Methodology and Guidelines for Cal/Val

Chair: Nigel Fox (NPL)

CEOS IVOS subgroup chair

Session support; Pierre Femenias (ESA)

Context

• **Product "quality" and the means to achieve it requires little explanation for most of the worlds manufacturing or service sectors. The concept of quality systems and traceability are so embedded that the system works, as it should, transparent to the end user- the consumer.**

This is not the case for EO although the consumer/user often thinks it is!

• **For some applications only localised "resolution" is required, but for long-time base measurements such as those needed for climate, high accuracy & reliability is demanded & can only be achieved through traceability to international stds e.g SI**

• **The vision and necessity of interoperability in order to meet the increasing demands for EO data products whilst in an environment of limited resources is reliant on the ability to robustly combine data from many sources. This can only happen if each data source has associated with it a means to unequivocally ascribe its "fitness for purpose".**

• **The benefits of implementing an international framework to facilitate the demonstration of traceability and product quality will be reaped by all stakeholders. The ability to be able to accept data and services at "face value" will bring both financial and time benefits.**

Overview

To consider a framework for carrying out "calibration and validation" to underpin a strategy for establishing "quantifiable data quality"

- **Consistent use of terminology**
- **Establishing and using "best practise" methodologies/protocols**
- **Demonstrating Traceability and Identification of "standards"**
- **Handling biases**
- **Facilitating the presentation/review of "quality" information**
- **Interoperability/user confidence without "checking"**
- **QA management: ISO, CEOS, Agencies …**

Scope

All activities which contribute to the delivery of an end product derived from an input measurand

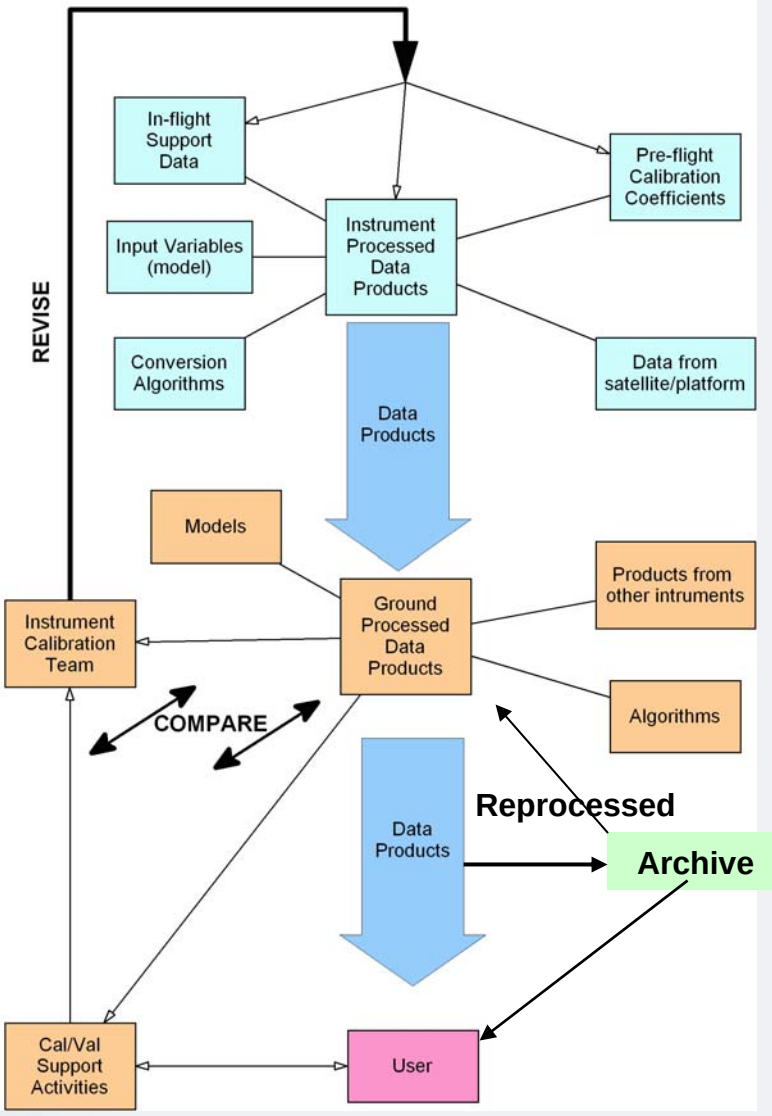
Pre-Flight

- Requirement/Design Specification
- Instrument build: characterisation/calibration
- Data processing: algorithms, ref/support data,

Post-Launch

- Instrument performance
- Output data quality characteristics:
  - accuracy
  - equivalence to others (sensors/in-situ)
- Processing – high level products
- Data distribution/archive …

