

# CTA-US Participation in DATA-WP: A proposal

Lucy Fortson, Michael Briggs, Valerie Connaughton, Brian  
Humensky

# Purpose: Increase CTA-US participation in the DATA WP

---

## ▶ Outline

- ▶ Context: review from Michael Briggs on Paris DAFA-CEIN meeting Feb 2-3
- ▶ Motivation for why CTA-US needs to be involved in the DATA WP process
- ▶ NSF Funding opportunities
- ▶ Straw-man plan to increase CTA-US participation in DATA-WP

# CTA DAFA Paris Meeting Report Feb 2-3

M. Briggs

---

- ▶ Goal: understand end-user calibration files
- ▶ 20+ attendees, France, Germany, Italy, Spain (lots), US (1)
- ▶ Led by Catherine Boisson, DAFA WG leader
- ▶ Several presentations, extensive discussions, group generation of an issues list.
  - ▶ Jurgen Knödseder: overview of “CTA Response Formalism”
  - ▶ Arache Djannati-Atai: Excellent discussion of background est. techniques → need for instrument response function??
  - ▶ Karl Kosack: Science Analysis Toolset Requirements
  - ▶ Overview of first data challenge: HESS, MAGIC data
  - ▶ Catherine Boisson: what calibrations to provide end-user?
  - ▶ All: development of issues document; some decisions taken (e.g. not to provide CR data to GI end-users)

# Why we need CTA-US participation in DATA-WP now!

---

- ▶ CTA-EU has produced the entire data management architecture and is fleshing this structure out as we speak
  - ▶ Key decisions are being made
    - ▶ Fits file format over Root
    - ▶ CTA response formalism → currently generic enough to allow SC integration, but:
      - Investigated XSPEC / OGIP representation and decided to support XSPEC and its format as secondary format.
      - Is this the right decision from US perspective? Large US community (HEASARC) familiar with XSPEC formats
      - Primary format is a 9-dim data cube similar to Fermi-LAT
  - ▶ Key “experience” being gained by EU groups
    - ▶ Producing and participating in data challenges
      - Software tools developed that could become de facto CTA analysis tools
        - If we don’t at least participate in challenge, we lose opportunity to contribute to development and experience with tools

# Why we need CTA-US participation in DATA-WP now!

---

- ▶ 1/23/2012 “Repartitioning Document” has assigned specific tasks within data management architecture to institutions. DESY-Zeuthen only CTA-US institution.
  - ▶ Data Model WGs (subtasks): Data (4), HLOD (3), Archive (3)
  - ▶ Data Pipelines WGs: MC-Analysis (2), Reconstruction (2), Real-time analysis (3)
  - ▶ Data Access WG: 8 sub-tasks
  - ▶ Example: Data Model definition, including low level (L0 and L1A) and mid-level (L1B to L2C) data products. Deliverables: DAWG.1: Raw data format including headers and full camera information, and list of calibration parameters for TCM. DAWG.2: Comparative studies on FITS format vs other formats. DAWG.3: Define common software to be applied to MC to produce image moment Hillas parameters to be included in the data, thus producing L1B. DAWG.4: Common format for DST, Root trees, etc. Definition of L2C data format.
    - ▶ 10 institutions already signed up for DAWG subtasks

