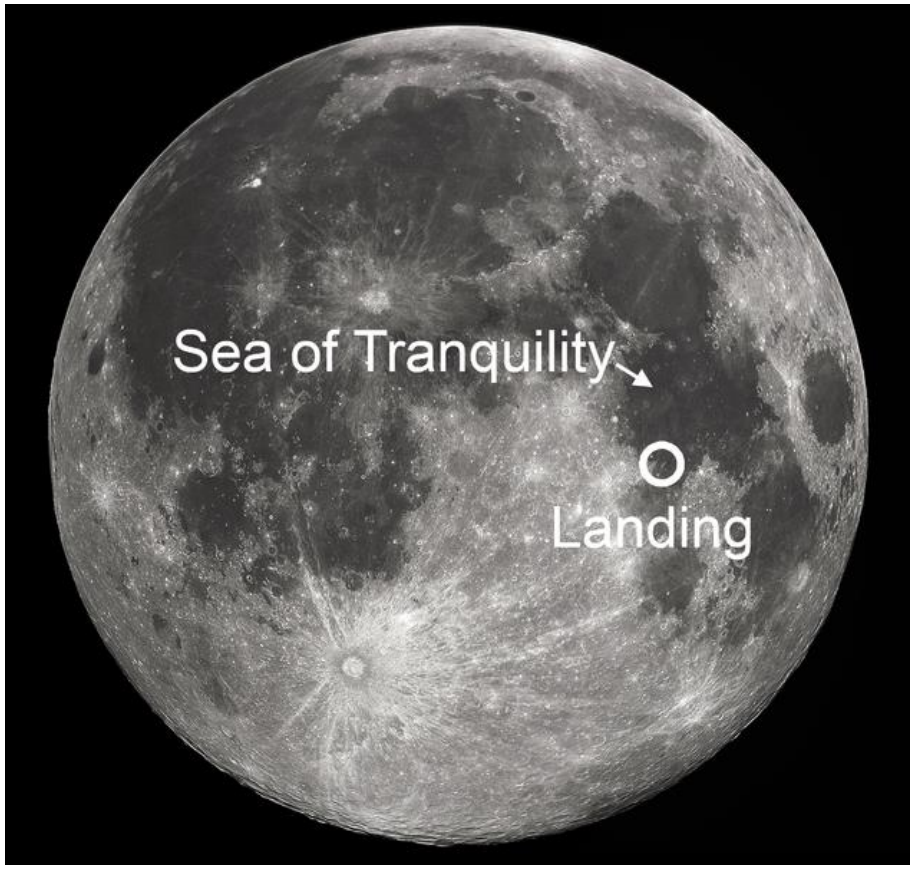
Conferencia Internacional de Archivistica

Mario M. Arreola Santander

@mario_eduspacio

¿Llegó el hombre a la Luna?





Problema común: Crecimiento de Archivos.

- Soportes físicos y digitales
- Libros Digitales
- Multiformato
- Multiplataforma
- Audio
- Video

Consulté a mis amigos en NASA

- Estación Espacial Internacional (Houston, Tx)
- Misiones Robóticas en Marte (JPL, California)
- Finalmente una búsqueda en
  - La Web

