Jet Propulsion Laboratory
California Institute of Technology

Open Source Science for ESO Mission Processing Study

Identify a system architecture that meets the ESO mission processing objectives, supports open science, enables system efficiencies, and promotes earth-system science.

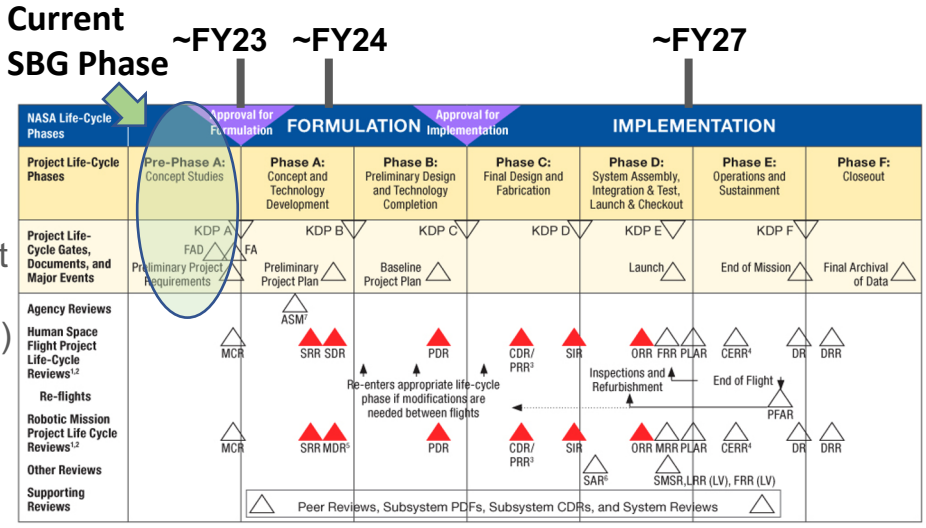
Workshop #1
October 19-20, 2021

SBG Mission Data Processing System Perspective

Jeff Pon

Overview

- SBG is in pre-Phase A lifecycle so SBG's mission concept and specific open science use-cases needs are evolving and subject to change
  - The SDS material in this briefing is a current SBG snapshot in time and is intended to help inform the Open Source Science (OSS) Steering Group's Study
    - SBG project is relying on OSS Study to help flush out future SBG SDS design/implementation details to support OSS goals and what OSS provided capability can be leveraged from other platform providers (MAAP, ImgSPEC, etc)
  - The material presented should **not** be interpreted as SBG's final plans



FOOTNOTES

1. Flexibility is allowed as to the timing, number, and content of reviews as long as the equivalent information is provided at each KDP and the approach is fully documented in the Project Plan.
2. Life-cycle review objectives and expected maturity states for these reviews and the attendant KDPs are contained in Table 2-5 and Appendix D Table D-3 of this handbook
3. PRR is needed only when there are multiple copies of systems. It does not require an SRB. Timing is notional.
4. CERRs are established at the discretion of program .
5. For robotic missions, the SRR and the MDR may be combined.
6. SAR generally applies to human space flight.
7. Timing of the ASM is determined by the MDAA. It may take place at any time during Phase A.

