# Satellite and *in situ* Cal/Val data access

Chair: Bojan R. Bojkov (NASA)

Session support: Guiseppe Ottavianelli (ESA)

Moderator: Pascal Lecomte (ESA)

# Objective

Reach a consensus on cal/val data, including:

- *Availability and accessibility*
- *Reporting, usability and interoperability*

to address the GEOSS SBAs and meet national and international commitments

# Outline

- ✓ Overview
  - *Example - atmospheric composition*
- ✓ Data issues
  - *Cal/val access: What is required?*
  - *Exchange protocol: How to balance the cal/val needs and data ownership*
  - *Data content and format*
  - *Archiving: short and long-term requirements*
- ✓ Discussions and wrap-up

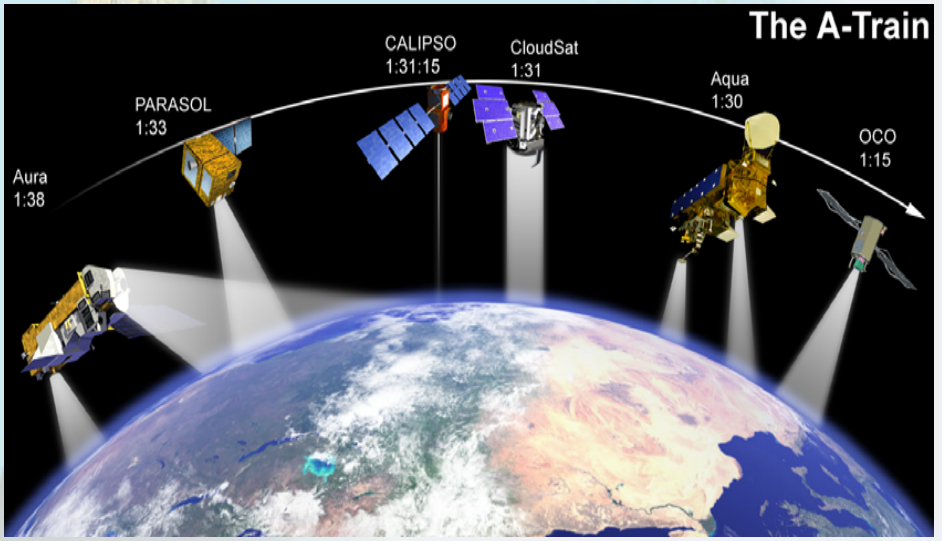
# Current state of affairs

*Atmospheric composition*

- ✓ All missions have cal/val plans and dataset requirements
  - *Mainly at the instrument level with some platform coordination*
  - *Limited coordination across missions*
  - *Usually local handling of cal/val datasets (at PI or cal/val investigator level)*

- ✓ Numerous existing and future space missions with similar cal/val needs
  - *NASA-EOS and ESA-Envisat*
  - *NASA OCO and JAXA GOSAT ('08/09)*
  - *MetOp2 and NPP/NPOESS ('09-)*
  - *A-train, etc.*

- ✓ Cal/Val activities reaching **across disciplines**
  - *For example the impact of aerosols, water vapor, tropospheric pollutants on retrieval algorithms*

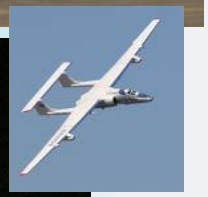


# Current state of affairs (2)

*Atmospheric composition*

- ✓ Numerous mission specific cal/val campaigns
  - *Ground, balloon, aircraft, ship*
- ✓ Special cal/val efforts
  - *through direct funding*
  - *through projects - for example NASA NRAs or ESA AOs, etc.*
- ✓ Not necessarily coordinated outside of the mission's agency
  - *Data available only to the participants*



