ApacheCon North America
Portland, Oregon 26th – 28th February 2013

Detecting radio-astronomical "Fast Radio Transient Events" via an OODT-based metadata processing pipeline

Chris Mattmann, Andrew Hart , Luca Cinquini
David Thompson, Kiri Wagstaff, Shakeh E. Khudikyan
NASA Jet Propulsion Laboratory,
California Institute of Technology

Radio Initiative: Archiving

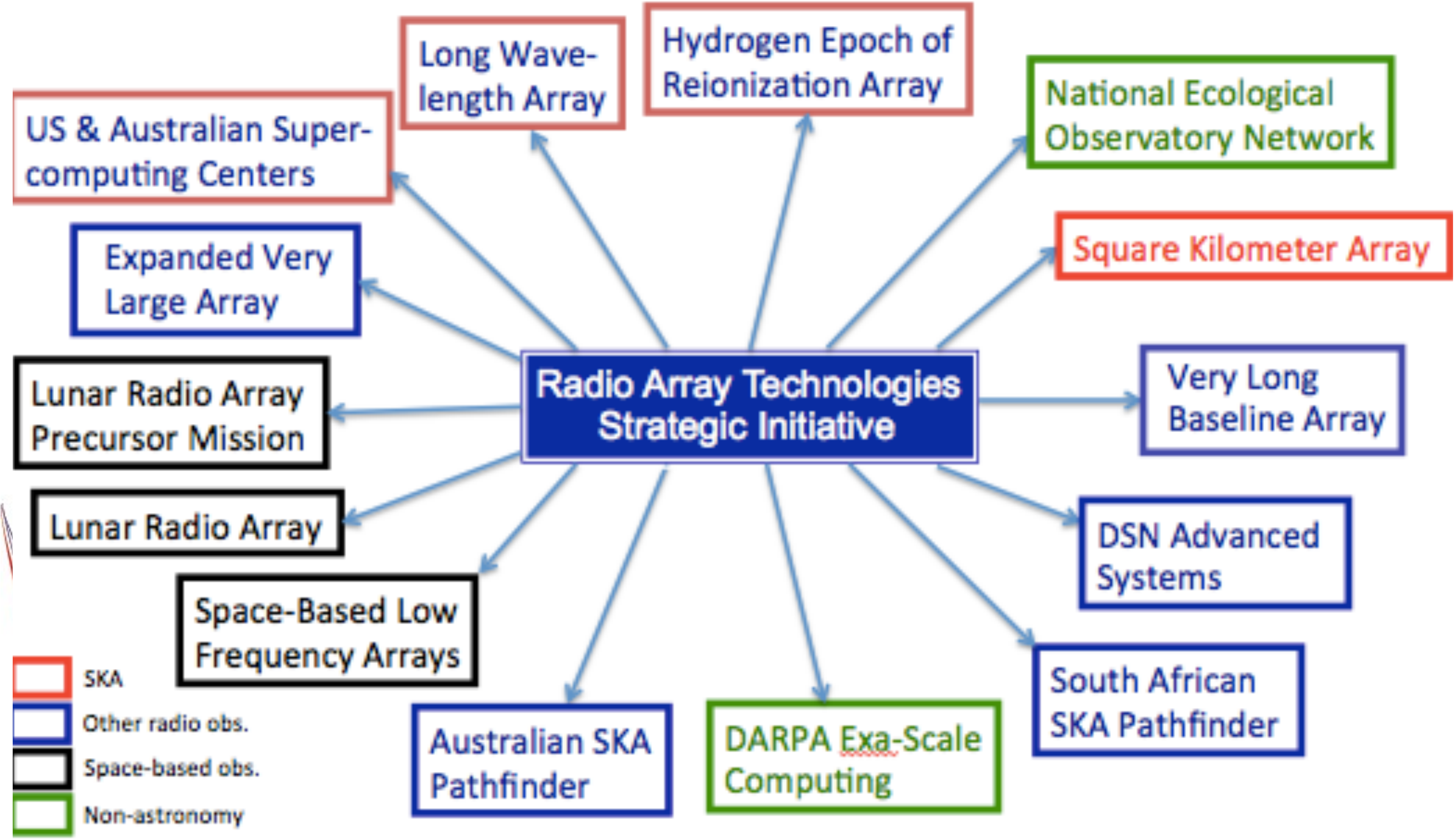
- **<u>Initiative Lead: Dayton Jones; Champion: Robert Preston</u>**
- **We will define the necessary data services and underlying substrate to position JPL to compete for and lead "big data" management efforts in astronomy, specifically, SKA, HERA, SKA precursors, and NRAO.**

- **Perform prototyping and deployment to demonstrate JPL's leadership in the "big data" and astronomy space.**

- **Collaborate on Data Products and Algorithms from Adaptive Data Processing task**

- **Establish partnerships with major SKA potential sites and pre-cursor efforts (South Africa, Australia)**

JPL Radio Array Technologies



Radio Array Technologies Strategic Initiative

US & Australian Super-computing Centers

Long Wave-length Array

Hydrogen Epoch of Reionization Array

National Ecological Observatory Network

Square Kilometer Array

Expanded Very Large Array

Very Long Baseline Array

Lunar Radio Array Precursor Mission

Lunar Radio Array

DSN Advanced Systems

Space-Based Low Frequency Arrays

South African SKA Pathfinder

Australian SKA Pathfinder

DARPA Exa-Scale Computing

🟥 SKA

🟦 Other radio obs.

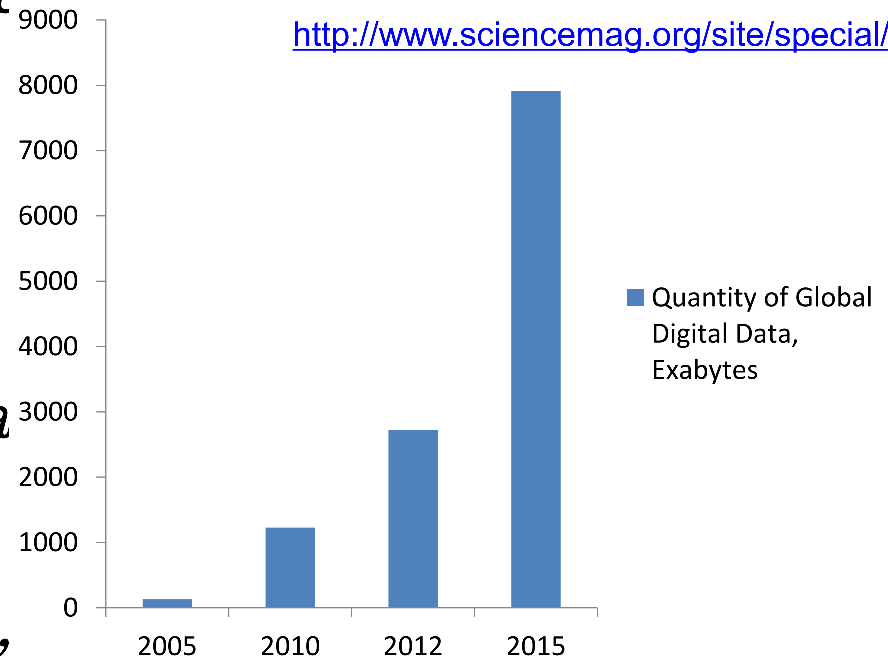
⬛ Space-based obs.

🟩 Non-astronomy

JPL "Big Data" Initiative

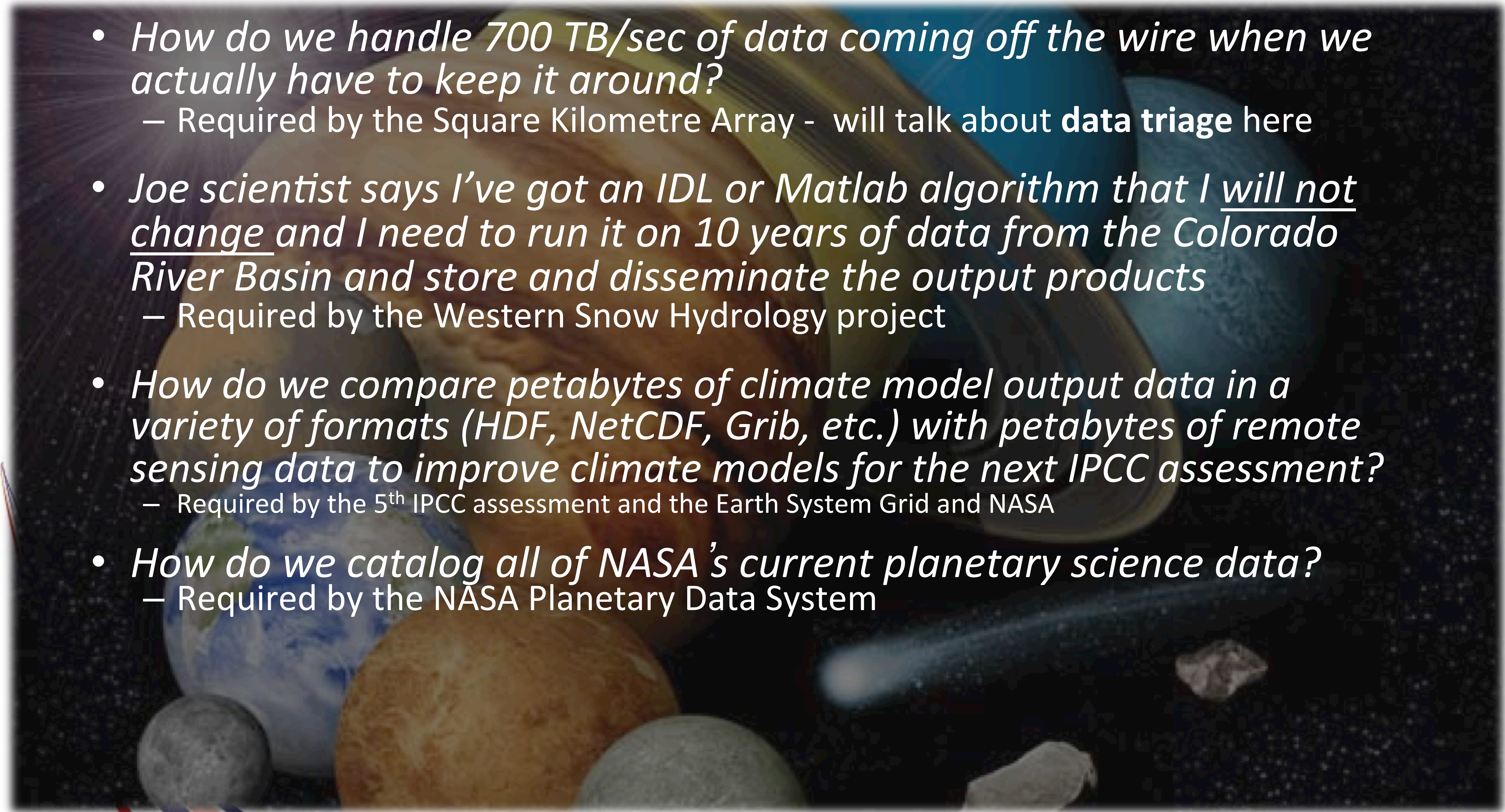
- The Big Picture
  - Astronomy, Earth science, planetary science, life/physical science all *drowning in data*
  - Fundamental technologies and emerging techniques in archiving and data science
    - Largely center around *open source* communities and related systems
- Research challenges (adapted from NSF)
  - *More data is being collected than we can store*
  - *Many data sets are too large to download*
  - *Many data sets are too poorly organized to be useful*
  - *Many data sets are heterogeneous in type, structu*
  - *Data utility is limited by our ability to use it*
- Our Focus: Big Data Archiving
  - *Research methods for integrating intelligent algorithms for data triage, subsetting, summariza*
  - *Construct technologies for smart data movement*
  - *Evaluate cloud computing for storage/processing*
  - *Construct data/metadata translators "Babel Fish'*

http://www.sciencemag.org/site/special/data/



Source: EMC/IDC Digital Universe Study, 2011

Some “Big Data” Grand Challenges



- *How do we handle 700 TB/sec of data coming off the wire when we actually have to keep it around?*
  - Required by the Square Kilometre Array - will talk about **data triage** here
- *Joe scientist says I’ve got an IDL or Matlab algorithm that I will not change and I need to run it on 10 years of data from the Colorado River Basin and store and disseminate the output products*
  - Required by the Western Snow Hydrology project
- *How do we compare petabytes of climate model output data in a variety of formats (HDF, NetCDF, Grib, etc.) with petabytes of remote sensing data to improve climate models for the next IPCC assessment?*
  - Required by the 5th IPCC assessment and the Earth System Grid and NASA
- *How do we catalog all of NASA’s current planetary science data?*
  - Required by the NASA Planetary Data System

Image Credit: http://www.jpl.nasa.gov/news/news.cfm?release=2011-295

VFASTR

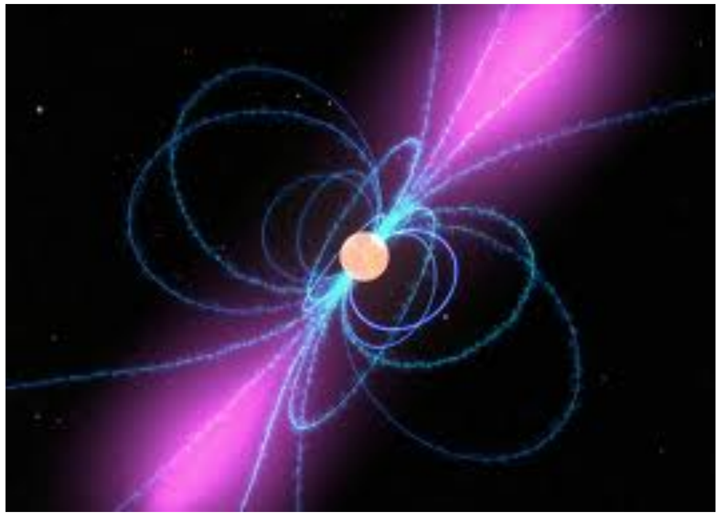
VFASTR ("VLBA Fast Radio Transients") is a project that aims at detecting short radio pulses (approx. a few milliseconds) from extra-terrestrial sources within the large amounts of data collected by the VLBA ("Very Large Baseline Array")

Fast Radio Transients may be generated by known and yet unknown sources:

- Pulsars
- Intermittent pulsars
- X-Ray binaries
- Supernovae

- Merging neutron stars
- Annihilating black holes
- ET signals ?
- New deep space objects

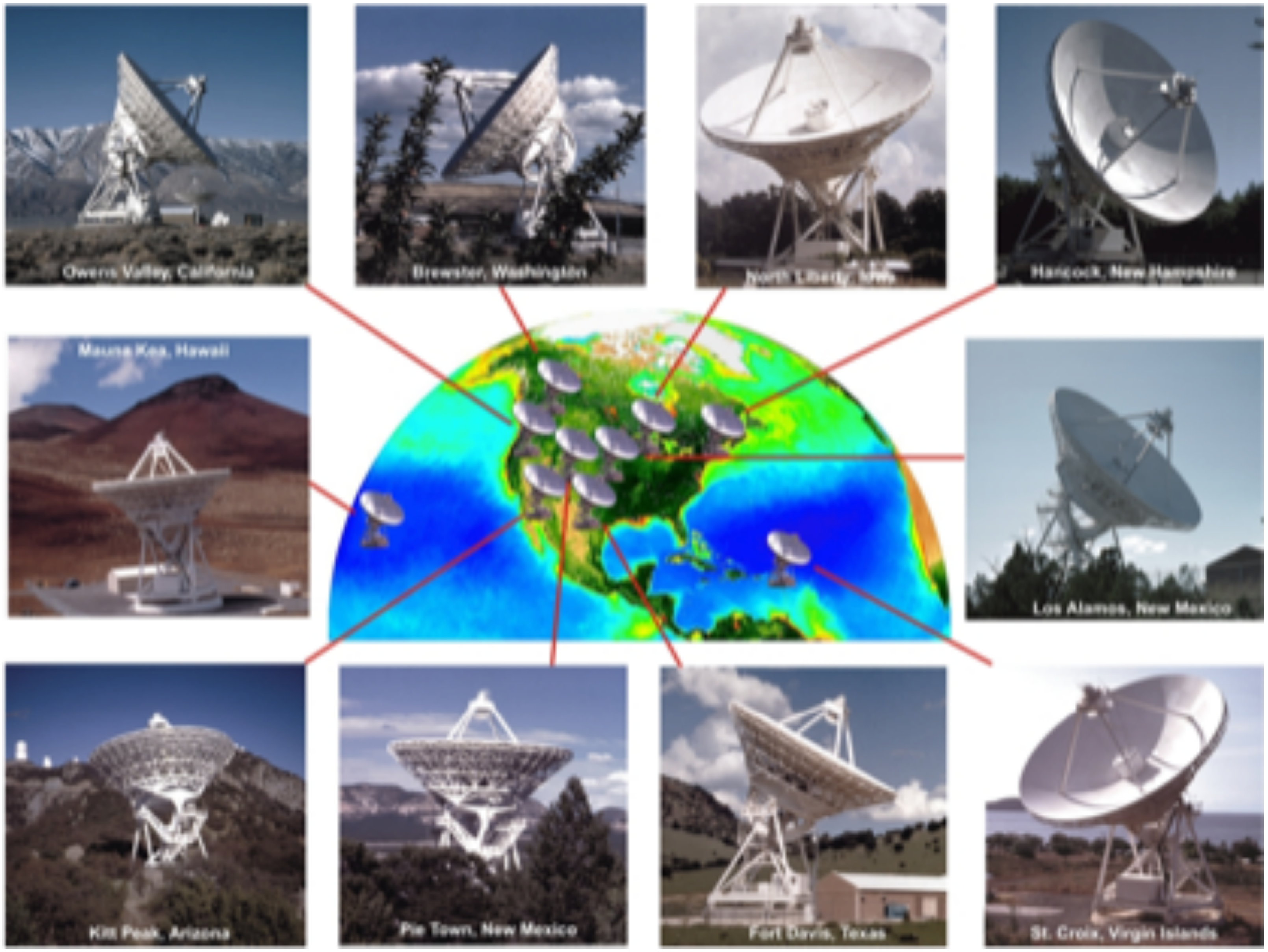
VFASTR is one of a new generation of Radio Astronomy experiments that aim at analyzing the "dynamic radio sky" as opposed to mapping and inspecting known static sources



VLBA

VLBA (“Very Large Baseline Array”) is a group of 10 large radio-telescopes (25m diameter) distributed across the U.S.A. from Hawaii to the Virgin Islands.