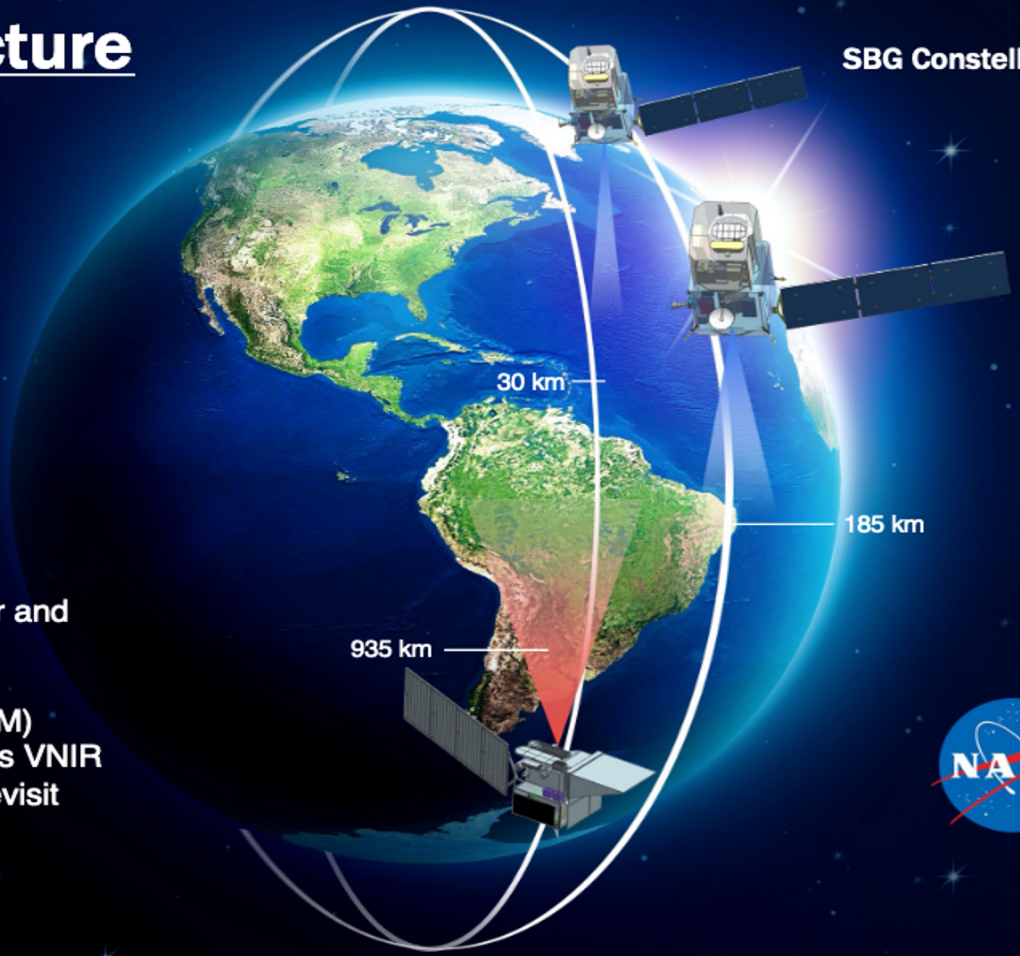
▲ Red triangles represent life-cycle reviews that require SRBs. The Decision Authority, Administrator, MDAA, or Center Director may request the SRB to conduct other reviews.

ACRONYMS

ASM – Acquisition Strategy Meeting
CDR – Critical Design Review
CERR – Critical Events Readiness Review
DR – Decommissioning Review
DRR – Disposal Readiness Review
FA – Formulation Agreement
FAD – Formulation Authorization Document
FRR – Flight Readiness Review
KDP – Key Decision Point
LRR – Launch Readiness Review
LV – Launch Vehicle
MCR – Mission Concept Review
MDR – Mission Definition Review
MRR – Mission Readiness Review
ORR – Operational Readiness Review
PDR – Preliminary Design Review
PFAR – Post-Flight Assessment Review
PLAR – Post-Launch Assessment Review
PRR – Production Readiness Review
SAR – System Acceptance Review
SDR – System Definition Review
SIR – System Integration Review
SMSR – Safety and Mission Success Review
SRB – Standing Review Board
SRR – System Requirements Review

FIGURE 3.0-1 NASA Space Flight Project Life Cycle from NPR 7120.5E

SBG Architecture



agenzia spaziale italiana

SBG Light

Wide-swath VSWIR spectrometer

Sun-sync orbit (late AM)
185 km swath
16 day revisit
10 nm, 200+ bands
30 meter GSD
High SNR and radiometric performance