Collection – Processing – Validation - Delivery



Process

- Session is broken into 2 sections
  - Wrap up at the end
- Focus on answering questions via team process
  - If we have multiple suggestion and solutions, note them for further discussion after all questions are addressed
  - Do not get carried away on insolvable issue
    - Establish a “parking lot” of issues for discussion in wrap up
- Note taker and facilitator

Objectives (Session 1)

- **Agree terminology and consequential implications**
- **Benefit/need for some commonly used methodologies/protocols/ best practise**
- **Identify example methodologies etc that could benefit from being “endorsed”**
- **Will agencies/organisations be prepared/able to share experience**
- **How do we agree what is “best practise” or “acceptable practise”**
- **What information should be contained in describing/defining a methodology / protocol**
- **Is there a benefit in the use of a common format, template, contents list for methodologies/protocols or at least some minimum contents?**
- **What sort of evidence is acceptable to support a result or process?**

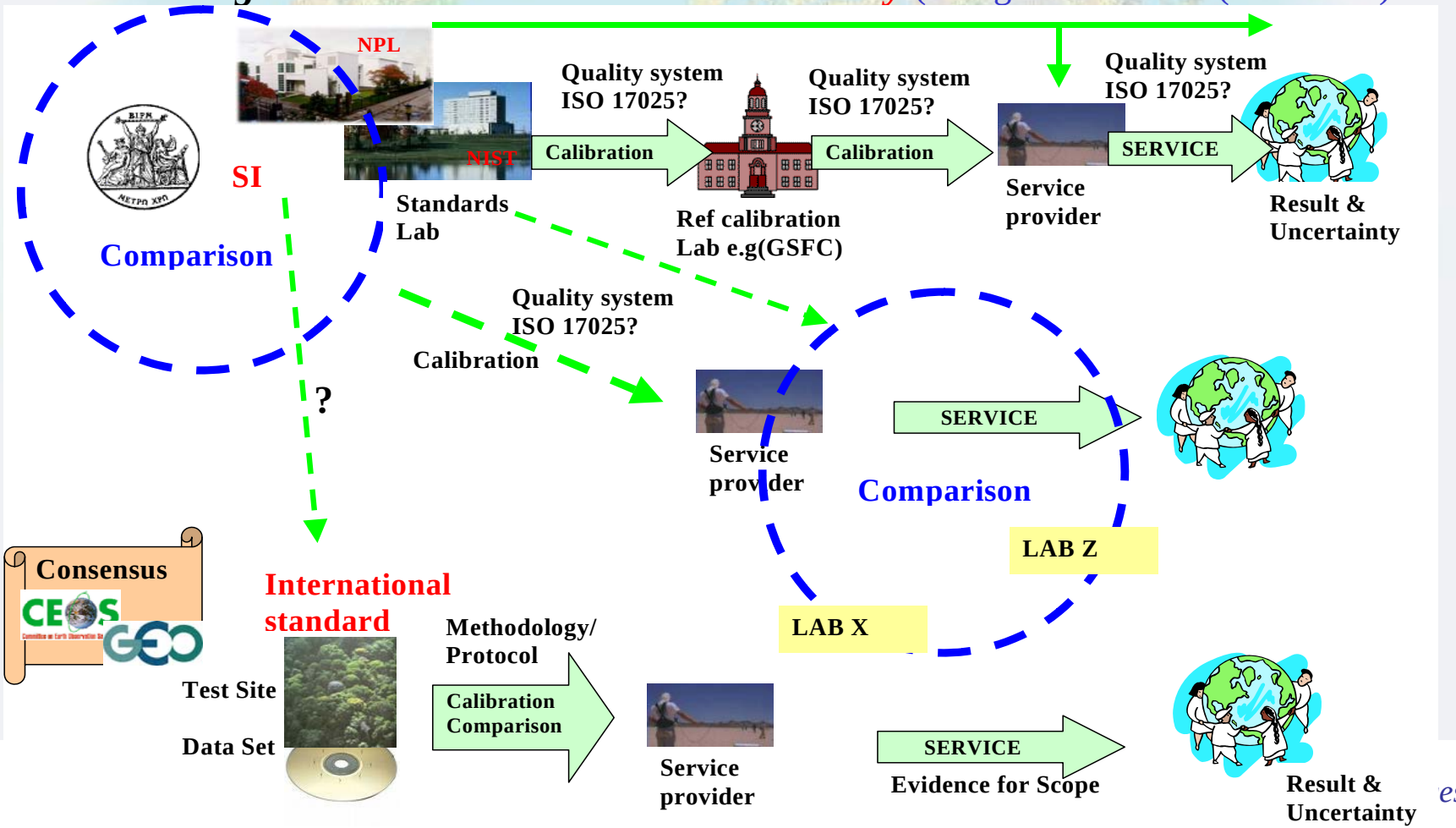
Objectives (Session 2)