Comunicaciones desde el espacio profundo

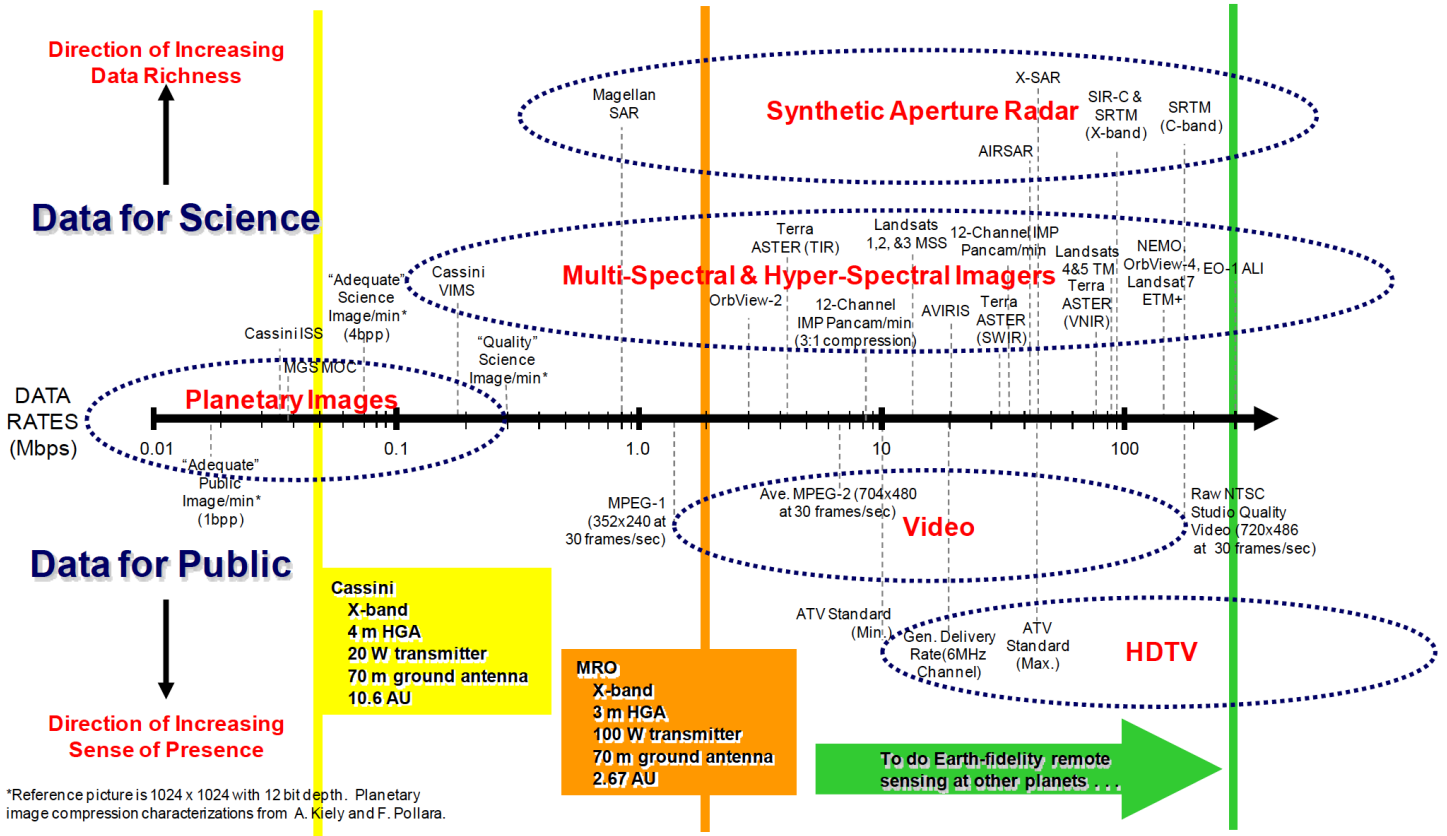


Figure 1. Required data rates as a function of data type.

Retos de NASA

- 1 PetaByte de archivos 1e15
- 1990’s Comunidad científica= miles de sistemas heterogéneos distribuidos. Unix, Linux, DOS, Win3.x, Mac, VMS…
- Diversas plataformas, bases de datos, formato de datos, RMS, ODBS, interfaces y volúmenes de datos de KiloBytes a TeraBytes.
- Múltiple de todo: tipos de objetos, interpretaciones, software.
- Metadatos incompatibles y/o perdidos
- Si, ¡todo un caos!

Respuesta: Object Oriented Data Technology

- Proyecto iniciado en 1998 en JPL.
  - Aplicación de IT a Ciencias Espaciales
  - Proporcionar infraestructura para gestión de datos distribuidos.
  - Métodos de Interoperabilidad, gestionar y descubrir conocimiento.
  - Marco de Software para gestión de los datos, reusar software, reducir costos, administrar el riesgo y apalancar la experiencia IT.
- OODT enfoque inicial:
  - Archivos de datos; gestionar productos de datos y recursos distribuidos.
  - Ubicación y descubrimiento de datos.
  - Recolectar e integrar datos.

Que resuelve OODT

- Amarrar datos heterogéneos distribuidos en una red virtual.
- Arquitectura que proporciona una base para el trabajo.
- Infraestructura de flujo de trabajo para manejar los productos de datos.
- Escalabilidad al enlazar grandes grupos de datos.

Arquitectura Distribuida

1. Las Aplicaciones científicas usan “APIs” para conectarse con un almacen virtual de datos

2. El “Middleware” crea la infraestructura en malla o red de datos que conecta los sistemas heterogeneous distribuidos y los datos

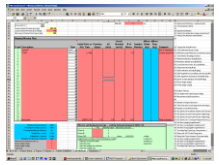
3. Almacenes para archivar y recuperacion de muchos tipos de datos.