# Why we need CTA-US participation in DATA-WP now!

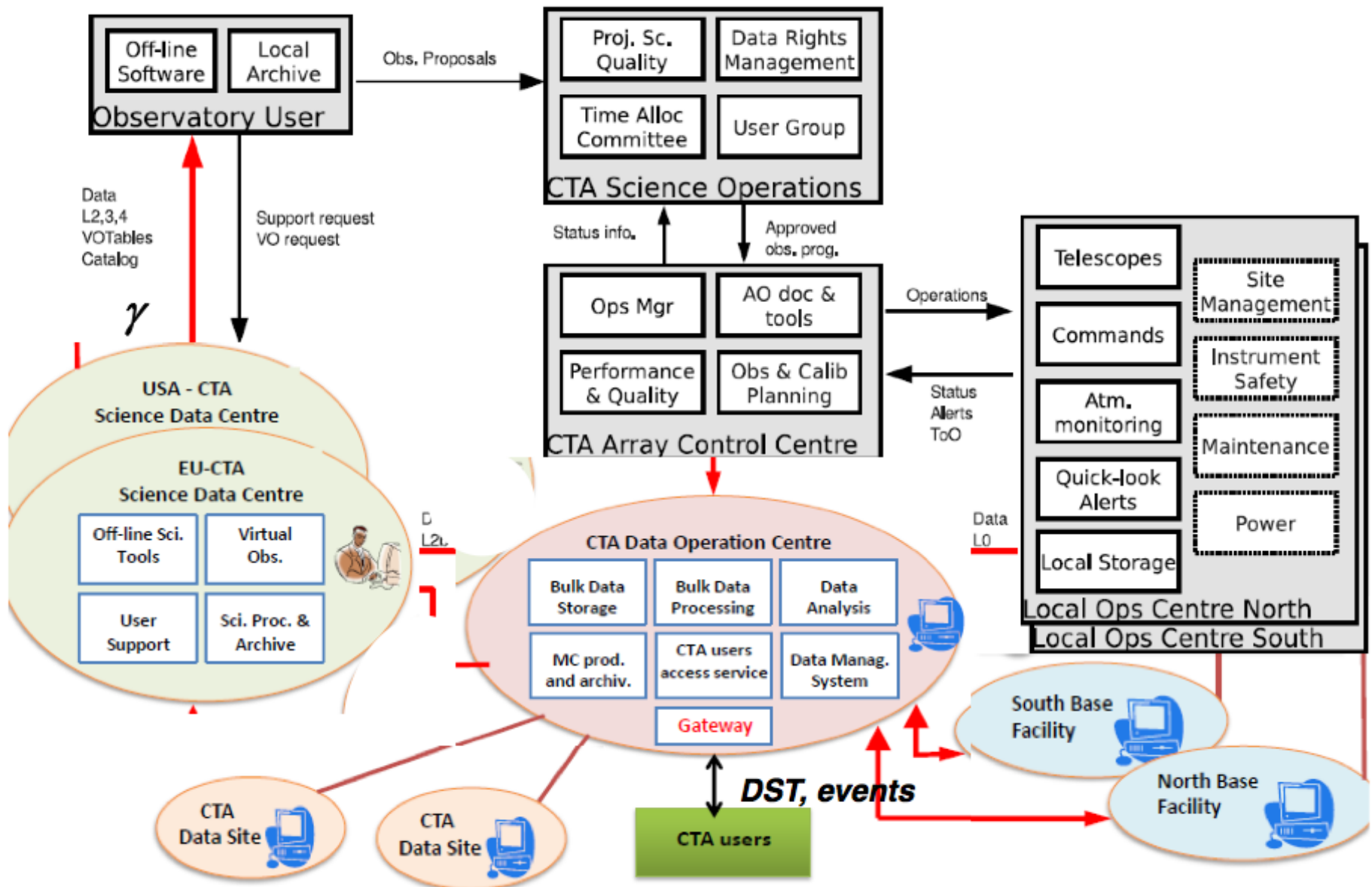
---

- ▶ Metadata meeting March 8-9 Montpellier
- ▶ address the following main topics:
  - ▶ MetaData General Structure
  - ▶ MetaData Management Infrastructure
  - ▶ Link with Data fields defined by DAFA WP
  - ▶ Link with Data-driven software development framework

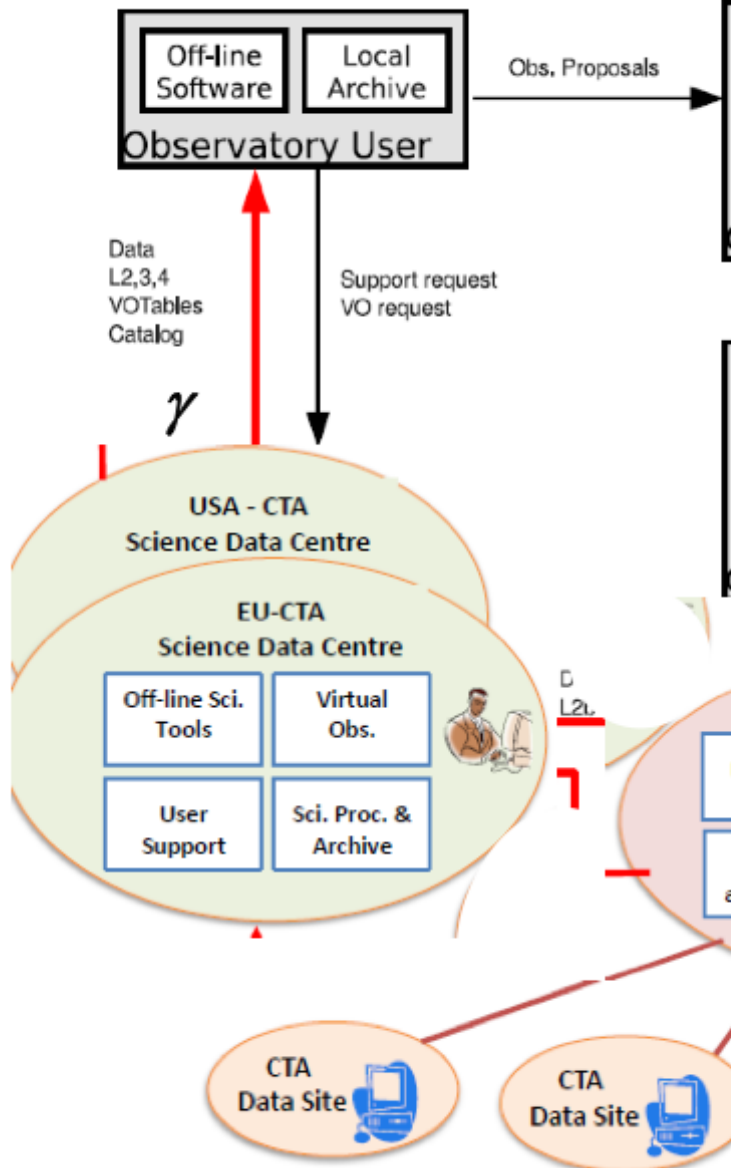
The idea is to have a few dedicated talks and a significant fraction of time devoted to round-table discussions. The meeting will take place at the Université Montpellier 2.