- **Types of comparisons, organisation of, visibility of results**
- **Defining and agreeing acceptable “reference standards” to allow “traceability”**
- **How to deal with biases and uncertainties and assign information to a “data product”/result**
- **How to facilitate user confidence, interoperability (without rechecking)**
- **Endorsement of results/processes – need, by whom obtaining international acceptance**
- **Infrastructure/tools to aid “peer review”**
- **Inclusion for all - Training, advice, support for review, comparisons, international coordination**
- **ISO – need for an EO specific standard?**
- **Next steps/roadmap? – timescales, community buy-in, early wins, transition**

Traceability: implications

Requirement: "all EO measurements (data) should be **traceable** to international (SI) standards where appropriate" (GEOSS implementation plan, CEOS plenary 14 and 19, WMO GSICS)

Traceability: ***property of a measurement result relating the result to a stated metrological reference through an unbroken chain of calibrations of a measuring system or comparisons, each contributing to the stated measurement uncertainty*** (ISO guide 99: VIM (2004 draft)



esses

Uncertainties

- A result has no meaning without an associated uncertainty

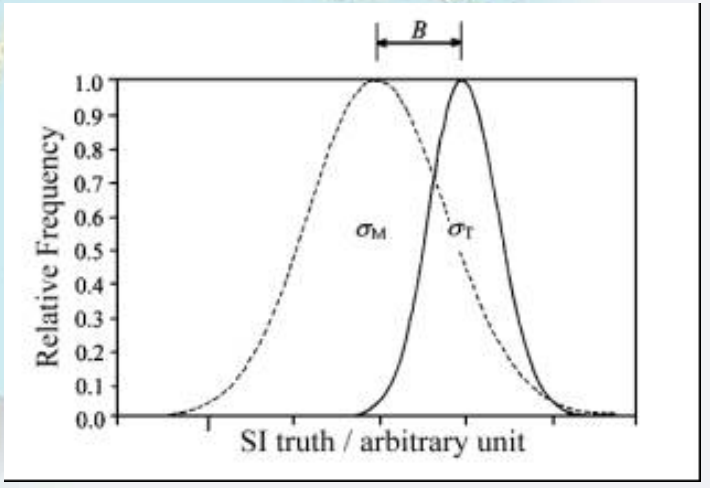
(but must be realistic, not over or under estimated)

- Uncertainties should be evaluated according to ISO GUM

- Establish measurement equation (consider ALL components)
- Separate "statistical" (Type A) and "other" (Type B) uncertainties
- Estimate uncertainty boundaries if necessary
- Make clear statistical confidence level of uncertainty

Uncertainties allow:

- data from different sources to be properly combined
- ability to identify biases
- realistic expectations and appropriate conclusions
- demonstration of understanding of the measurement



Should be expressed unambiguously

- as a range
- +/-
- NOT a single number

Not an assessment of right or wrong or relative merit

Methodologies / Protocols

- **Methodology** – generalised description of an activity

  e.g. use of test sites to evaluate sensor biases

- **Protocol** – prescriptive "recipe" for carrying out an activity

  - e.g. a calibration of an instrument against a reference standard or carrying out a comparison.

  - Agreement on use of a particular reference source e.g solar irradiance spectrum or a specific absorption cross-section

Sharing/Standardising

- Reduces costs overall

- Facilitates comparison of results and leads to improved consistency

- Enables newcomers to participate quickly

- Ideal for "mature" regular/common activities (adoption of best practise)

- Reduces scope for innovation/improvement

- Financial burden on a few

- Not necessarily "traceable" or "true value"

- one off tasks!

Examples
AERONET, (SIMBIOS/IOCCG), NDACC, WMO-GAW, WRR

Possibles: discussion topic

Test site characterisation, Atmospheric correction?, Linearity, cross-calibration,

Methodologies / Protocols common content

Commonality of Content (minimum)/ format (CEOS TEMPLATE?)

- simplifies development and transferability
- ensures a minimum set of key information (quality parameters) a “checklist”
- Enables easy review and development into “best practise”

Content –

Identifier

Description of task (Scope)

- inputs
- outcomes (including performance indicators e.g.accuracy
- means to evaluate results (test plan with identified ref “standards”)

Method to carry out activity (greater detail for protocol)

Requirement to record results and evidence of traceability

Uncertainty table (key components in a template)

How do we agree what general or detailed content should be ?

