



# NPP Science Data Segment (SDS) Data Depository and Distribution Element (SD3E) Overview

February 28, 2008  
NASA/Langley  
Evelyn.L.Ho@nasa.gov



# Agenda

- Overview
- Interfaces
- Architecture
- Requirements
- Contingency Plans
- Schedule
- Issues/Concerns

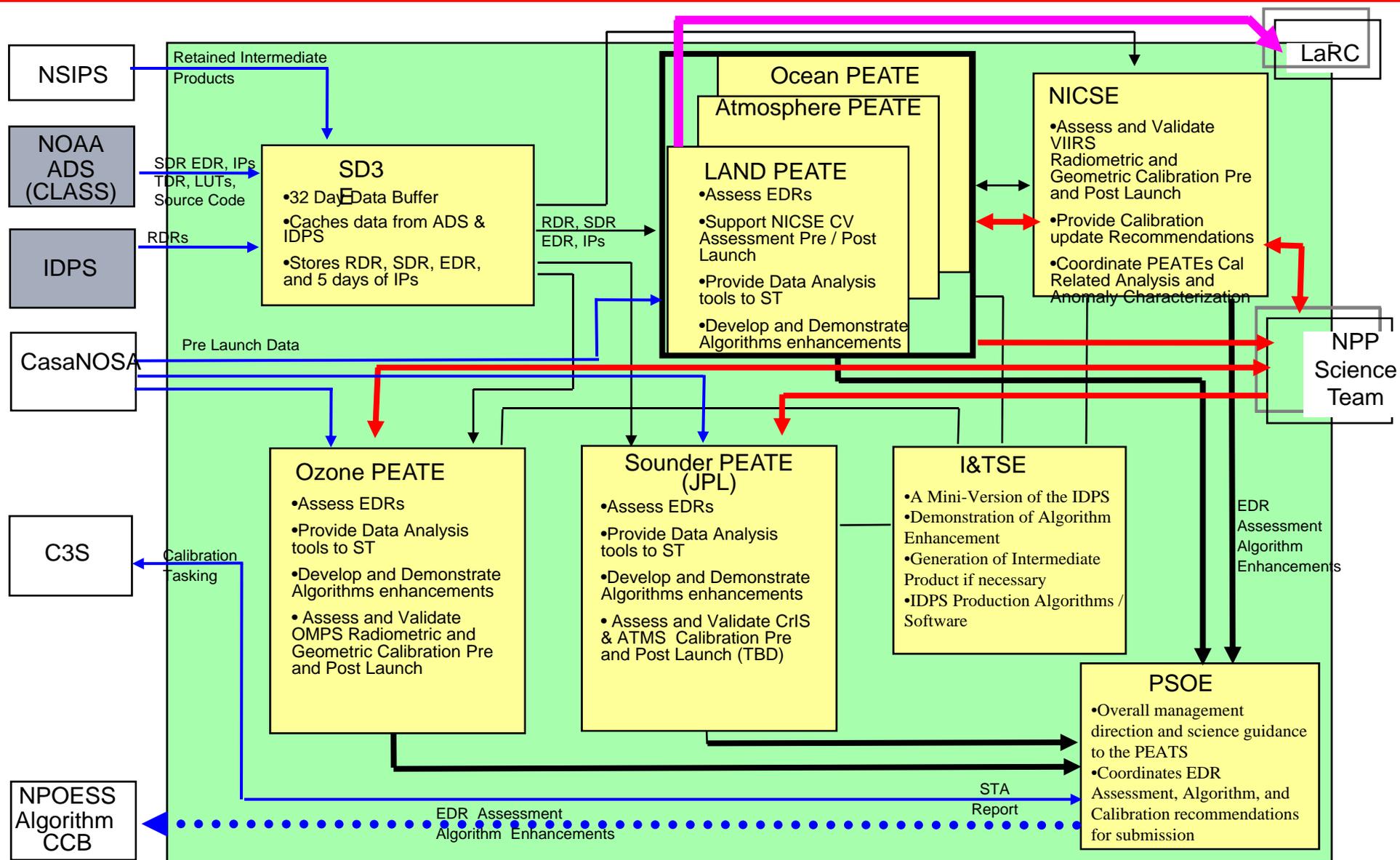


# SD3E Overview

- Primary SD3E Tasks:
  - Central system dedicated to
    - Request and acquire data from major data providers
    - Store data temporarily up to 32 days of NPP xDR data and 7 days of retained IPs
    - Provide data to users
- No data production capabilities
  - No reformatting
  - No aggregation
  - No subsetting