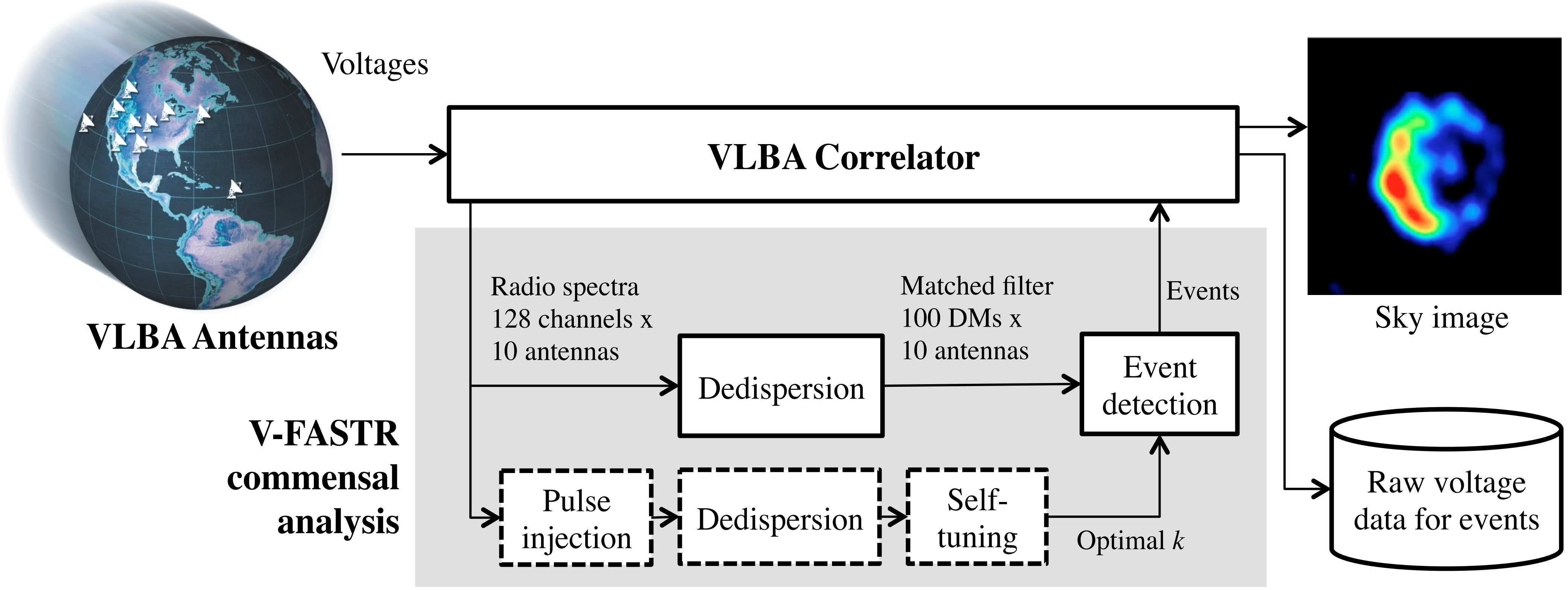
- No two antennas are within each other’s local horizon
- The overall array has a baseline of 800 km => resolution of milliarcsecond



VFASTR Commensal Approach

VFASTR employs a commensal (a.k.a. “passive”) approach by analyzing data that is collected during normal VLBA operations for other scientific purposes:

- Raw voltages from VLBA antennas are transferred to NRAO, time-correlated, corrected for dispersion through the interstellar medium (“de-dispersion”) and separated from instrument noise
- Candidate events are staged on disk and remain available for limited time
- VFASTR team must review tens of candidates daily, archive the promising ones and disregard the others

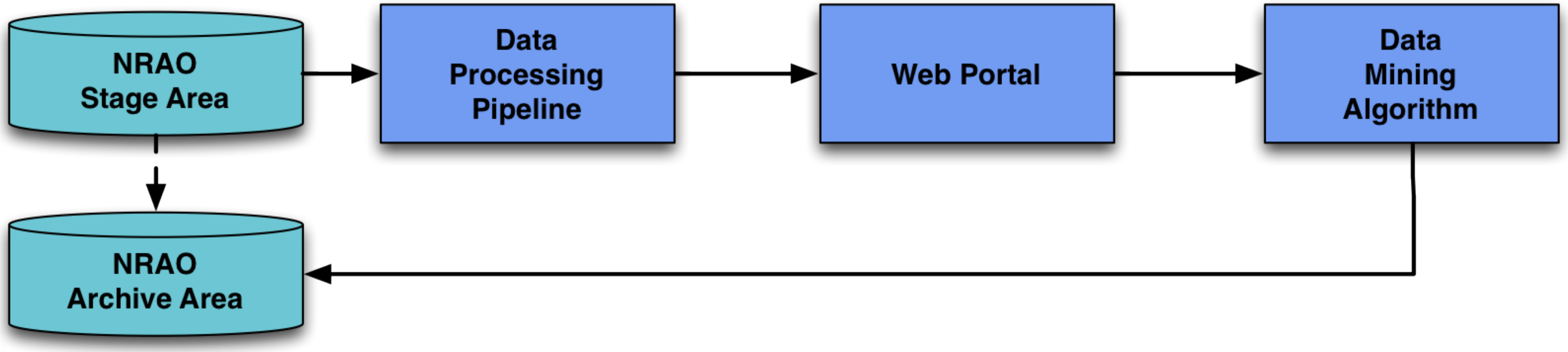


VFASTR Science Team: Astron (The Netherlands), ICRAR (Australia), JPL (U.S.A.), NRAO (U.S.A.)

VFASTR Data System Overview

- The software engineering team at JPL has developed an end-to-end data system in support of VFASTR activities with two major goals:
  - provide a web-based platform for easy and timely review of candidate events by the science team
  - enable the automatic identification of interesting events by a self-trained machine agent

- The system is composed of three major components:
  - Data processing pipeline: responsible for data transfer from NRAO and archiving at JPL, and for metadata extraction and cataloging
  - Web portal: easy accessible application for display of product data and metadata, and selection and tagging of interesting events
  - Data mining algorithm: analyzes the events pool and tags candidates with characteristics similar to sample interesting events



Apache OODT Overview

The VFASTR data processing pipeline was built by using Apache OODT in combination with other Apache and Open Source technologies

1 Ob Object
2 Or Oriented
3 D Data
4 T Technology
Catalogs, archives, metadata, & more
Data grid framework for transparent search and discovery of disparate science resources

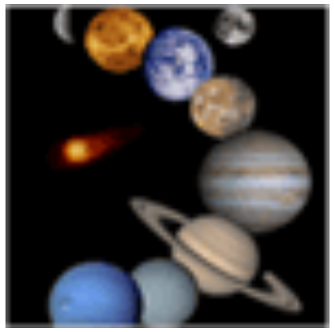
OODT: Object Oriented Data Technology: framework for management, discovery and access of distributed data resources. Main features:

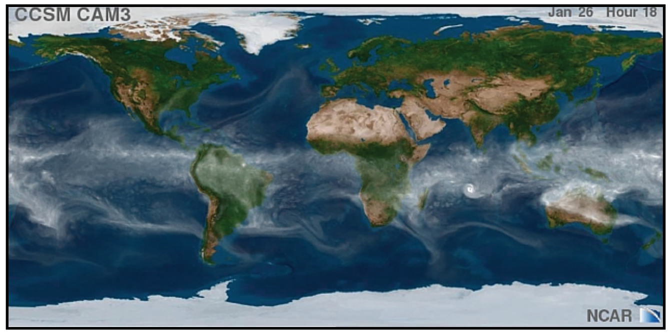
- Modularity: eco-system of standalone components that can be deployed in various configurations to fulfill a project specific requirements
- Configurability: each component can be easily configured to invoke alternate out-of-the-box functionality or deployment options
- Extensibility: components can be extended by providing alternate implementations to its core APIs (expressed as Java interfaces) or configuring custom plugins

Apache OODT Adoption

OODT is used operationally to manage scientific data by several projects in disparate scientific domains:

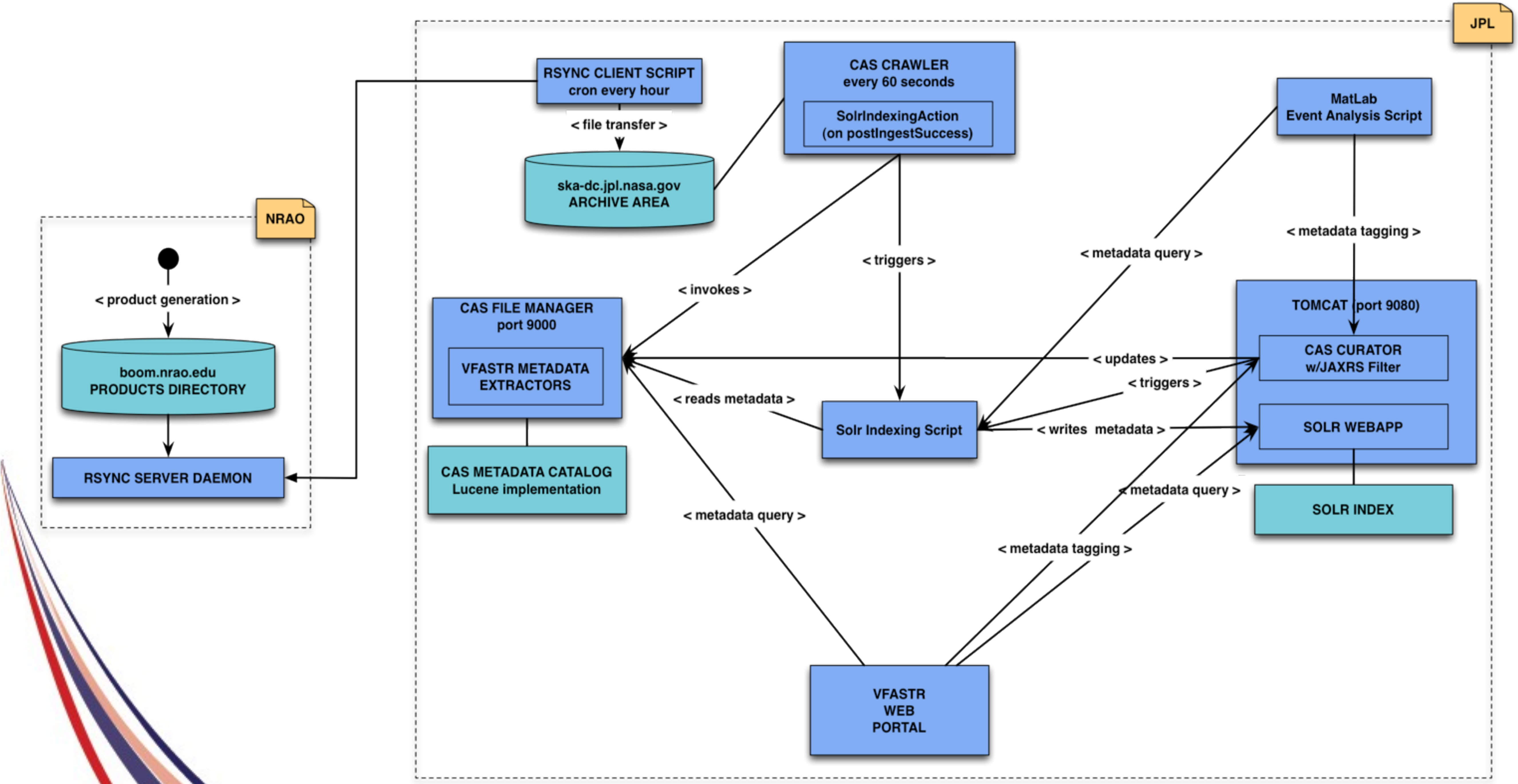
- Earth Sciences:
  - NASA satellite missions (SMAP,...) are using OODT components as the base for their data processing pipeline for generation and archiving of products from raw observations
  - ESGF (Earth System Grid Federation) used OODT to build and publish observational data products in support of climate change research
- Health Sciences: EDRN (Early Detection Research Network) uses OODT to collect, tag and distribute data products to support research in early cancer detection
- Planetary Science: PDS (Planetary Data System) is developing data transformation and delivery services based on OODT as part of its world-wide product access infrastructure
- Radio Astronomy: several projects (ALMA, Haystack,...) are adopting OODT based on successful VASTR example





VFASTR Data System Architecture

Data products are continuously generated by the VLBA ground and processing system and stored on temporary cache at NRAO.



Data products are transferred to JPL where metadata is extracted, products are made available for review by scientists, sub-selected for further analysis, and tagged.

Data System Design Considerations

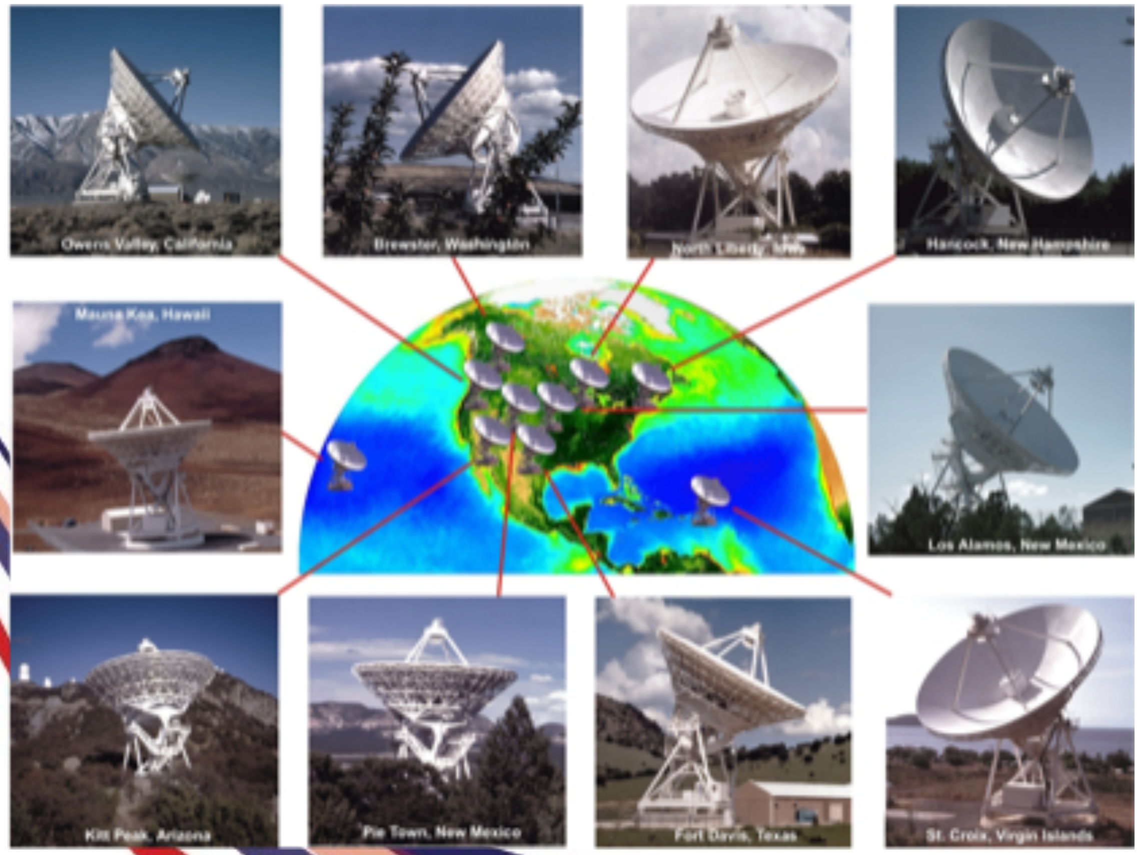
- Some of the architectural decisions that factored in the data system design were motivated by specific project constraints:
  - Minimize impact on NRAO resources: because VFASTR is a "guest" project at NRAO, attention had to be paid to limit use of disk storage, network bandwidth and CPU resources
  - Security: all NRAO resources were exposed as "read-only": no action initiated at JPL could result in any modification of the original products (or compromise the NRAO system)

- Architecture evolved over time as a result of new requirements such as increased data volumes and higher frequency updates
  - Use of different OODT and Apache components (Lucene vs MySQL data store back-ends, Solr for fast metadata retrieval)
  - Development of new OODT functionality (RESTful API for metadata updates)

VFASTR Data Products

VFASTR data is logically organized into three levels: jobs, scans and events

- Job: a batch of data that is processed at one time, and stored together on physical disk. They are associated with a specific investigator scientist. Each contains 1-100+ scans.
- Scan: a physical telescope pointing, e.g. a period where the antennas are all directed to a common point on the sky. They have durations of 1-100+ seconds.
- Event: a time segment that the system thinks is interesting. Duration is usually about 1-2 seconds. Most scans have no such events, but some have a dozen or more. The “interesting” part of an event is much shorter: 5-50 milliseconds.