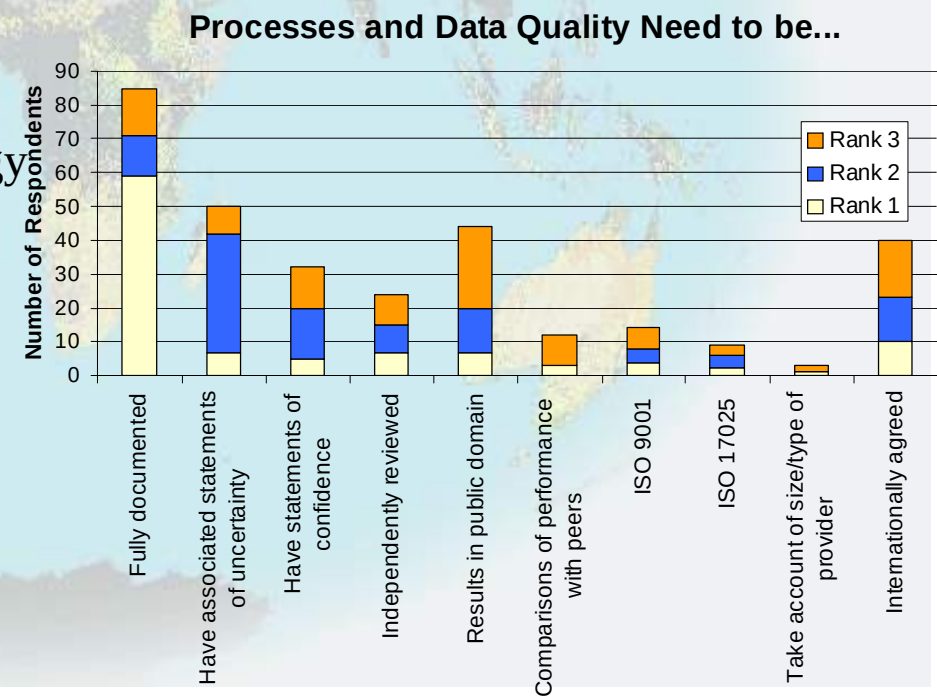
CEOS WGCV? (sub-groups), workshops of experts, mission driven,

Need to review/endorse specific procedures / outcomes?

Evidence of traceability and performance

- **How do we ensure traceability?**

- **What evidence is acceptable to allow automatic acceptance (no need to check)?**
  - Trust (based on experience)
  - Formal accreditation to an appropriate quality system e.g ISO 17025
    (ISO 9001 on its own probably inadequate)
  - Adherence to quality system (providing evidence e.g peer review)
  - Self-declaration (establishing and maintaining evidence "in-house")
  - Participation in a comparison
  - Fully justified uncertainty budgets
  - Following an (agreed) protocol or methodology
  - Evidence of repeatability
  - Evidence of reproducibility
  - Records of testing
  - Validation against "real world"
  - Use of multiple methods



- **Not an evaluation of right or wrong,**
- **Not "policing"**
- **Is a means to demonstrate confidence**
- **Benefits all stakeholders**

Discussion topics (Session 1)

- **Agree terminology and consequential implications**
- **Benefit/need for some commonly used methodologies/protocols/ best practise**
- **Identify example methodologies etc that could benefit from being “endorsed”**
- **Will agencies/organisations be prepared/able to share experience**
- **How do we agree what is “best practise” or “acceptable practise”**
- **What information should be contained in describing/defining a methodology / protocol**
- **Is there a benefit in the use of a common format, template, contents list for methodologies/protocols or at least some minimum contents?**
- **What sort of evidence is acceptable to support a result or process?**

Agreed conclusions of session 1

- Need to have some best practises
- Need to have some common information to document an activity
- Evidence needed to support a claim needs to contain a description and evidence of traceability

Comparisons

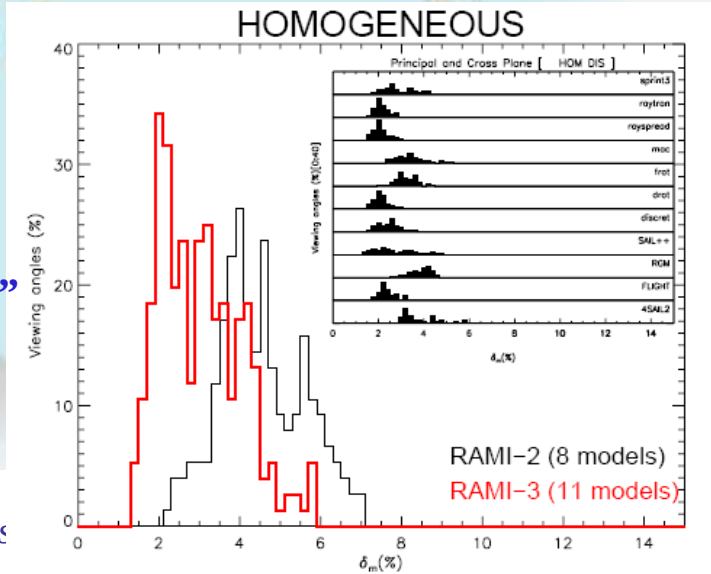
- **Types of comparison**
  - Of equals
  - Against an agreed reference “standard”
  - To obtain consistency or the “true” value
  - Means of evaluating biases or new methods
  - Linking data sets/sensors
  - Blind?