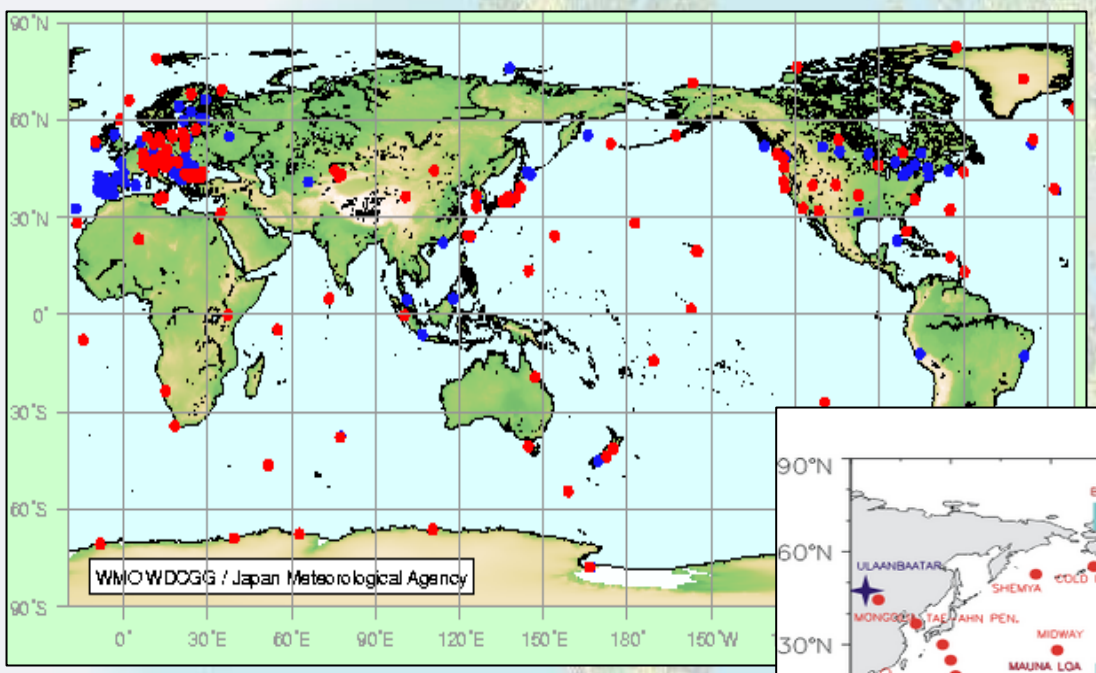
# Current state of affairs (3)

*Atmospheric composition*

- ✓ Cal/val teams are using/accessing the same independent datasets
  - *Groundbased/ocean networks: WMO/GAW, NDACC, etc.*
  - *World data centers (for example WOUDC, WDCGG, etc.)*
  - *National data centers (CNES-ETHER, NOAA-NCDC, etc.)*
  - *Network repositories (Aeronet, EMEP, etc.)*
- ✓ However with very different requirements
  - *Both for satellite and correlative systems temporal/content requirements , formats, timeliness, etc.*

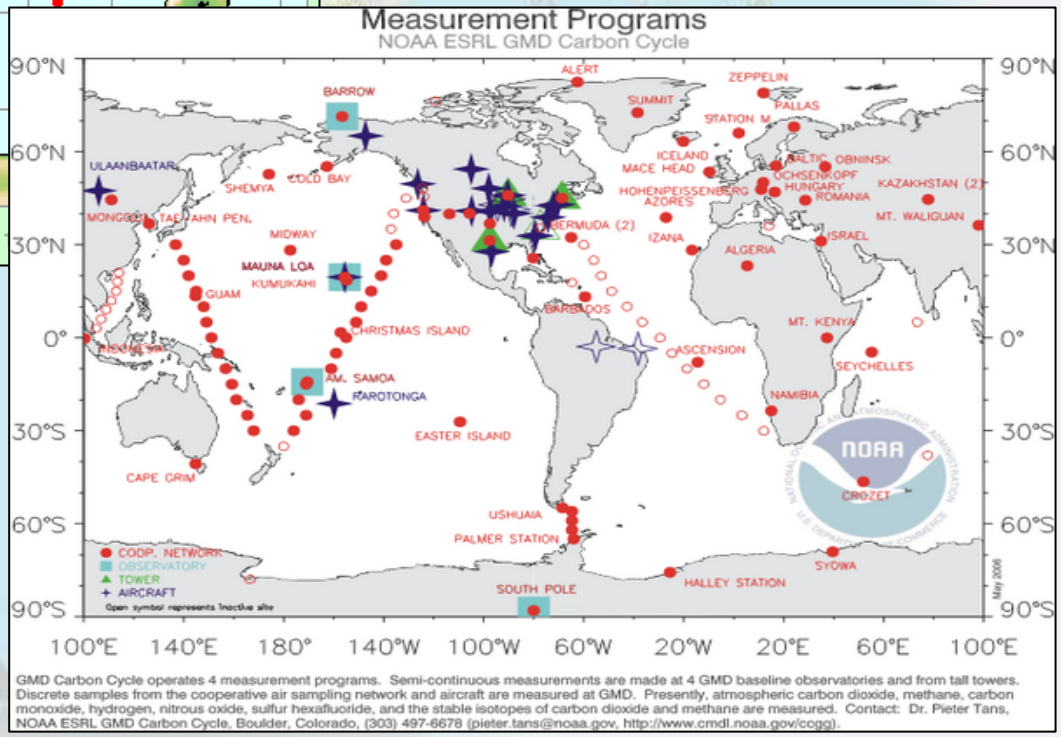
# Current state of affairs (4)