# SDS Block Diagram



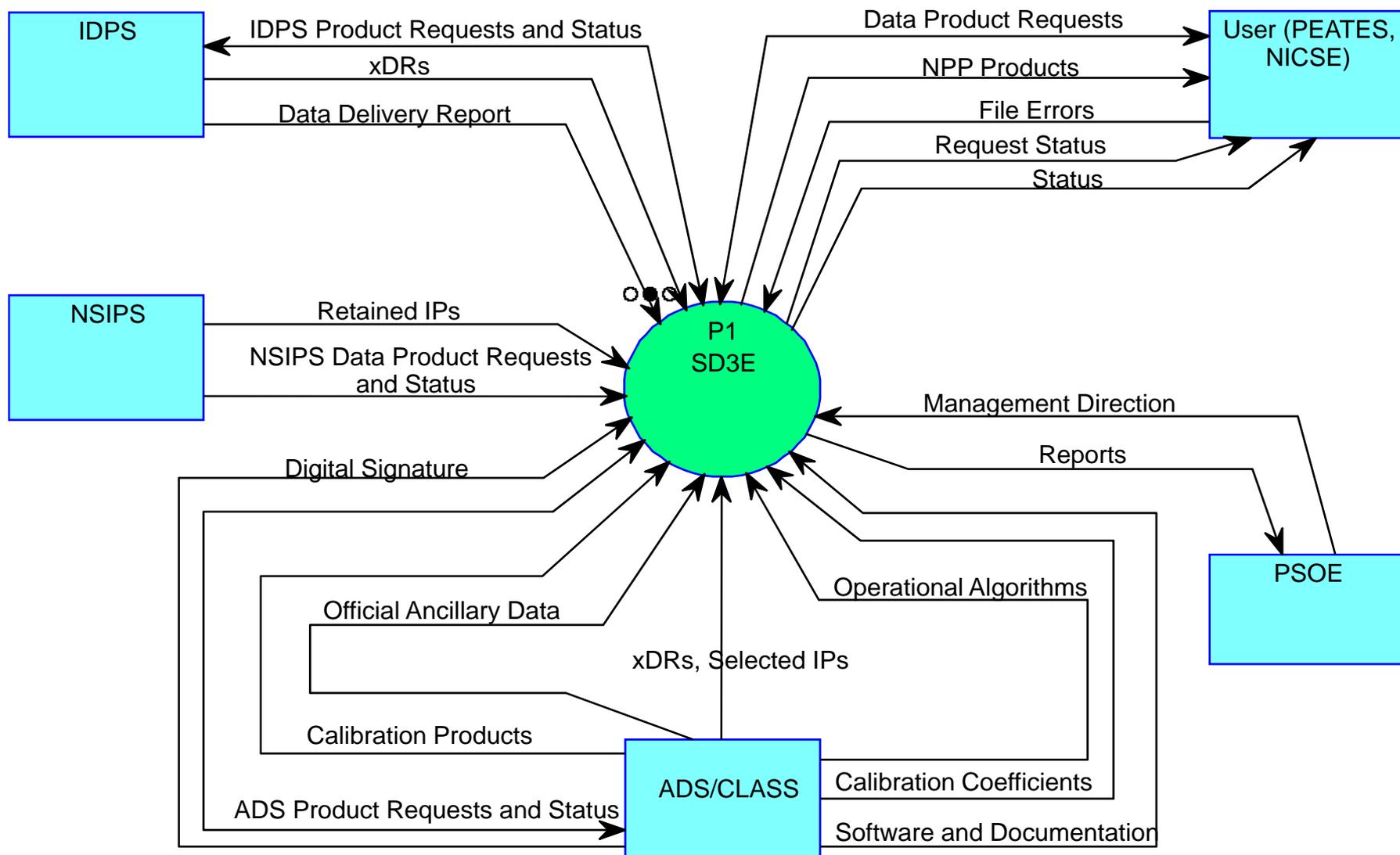


# SDS Elements

- SD3E – SDS Data Delivery & Depository Element
  - Acquires data from NOAA/CLASS and NESDIS/IDPS (Now to handle CERES RDRs)
  - Provides ~32 days “rolling storage” for pick-up by PEATEs and the NICSE
- PEATEs – 5 Product Evaluation and Algorithm Test Elements
  - Performs EDR Evaluation / Characterization and algorithm improvements coordinated with External Science Investigators
  - Ocean (VIIRS) leveraged off Ocean Data Processing System (ODPS)
  - Land (VIIRS) leveraged off MODIS Adaptive Processing System (MODAPS)
    - Re-sample and forward selected VIIRS SDRs and EDRs. Also Forward CERES RDRs.
  - Atmosphere (VIIRS) leveraged off University of Wisconsin-Madison’s LEOCAT
  - Ozone (OMPS) leveraged off OMI Data Processing System (OMIDAPS)
  - Sounder (CrIS & ATMS) leveraged off JPL Atmospheric Infrared Sounder (AIRS) Project’s Team Leader Science Computing Facility
- NICSE - NPP Instrument Calibration Support Element
  - VIIRS radiometric and geometric calibration and Instrument characterization support
- I&TSE – Integration and Test System Element
  - Mini-IDPS for demonstrating algorithm enhancements and / or calibration improvements
- PSOE - Project Science Office Element
  - Submits algorithm and calibration recommendations to NPP/NPOESS Algorithm CCB