International Pyroheliometer IPC X comparison viewing the sun at Davos

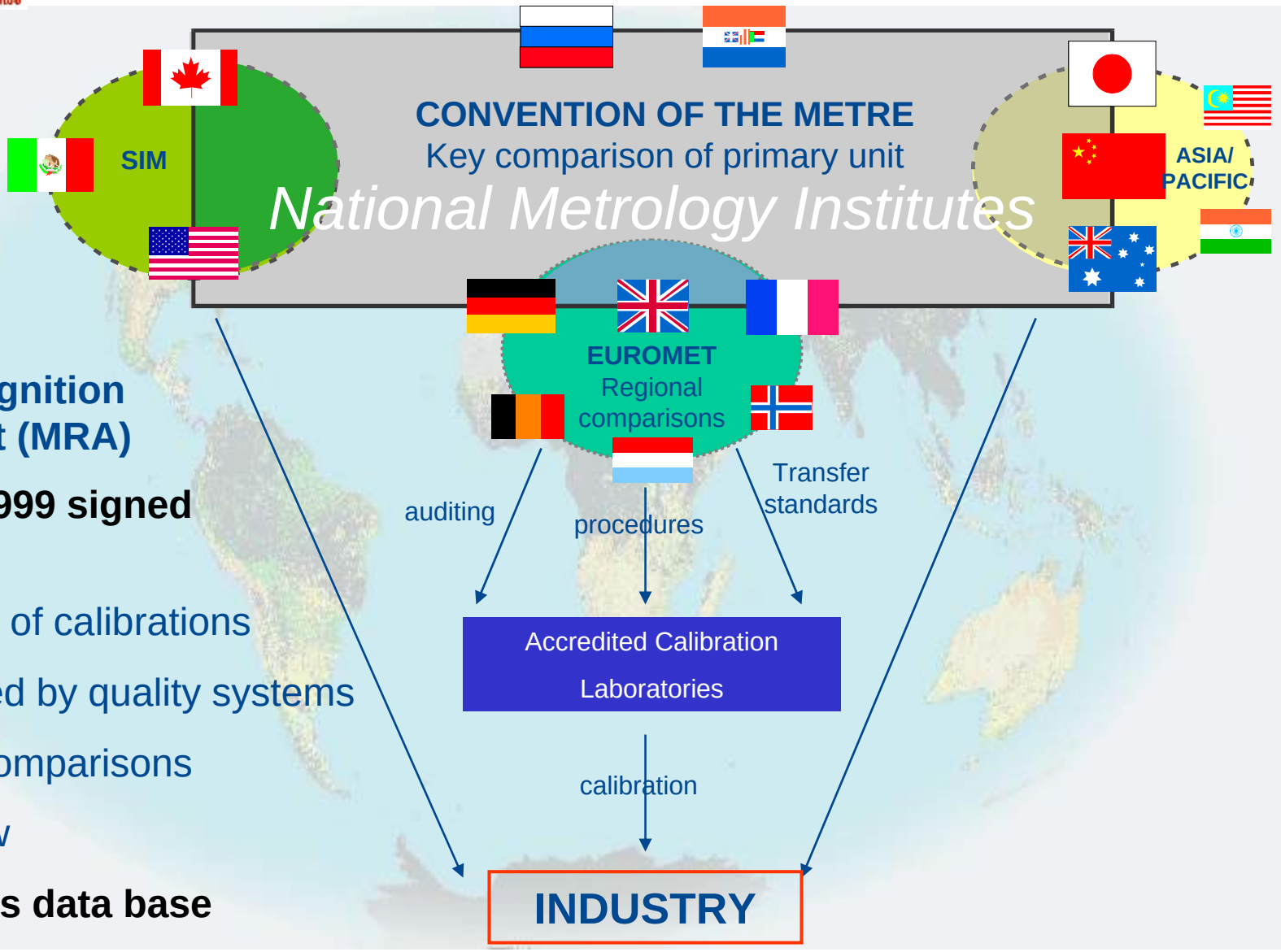
- **Standardised methods of carrying out, reporting and analysing**
  - National standards labs have established “best practise guidance”
  - Unless a-priori a “true” value exists comparisons usually establish a “comparison reference value” or “consensus” derived from the results

    the exact method for this is in debate
- **Types of standard**
  - Test sites (Natural and man-made)
  - Artifacts
  - Data sets
- **Regularity**
- **Inclusion of all**
- **Geographical linking**
- **Costs (organisation / participation)**

JRC use a “comparison average” to compare radiative transfer models in RAMI-3



Traceability to SI and evaluating equivalence



Mutual Recognition Arrangement (MRA)

Initiated in 1999 signed by 44 states

- recognition of calibrations
- underpinned by quality systems
- selective comparisons
- peer review