Visualization Tools

OODT API

Web Search Tools

OODT API



Analysis Tools

OODT API

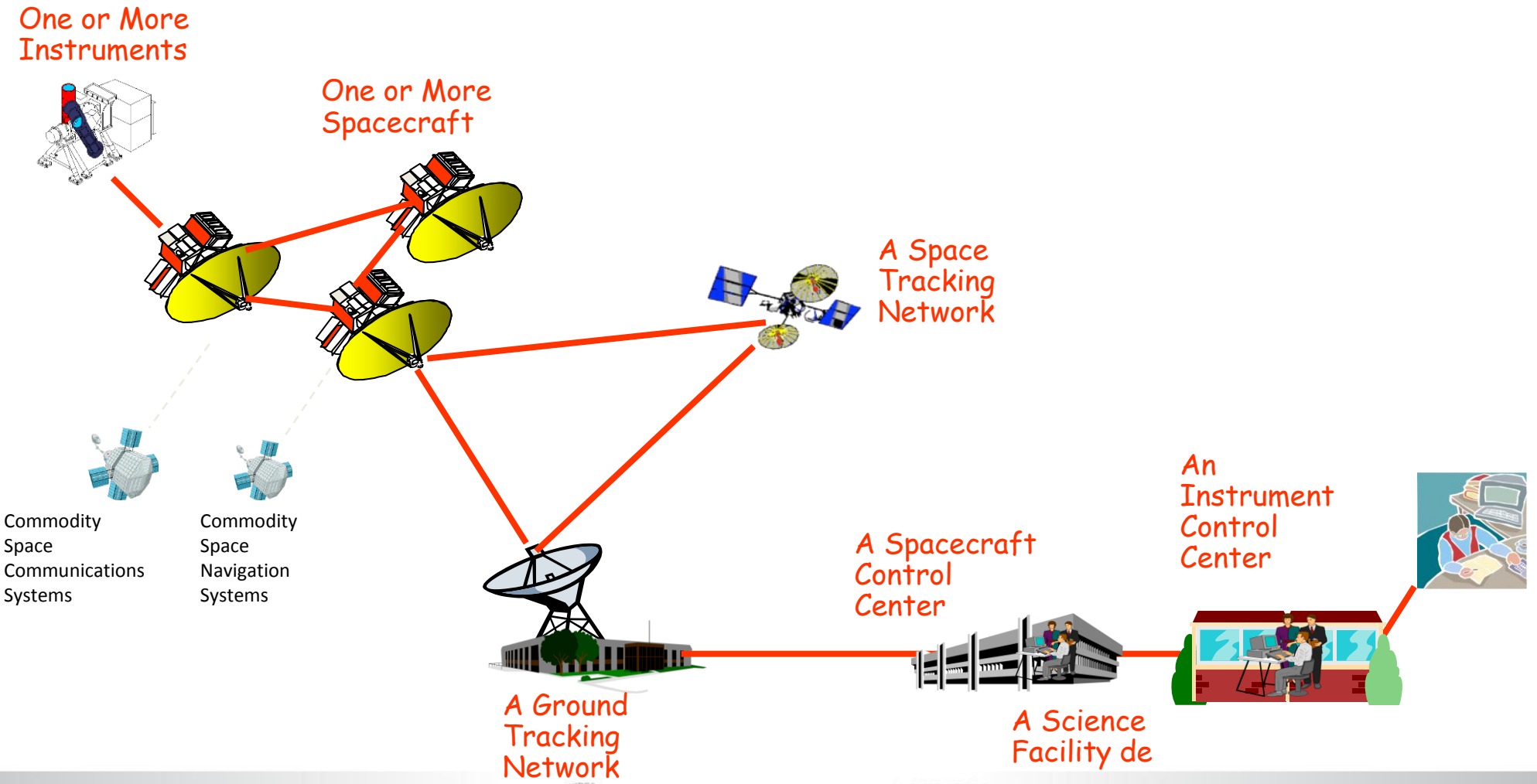
OODT Reusable Data Grid Framework

Mission Data Repositories

Biomedical Data Repositories

Engineering Data Repositories

Arquitectura espacial de principio-a-fin