# DATA Architecture includes US Science Data Center



# DATA Architecture includes US Science Data Center



➤ Are we happy to take what decisions are made for us?

- What level data products do we want, need for US community interests?

➤ We need to participate in planning for data centers.

- So far no major decisions taken in terms of distribution of centers and level of products at which types of centers.
- Coming soon after site selection!



# CTA-US Data Management?

---

- ▶ Clear need for US to bring more resources to allow better integration with DATA-WP
  - ▶ Eventual need to integrate SC into data management structure
- ▶ US involvement in developing data management/processing/support center(s)
  - ▶ Gain/share expertise at all levels of data product development so we are not left as “end-users”
    - Utilize extensive US experience with public data products
  - ▶ US agency funding => bigger/better return to US community on US investment if we are efficiently integrated with CTA-EU

## Potential Funding opportunity with NSF (1):

- ▶ SAVI program: Science Across Virtual Institutes
- ▶ Must have international component
- ▶ Requires PI to be from currently funded NSF project
- ▶ This is a concept intended to foster interaction among scientists, engineers and educators around the globe. It is based on the principle that scientific advances can be accelerated by working together via virtual institutes that connect researchers with complementary strengths for greater impact on an identified problem. **The primary objective of SAVI is to facilitate collaboration among scientists and engineers from the U.S. and other countries by building on mutually beneficial partnerships that are initiated by NSF-supported teams of researchers, research institutes, and universities. Such virtual institutes are to serve as the creative hubs for innovative research and education activities across borders.**
- ▶ Max budget request \$400K/year for five years
- ▶ Submission through NSF-Physics Oct 2012 – discussion with Whitmore very positive. Potential for supplement (small!) to an NSF institution to get started this year.



## Potential Funding opportunity with NSF (2):

- ▶ RCN program: Research Coordination Networks
- ▶ Encouraged to have international component
- ▶ The goal of the RCN program is to advance a field or create new directions in research or education. Groups of investigators will be supported to communicate and coordinate their research, training and educational activities across disciplinary, organizational, geographic and international boundaries. RCN provides opportunities to foster new collaborations, including international partnerships, and address interdisciplinary topics. **Innovative ideas for implementing novel networking strategies, collaborative technologies, and development of community standards for data and meta-data are especially encouraged.**
- ▶ Max Budget: 500K total for expected 5 years
- ▶ Submission through OCI (Office of Cyberinfrastructure)



# Strawman plan

---

Put together group of interested parties from CTA-US that can:

- ▶ begin the process of documenting and promoting US needs and interests to DATA-WP
  - ▶ Could some of this be taken care of in PTDR for SC design??
  - ▶ list out US groups with interest and resources
- ▶ begin contributing to DATA-WP
  - ▶ promote Data Challenge participation
    - ▶ shows US is contributing to design through feedback
    - ▶ US teams will get used to using data products and hopefully contribute to development
    - ▶ Potential use of VERITAS data in challenges (Gernot already doing??)
  - ▶ Define requirements for US data center
    - ▶ What level of data product, end user support, archiving, etc
    - ▶ What level of development contributions are we interested in or are we just going to “mirror”/host what we’re given?
- ▶ Select tasks from “Repartitioning” document

# Strawman plan

---

Put together group of interested parties from CTA-US that can:

- ▶ obtain funding for US involvement in DATA-WP
  - ▶ start with plan to make VERITAS data (some subset!) public
    - ▶ develop in-house experience with and understanding of what it means to provide public access to even the highest level data products e.g. formats, metadata standards, level of support required
    - ▶ agencies will like (new DMP focus on open data)
    - ▶ development of metadata standards
  - ▶ in parallel work towards agreement with HESS, MAGIC, VERITAS to archive and provide public access to archived data products
    - ▶ provide historical data for CTA GI program
    - ▶ exercise will
      - Provide critical open-data expertise to US groups
      - Develop culture of open-data products not yet embraced by community

# CTA-US Data Management Group??

---

- ▶ **Groups already involved/contributing:**
  - ▶ Huntsville – (Michael Briggs)
  - ▶ SLAC – (Richard Dubois)
  - ▶ DESY – (Gernot Meier)
- ▶ **Groups with expressed interest:**
  - ▶ UMN - Columbia
  - ▶ UCSC - Iowa State
  - ▶ CfA - Penn State ?
- ▶ **Start round of telecons with interested parties.**
  - ▶ Include representative from CTA-EU?