Public access data base

http://kcdb.bipm.org/

Activities benefiting from international coordination/consensus

- Terminology
- Inter-regional comparisons
- Selecting standards to use for performance (accuracy) evaluation
- Assigning “values” to standards
- Acceptable evidence of performance/traceability
- Format/content of protocol/methodologies
- Providing access to evidence and methods?
- Presentation of results – means to compare/correct
- Identification and promotion of best practise
- Training and education

Facilitating coordination

- **Existing structures of CEOS / GEOSS**

  (Supported by topical workshops of community)

  - **Joint funding of cal/val activities (e.g. Miami SST)**
  - **Development and use of common tools**

    e.g WGISS/WGCV test facility and Web portal
  - **Establish training and education packages**
  - **Establish international reference standards e.g Amazon rain forest**
  - **Seek to establish and encourage "best practise" through exchange of information**
  - **Forum and framework to obtain calibration and validation data from agencies**

- **Issues:**

  - **Many activities are relatively ad-hoc and funding/resourcing difficult**
    - **Needs to meet specific agencies "current need"**
    - **Concept of "long-term sharing" at infrastructure level is critical for GEOSS**
  - **Does not (as yet) address key interoperability requirement: data quality**

Quality Assurance framework to facilitate interoperability

Specification: ***to allow data/models/services with their associated set of "quality parameters" to be readily distributed and used throughout the world without the need for additional checks.***

- **Must encompass all stakeholders: small academic groups to large space agencies**
- **Transparent, rigorous, internationally accepted but not overly bureaucratic**
- **Add value and encourage innovation and improvement**
- **Build on "best practise" of EO and wider community**

Solution: ***Require implementation and adherence to formal ISO quality systems: e.g. ISO 9001 for management and general process and ISO 17025 for specific measurement and testing with others for software etc***

- **Would meet requirements and should be considered acceptable BUT**
  - Difficult to implement for small organisations (and large)
  - Resistance from many (considered bureaucratic, restrictive, too general)
- **Could extract and implement "guiding principles" and key elements tailored to community needs essentially similar to existing examples, e.g. WMO, IOCCG ...**

  (Could initiate writing of an EO specific ISO standard)
  - Document process
  - Collect and **demonstrate evidence** to support "traceability" of quality parameters (uncertainty/reliability of result)

Establishing and presenting evidence

What evidence is required? To be decided by the community

- **comparison at some level is probably a minimum**

Access to evidence, transparency, facilitating review and bias correction

- **Data base/s containing agreed results of comparisons and subsequent endorsement of "service" c.f MRA database of BIPM**
  - linked by cal/val portal, e.g. GECA (ESA), (NILU), Cal/Val test sites USGS), AVDC (NASA)
  - Some form of "peer review" for scopes beyond comparison
- **Data base of "other" evidence: methodologies, protocols followed …..**
  - "peer reviewed"? Accepted "on trust"
  - Details supplied, summary, reference I.d. ?