Source: A. Hooke, NASA/JPL

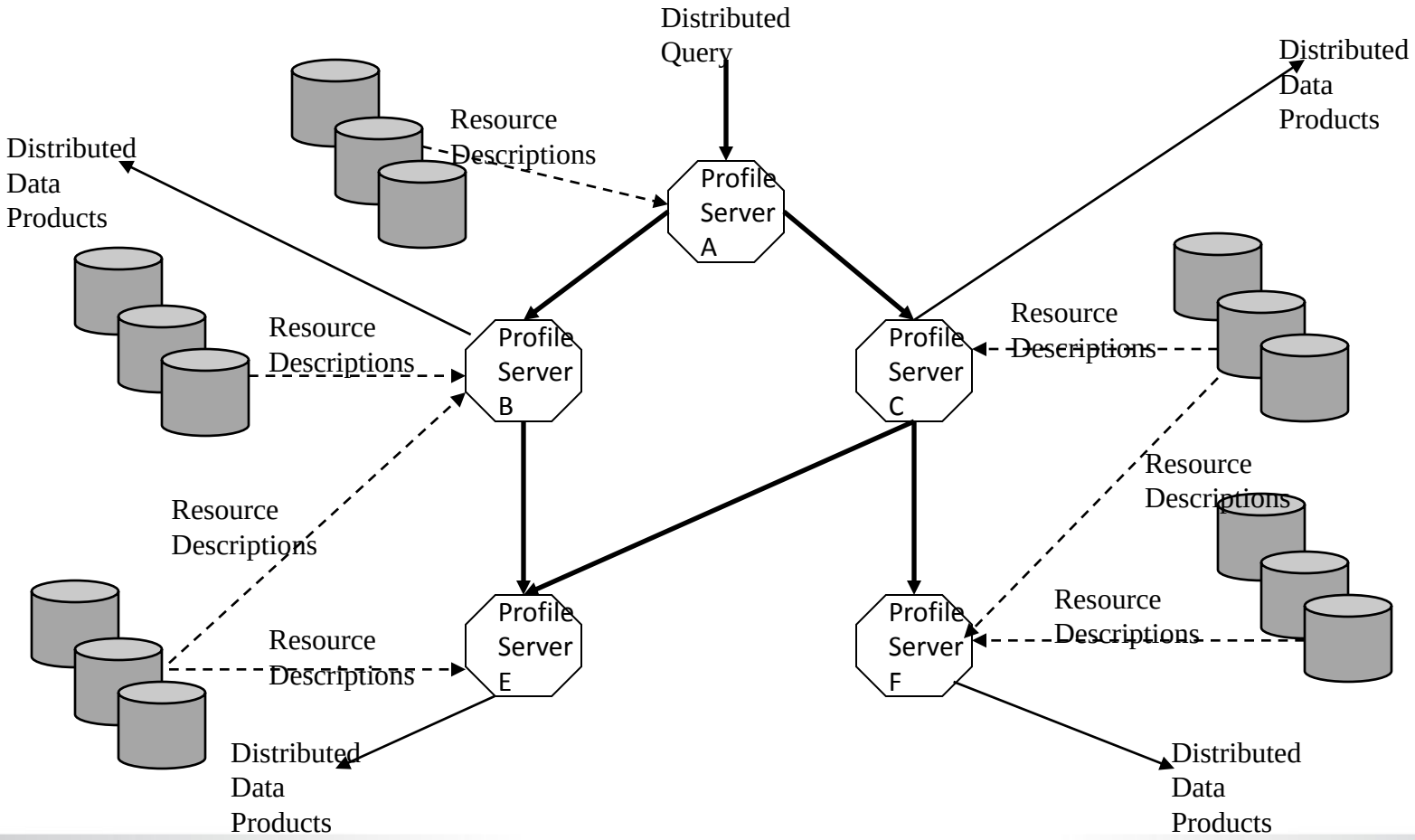
Implantación del software

- Desarrollo de un algoritmo de búsqueda.
- OODT esta basado en software libre: www.openchanelsoftware.com
- Basado en estandares: ISO/IEC 11179
- Esquemas XML comunes para intercambio de datos.