# Context Diagram (1 of 2)



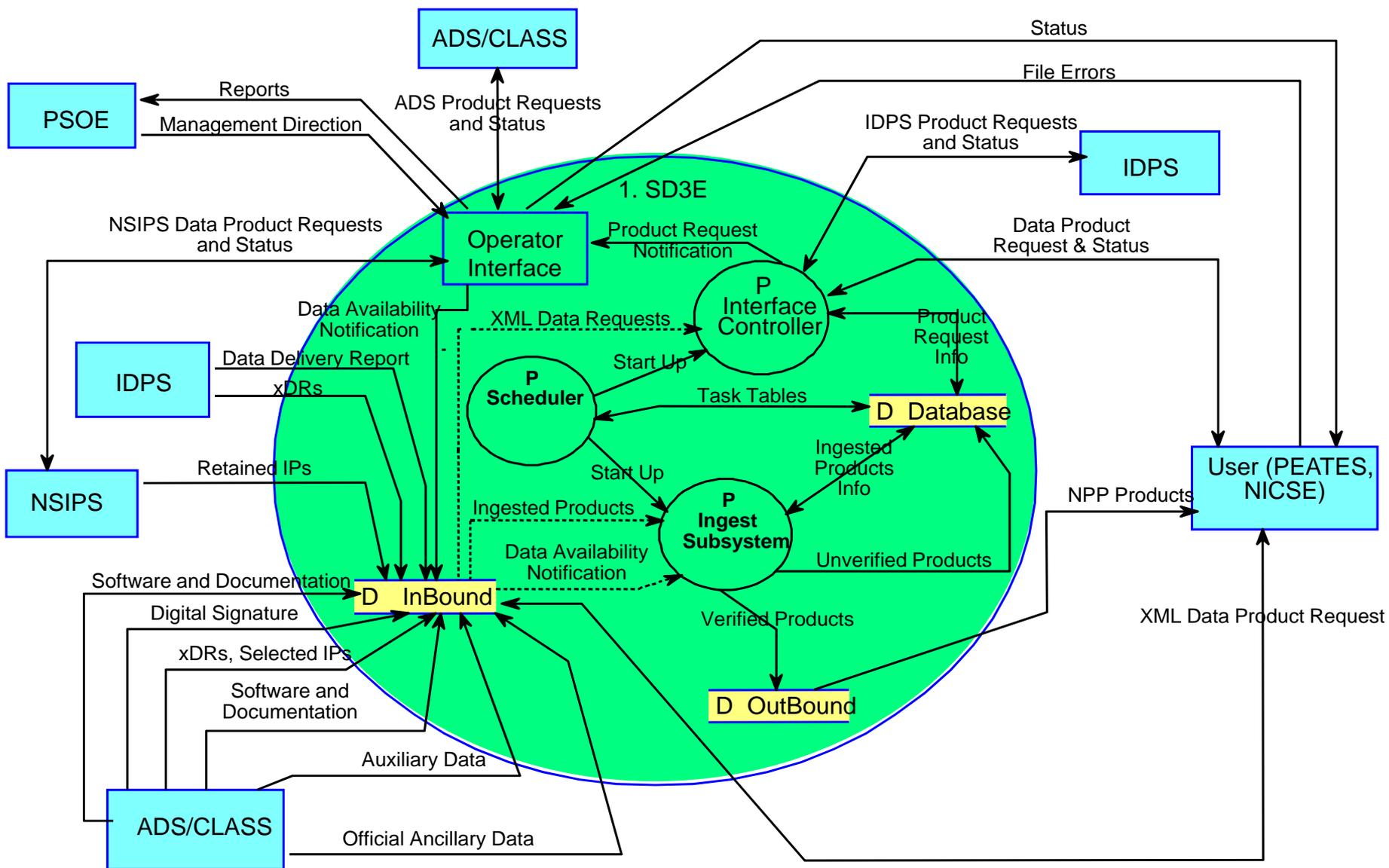


# Context Diagram (2 of 2)

- IDPS
  - Primary source for RDRs
- NSIPS
  - Primary source for retained intermediate products (RIPs)
- ADS/CLASS
  - Primary source for archived data (SDRs, EDRs, TDRs, DIPs, calibration products/coefficients, official ancillary/auxiliary, operational IDPS source code and documentation)
- Product Evaluation and Analysis Tool Elements (PEATEs)
  - Primary recipients of RDRs, SDRs, EDRs, ARP, TDRs, IPs, Official Ancillary/Auxiliary data, Operational algorithms and source code, and calibration products
- NPP Instrument Calibration Support Element (NICSE)
  - Primary recipients of VIIRS-related xDRs and calibration products for independent cal/val support
- Project Science Office Element (PSOE)
  - Primary source for data acquisition direction for subscription orders, etc.
  - Data ingest volume reports, data transfer reports