Recognition/Guarantee of acceptance

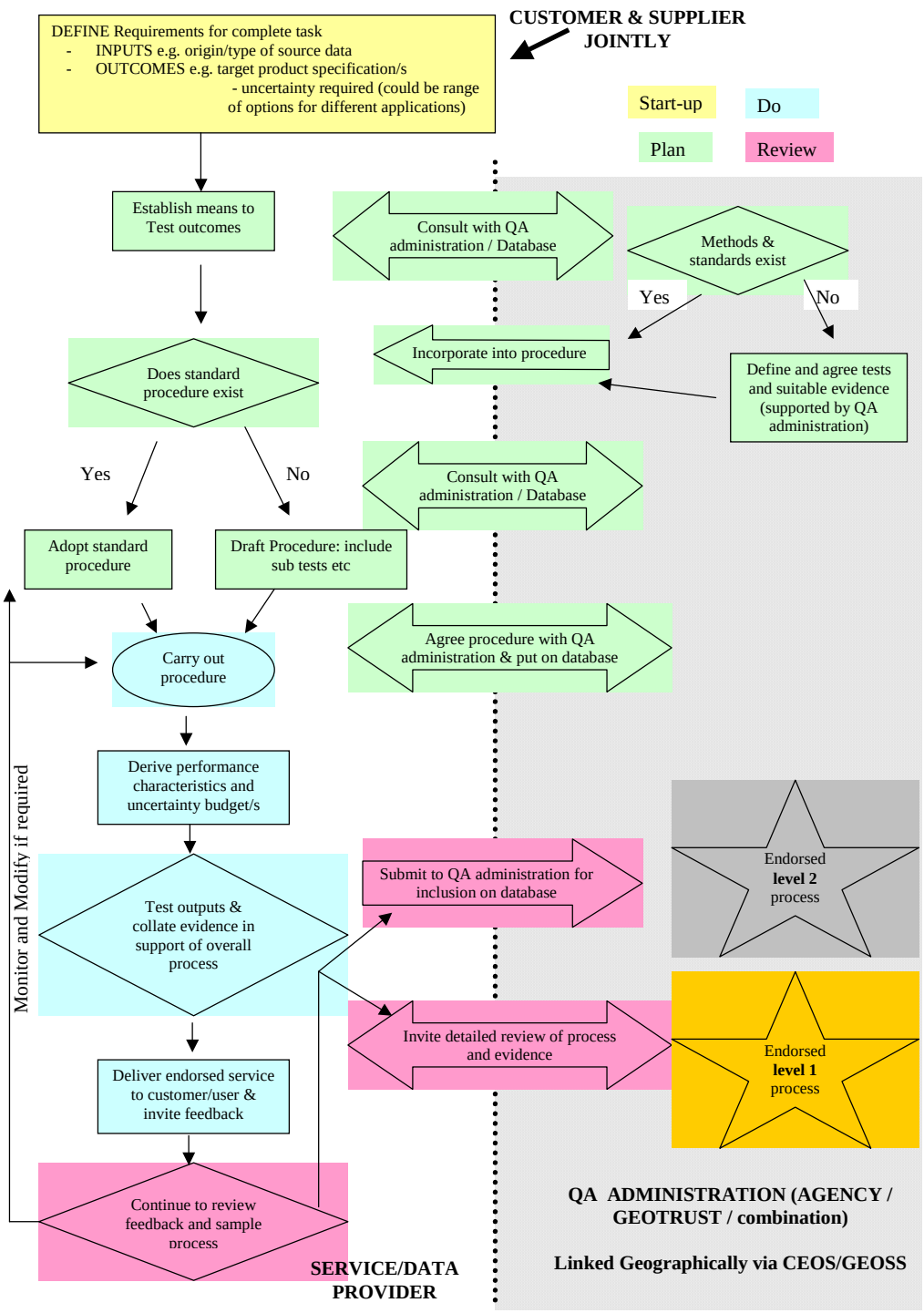
- **Endorsement/"certification"**
  - Based on published scope (data base) supported by agreed evidence
  - Gradings of endorsement based on type of evidence/rigour of review process/confidence
- **Administration**
  - Under auspices of GEO/CEOSS (supported by Std labs, ISO)
  - Through national authorities e.g. space agencies,
  - Through regional authorities e.g. JCSIC (US), ESA+? (Europe) ….
  - Independent of activities

Endorsement classification

REQUIREMENTS FOR ENDORSEMENT	
Endorsement Level 1	**Endorsement Level 2**
- Defined scope (activity/service/product) - Written operational procedures for all processes - Uncertainty breakdown - Evidence of QA of all sub-contractors (level 2 or 1) + equipment used - Review of above by technical experts - Evidence of performance/uncertainty (participation in comparison/calibration of artefact or test data) - Visit by Technical experts (review of end to end process) - Scope and Endorsement classification on data base	- Defined activity - Written operational procedure - Uncertainty breakdown - Evidence of QA of equipment used - Review of above by technical experts - Evidence of performance/uncertainty (participation in comparison/calibration of artefact or test data) - Results and procedures on data base

Implementation

- **Operating principles established and maintained by consensus through CEOS on behalf of GEOSS**
- **Administered within regions by National agencies or grouping of agencies**
  - significant benefits in infrastructure from establishment of "centralised" geographical bodies at least for support
- **Local infrastructure to provide guidance and advice to community on implementation and provide access to the means of obtaining evidence**
- **Peer review is a means to demonstrate "quality parameters" experts need to have degree of independence.**
- **Training and education will be essential**
- **During transition, costs may need to be absorbed by large agencies**



Geographical node implementing and/or supporting national agencies

- Transfer standards
- Comparisons
- Innovation on techniques
- Measurement & test protocols
- International link
- Independence