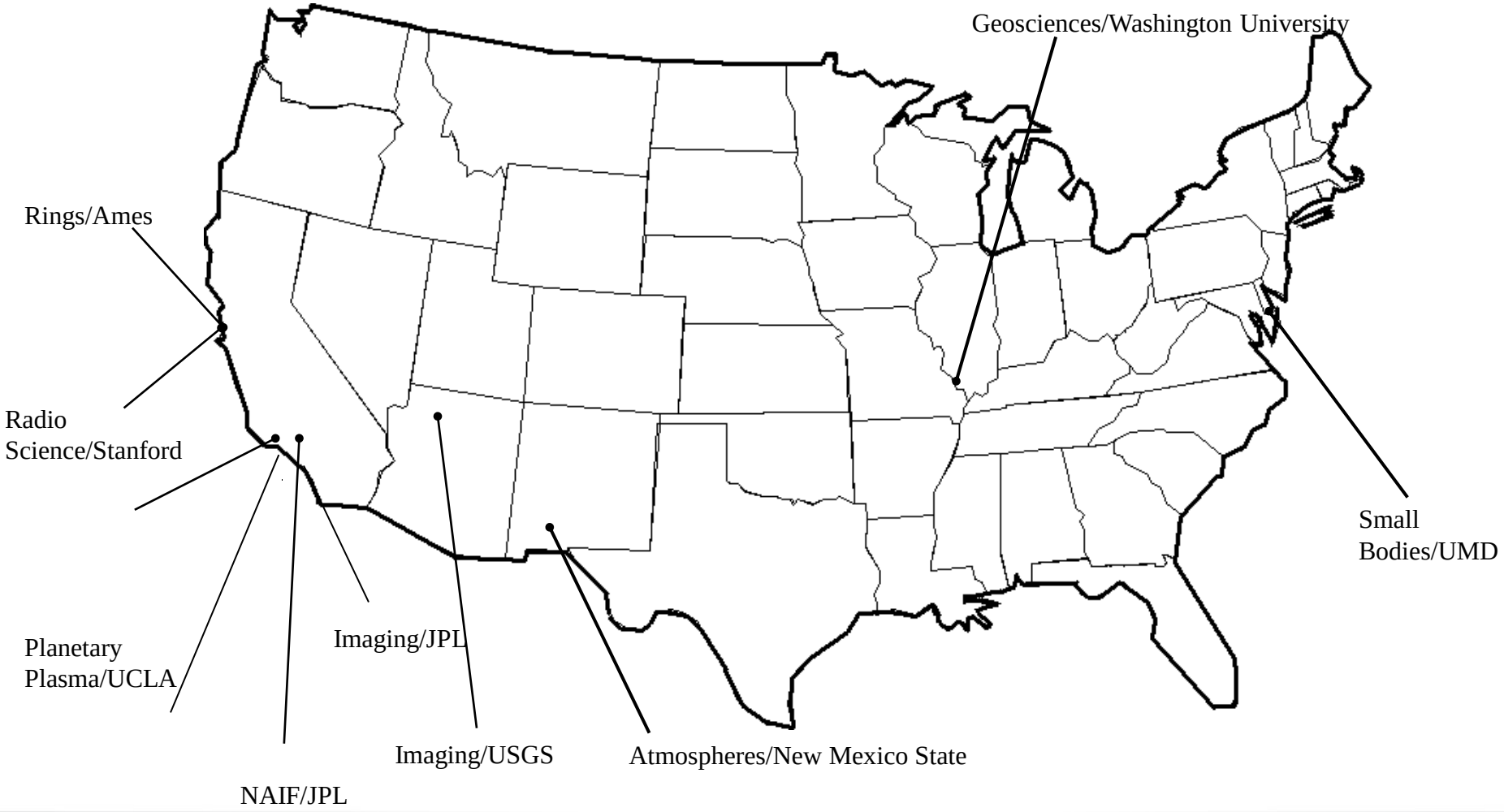
Planetary Data System (PDS)

- Es el archivo de datos de ciencias planetarias de NASA
- Disponible bajo contrato, asegura su resguardo y disponibilidad a la comunidad científica.
- PDS es un sistema distribuido diseñado para optimizar el descuido científico en el proceso de archivo.

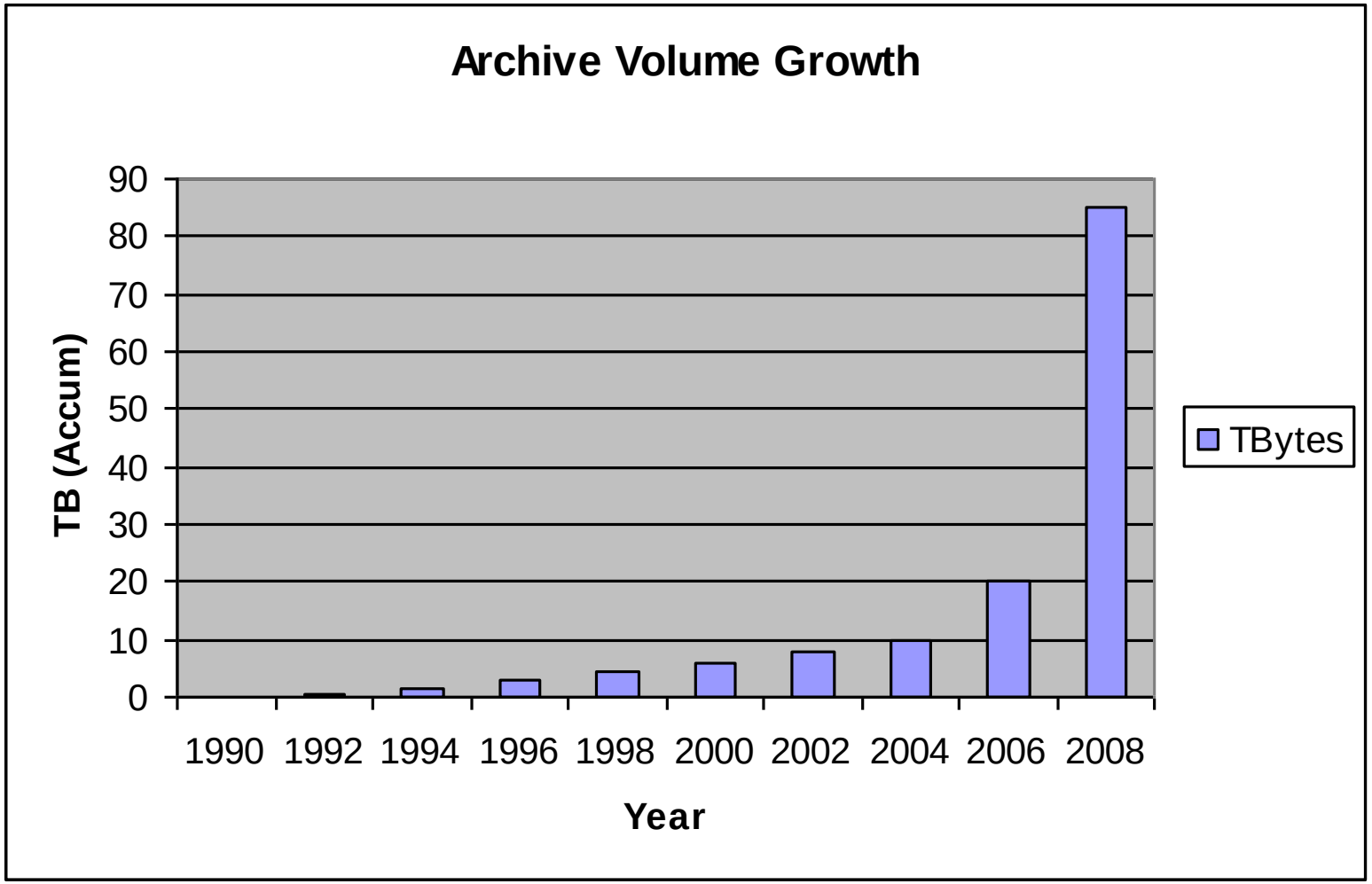
escalabilidad para buscar Recursos Distribuidos