# SD3E S/W Architecture (1 of 2)





# SD3E S/W Architecture (2 of 2)

- Scheduler
  - Directs/Controls or schedules program execution
- Interface Controller
  - Accepts and checks PEATEs'/NICSE's request for data
- Ingest Subsystem
  - Ingests and verifies NPP products
- Inbound Data Store
  - Provides location for data providers to push data products
  - Stages data temporarily
- Outbound Data Store
  - Organizes and stores data products into logical hierarchies
- Database
  - Provides mechanism for communication between tasks and data accounting
- Operator Interface
  - Submits subscription/ad-hoc request to the ADS via a public web interface
  - Submits subscription/ad-hoc requests to NSIPS via the NSIPS web interface
  - Resolves data, access, or system issues
  - Resolves file error issues indicated by PEATEs



# SD3E to PEATE Interface (1 of 2)



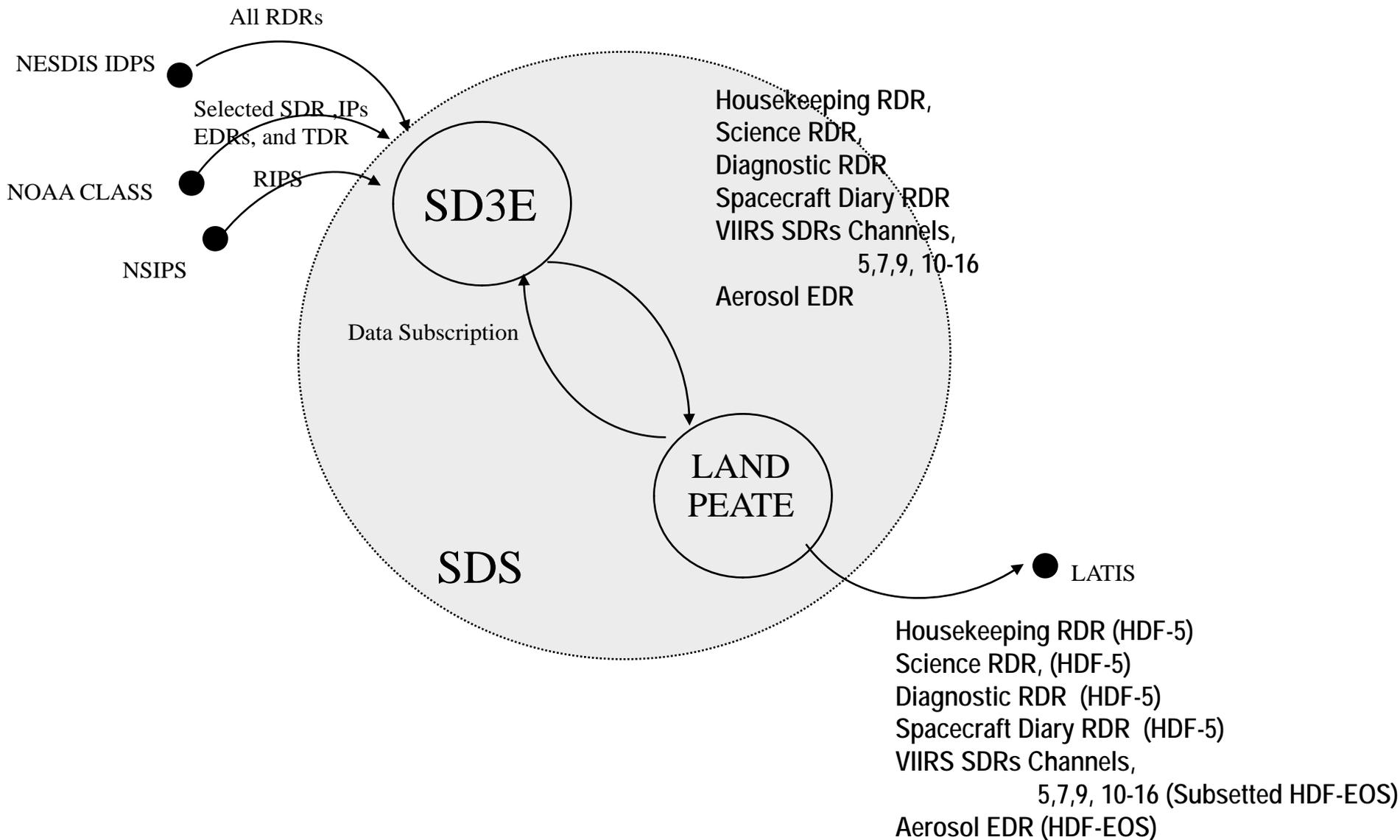


# SD3E to PEATE Interface (2 of 2)

- Web Interface
  - Subscription/ad-hoc requests
  - Provides request status, system status
- Machine-to-Machine Interface
  - Anonymous FTP with IP restriction
  - Subscription/ad-hoc requests using XML request
  - Data requests FTPed to SD3E inbound location
- Operator
  - Monitor ingest and data requests
  - Resolve possible data, access, or system issues
  - Resolve file error issues indicated by the PEATEs
- Mantis
  - Bug Tracker/Problem reporting tool
- Storage
  - Anonymous FTP pull by PEATEs
- ***NPOESS Preparatory Project (NPP) Science Data Segment (SDS) Data Depository and Distribution Element (SD3E) to Product Evaluation and Analysis Tool Element (PEATEs) and NPP Instrument Calibration Support Element (NICSE) Interface Control Document (ICD)***



# SD3E Data Flow to LaRC





# Subscription/Ad-hoc Requests

- Data requests
  - Subscriptions
    - Through Land PEATE
    - VIIRs Aggregation -- 5 minutes
    - VIIRs Granule size -- 85.74 seconds
  - Ad-hoc
    - Re-request of products that are corrupted, lost, or not part of the subscription
      - Aggregation that is different from VIIRs' subscription
    - Retransmission requests
    - xDR products can be re-requested if within 32 day window. IP products can be re-requested if within 7 day window
    - xDR products outside of 32 day window must be requested by the PEATEs from ADS/CLASS. IP products are no longer available after 7 days
    - No cancellations or modifications to request once submitted
  - Products
    - CERES RDRs
- Data Integrity files
  - Products from IDPS
    - RDRs will have a SD3E generated integrity file with the IDPS checksum included
  - Products from ADS/CLASS
    - xDRs, ancillary/auxiliary, calibration products, the SD3E will send the digital signature file to the PEATEs



# Inbound Data Store

- FTP push of products from external data providers for subscription/ad-hoc requests