*Atmospheric composition*



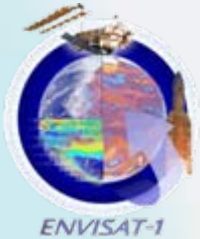

Hourly/monthly averages

Continuous, flask, campaigns



# Current state of affairs (5)

*Atmospheric composition*


✓ Promising collaboration:

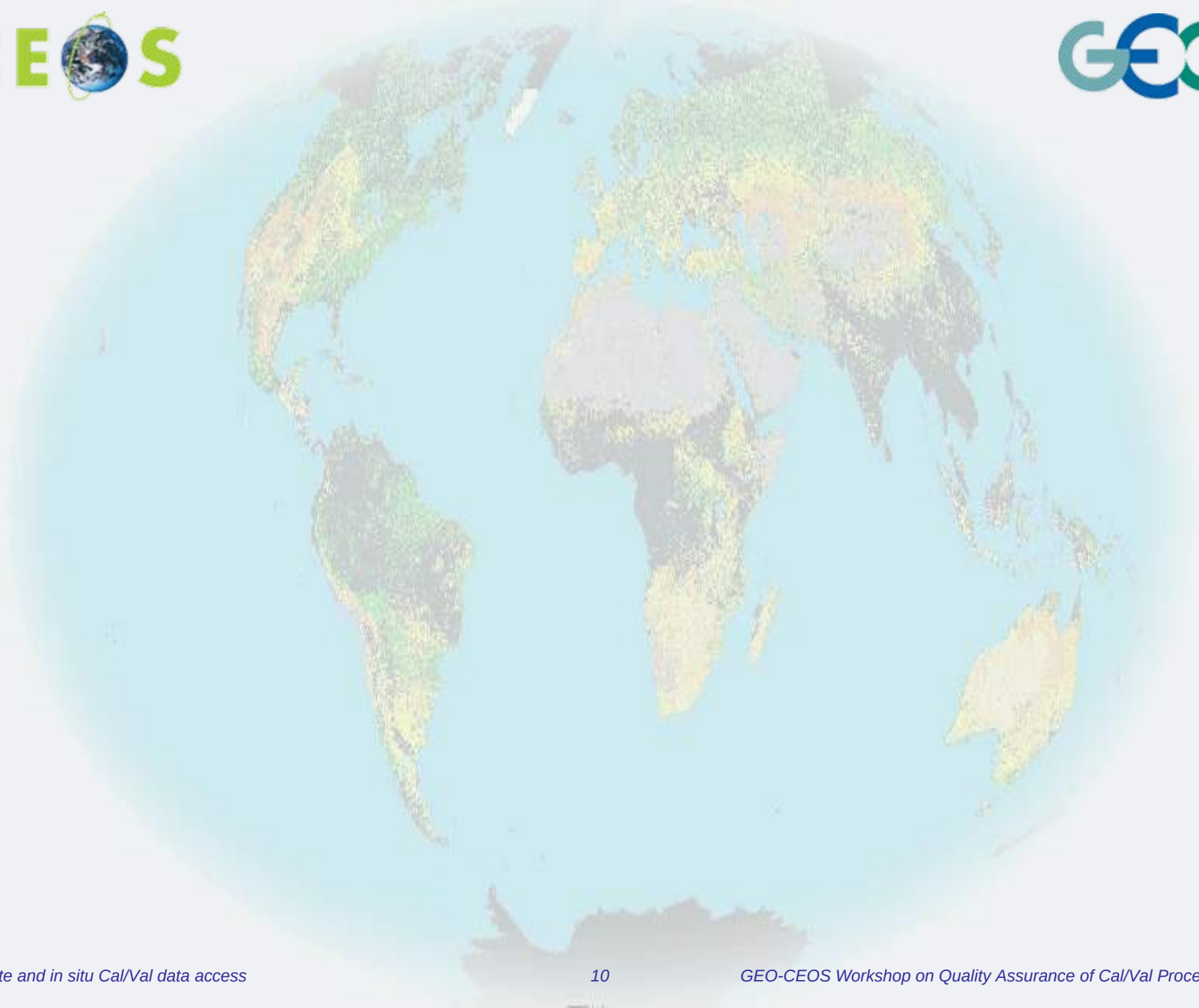
- *ESA Envisat/GMES and NASA Aura*
  - Data sharing, campaign planning/coordination, satellite data access
  - Synchronization of cal/val data requirements