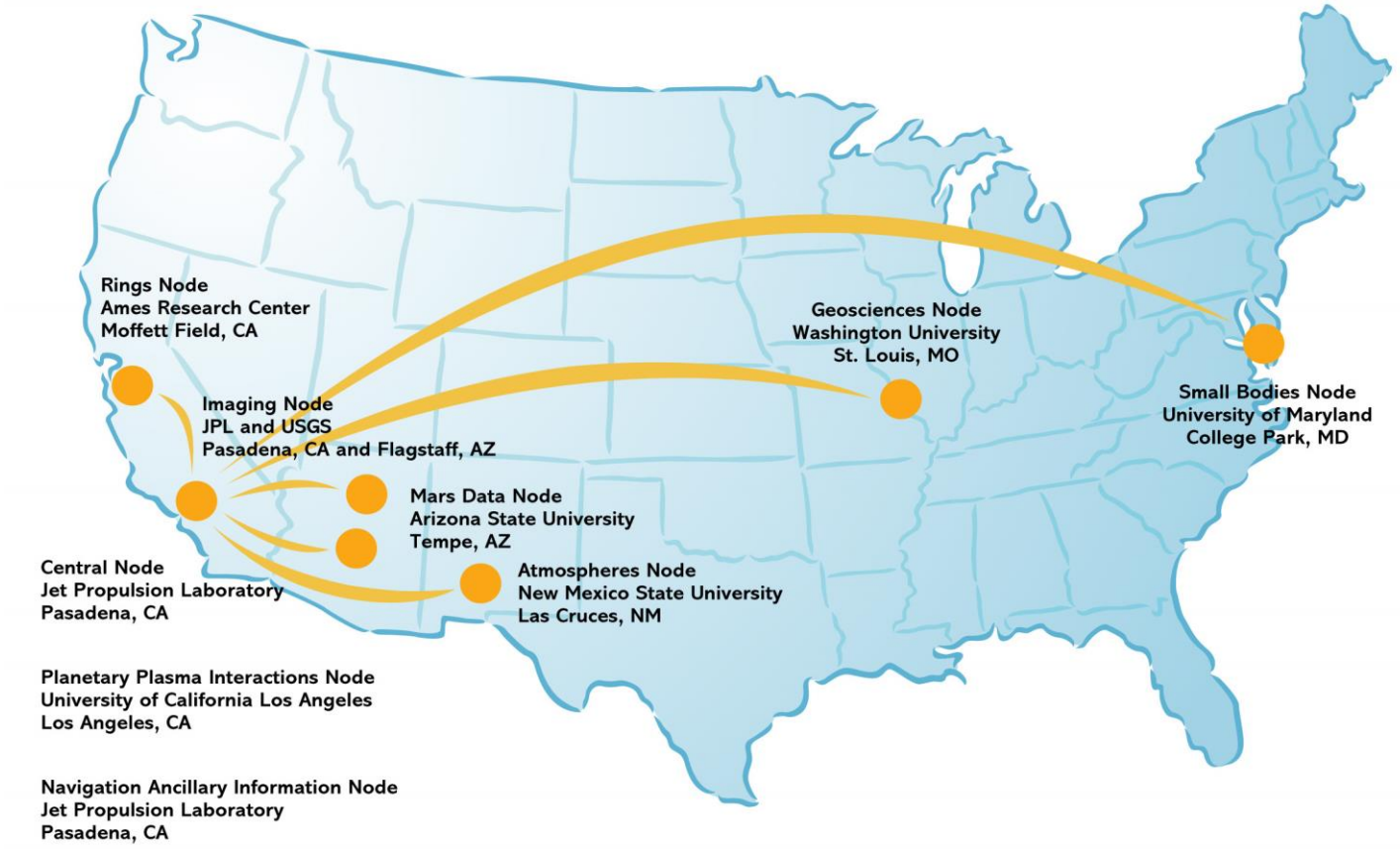
PDS Nodos e Instituciones (Silos)