- Stages Data Temporarily
  - 0.017 to 0.020 TB/5 minutes (5 - 6 TB/day)
  - 1.053 TB for 6th hour spike (ADS data is 6 hours old)
- Anonymous FTP push of data requests from PEATEs/NICSE
  - IP restriction



# Outbound Data Store

- Organizes and stores data products into logical hierarchies
  - DailyIngest directory
    - Date ingested
  - NPP\_Products directory
    - Date of data
  - NPP\_Closed (closed data day) directory
    - Date of data
- Provides 32-day storage for all except Intermediate Products (7 days)



# Outbound - Daily Ingest

- DailyIngest directory
  - Provide directory of requested products by PEATE
  - Directory name uses calendar date (date SD3E ingests product) — Eastern local time
  - Subdirectory provides a listing of all files ingested
  - A data integrity file is provided with each product
    - Checksum file includes:
      - filename, checksum, file size and timestamp
    - Digital Signature File:
      - SD3E will pass the same digital signature file it receives from ADS/CLASS to the PEATEs to use for verification
  - PEATEs/NICSE must tell us what products they would like for us to report under the daily ingest directory. By default, the directory will include products requested (by subscription or ad-hoc request) by the particular PEATE



# Outbound - Daily Ingest

/pub/DailyIngest

/YYYYMMDD

/Atmosphere

/RDR

/SDR

....

/Land

/NICSE

/Ocean

/Ozone

/Sounder

... /YYYYMMDD



# Outbound - NPP\_Products

## NPP Products directory structure

- Provide 32-day directory structure, except for packaged deliveries
- Directory name will use data date (date data acquired by instrument) except for packaged deliveries
- Subdirectories for xDRs and IPs will use product type and next lower level will use data product ID
  - ARPs will reside in the EDR directory
- Subdirectories for ancillary data products will use the collection short name. /Ancillary contains DQM products
- Package directory for bundled files
  - SD3E will not unbundle (i.e., extract the various products and place in their corresponding directories)
  - Software, Documentation, Support Data, and Test Data as package bundles
- Calibration products are part of the Auxiliary directory (based on CDFCB-X Vol I)
- Other directories under sensor directory is for aggregated products (product packaging on) or other requested files that do not fall into a specific category
- A checksum file or digital signature file is provided with each product
  - Checksum file for products from IDPS and NSIPS includes:  
filename, checksum, file size, timestamp
  - Digital Signature File for products from ADS/CLASS:
    - SD3E will pass the same digital signature file it receives from ADS/CLASS to the PEATEs to use for verification