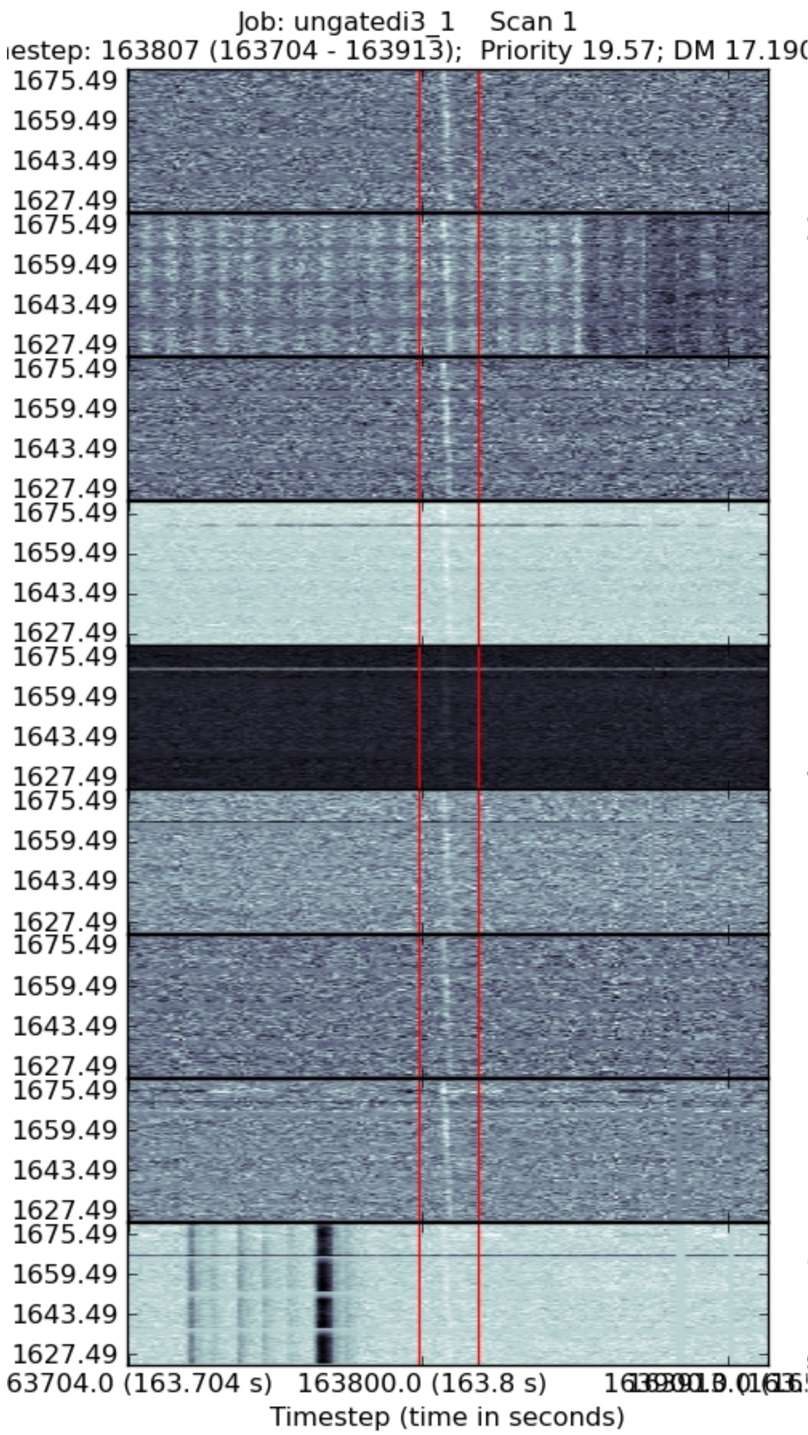
VFASTR Data Product: directory tree containing all data recorded for a single job (“tns_bmp360p2_44”)

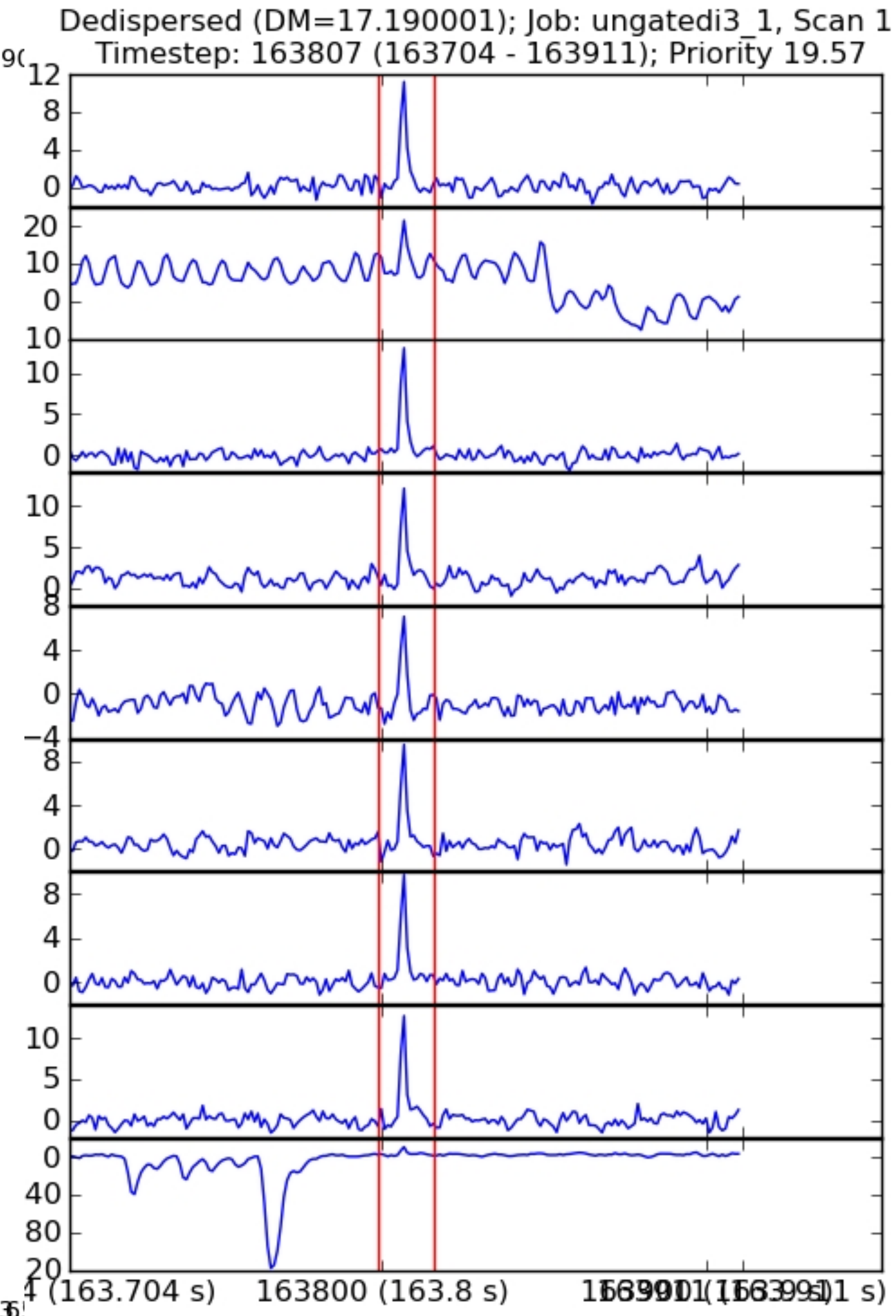
- Job calibration files
- Scan output files
- Event raw voltages
- Event reconstructed images
- ...and other files....
- Approximately 1-100 GB

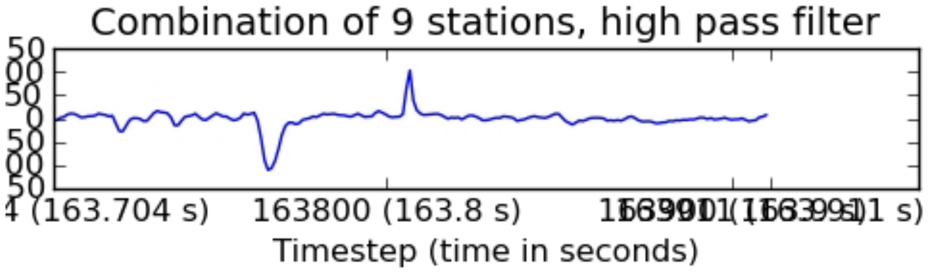
Pulsar PSR J0826+2637 Signal

Telescope signal processing:

- Time correlation
- “De-dispersion” (i.e. corrected for dispersion in interstellar medium)
- “Adaptive excision” (some telescopes are disregarded based on self-learning algorithm)

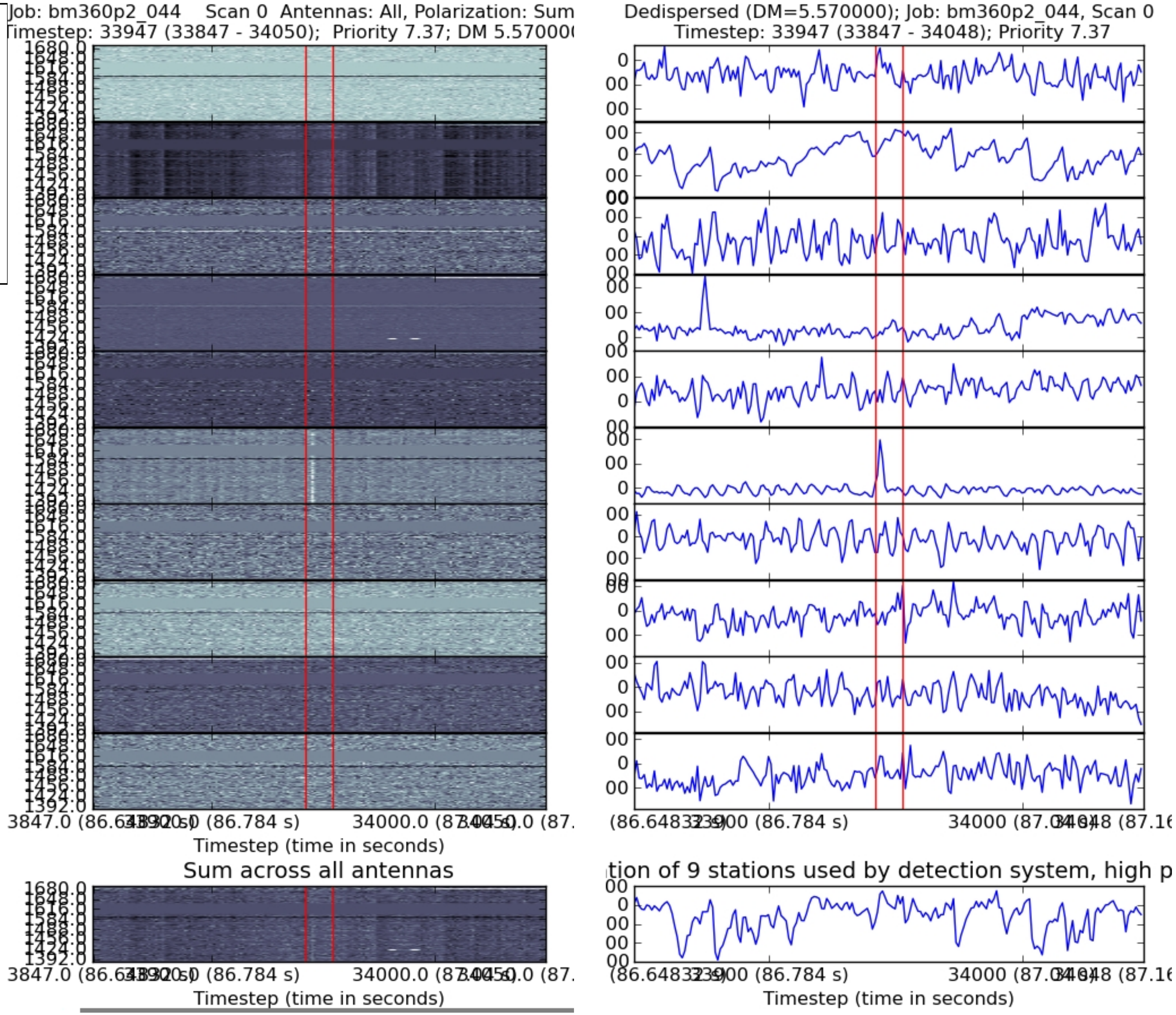




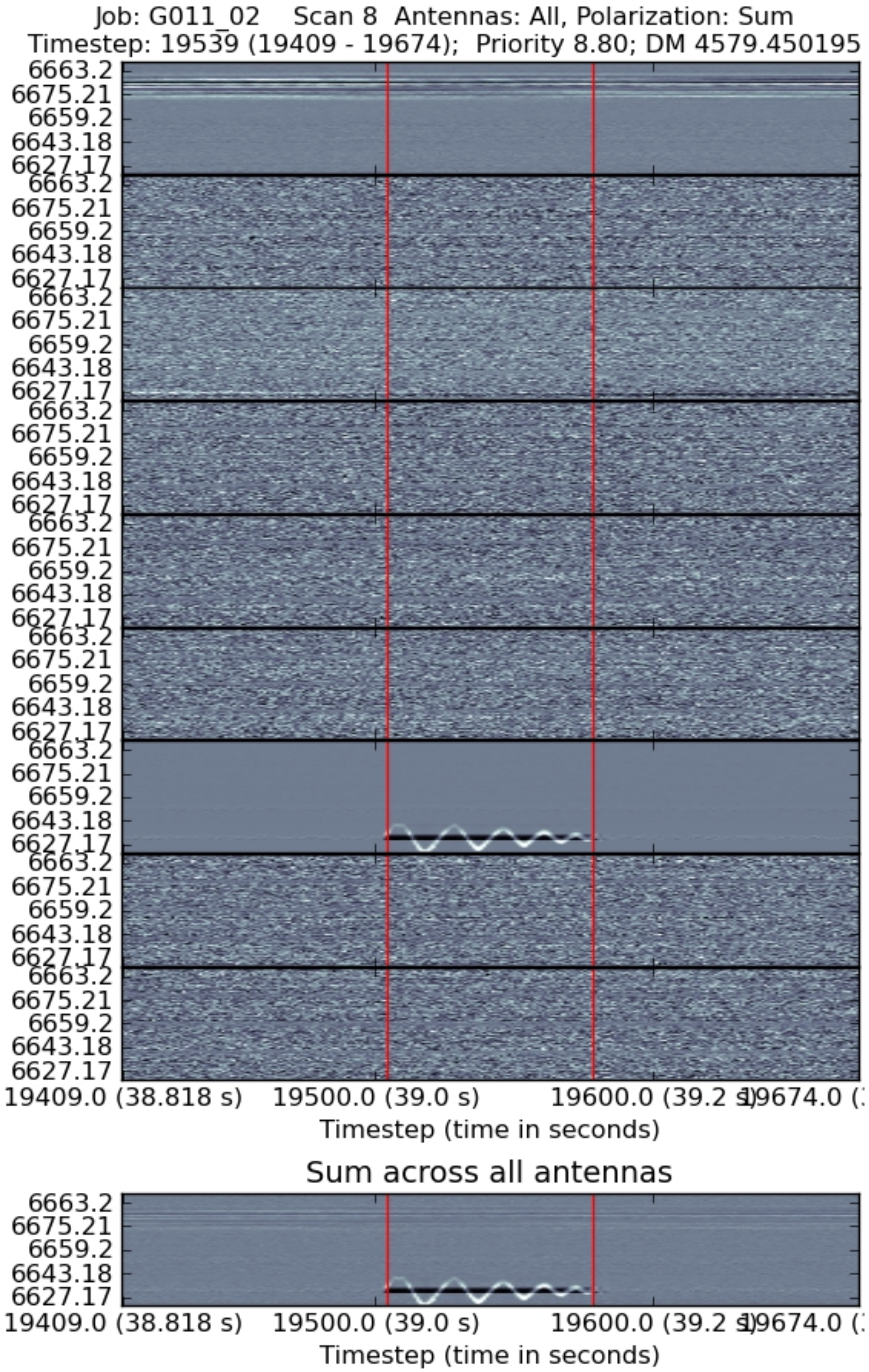


Radio Frequency Interference (RFI)