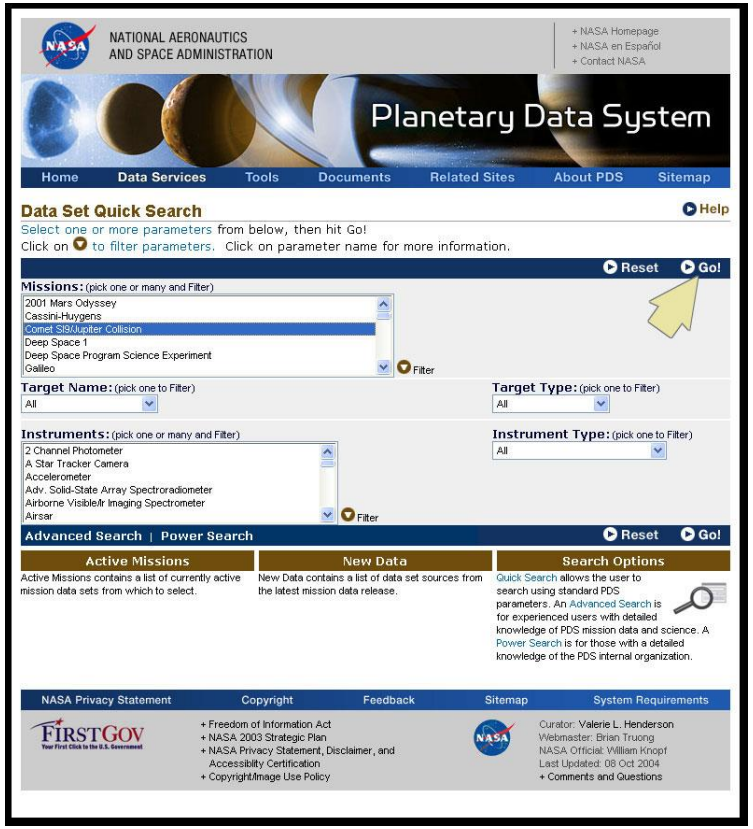
Crecimiento del Archivo por MisionesPlanetarias