Modelling & Data processing

QA

Traceability

Audit

Calibration

Validation

Advice

Public sector

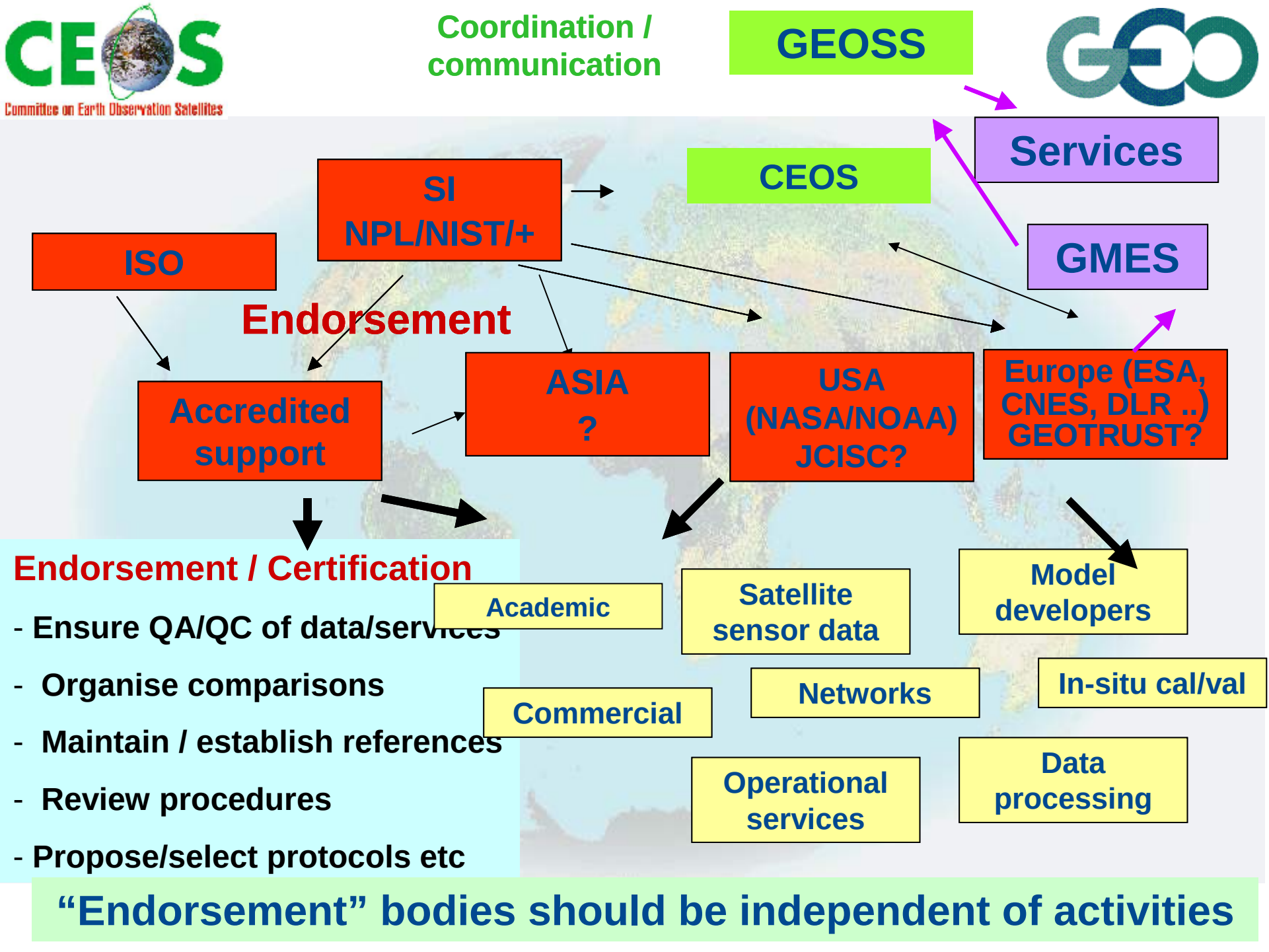
Private Industry

Academia

Manages cal/val databases (policies)

Coordinates interface to other regions

Provides support to national agencies wishing to administer QA process


Coordination / communication
GEOSS
Services
CEOS
SI NPL/NIST/+
ISO
GMES
Endorsement
Accredited support
ASIA ?
USA (NASA/NOAA) JCISC?
Europe (ESA, CNES, DLR ..) GEOTRUST?
Endorsement / Certification
- Ensure QA/QC of data/services
- Organise comparisons
- Maintain / establish references
- Review procedures
- Propose/select protocols etc
Academic
Satellite sensor data
Model developers
Commercial
Networks
In-situ cal/val
Operational services
Data processing
"Endorsement" bodies should be independent of activities

Restrictions, challenges & opportunities

- Language
- Confidentiality / data/procedure access /
- Change - developing momentum
- Concerns regarding process of review (policing) right and wrong
- Diversity (size and type) of stakeholders
- GEO (GMES) presents opportunity
- Climate monitoring is critical driver
- Potential cost savings
- Many new operating agencies
- Tools are under development
- Drive towards commercial operations requires a means to monitor quality and discriminate between data products
- Science data often becomes used in operational situations
    - No reason why science should not want quality data
- ISO (TC20) are ready to work with CEOS/GEOS to develop a new EO specific standard based on community practices and on community agreed “standards”

Discussion topics (Session 2)

- Types of comparisons, organisation of, visibility of results
- Defining and agreeing acceptable “reference standards” to allow “traceability”
- How to deal with biases and uncertainties and assign information to a “data product”/result
- How to facilitate user confidence, interoperability (without rechecking)
- Endorsement of results/processes – need, by whom obtaining international acceptance
- Infrastructure/tools to aid “peer review”
- Inclusion for all - Training, advice, support for review, comparisons, international coordination
- ISO – need for an EO specific standard?
- Next steps/roadmap? – timescales, community buy-in, early wins, transition

Agreed conclusions of session 2