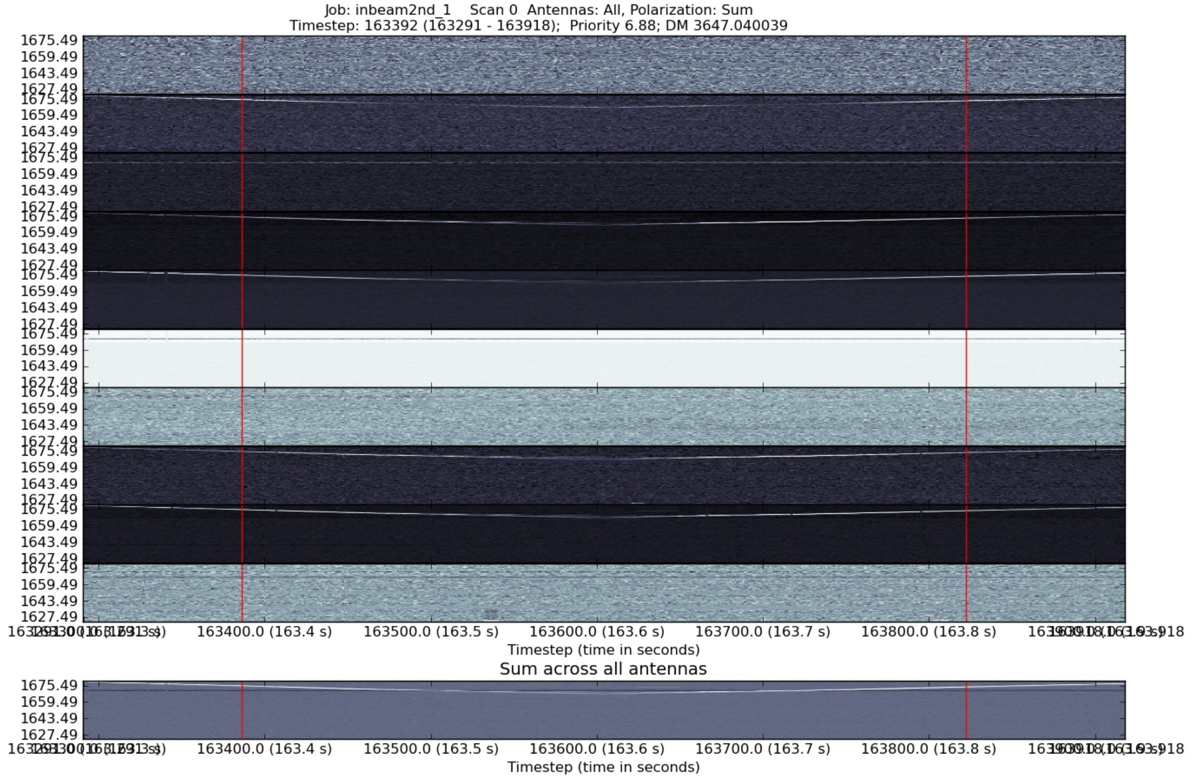
- RFI events only detected by one or two antennas, have terrestrial origin

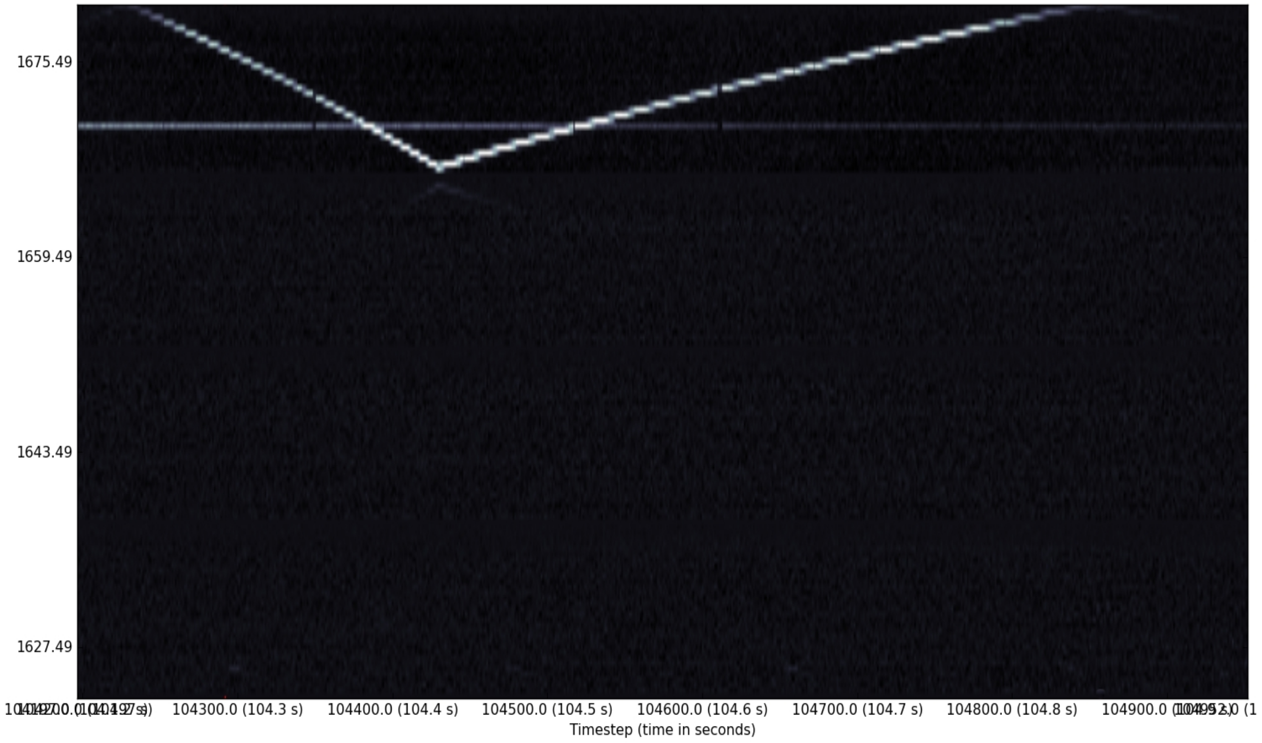


“Mystery” Events



“V-chirp”



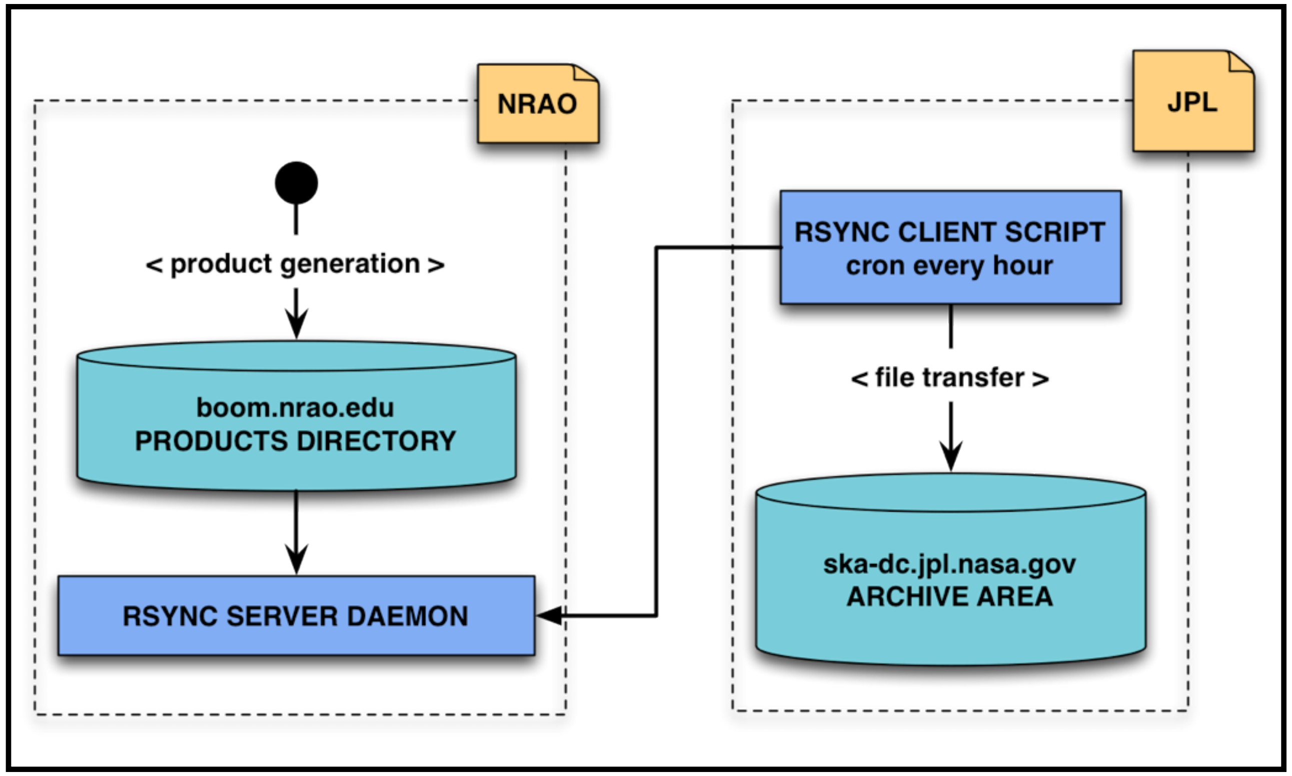


Rsync

Rsync: freely available utility for Unix systems that can be used to synchronize the content of directory trees between two hosts with minimal human intervention.

Features:

- Easy deployment
- Extensive range of configuration options
- High performance: only file changes are transferred ("delta encoding") between sub-sequent invocations, + optional compression
- Optional recursion into sub-directories
- Reliability: "turn-key" toolkit



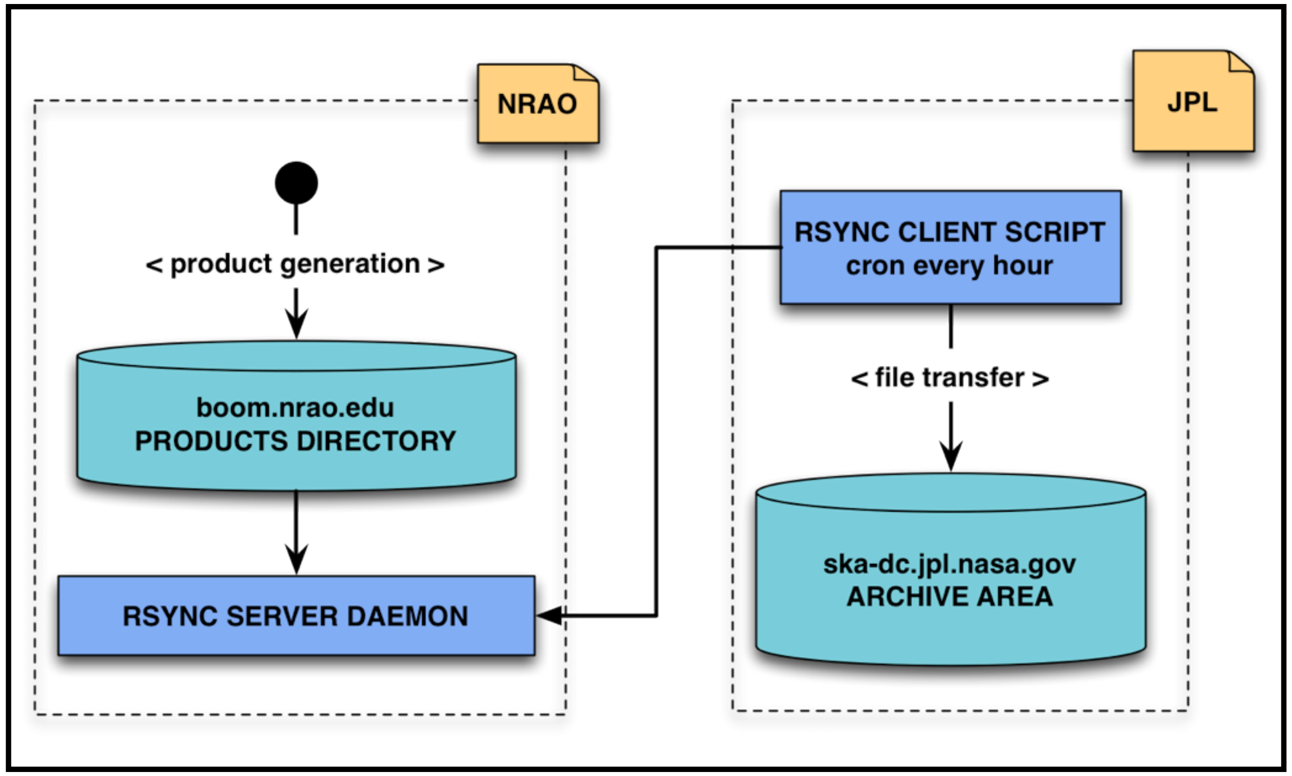
Rsync

VFASTR deployment:

- rsync server daemon was deployed at NRAO to make VFASTR products available for download
  - Configured for read-only mode
  - Limited to requests coming from JPL IPs
- rsync client running at JPL as system cron job to pull data every hour
  - Configured to only transfer a subset of the product files (images, output, calibration files)

Measured Data Transfer Rates:

- ~ 2MB/sec between NRAO and JPL
- Approximately 10-20 products per day
- Average volume for transferred product: ~50MB (reduced from 50GB)
- Can transfer all (reduced) daily products in a few minutes!

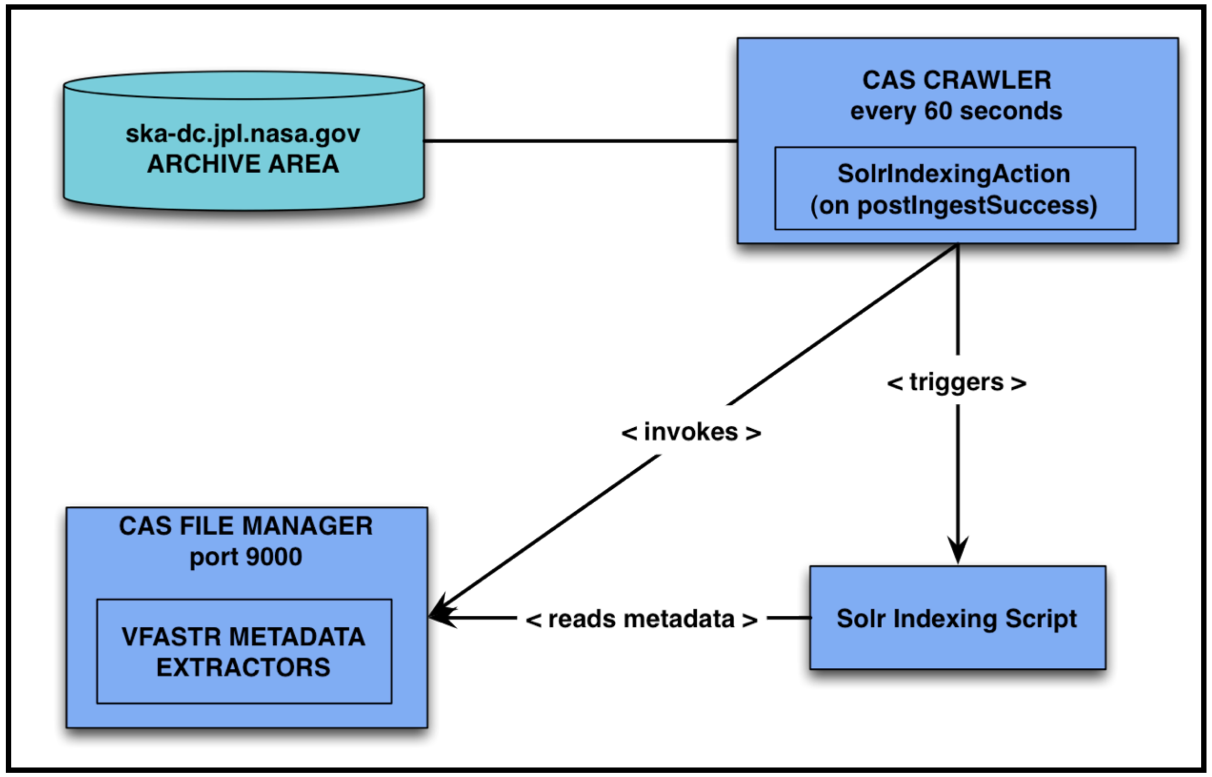


CAS Crawler

The CAS Crawler is an OODT component that can be used to list the contents of a staging area and submit products for ingestion to the CAS File manager. Typically used for automatic detection of new products transferred from a remote source.