- *NASA OCO and JAXA GOSAT for carbon budget*
  - Tentative cal/val data sharing and planning
  - Collaboration/measurement planning with ground sites/networks (should also include Sciamachy)

# To summarize

*Atmospheric composition*

- ✓ Mostly individual (and uncoordinated) mission efforts
- ✓ Distributed datasets with limited or no access
- ✓ Different and/or diverging data reporting requirements
- ✓ Repeated and differing demands on independent providers
- ✓ Similar issues across disciplines



# Cal/val data issues

# Data access

- ✓ Availability
  - *It should include all types of correlative data, including satellite datasets*
  - *Datasets should not be limited to mission driven efforts only*
  - *Cal/val should also include voluntary contributions, campaigns and network efforts*
  - *Data should be made available in timely manner (even if only in preliminary form) to the validation teams*

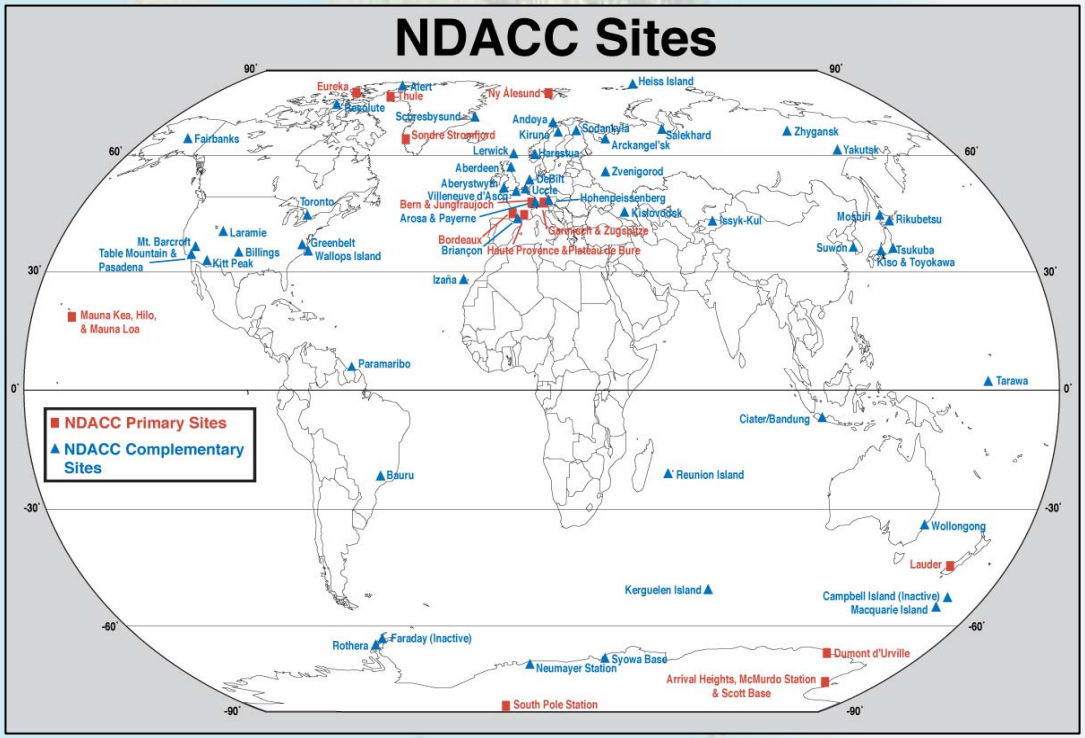
- ✓ Good examples are:
  - *NDACC preliminary data access for satellite validation*
  - *CSA SciSat mission and EOS-MLS exchanges*
  - *EOS-OMI and Envisat Sciamachy*
  - *EOS-Aura station/regional data subsets in NRT*
  - *etc.*


