SWAMP / AQUA Sub- Group Workshop on Data System Capacity (Volumes and Loads): needs and capabilities

Chris Justice

(MODIS Land Lead)

UVa

Background to the Meeting

- **SWAMP Meeting (GSFC, March 16/17th)**
  - Questions arose on:
    - **Current system performance (expected c. 3 months for the production system to stabilize)**
    - **System capacity for Terra Reprocessing and Aqua Processing**
    - **Need to revisit the '96 baseline numbers – which are being used to scope the system and are almost 5 years old – attempt to get the issue addressed by the Reber /Barron Comm.**
  - Yoram Kaufman requested a SWAMP sub-group meeting to work through these issues and to draft a letter to the EOS project and HQ
  - Dolly Perkins (ESDIS) suggested that for this to be useful in terms of making the case for resources, that input would be needed by the end of May
  - Recognition that the production system is not yet stable and that the distribution system is only just starting to be exercised
  - There is a need to revisit system performance periodically at working sub-group level as the system becomes more stable and sustained performance is established

Background Cont'd.

- **Current Status**
  - **6 months after launch EC system remains unstable - finding it hard to meet and sustain our production targets – but slowly fixing problems as they occur – some resource sharing is happening**
  - **Some system improvements planned /scheduled by ECS**
  - **Data are being distributed e.g. L7, MODIS L1, CERES**
  - **Instrument teams busy refining algorithms based on instrument performance – hampered by incomplete data coverage moving towards product distribution**
  - **Reprocessing is increasingly important – need to look at reprocessing turn-around time (1 year in 1 year is too slow, 3- 4 months needed)**
  - **No science advisory process in place to help advise/guide ESDIS Project on resource decision-making – AHWGP last concerted effort – a steady decoupling of data system capabilities and science needs**
  - **PI processing has helped reduce costs and share responsibilities**
  - **Instrument teams/data production groups willing to develop a closer partnerships with ESDIS to solve problems, taking on more responsibility for services**
  - **Budget shortfalls for ESDIS w. requests for over-guide – "no new money" message - need to help ESDIS make the case for resource increases and cost effective reallocation of resources – new solutions needed**

Objectives of the Meeting

- **Review the ’96 baseline in terms of instrument/science needs – draft a revised baseline based on better understanding**
- **Review current system performance re. 96 baseline and the revised baseline**
  - **Production (at each stage in the chain)**
  - **Ingest from SIPS**
  - **Archive**
  - **Distribution**
- **Identify current obstacles/bottlenecks and suggest what needs to be done to alleviate them**
- **Identify mismatch between current and planned performance and needed capacity**
- **Review what is planned in terms of increased capacity (this coming year and a view to 2003)**
- **Identify anticipated challenges and improvements**
- **Suggest practical options for meeting our goals: resource needs, efficiencies, different approaches, areas for cost saving, ways of doing more for the same**
- **Identify possible next steps**

Approach to the Meeting

Pre meeting - submission of requested material on volumes and loads

---

One day open meeting

- **Compiling improved information**
    - **Developing a revised database on needs versus current and planned capacity - devise common reporting /display of material**
    - **Identifying any major mismatches**
    - **Identifying current bottlenecks**
  - **Suggested solutions to current bottlenecks (near-term)**
- **Identifying anticipated challenges and improvements**
  - **Suggested approaches to challenges**
- **Suggested options for meeting our goals**
  - **What additional resources are needed by ESDIS?**
  - **Are there cost saving approaches and solutions that could be adopted?**
- **Identify next steps – additional issues that need to be addressed**

Half day closed door session for instrument teams

- **To draft a formal letter and recommendations from the sub group meeting to AM Project Scientist to send to the Project**

SWAMP Input received

- - DAACs: GSFC, LaRC, EDC, NSIDC
- - Teams: MISR, MODIS (DAAC / SIPS), CERES, MOPITT (SIPS), AIRS (unsigned DIPS ICD)
- - Not ASTER Team – phone call– no new processing needs
- - EDOS input provided by the Project
- - L7 Team (like a SIPS) not asked – shared DAAC Distribution

Group developed a template for reporting needs v. capabilities

Major mismatches between needs and capabilities

- **General comments**
  - **At 1x effective throughput – keeping up is a problem (currently production issue L0-L3)**
  - **Distribution mismatches - TBD**
  - **Appear to have capacity for .5 of the 96 Baseline (MISR excepted ?) – however updated science needs ’00 draft baseline c. 3 times larger than 1X ’96 baseline**
  - **Reprocessing is an emerging tall pole**
- **MODIS**
  - **Revised baseline numbers will increase ingest at EDC/GSFC/NSIDC, Archive and distribution flow implications – need creative solutions**
  - **Reprocessing suggested at MODAPS – return products into archive - 1A needed**
- **CERES**
  - **Processing for timely validation (3x-4x??)**
  - **CERES subset of MODIS Level 1B not happening at GDAAC currently**
- **MISR**
  - **Processing and reprocessing capacity**
  - **PDPS efficiency**
  - **Memory per processor**
  - **Bandwidth to ST**
- **MOPITT**
  - **RAID**
  - **Processors sharing functions**
  - **Ancillary data reliability and cost**

Options on how to proceed (cost savings)

- **No new money – or small changes – work within 96 baseline $ - However real need for revised baseline numbers**
- **Current plans should be made with a view to transitioning to '03 and NEWDIS**
- **Move away from 1 size fits all approach - work by instrument – keep the data solutions close to the science team – each instrument to suggest how it would get to its '00 goal within an allocated budget**
- **Creative / innovative solutions needed**
  - **Build on ECS and SIPS Capabilities**
  - **New low cost solutions e.g. Linux developments GSFC/Langley**
- **Proactive DAAC activities to keep the users satisfied over the next year– innovative and creative approaches needed on the distribution side**
  - **e.g. additional media types, data set lending library, regional distribution points, regional subsets, on-line analysis, on-demand subsets, browse, visiting scientist facility**
- **Revisit the product suites – revised schedule (c. 1 year after launch - what products are truly operational – could we develop an acceptable experimental product category)**
- **More interaction between the ST and the archive for first year – flexibility as to what needs to be archived**
- **Determine how much risk is acceptable – move the risk closer to where it is is likely to have an impact – better assessment of what is non-negotiable**
- **Mechanism needed for working these issues – continuing SWAMP Data Working group**