VFASTR deployment:

- Run as daemon every 300 seconds
- In-place archiving of products (no movement)
- Preconditions:
  - Product must be complete
  - Product must be no older than 10 days
  - Product must not exist in catalog already
- Post-ingest action on success:
  - Trigger metadata harvesting by Solr script

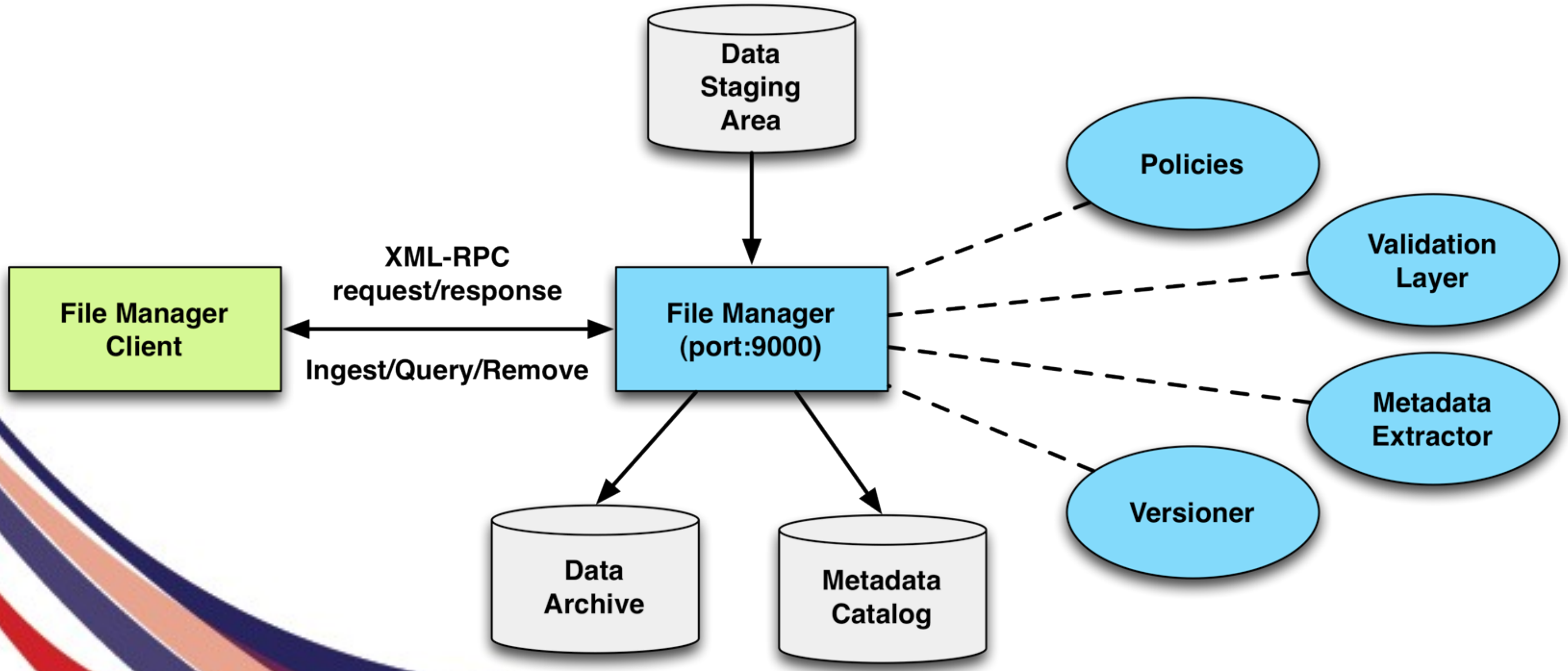


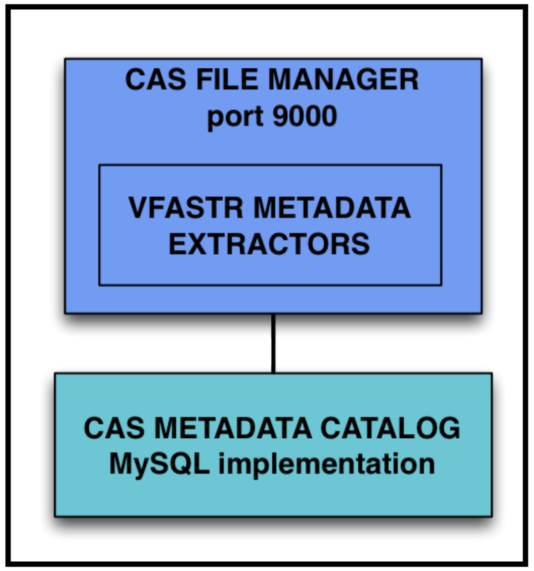
CAS File Manager

The CAS File Manager is an OODT service for cataloging, archiving and delivery of data products (files and directories) and associated metadata. It is used as core data management component in most OODT-based data systems.

VFASTR Deployment:

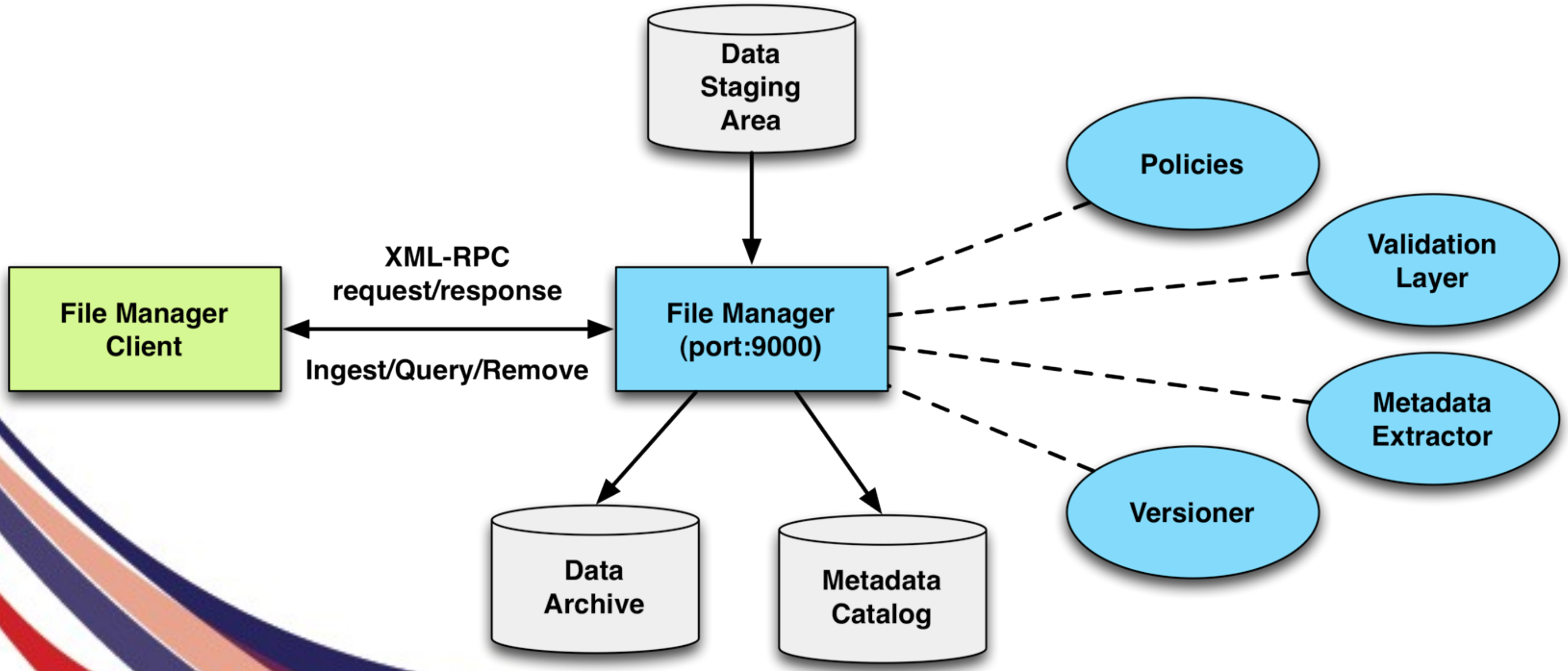
- Policy Files: define a single VFASTR metadata type to capture ALL information associated with a single product (job, scans and events)
  - The full metadata for a product can be retrieved by a client with one request
  - Metadata keys must be named dynamically to capture job-scan-event references
  - Example: key=EventStartDateTime_s6 values=2013-01-12T15:48:21.800-0800, 2013-01-12T15:48:22.830-0800 (scan 6 contains 2 events)

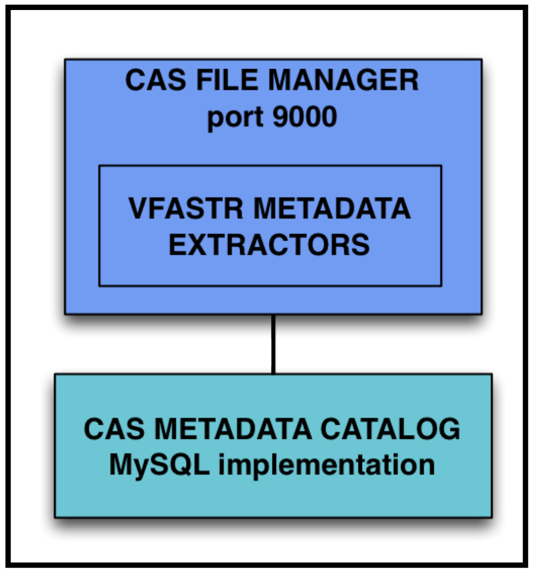



CAS FILE MANAGER port 9000
VFASTR METADATA EXTRACTORS
CAS METADATA CATALOG MySQL implementation

CAS File Manager

- Validation Layer: no validation applied as metadata fields are not known a-priori
  - Back-end catalog implementations had to be extended to allow for optional "lenient" behavior for ingested metadata
- Metadata Extractors: custom metadata extractors written to parse information for job, scans, events from directory structure, calibration and output files, and to assign detection images to the events that generated them
- Metadata Catalog: used both Lucene and MySQL back-ends
  - Switched to MySQL to support high-frequency updates
  - Lucene File Manager implementation now fixed to support high frequencies
- Data Transfer Protocol: archive products in place
  - Otherwise they would be re-transferred by rsync

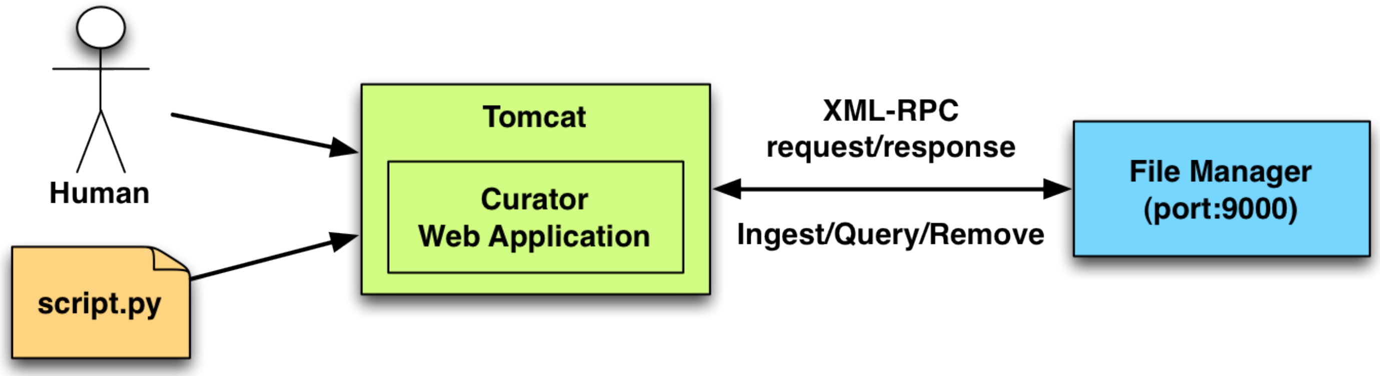


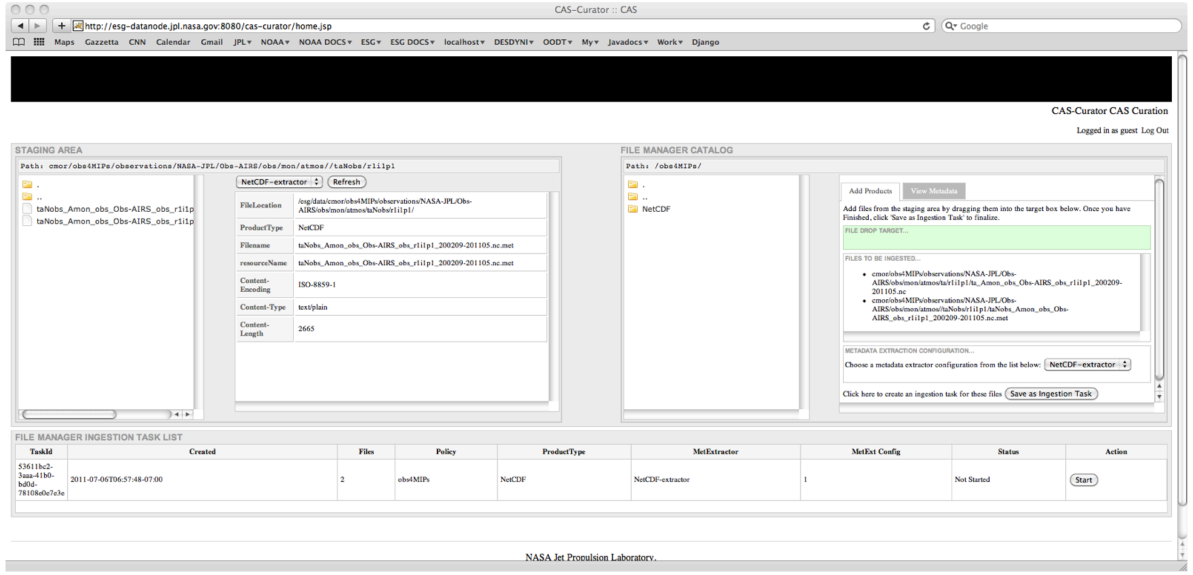

CAS FILE MANAGER port 9000
VFASTR METADATA EXTRACTORS
CAS METADATA CATALOG MySQL implementation

CAS Curator

The CAS Curator is a web application for interacting with the File Manager (i.e. web-based client for File Manager service):

- Submit data product ingestion jobs
- Inspect, add and update product metadata ("curation")




CAS-Curator :: CAS
http://esg-datanode.jpl.nasa.gov:8080/cas-curator/home.jsp
Google
Maps Gazzetta CNN Calendar Gmail JPL NOAA NOAA DOCS ESG ESG DOCS localhost DESDYNI OODT My Javadocs Work Django

CAS-Curator CAS Curation

Logged in as guest Log Out

STAGING AREA

Path: cmor/obs4MIPs/observations/NASA-JPL/Obs-AIRS/obs/mon/atmos//taNobs/r1i1p1

.
..
taNobs_Amon_obs_Obs-AIRS_obs_r1i1p
taNobs_Amon_obs_Obs-AIRS_obs_r1i1p

NetCDF-extractor | Refresh

FileLocation	/esg/data/cmor/obs4MIPs/observations/NASA-JPL/Obs-AIRS/obs/mon/atmos/taNobs/r1i1p1/
ProductType	NetCDF
Filename	taNobs_Amon_obs_Obs-AIRS_obs_r1i1p1_200209-201105.nc.met
resourceName	taNobs_Amon_obs_Obs-AIRS_obs_r1i1p1_200209-201105.nc.met
Content-Encoding	ISO-8859-1
Content-Type	text/plain
Content-Length	2665

FILE MANAGER CATALOG

Path: /obs4MIPs/

.
..
NetCDF

Add Products | View Metadata

Add files from the staging area by dragging them into the target box below. Once you have Finished, click 'Save as Ingestion Task' to finalize.

FILE DROP TARGET...

FILES TO BE INGESTED...

- cmor/obs4MIPs/observations/NASA-JPL/Obs-AIRS/obs/mon/atmos/ta/r1i1p1/ta_Amon_obs_Obs-AIRS_obs_r1i1p1_200209-201105.nc
- cmor/obs4MIPs/observations/NASA-JPL/Obs-AIRS/obs/mon/atmos//taNobs/r1i1p1/taNobs_Amon_obs_Obs-AIRS_obs_r1i1p1_200209-201105.nc.met

METADATA EXTRACTION CONFIGURATION...

Choose a metadata extractor configuration from the list below: NetCDF-extractor

Click here to create an ingestion task for these files Save as Ingestion Task

FILE MANAGER INGESTION TASK LIST

TaskId	Created	Files	Policy	ProductType	MetExtractor	MetExt Config	Status	Action
53611bc2-3aaa-41b0-bd0d-78108e0e7e3e	2011-07-06T06:57:48-07:00	2	obs4MIPs	NetCDF	NetCDF-extractor	1	Not Started	Start

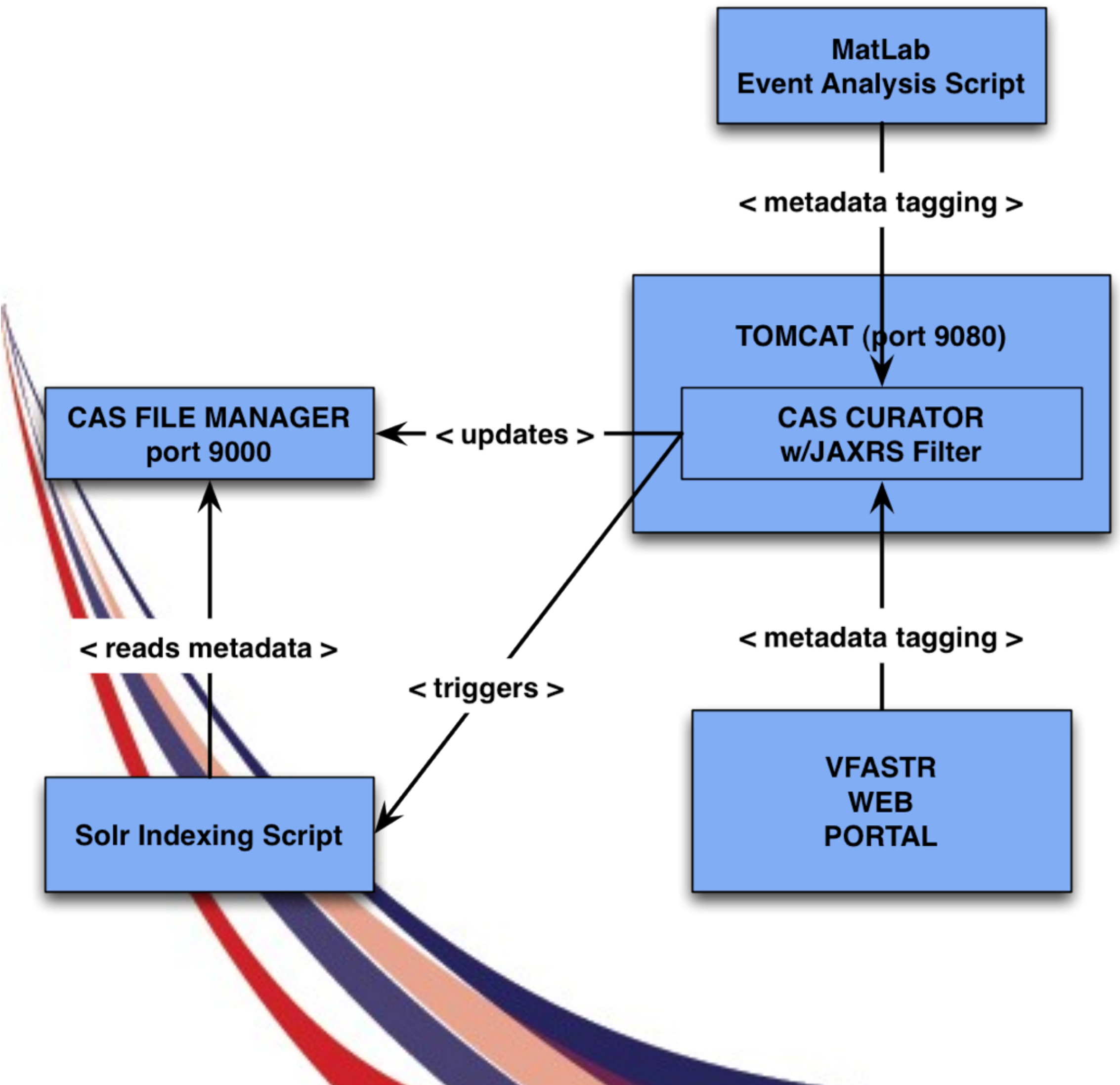
NASA Jet Propulsion Laboratory.