# Data access (2)

- ✓ Campaign and network datasets
  - *Datasets should be readily available/shared across agencies and investigators*
  - *Efforts should be coordinated, especially as needs overlap for missions and across disciplines → minimize duplicate reporting efforts*
- ✓ Version control
  - *Timely updating/revisions essential*
  - *Version histories need to be maintained (sub-orbital and satellite)*
  - *Towards maintaining/tracking "data quality"*

# Data access (3)

- ✓ Regarding independent data providers
  - *Individual investigators which may or may not be part of "networks" or part of "super-sites"*
  - *Dedicated long-term efforts, fostering innovation*
  - *BUT mostly outside of space (support) infrastructure - particularly between missions*
  - *Special arrangements may be required to facilitate access*



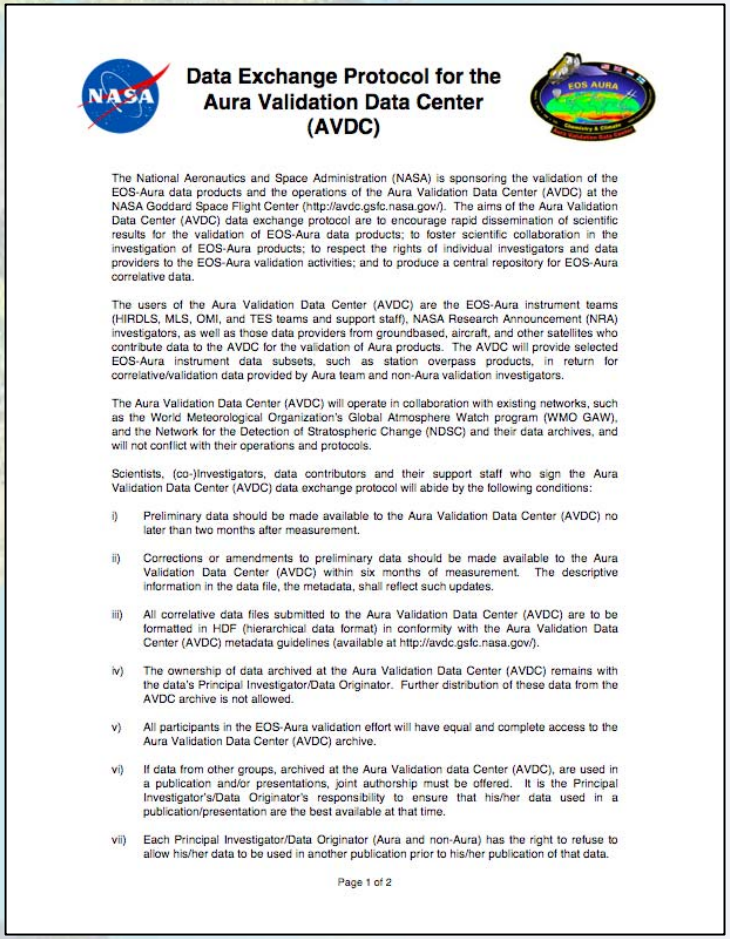

*e.g.* almost all NDACC PIs are outside the space realm

# Data policy

- ✓ A limiting factor with cal/val data access is the lack of clear data sharing protocols and policies
- ✓ Concern with potential “runaway effect” and “data poaching”
  - *particularly as data is shared across agencies*
  - *“loss of control”*
- ✓ Unfortunately diverging mission requirements (formats)