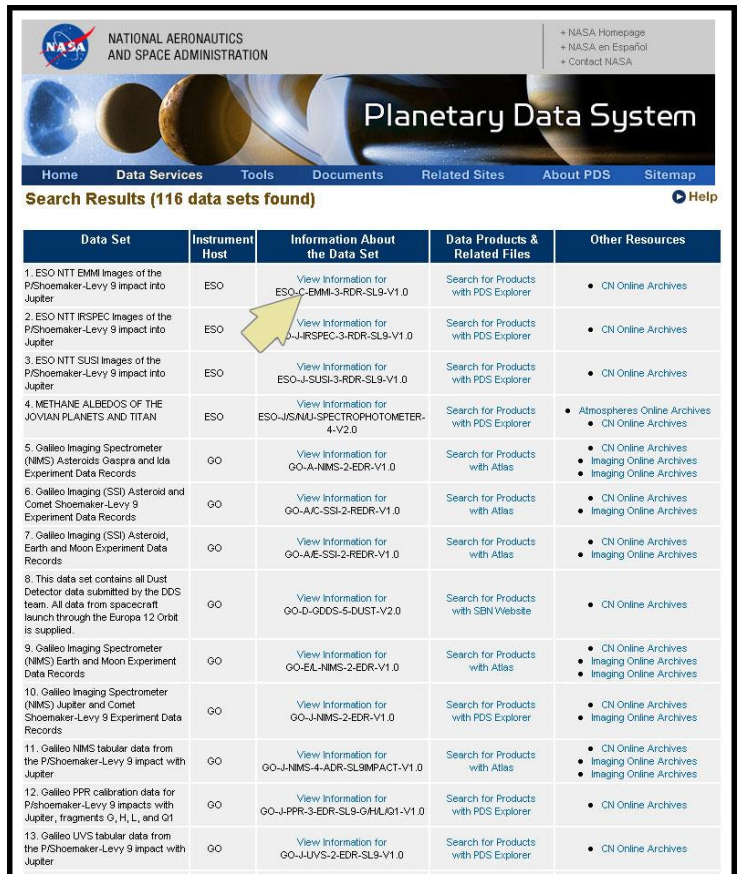


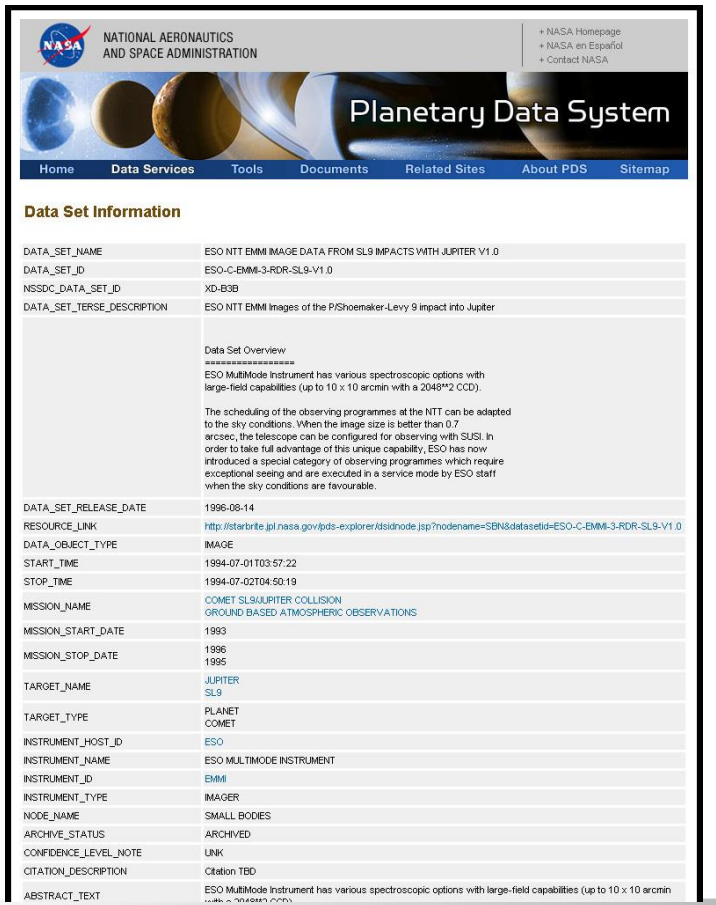
OODT Infrastructura de Datos de Ciencias Planetarias emplazada



PDS interfaz de usuario para busquedas







Juno >> WAV

VOLUME 0000: JUNO WAVES STANDARD RAW PRODUCTS

JNO-E/J/SS-WAV-2-EDR-V1.0

Start Time: 2011-08-09 00:00:00 - Stop Time: 2017-04-10 00:00:00

This volume contains documentation and all available Juno Waves telemetry for spacecraft event times (SCET) 2011-08-05 to 2017-04-10. These are raw data and are not intended for use by the wider community. The companion volume JNOWAV_1000 contains calibrated, full resolution, electric and magnetic field data of greater interest to most more ...

Top

- CATALOG
- DATA
- DOCUMENT
- EXTRAS
- INDEX
- AAREADME
- ERRATA.TXT
- VOLDESC.CAT

Juno >> WAV

VOLUME 0000: JUNO WAVES STANDARD RAW PRODUCTS

JNO-E/J/SS-WAV-2-EDR-V1.0

Start Time: 2011-08-09 00:00:00 - Stop Time: 2017-04-10 00:00:00

This volume contains documentation and all available Juno Waves telemetry for spacecraft event times (SCET) 2011-08-05 to 2017-04-10. These are raw data and are not intended for use by the wider community. The companion volume JNOWAV_1000 contains calibrated, full resolution, electric and magnetic field data of greater interest to most investigators. less ...

Top/ DATA/ 2011_2XX/ 221 /

WAV_2011221T165656_HRS_REC_V01

Start Time:	2011-08-09T16:56:56.796	Stop Time:	2011-08-09T16:59:53.171

WAV_2011221T235924_HRS_REC_V01

Start Time:	2011-08-09T23:59:24.796	Stop Time:	2011-08-10T00:00:07.171

WAV_2011221_HSK_V01

Start Time:	2011-08-09T00:00:00.000	Stop Time:	2011-08-10T00:00:00.000

WAV_2011221_LRS_V01

Start Time:	2011-08-09T00:00:00.000	Stop Time:	2011-08-10T00:00:00.000

Gracias