Features: provides two interfaces for interacting with the File Manager:

- Web User Interface: used by humans to manually interact with the system
  - Drag-and-drop selection of files from the staging area
  - Selection of metadata extractor, versioner from available pool
  - Submission of job for bulk ingestion to File Manager
  - Widget for display and update product metadata

CAS Curator

- Web RESTful API: used by programs and scripts for machine-machine interaction
  - Based on Apache JAX-RS project (project for RESTful web services)
  - Allows to annotate existing products with enhanced metadata
  - Example HTTP/POST invocation:
  - curl --data “id=<product_id>&metadata.<name>=<value>” http://<hostname>/curator/services/metadata/update



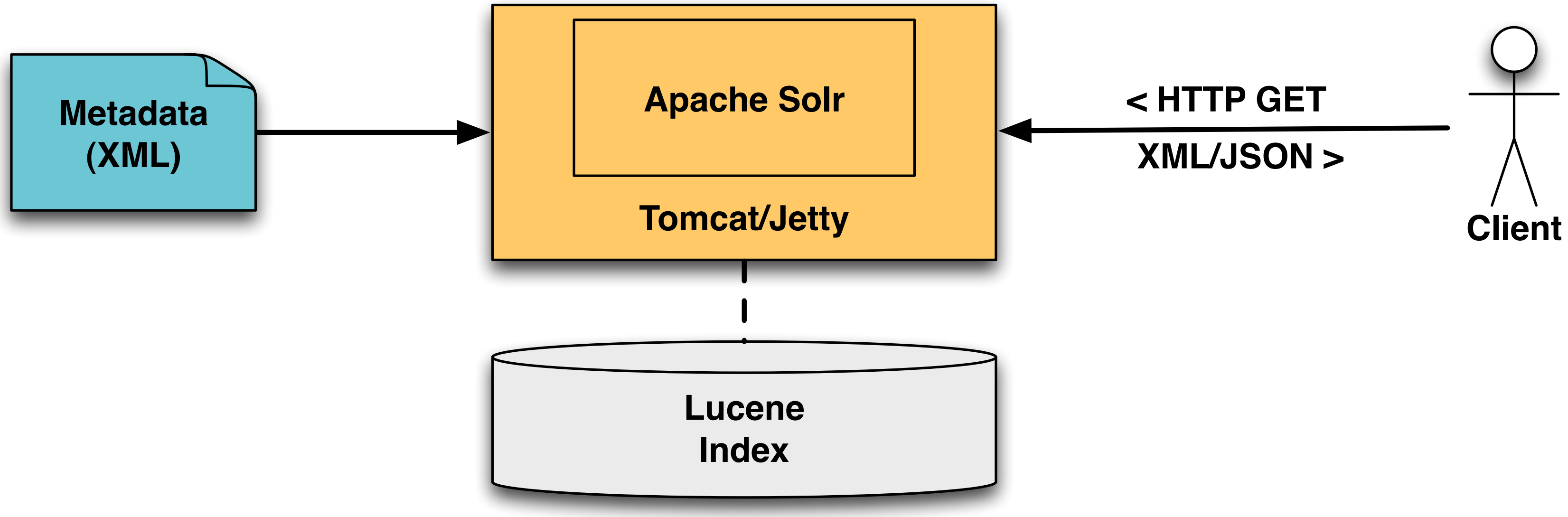
VFASTR deployment:

- REST API used by Web Portal and MatLab script to tag interesting events
- Updated product metadata submitted to FM via XML/RPC request
  - VFASTR Curator was wired with JAXRS ResponseHandler (servlet filter invoked before response is sent back to client) to invoke the script for updating the Solr metadata
- metadata.event_s0_e1="pulsar|machine|date"

Apache Solr

Solr is a high-performance web-enabled search engine built on top of Lucene

- Used in many e-commerce web sites
- Free text searches (w/ stemming, stop words, ...)
- Faceted searches (w/ facet counts)
- Other features: highlighting, word completion,...
- Flat metadata model: (key, value+) pairs



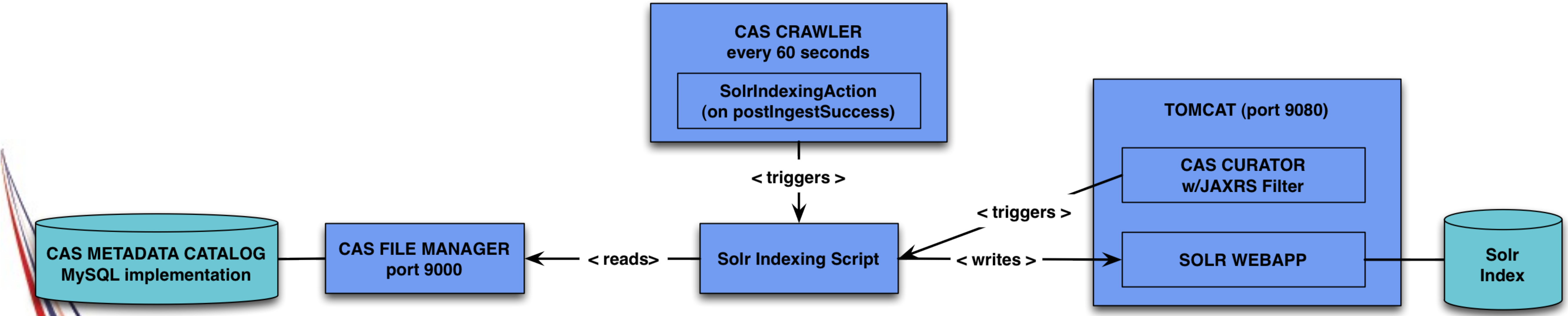
Scalability: Solr includes features to scale to 10-100 M of records:

- Multiple Cores to partition records into distinct Indexes
- Multiple Shards to distribute the query across complementary Indexes
- Replicated Indexes for high availability and low latency

Solr Indexing Script

VFASTR deployment: Solr is used to enable high performance metadata querying by clients: Web Portal and MatLab script

- Solr web application deployed within same Tomcat container as CAS Curator
- Python Indexing script harvests metadata from CAS Catalog to Solr Index
  - Triggered by CAS Crawler when a product is first ingested
  - Triggered by CAS Curator when the product metadata is updated
- VFASTR Solr schema specifies name and data type for all metadata fields
  - "Type=job/scan/event" field used to discriminate among records

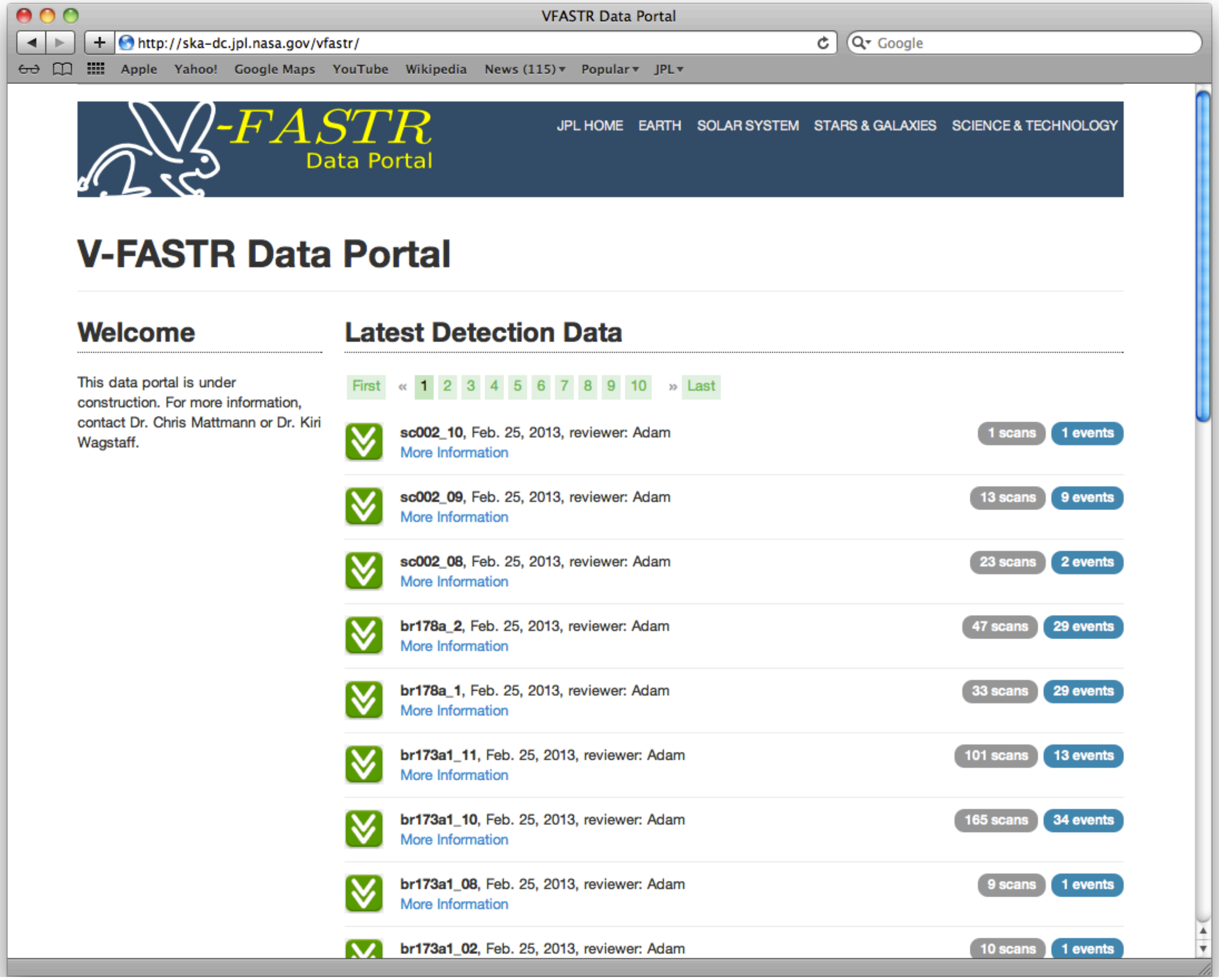


Examples of VFASTR queries to Solr Index:
- List of latest products by date
- Full metadata for a given job, scan or event
- All events that were assigned a given tag
- All tags assigned to all events

Review Data Portal

What is it?

- Web-based view of the metadata associated with nightly observations
- Collaborative environment for review by V-FASTR science team


VFASTR Data Portal
http://ska-dc.jpl.nasa.gov/vfastr/
Google
Apple Yahoo! Google Maps YouTube Wikipedia News (115) Popular JPL
V-FASTR Data Portal
JPL HOME EARTH SOLAR SYSTEM STARS & GALAXIES SCIENCE & TECHNOLOGY
First « 1 2 3 4 5 6 7 8 9 10 » Last

V-FASTR Data Portal

Welcome

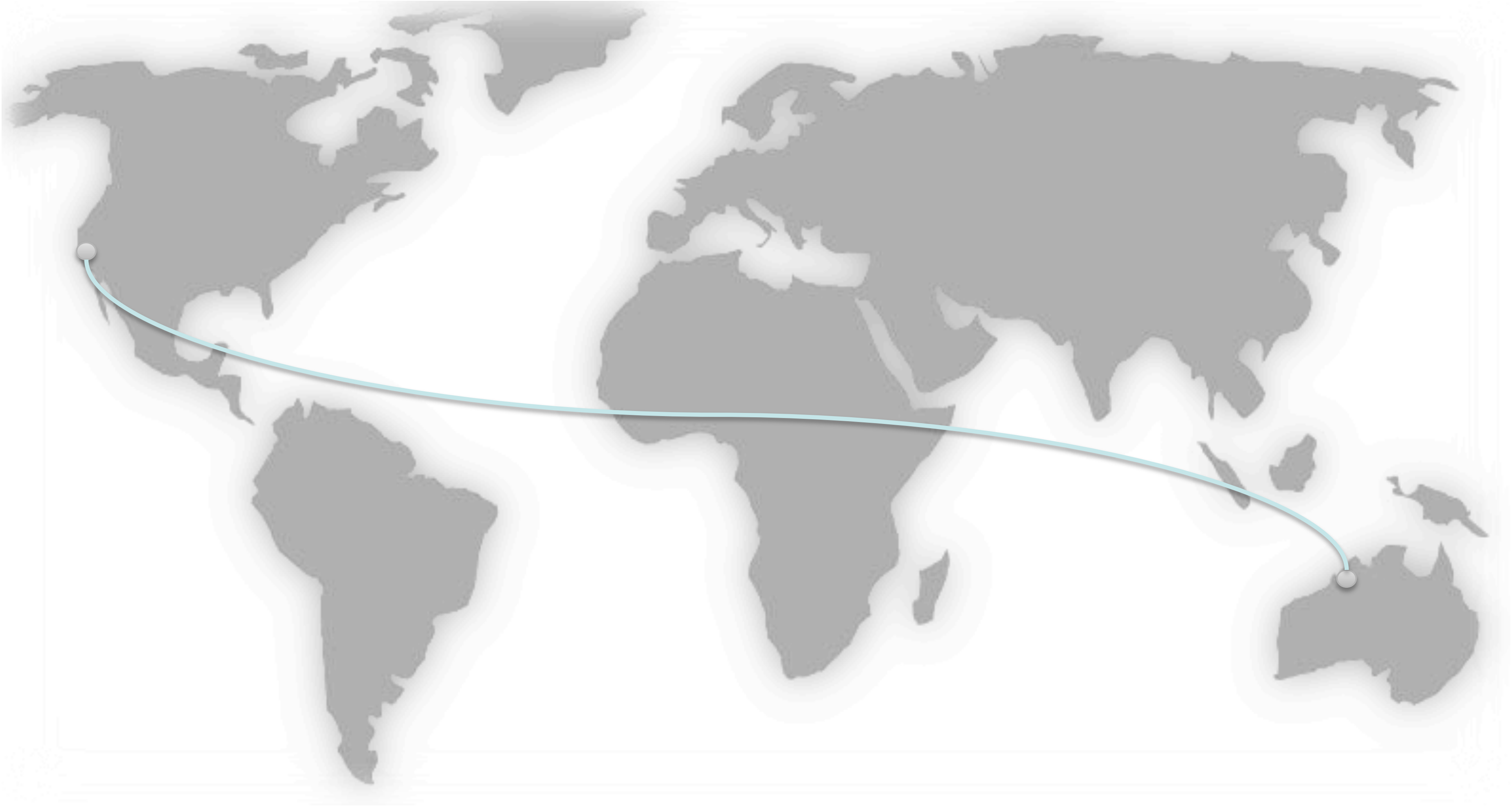
This data portal is under construction. For more information, contact Dr. Chris Mattmann or Dr. Kiri Wagstaff.

Latest Detection Data

- **sc002_10**, Feb. 25, 2013, reviewer: Adam — More Information — 1 scans, 1 events
- **sc002_09**, Feb. 25, 2013, reviewer: Adam — More Information — 13 scans, 9 events
- **sc002_08**, Feb. 25, 2013, reviewer: Adam — More Information — 23 scans, 2 events
- **br178a_2**, Feb. 25, 2013, reviewer: Adam — More Information — 47 scans, 29 events
- **br178a_1**, Feb. 25, 2013, reviewer: Adam — More Information — 33 scans, 29 events
- **br173a1_11**, Feb. 25, 2013, reviewer: Adam — More Information — 101 scans, 13 events
- **br173a1_10**, Feb. 25, 2013, reviewer: Adam — More Information — 165 scans, 34 events
- **br173a1_08**, Feb. 25, 2013, reviewer: Adam — More Information — 9 scans, 1 events
- **br173a1_02**, Feb. 25, 2013, reviewer: Adam — 10 scans, 1 events

Why does it exist?