# Outbound - NPP\_Products

```

/pub/NPP_Products
  /Ancillary
    /YYYYMMDD
      /CollectionShortName ... /CollectionShortName
    ... /YYYYMMDD
  /Auxiliary
    /YYYYMMDD
      /CollectionShortName ... /CollectionShortName
    ... /YYYYMMDD
  /Package
    /Software /Documentation /Support-Data /Test-Data
      /Version0 ... /Version5
  /ATMS /CrIS /CrIMSS /OMPS /VIIRS
    /YYYYMMDD
      /Others
      /RDR
        /DataProductID ... /DataProductID
      /SDR
        /DataProductID ... /DataProductID
      /EDR
        /DataProductID ... /DataProductID
      /TDR
        /DataProductID ... /DataProductID
      /IP
        /DataProductID ... /DataProductID
      /GEO
        /DataProductID ... /DataProductID

```



# Outbound - NPP\_Closed

- Closed directory
  - Provide a 32-day directory structure
  - Used to check if data product ID level of products is “closed”
  - Users will look in NPP\_Product directory for products
  - Directory name will use data date (date data acquired by instrument)
  - Subdirectory level
    - Lower level directory structure matches “NPP\_Products” data product directory structure



# Outbound - Closed Day

```
/pub/NPP_Closed
  /Ancillary
    /YYYYMMDD
      /CollectionShortName ... /CollectionShortNam
    ... /YYYYMMDD
  /Auxiliary
    /YYYYMMDD
      /CollectionShortName ... /CollectionShortName
    .... /YYYYMMDD
  /ATMS /CrIS /CrIMSS /OMPS /VIIRS
    /YYYYMMDD
      /RDR
        /DataProductID ... /DataProductID
      /SDR
        /DataProductID ... /DataProductID
      /EDR
        /DataProductID ... /DataProductID
      /TDR
        /DataProductID ... /DataProductID
      /IP
        /DataProductID ... /DataProductID
      /GEO
        /DataProductID ... /DataProductID
    ... /YYYYMMDD
```



# Catch-up Mechanism

- SD3E down < 3 hours
  - ADS and IDPS will retransmit products
  - NSIPS is TBD
  - Approximately 24 hours to acquire and ingest missing 0.5 TB
- SD3E down > 3 hours but < 24 hours
  - Acquire missing RDRs (180 GB if 24 hours). Will take approx. 8 hours to catch-up
  - Acquire missing retained IPs (1.3 TB if 24 hours). Will take 3 days to catch-up.
  - Will lose ~0.5 TB for every 3 hours of down time
  - Will lose ~4.5 TB of data if down for 24 hours
  - Within 24 hours, RDRs requested from IDPS
  - Greater than 24 hours, RDRs requested from ADS/CLASS
  - PEATEs/NICSE will need to request missing products (e.g. SDRs, EDRs, TDRs, ancillary, etc.) from ADS/CLASS
- SD3E down > 24 hours
  - Will obtain full set of RDRs from ADS/CLASS
  - PEATEs/NICSE will need to request missing products (e.g. xDRs, ancillary, etc.) from ADS/CLASS



# New SD3E Requirement

- The SD3E shall be capable of requesting and accepting CERES RDRs generated by the IDPS.



# CERES Data Volume Land PEATE to LaRC

| Product              | Files    | GB   | Comment  |
|----------------------|----------|------|--|
| Subsetted SDR        | 288/day  | 16.3 | SDR 750m bands, every 3 <sup>rd</sup> pixel by every 3 <sup>rd</sup> scan              |
| Geolocation in SDR   | 288/day  | 21.8 | ellipsoidal geolocation, every 3 <sup>rd</sup> pixel by every 3 <sup>rd</sup> scan     |
| Aerosol EDR          | 160/day* | 50.2 | Aerosol EDR, 1.6km at end of swath Ocean, 9.6 km x 9.6km grid Land                     |
| CERES RDR            | 6/day    | .2   | Per ESDIS S/E (includes aft and fore) based on Terra CERES                             |
| Spacecraft telemetry | N/A      | 0.0  | Included in RDR  |
| VIIRS DDR            | 1/week   | 3.5  | Snow/Ice cover Map gridded Level 3 8-day   |
| Total                | 748      | 92GB | Peak volume and files on the day that the weekly file is shipped, average vol. is 89GB |

\* Day granules 55% of orbit includes day-only and mixed granules



# Schedule

- Build 3 delivery to I&T 07/2008
- Build 3 I&T complete 09/2008
- Build 3 delivery 09/2008
  