# Data policy (2)

- ✓ European Commission (1991-on), ESA Envisat, NASA aircraft missions, etc. have long standing data protocols
  - *Ownership and "intellectual" property issues*
  - *Carried over by space agencies - refer to Aura Validation Data Center Data Exchange Protocol which incorporates key items w.r.t. ownership, publication, updating, status, reporting requirements, …*



**Data Exchange Protocol for the Aura Validation Data Center (AVDC)**

The National Aeronautics and Space Administration (NASA) is sponsoring the validation of the EOS-Aura data products and the operations of the Aura Validation Data Center (AVDC) at the NASA Goddard Space Flight Center (http://avdc.gsfc.nasa.gov/). The aims of the Aura Validation Data Center (AVDC) data exchange protocol are to encourage rapid dissemination of scientific results for the validation of EOS-Aura data products; to foster scientific collaboration in the investigation of EOS-Aura products; to respect the rights of individual investigators and data providers to the EOS-Aura validation activities; and to produce a central repository for EOS-Aura correlative data.

The users of the Aura Validation Data Center (AVDC) are the EOS-Aura instrument teams (HIRDLS, MLS, OMI, and TES teams and support staff), NASA Research Announcement (NRA) investigators, as well as those data providers from groundbased, aircraft, and other satellites who contribute data to the AVDC for the validation of Aura products. The AVDC will provide selected EOS-Aura instrument data subsets, such as station overpass products, in return for correlative/validation data provided by Aura team and non-Aura validation investigators.

The Aura Validation Data Center (AVDC) will operate in collaboration with existing networks, such as the World Meteorological Organization's Global Atmosphere Watch program (WMO GAW), and the Network for the Detection of Stratospheric Change (NDSC) and their data archives, and will not conflict with their operations and protocols.

Scientists, (co-)Investigators, data contributors and their support staff who sign the Aura Validation Data Center (AVDC) data exchange protocol will abide by the following conditions:

i) Preliminary data should be made available to the Aura Validation Data Center (AVDC) no later than two months after measurement.

ii) Corrections or amendments to preliminary data should be made available to the Aura Validation Data Center (AVDC) within six months of measurement. The descriptive information in the data file, the metadata, shall reflect such updates.

iii) All correlative data files submitted to the Aura Validation Data Center (AVDC) are to be formatted in HDF (hierarchical data format) in conformity with the Aura Validation Data Center (AVDC) metadata guidelines (available at http://avdc.gsfc.nasa.gov/).

iv) The ownership of data archived at the Aura Validation Data Center (AVDC) remains with the data's Principal Investigator/Data Originator. Further distribution of these data from the AVDC archive is not allowed.

v) All participants in the EOS-Aura validation effort will have equal and complete access to the Aura Validation Data Center (AVDC) archive.

vi) If data from other groups, archived at the Aura Validation data Center (AVDC), are used in a publication and/or presentations, joint authorship must be offered. It is the Principal Investigator's/Data Originator's responsibility to ensure that his/her data used in a publication/presentation are the best available at that time.

vii) Each Principal Investigator/Data Originator (Aura and non-Aura) has the right to refuse to allow his/her data to be used in another publication prior to his/her publication of that data.

Page 1 of 2

# Data policy (3)

- ✓ Feedback mechanism
  - *Involves directly the cal/val data provider and investigators and their data users*
  - *Towards a better understanding of data*
- ✓ Enhanced data availability
  - *for example, Aura tropospheric ozone validation had access to twice the data through ESA/NASA/EC collaboration than through the conventional means*

# Data content and format

- ✓ Format, or dataset “packaging”, is an issue of **convenience**
- ✓ The format should be (ideally) the same across missions (and disciplines)
- ✓ Format consistency should be maintained (at the very least) across networks and similar measurement types
- ✓ There is a need to avoid conflicting requirements
  - *Increased burden on the providers, users and administrators*
  - *Unfortunately a frequent occurrence from the different data centers/missions - for example MetOp vs. Envisat vs. Aura*

# Data content and format (2)

- ✓ The content (including metadata, nomenclature, content) is a **usability issue**
- ✓ Is the information provided appropriate for physical interpretation and cal/val?
  - *Are uncertainties, resolution, caveats, etc. defined/reported?*
  - *Interoperability across measurements… for example, is a profile from X equivalent to a profile from Y?*
- ✓ Is ancillary data available for interpretation
  - *e.g. PTU, AOD, etc..*
- ✓ Is the nomenclature/meaning consistent across similar measurements
  - *Naming: O3 vs. ozon vs. OZ?*
  - *Meaning: is altitude resolution consistent across measurements?*