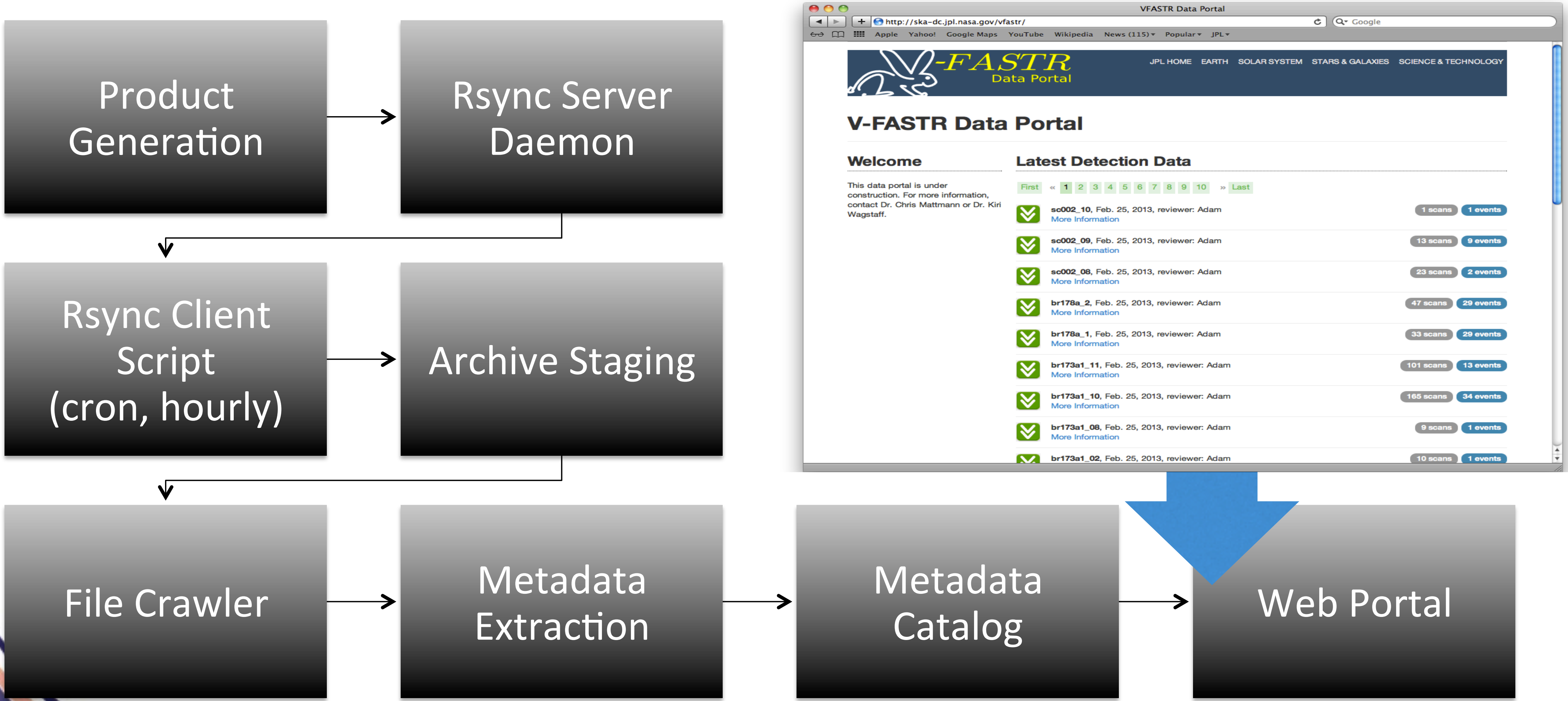
- Provide distributed science team with convenient access to metadata



Review Data Portal

How does it fit?

- Focal point for end-user access to the data pipeline



V-FASTR Data Portal

Welcome

This data portal is under construction. For more information, contact Dr. Chris Mattmann or Dr. Kiri Wagstaff.

Latest Detection Data

sc002_10, Feb. 25, 2013, reviewer: Adam — More Information — 1 scans — 1 events

sc002_09, Feb. 25, 2013, reviewer: Adam — More Information — 13 scans — 9 events

sc002_08, Feb. 25, 2013, reviewer: Adam — More Information — 23 scans — 2 events

br178a_2, Feb. 25, 2013, reviewer: Adam — More Information — 47 scans — 29 events

br178a_1, Feb. 25, 2013, reviewer: Adam — More Information — 33 scans — 29 events

br173a1_11, Feb. 25, 2013, reviewer: Adam — More Information — 101 scans — 13 events

br173a1_10, Feb. 25, 2013, reviewer: Adam — More Information — 165 scans — 34 events

br173a1_08, Feb. 25, 2013, reviewer: Adam — More Information — 9 scans — 1 events

br173a1_02, Feb. 25, 2013, reviewer: Adam — 10 scans — 1 events

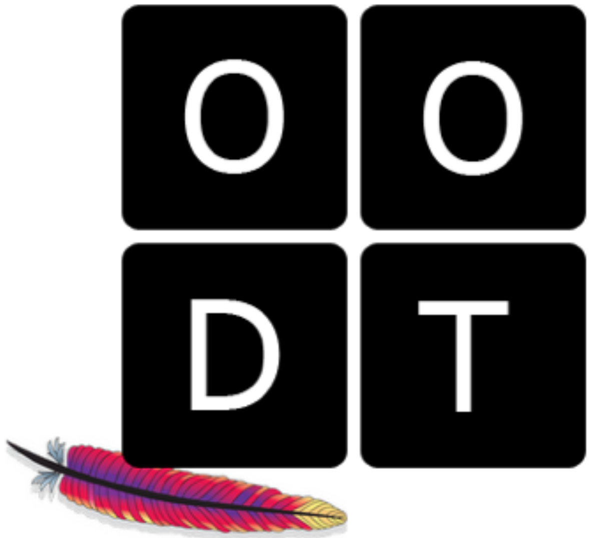
Review Data Portal

What is it built with?

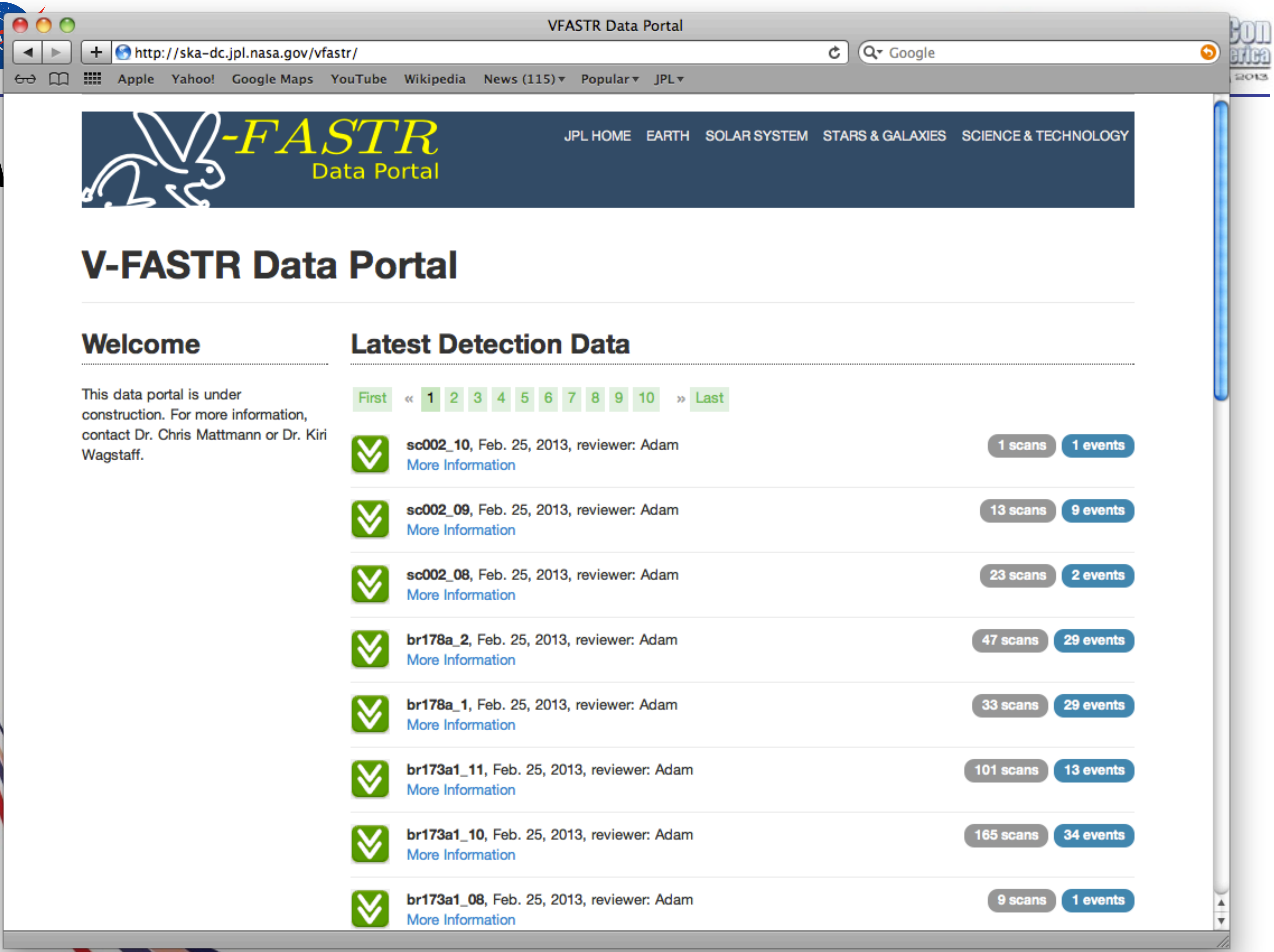
- Some of the technologies behind the data portal:

APACHE
HTTP SERVER

Apache
Solr



W



V-FASTR Data Portal

Welcome

This data portal is under construction. For more information, contact Dr. Chris Mattmann or Dr. Kiri Wagstaff.

Latest Detection Data

First « 1 2 3 4 5 6 7 8 9 10 » Last

sc002_10, Feb. 25, 2013, reviewer: Adam
More Information
1 scans 1 events

sc002_09, Feb. 25, 2013, reviewer: Adam
More Information
13 scans 9 events

sc002_08, Feb. 25, 2013, reviewer: Adam
More Information
23 scans 2 events

br178a_2, Feb. 25, 2013, reviewer: Adam
More Information
47 scans 29 events

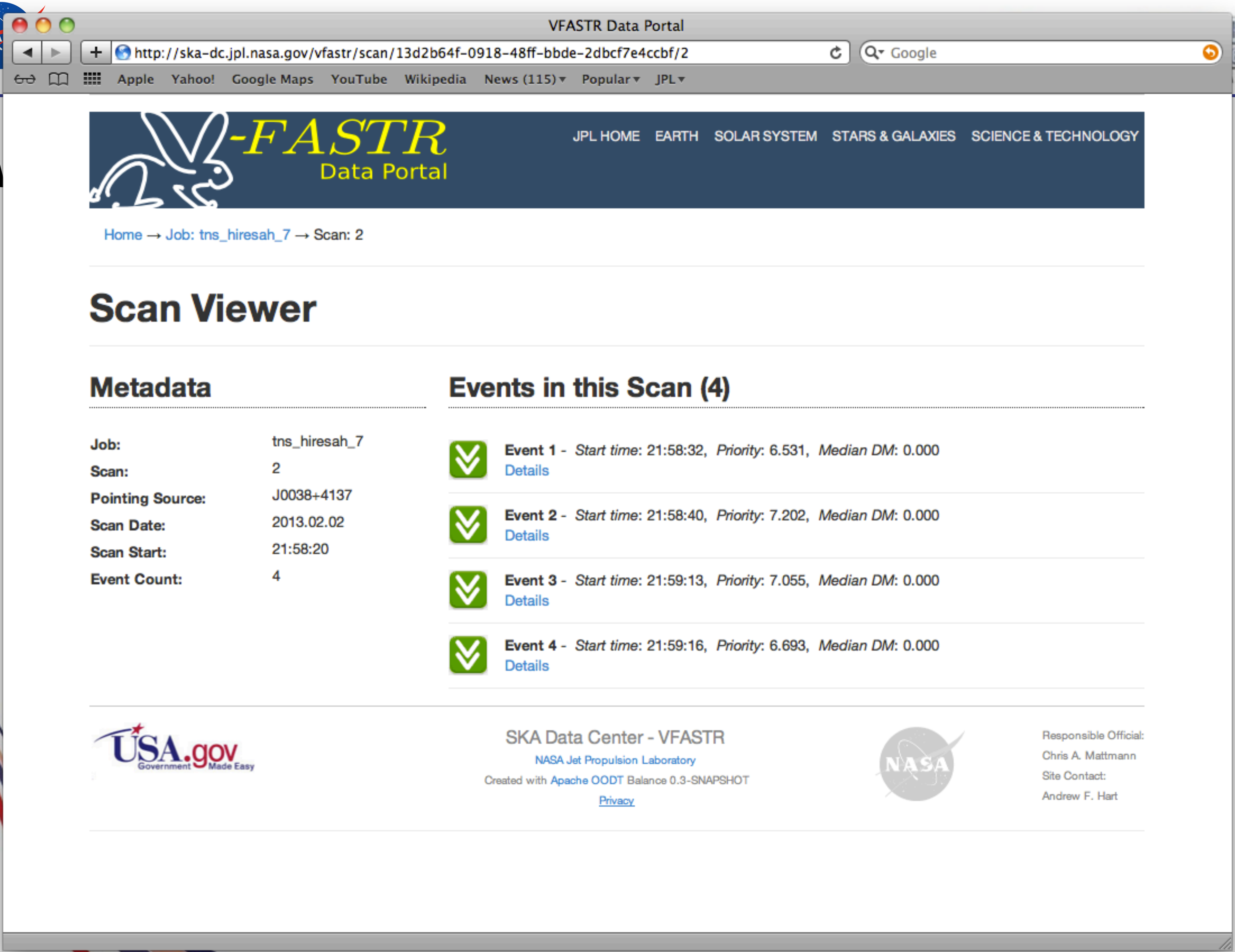
br178a_1, Feb. 25, 2013, reviewer: Adam
More Information
33 scans 29 events

br173a1_11, Feb. 25, 2013, reviewer: Adam
More Information
101 scans 13 events

br173a1_10, Feb. 25, 2013, reviewer: Adam
More Information
165 scans 34 events

br173a1_08, Feb. 25, 2013, reviewer: Adam
More Information
9 scans 1 events

W



Home → Job: tns_hiresah_7 → Scan: 2

Scan Viewer

Metadata

Job:	tns_hiresah_7
Scan:	2
Pointing Source:	J0038+4137
Scan Date:	2013.02.02
Scan Start:	21:58:20
Event Count:	4

Events in this Scan (4)

Event 1 - *Start time*: 21:58:32, *Priority*: 6.531, *Median DM*: 0.000
Details

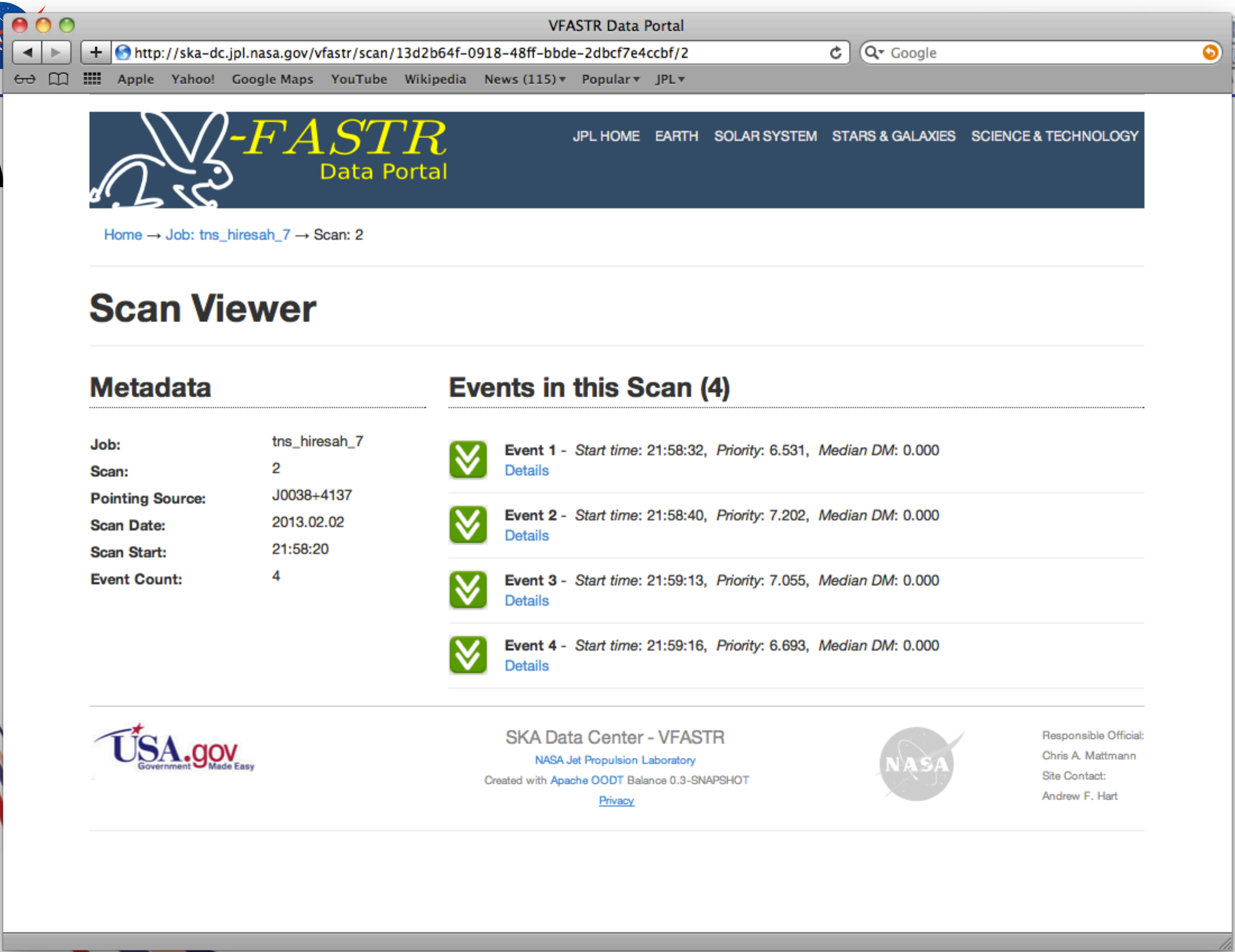
Event 2 - *Start time*: 21:58:40, *Priority*: 7.202, *Median DM*: 0.000
Details