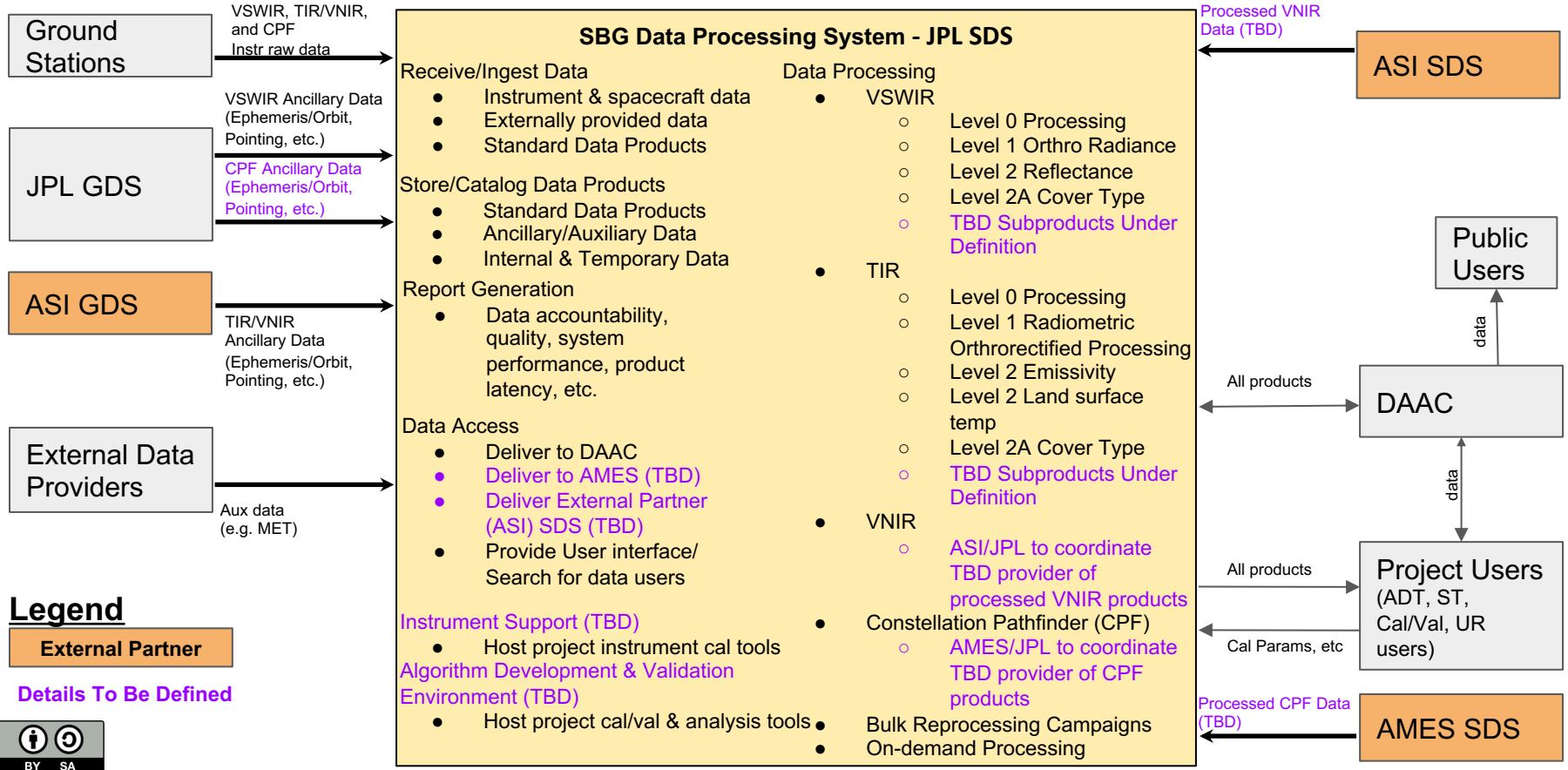
SBG Heat

Wide-swath TIR imager and ASI VNIR camera

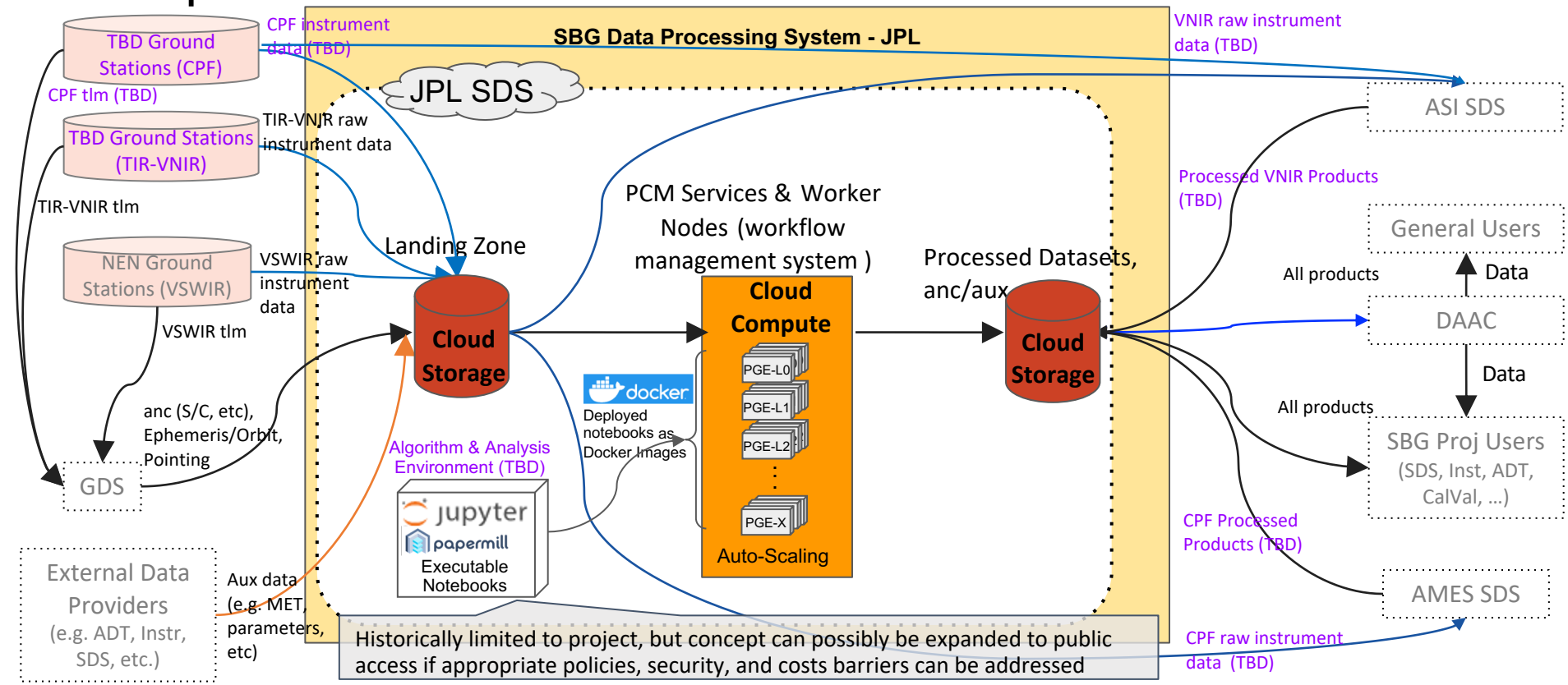
Sun-sync orbit (early PM)
5+ bands TIR, 2+ bands VNIR
935 km swath, 3 day revisit
60 meter GSD
0.2K NeDT

NASA

Data Processing System Architecture - SBG



Component and Infrastructure View - SBG



Supporting Earth System Science

- Does your MDPS use/share any data, algorithms, etc from other ESO projects to support Earth System Science?
  - AVIRIS, AVIRIS-NG, EMIT, ECOSTRESS, HyTES, Hyperion, HISUI, DESIS, and PRISMA
- What are your barriers to enabling collaboration to support Earth System Science within your ESO MDPS? (e.g., firewalls, access, schedule, developments costs, etc)
  - Can SBG science mission goals and open science goals fit within SBG's schedule/cost box?
    - Need to mature SBG's mission concept and implementation details
    - Need to flowdown high-level open science goals into actionable implementation steps
    - Allocation funding sources and OSS capability providers (e.g. differentiating which open science efforts should be covered by SBG project vs project agnostic funding sources)
  - Security: Cybersecurity controls and policies are in place to enable open science public access while protecting against malicious actors
  - ITAR: ITAR controls are understood and implemented to prevent ITAR violations while enabling open science public access and sharing with international partners
- What are the opportunities for improved support of Earth System Science?
  - Because SBG is early on in its lifecycle, SBG can potentially leverage new open science capabilities as they become available throughout SBG development
    - However, SBG should be prepared to provide any point-designs for its own SBG development wherever dependent open science platform capabilities will not be available in time
    - SBG will evaluate what open science capabilities are available over time and adjust plans accordingly while adhering to SBG development schedule and cost constraints

Supporting Open Science

SMD defines open science as a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding.