# Data content and format (3)

- ✓ Limited harmonization efforts on-going
  - *Mostly regarding format (Gaines & Hipkind, 1988)*
  - *Limited efforts regarding content, nomenclature and metadata (Envisat, Aura efforts, 2000-)*
- ✓ Commend the NDACC efforts to homogenize profile measurements
  - *Formats, nomenclature and content, including "quality indices" variables*
  - *Documentation available at:*

  *http://avdc.gsfc.nasa.gov/Documentation/Metadata/index.html*
- ✓ Quality information is rather limited in correlative datasets
  - *ESA GSE efforst such as PROMOTE and related efforts noted*

# Archiving: short-term

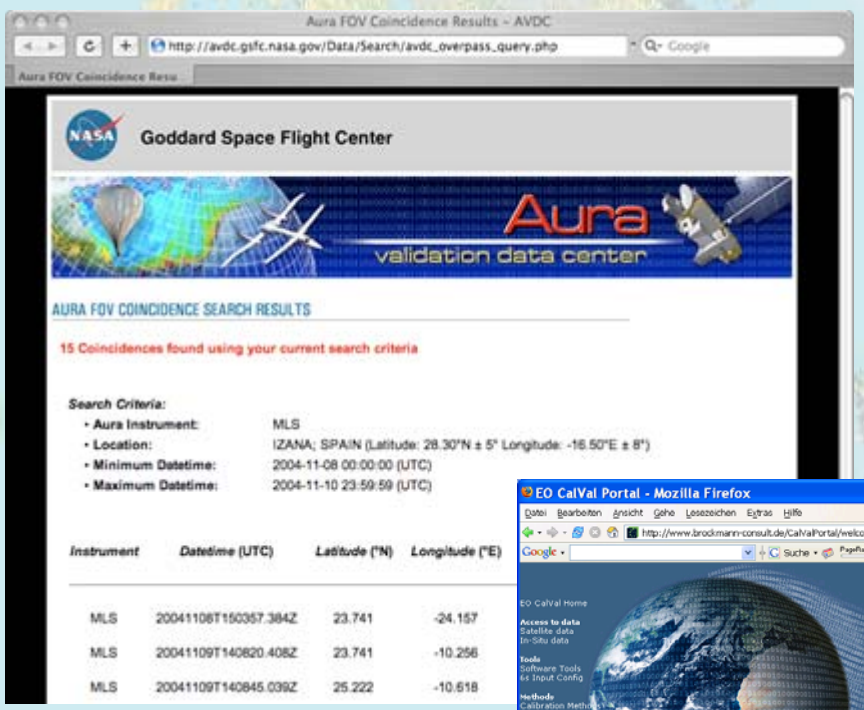
- ✓ Essential for the commissioning phase
- ✓ Data required in near real-time
  - *Include sub-orbital and satellite dataset*
  - *Processing permitting -"burden" for PIs - particularly when unfunded*
  - *Complete datasets (not limited to OVP time)*
  - *Should be broad and not exclude individual investigators*
- ✓ Provide satellite data subsets/information (satellite team responsibility)
  - *Routinely over key sites and regions, and for campaigns*
  - *Instrument FOVs - predictions for measurement planning; and actual for analyses purposes*