Event 3 - *Start time*: 21:59:13, *Priority*: 7.055, *Median DM*: 0.000
Details

Event 4 - *Start time*: 21:59:16, *Priority*: 6.693, *Median DM*: 0.000
Details

SKA Data Center - VFASTR
NASA Jet Propulsion Laboratory
Created with Apache OODT Balance 0.3-SNAPSHOT
Privacy

Responsible Official:
Chris A. Mattmann
Site Contact:
Andrew F. Hart



Home → Job: tns_hiresah_7 → Scan: 2

Scan Viewer

Metadata

Job:	tns_hiresah_7
Scan:	2
Pointing Source:	J0038+4137
Scan Date:	2013.02.02
Scan Start:	21:58:20
Event Count:	4

Events in this Scan (4)

✓ **Event 1** - *Start time*: 21:58:32, *Priority*: 6.531, *Median DM*: 0.000
Details

✓ **Event 2** - *Start time*: 21:58:40, *Priority*: 7.202, *Median DM*: 0.000
Details

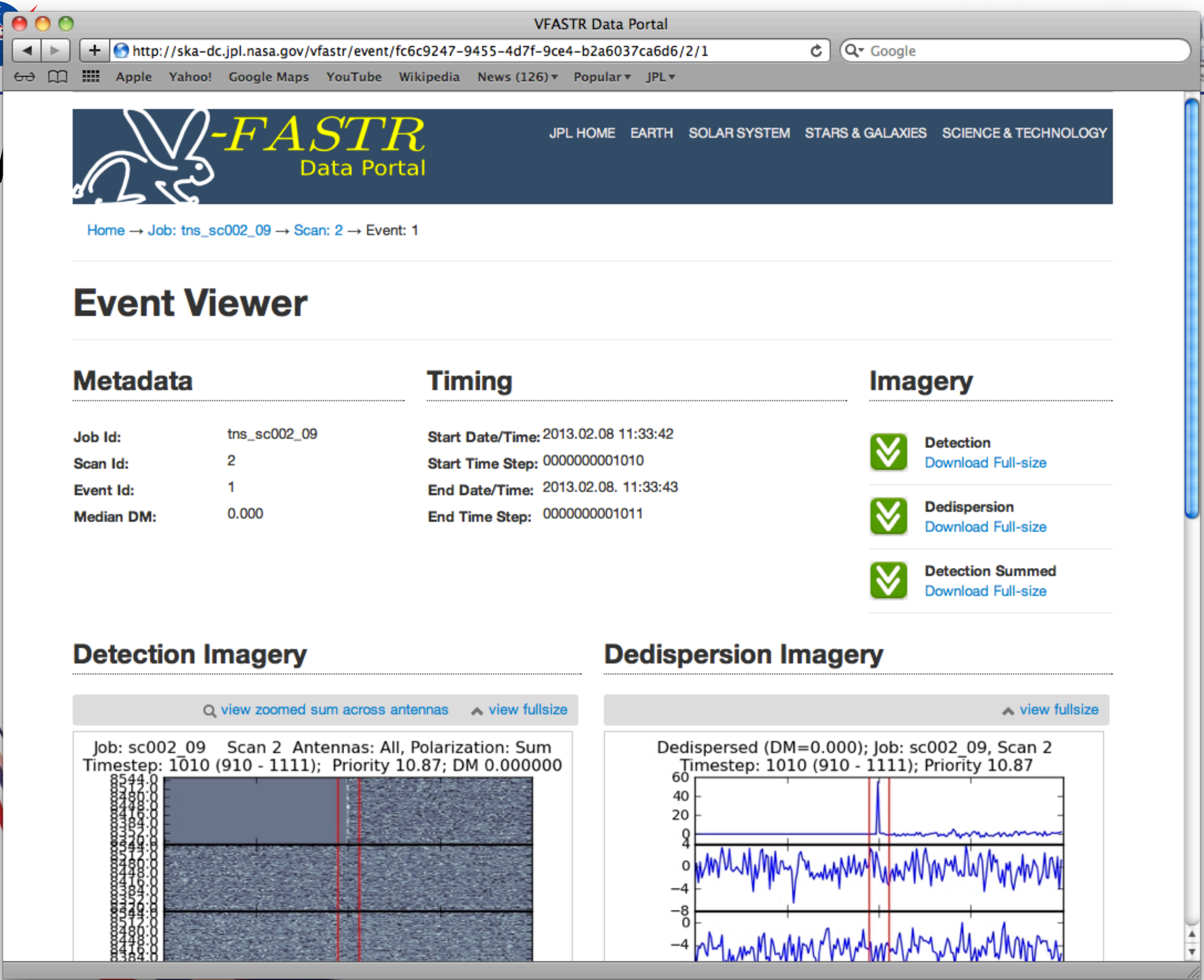
✓ **Event 3** - *Start time*: 21:59:13, *Priority*: 7.055, *Median DM*: 0.000
Details

✓ **Event 4** - *Start time*: 21:59:16, *Priority*: 6.693, *Median DM*: 0.000
Details

SKA Data Center - VFASTR
NASA Jet Propulsion Laboratory
Created with Apache OODT Balance 0.3-SNAPSHOT
Privacy

Responsible Official:
Chris A. Mattmann
Site Contact:
Andrew F. Hart

W



Home → Job: tns_sc002_09 → Scan: 2 → Event: 1

Event Viewer

Metadata

Job Id:	tns_sc002_09
Scan Id:	2
Event Id:	1
Median DM:	0.000

Timing

Start Date/Time:	2013.02.08 11:33:42
Start Time Step:	0000000001010
End Date/Time:	2013.02.08. 11:33:43
End Time Step:	0000000001011

Imagery

✓ **Detection**
Download Full-size

✓ **Dedispersion**
Download Full-size

✓ **Detection Summed**
Download Full-size

Detection Imagery

view zoomed sum across antennas
view fullsize

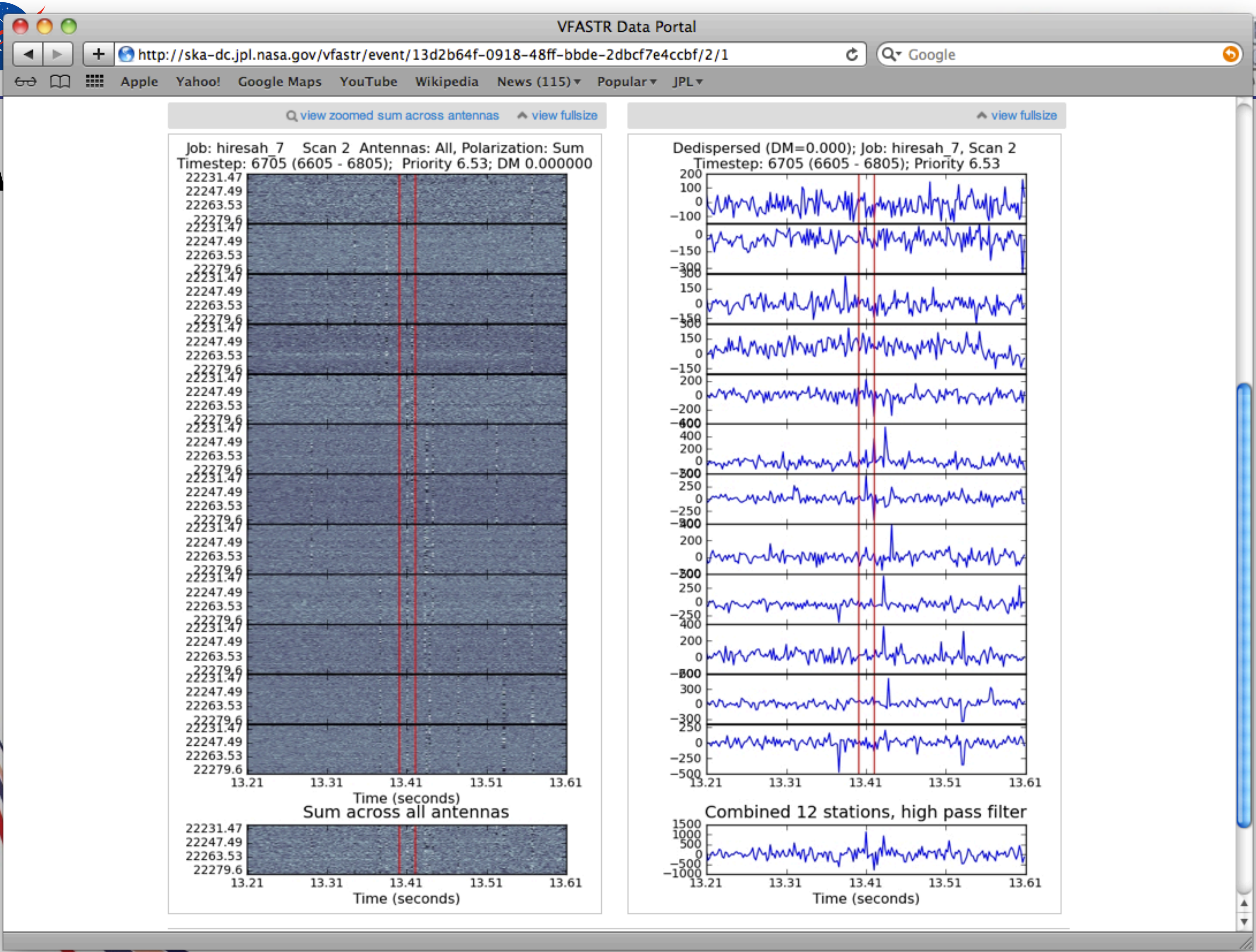
Job: sc002_09 Scan 2 Antennas: All, Polarization: Sum
Timestep: 1010 (910 - 1111); Priority 10.87; DM 0.000000

Dedispersion Imagery

view fullsize

Dedispersed (DM=0.000); Job: sc002_09, Scan 2
Timestep: 1010 (910 - 1111); Priority 10.87

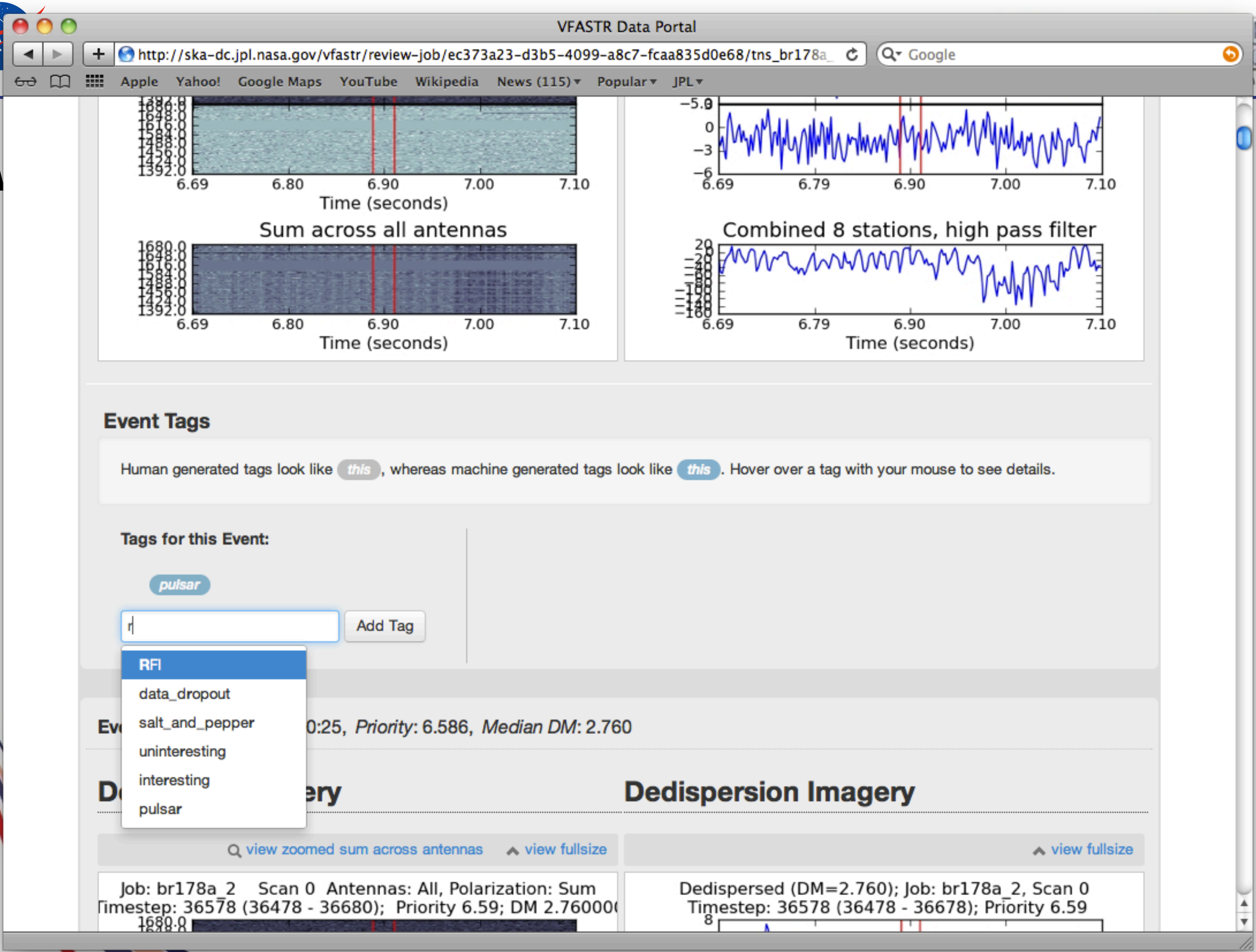
W



VFASTR Data Portal
http://ska-dc.jpl.nasa.gov/vfastr/event/13d2b64f-0918-48ff-bbde-2dbcf7e4ccbf/2/1
Google
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (115)
Popular
JPL
view zoomed sum across antennas
view fullsize
view fullsize
Job: hiresah_7 Scan 2 Antennas: All, Polarization: Sum
Timestep: 6705 (6605 - 6805); Priority 6.53; DM 0.000000
22231.47
22247.49
22263.53
22279.6
13.21
13.31
13.41
13.51
13.61
Time (seconds)
Sum across all antennas
Dedispersed (DM=0.000); Job: hiresah_7, Scan 2
Timestep: 6705 (6605 - 6805); Priority 6.53
Combined 12 stations, high pass filter
Time (seconds)

What are tags?

- Descriptive metadata associated with an event
- Enable classification and filtering
- Serve as training for AI (now)
- Serve as guide for what to archive (soon)



Event Tags

Human generated tags look like *this*, whereas machine generated tags look like *this*. Hover over a tag with your mouse to see details.

Tags for this Event:

pulsar

Ev... 0:25, *Priority*: 6.586, *Median DM*: 2.760

D...ery

Dedispersion Imagery

Job: br178a_2 Scan 0 Antennas: All, Polarization: Sum
Timestep: 36578 (36478 - 36680); Priority 6.59; DM 2.76000(

Dedispersed (DM=2.760); Job: br178a_2, Scan 0
Timestep: 36578 (36478 - 36678); Priority 6.59

What can users do?

- Event tagging (now)
    - Nightly classification, viewable by all users
    - Used as training input to the automatic candidate detection
- Job archiving (soon)
    - Initiate archival of job on NRAO hardware based upon the contents of the tags and/or other metrics

What have users thought?

- Interface with all event imagery on-screen is an improvement over command-line methods (more efficient)
- Organization of the interface should support rapid evaluation of an entire job (minimize clicks)
- Improved accessibility of the information a big plus (mobile access)

What is next?

- Provide more comprehensive job/event search capability
  - Facet by tag values, index metadata
- Continued efficiency improvements
  - Bulk tagging of all events in a job
- Implement front-end capability to initiate back-end archive process

Questions ?

Acknowledgments

Walter F. Brisken
Sarah Burke-Spolar
Catheryn Trott
Adam T. Deller
Dayton L. Jones
Walid A. Majid
Divya Palaniswamy
Steven J. Tingay
Randall B. Wayth