- Build 4 expected start 8/2008
- Build 4 I&T complete 1/2009
- Build 4 expected delivery 3/2009



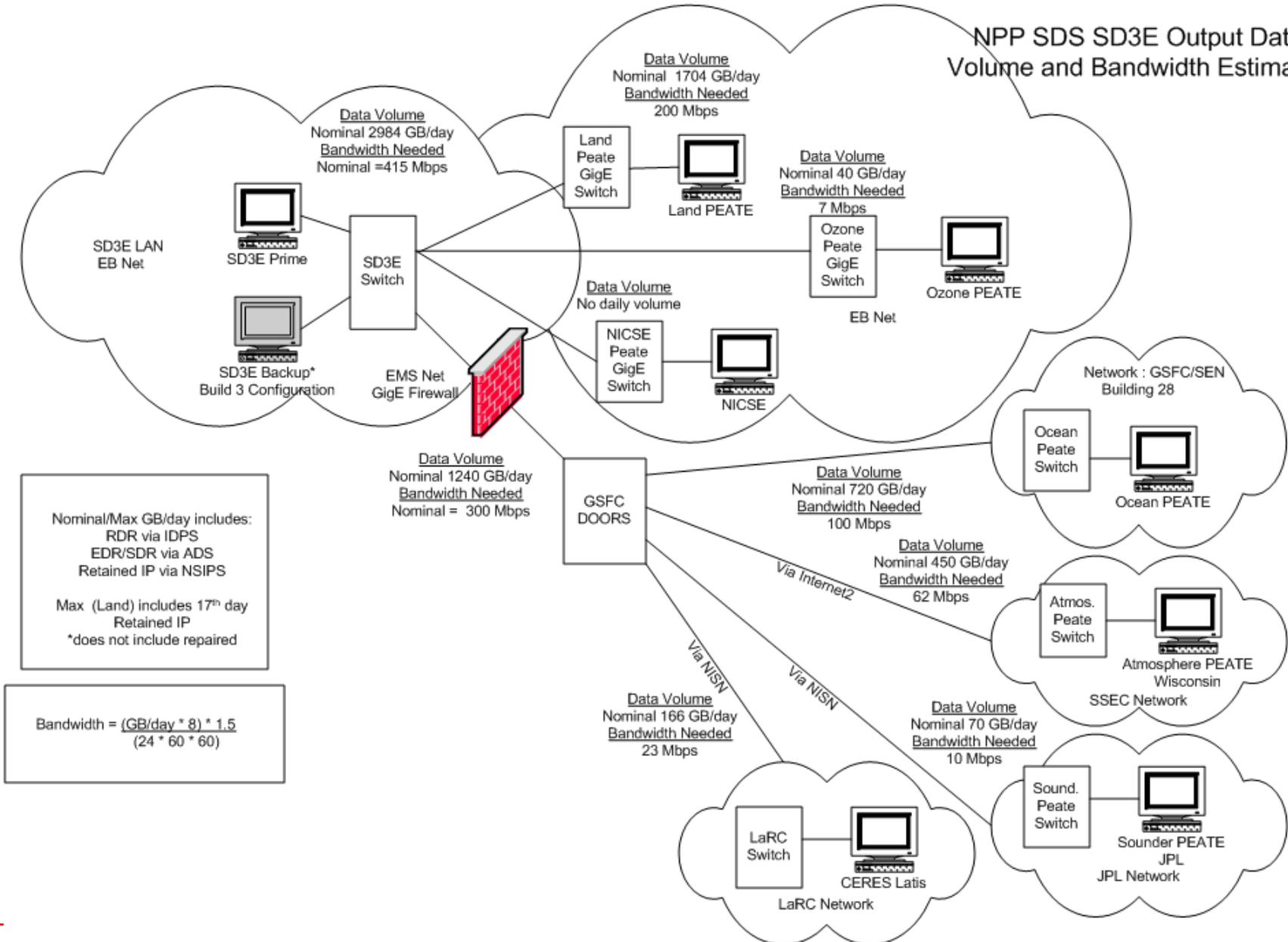
# Issues/Concerns

- **Reliability**
  - Background: SDS and associated elements have only been “Research Grade”, i.e., Ao of 0.95.
  - Recommendation: **Analyze, if this is sufficient.**
- **Latency**
  - Background: CERES desires RDRs and VIIRS SDRs within 24 hours
  - Recommendation: **Analyze, derive / model nominal latency**
    - (1) Land PEATE check SD3E within 24 hours
    - (2) CERES periodically check for available RDRs from SD3E
- **Network/bandwidth**
  - Background: Outbound (to PEATEs/NICSE) ~6 TB/day. Inbound (from external segments) ~5TB/day. CERES RDRs and VIIRS SDRs ~92 GB/day (TBD) will flow on same initial SDS circuits
  - Recommendation: : **Analyze** & Continue to test performance
    - Early tests with ADS/CLASS, IDPS, and NSIPS
    - Look at network configuration
    - Addition of CERES data should not be an issue