- What does this definition of Open Science mean in the context of data processing systems?
  - Data processing system can contribute toward open science objectives by improving public access and ease of:
    - Data Discovery
    - Data Sharing
    - Incorporating Data Processing Community Algorithm Contributions
- What do you feel are the most beneficial opportunities for improvement in the MDPS to support Open Science?
  - Ease usability for scientific community and public to filter through massive amounts of data
  - Enable flexibility in selecting algorithms or processing knobs in existing applications so users don't have to "reinvent the wheel"
  - Ease the logistical complexity of administration billing, configuration management, and platform deployments to enable the scientific community and wider public contributions to focus on creating new processing applications and discoveries
- What are the barriers to supporting open science?
  - Cost and Schedule: Every project has different cost budgets and schedule timelines so each project likely would prioritize certain open science goals based on their project needs differently
  - Culture and Policies: Changing mindset and changing policies to allow adoption of an open science system, while addressing cybersecurity concerns and complying with ITAR rules
- What components (Data system, PGEs, algorithms, data) of your system will be developed in the open (open source from the outset)?
  - Components and algorithms under definition
    - JPL is handling a subset of core SBG products, but SBG project intends to rely on community based algorithms to meet the sheer number of possible SBG subproducts which is much greater than JPL can provide by itself
    - SBG SDS considering community-based contributions to be provided in the form of Jupyter-based notebooks to fit within existing deployment frameworks leveraged from other ESO missions like NISAR's ADE env
  - Products will be published to DAACs for public distribution

Other

- What are your pain points in support of this mission?
  - SBG is in pre-Phase A lifecycle so SBG's mission concept and specific open science use-cases are evolving
  - SBG may have a subset of time-sensitive product requirements to achieve time-sensitive science community applications
  - SBG includes a hyperspectral sensor so raw data and resulting product volumes are **very large**
    - SBG Constellation Pathfinder still under formulation but may exponential increase data volume by hosting a whole smallsat constellation of spectral payloads in addition to the primary SBG-LIGHT(VSWIR) and SBG-HEAT (TIR-VNIR) spacecraft
- What does system efficiency mean in the context of an MDPS? (cost, data storage, processing time, etc.)
  - Optimized Compute and Storage Allocation
    - Reduce waste of duplicating data and unnecessary transfer hops especially with scale of SBG data volumes
    - Reduce waste of processing unnecessary products to save on storage and compute costs
      - Appropriate handling of cloud covered areas
      - Some SBG subproducts only need to be processed for certain geographic cover types
  - Optimizing Latency
    - SBG may require the ability to prioritize compute resources to minimize latency for time-sensitive products
    - Latency or accuracy optimized applications may be needed for different ~24 hr urgent response vs standard products needs
- Is there anything else you'd like to share that you feel would be helpful in our study?
  - Due to sheer volume of SBG data, it would be very useful if users had the ability to detect which products are not of interest to avoid 1) unnecessary standard product processing and 2) unnecessary adhoc product orders due to challenges of filtering data to ones of scientific value and the lack of community user ability to set up their own custom processing subscription triggers
    - If filter/custom processing subscription trigger capability developed for SBG, need training and social/administrative/billing policies to influence users to use those features to optimize the system to satisfy as many users request as possible
  - Due to evolving urgent response and standard product use-cases, SBG would benefit if public users had the flexibility to select between different combinations of algorithms or processing knobs to produce a latency vs accuracy optimized product
  - SBG is very interested in leveraging work from MAAP/IMGspec efforts to pathfind the development and usage of a common OSS platform that SBG SDS could adopt