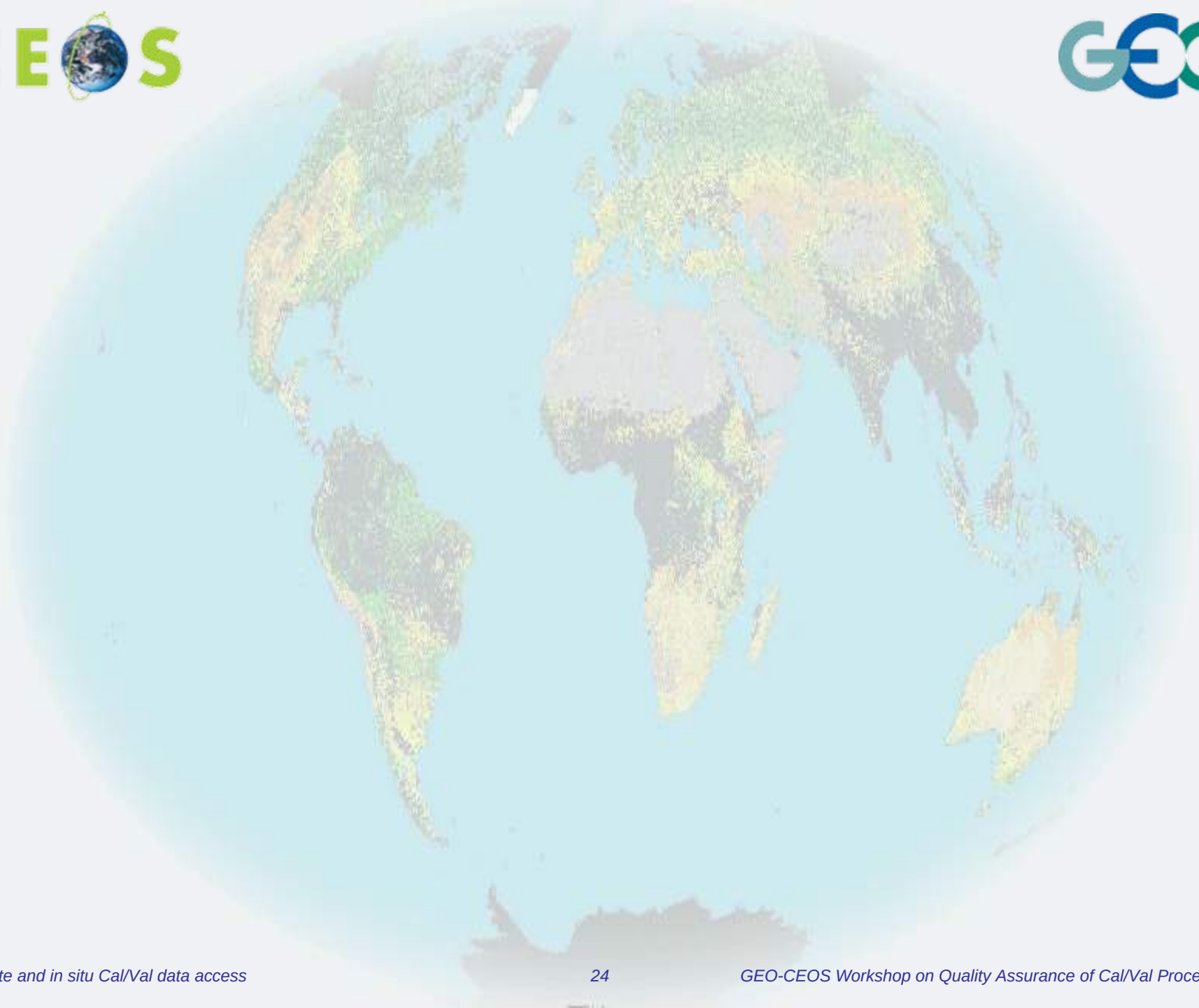
# Archiving: long-term

- ✓ Same requirements as for short-term archiving, but
- ✓ Additional information crucial essential
  - *Need to maintain "soft" calibration and platform/instrument "housekeeping" information*
  - *Need data completeness*
  - *Need interpretation information such as "quality information" and meteorological analyses*
  - *Maintain algorithm/processing documentation, version controls and histories (for satellite and correlative datasets)*

# Archiving: short/long-term

- ✓ An example of data integration is the Aura Validation Data Center (AVDC) or the ESA Cal/Val EO portal test implementation for the CEOS WGCV Infrared and Visible Optical Sensors (IVOS)



# Discussion

# Recommendations

- ✓ Unrestricted cal/val data (“cal/val process”) access - for cal/val purposes only
- ✓ The “user” is understood to be a WGCV and or GEO cal/val “member”
- ✓ Establish a unified “CEOS/GEO” cal/val data (“cal/val process”) exchange principles - *code of good behavior*
- ✓ Implement a CEOS/GEO cal/val support portal