# Outbound Network

## NPP SDS SD3E Output Data Volume and Bandwidth Estimates



Nominal/Max GB/day includes:  
RDR via IDPS  
EDR/SDR via ADS  
Retained IP via NSIPS

Max (Land) includes 17<sup>th</sup> day Retained IP  
\*does not include repaired

$$\text{Bandwidth} = \frac{(\text{GB/day} * 8) * 1.5}{(24 * 60 * 60)}$$



# Backup



# SD3E Failure < 3 Hours

- Issue: Analysis of SD3E data catch-up mechanism if the SD3E system is down
  - SD3E is down if Ingest, Scheduler, Database has failed
  - SD3E is down if FTP server, raid controller or Database server is failed
  - Interface controller failure does not affect subscription; user's will not be able to place ad-hoc or subscription modifications/additions/deletion
- SD3E down less than 3 hours
  - ADS and IDPS will retransmit data products
  - NSIPS will retransmit data products TBD
  - SD3E operator will verify data subscriptions are active
  - SD3E operator will verify data from 3 hour time period is being received from IDPS, ADS, and NSIPS
  - External interfaces detect failed FTP transfers for this time period and continuously attempt to retransmit
  - SD3E will need to ingest an additional 475 GBytes if down 3 hours
    - ADS 300 GByte
    - NSIPS 150 GByte
    - IDPS 25 GByte
    - Estimate approximately 24 hours to acquire and ingest the missing 1/2 TByte



# SD3E Failure < 24 Hours

- SD3E down time is greater than 3 hours, but less than or equal to 24 hours:
  - SD3E will first acquire missing RDRs (180 GB if 24 hours); may take 8 hours
  - SD3E will then acquire missing Retained IPs (1.3 TB if 24 hours); may take 3 days
    - SD3E does not have the processing power to recover SDR or EDR data products
  - SD3E will lose ~0.5 TB of data for every 3 hours of down time
  - SD3E loss of ~4.5 TB of data if down for 24 hours
    - 3 TB from ADS (SDRs, EDRs, TDRs, ancillary, auxiliary)
    - 1.3 TB or 1.5 TB from NSIPS (Retained IPs)
    - 180 GB from IDPS (RDRs)
    - 240 GB from IDPS (Ocean SDR and EDR products)
- Retrieving RDRs
  - Some RDRs will be requested from IDPS (SD3E API will **automatically send ad-hoc request to IDPS**)
  - Older RDRs will be requested from ADS (**script to FTP pull from ADS** specified location)
  - SD3E operator will have to suspend Ocean SDR and EDR products from IDPS until SD3E retrieves missing RDRs; and missing Retained IPs (manual)
- PEATEs/NICSE will need to request missing SDRs, EDRs, TDRs, ancillary, auxiliary, and delivered IPs from ADS (Web will display SD3E down time)



# SD3E Failure > 24 Hours

- SD3E down time is greater than 24 hours:
- SD3E will obtain full set of RDRs for all of the missing days from ADS (**via ad-hoc request to ADS**)
  - Total data product loss of ~4.5 TB per day
  - SD3E Operator will retrieve missing RDRs only (ad-hoc this request to ADS using start and end times); 180 GB each day
  - SD3E operator will verify/check ADS subscription going forward
  - SD3E operator will verify NSIPS subscription going forward
  - SD3E operator will receive email from ADS for FTP pull opportunity of missing subscription data (will publish this to the PEATEs/NICSE)
- PEATEs/NICSE will need to request missing SDRs, EDRs, TDRs, ancillary, auxiliary, and delivered IPs for missing hours from ADS



# External Segment Downtime

- IDPS down
  - No data from IDPS, SD3E API will disconnect if DDS is down
  - SD3E may receive ADS xDRs for additional 6 hours
  - SD3E may receive NSIPS Retained IPs for additional 6 hours
  - Display system notice on SD3E Request Web site concerning data not ingested
- ADS distribution servers down over 6 hours
  - SD3E will not receive/ingest xDRs for down time period
  - SD3E will receive Retained IPs/RDRs from NSIPS/IDPS
  - PEATES can FTP pull missing SDRs/EDRs/DIPS/AUX/ANC from ADS if available
  - Display system notice on SD3E Request Web site concerning data not ingested
- NSIPS down over 12 hours
  - SD3E will not receive/ingest Retained IPs for down time period
  - SD3E will receive xDRs/RDRs from ADS/IDPS
  - Operator cancels subscription and re-subscribe
  - Display system notice on SD3E Request Web site concerning data not ingested