Roadmap towards an Assessment Framework for CEOS-Fiducial Reference Measurements (FRMs)

Version 2.0 – July 2025

Table of Content

1. Introduction

1.1 Purpose of the Document

The purpose of this document is to propose a roadmap towards an assessment framework to endorse a measurement of a specified measurand as a CEOS-Fiducial Reference Measurement (FRM).

1.2 Acronyms and Abbreviations

Acronyms	Full Version
ATBD	Algorithm Theoretical Basis Document
CEOS	Committee on Earth Observation Satellites
CF	Climate and Forecast
CO2	Carbon dioxide
DOI	Digital Object Identifier
EDAP	Earthnet Data Assessment Pilot
ESA	European Space Agency
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAIR	Findable, Accessible, Interoperable, Reusable
FIDUCEO	Fidelity and Uncertainty in Climate Data Records from Earth Observation
FRM	Fiducial Reference Measurement
FRM4SST	Fiducial Reference Measurements for Sea Surface Temperature
FRM4STS	Fiducial Reference Measurements for validation of Surface Temperatures
FTIR	Fourier Transform Infrared Spectroscopy
GEO	Group on Earth Observations
GHG	Greenhouse gas
GUM	Guide to the Expression of Uncertainty in Measurement
HCHO	Formaldehyde
ISFRN	International SST FRM Radiometer Network
ISO	International Organization for Standardization
MM	Maturity Matrix

OC	Ocean Colour
PoC	Point of Contact
QA4EO	Quality Assurance Framework for Earth Observation
RADCALNET	Radiometric Calibration Network
ROI	Return on Investment
SAR	Synthetic Aperture Radar
SARCalNet	SAR Calibration Network
SI	International System of Units
URL	Uniform Resource Locator
VIM	Vocabulary of Metrology
WGCV	Working Group on Calibration and Validation

1.3 References

[1] Goryl, P.; Fox, N.; Donlon, C.; Castracane, P. Fiducial Reference Measurements (FRMs): What Are They? Remote Sens. 2023, 15, 5017. https://doi.org/10.3390/rs15205017
[2] QA4EO website: http://qa4eo.org/
[3] SITSat webpage on CEOS Cal/Val portal: https://calvalportal.ceos.org/sitsat
[4] CEOS Cal/Val Portal: https://calvalportal.ceos.org/
[5] EDAP project website: https://earth.esa.int/eogateway/activities/edap
[6] QA4EO Learners - https://qa4eo.org/users/learners/
[7] QA4EO Tools - https://qa4eo.org/tools.php
[8] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[9] CF Metadata Convention: https://cfconventions.org/
[10] The EU's infrastructure for spatial information (Inspire): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:l28195
[11] Guide to the Expression of Uncertainty in Measurement (GUM): (https://www.bipm.org/documents/20126/194484570/JCGM_GUM-1/74e7aa56-2403-7037-f975-cd6b555b80e6)
[12] QA4EO Training https://qa4eo.org/training.php
[13] QA4ECV guidance: https://www.copernicus.eu/en/quality-assurance-essential-climate-variables
[14] Fiduceo Project: https://www.fiduceo.eu/
[15] FRM4STS Project: http://www.frm4sts.org/

2. Background and Objectives

The need for post-launch calibration and validation (hereafter Cal/Val) of satellite sensors and derived data products is well-established. For the primary Level-1 observations, post-launch calibration against independent references external to the spacecraft ensures that unforeseen changes, and changes that cannot be corrected by any on-board calibration systems, are accounted for. Post-launch validation supports the accurate delivery of Level-1 and higher-level data products, confirming expected performance and including processors and any retrieval algorithms. However, the cost, complexity and robustness of establishing mission specific exercises have led to space agency efforts to improve coordination and generalisation of methods and infrastructure, through bodies like CEOS. This coordination not only reduces cost but also leads to improved interoperability between sensors.

In parallel, driven in part by the critical needs of climate and environmental monitoring and research, there has been a significant move towards a more coordinated and comprehensive assessment and reporting of the quality, biases and residual uncertainty in the observations made by different satellite sensors. This led to the endorsement by the Group on Earth Observations (GEO) of a CEOS led community initiative to create a Quality Assurance Framework for Earth Observation (QA4EO). QA4EO provides a set of principles and associated guidance to encourage the provision of internationally consistent quality indicators on their delivered data. More recently this need has grown with the emergence of new commercial satellite operators and a desire to ensure that users have a transparent and fair means to judge the adequacy of data products for their needs.

One key element of QA4EO relates to post-launch Cal/Val of satellite data through an independently derived dataset that can be correlated to that of the satellite. This independent dataset fit for satellite Cal/Val purposes refers to the concept of Fiducial Reference Measurement (FRM). Of course, for this process to be robust, the independent dataset itself must be fully characterised in a manner consistent with the QA4EO process, i.e., it has documented evidence of its level of consistency with a suitable reference (nominal "truth") within an associated uncertainty. The uncertainty associated with the comparison process itself, which almost always involves a transformation of the FRM measurand to be comparable/representative of that observed by the satellite, must also be characterised and be sufficiently small in comparison to the uncertainty of the satellite data. In practice, the "reference" should be agreed by the community and ideally tied to SI. Ideally satellites need to be validated under a range of potential operating conditions and geographical locations to account for potential effects due to orbital location and/or atmospheric or environmental conditions. They also need to span the full range of values of the parameter that the satellite sensor is likely to measure and of the influencing parameters not measured by the satellite but influencing the measurand and its associated uncertainty. This generally means that multiple geographically distributed and temporally sampled FRMs are likely to be required.

Satellite Cal/Val has thus traditionally relied on a wide range of datasets collated by different teams and based on, for example, in situ and aircraft campaigns, (semi)-permanently deployed ground-based instrumentation, comparisons with measurements from other satellite sensors, opportunistic measurements (e.g., from commercial shipping and aviation) and modelled natural phenomena (e.g., Rayleigh scattering of the atmosphere over dark ocean surfaces). These datasets have differing levels of adherence to the critical QA4EO principles, and while all provide information, it is difficult to distinguish high quality satellite optimised datasets from those that need more careful treatment. Consistency between datasets for comparison with a single sensor is not sufficient to ensure the quality of all comparisons.

In an attempt to create a more robust validation framework and encourage, where appropriate, teams to specifically tailor their observations to address the needs of satellites and to help funding bodies distinguish and allocate resources, ESA coined the concept and label of Fiducial Reference Measurements, or FRM. In creating the FRM label and associated criteria, see below, the objective has been to create a distinct class of observations that are optimised to meet the needs of satellite Cal/Val. In particular, to have some authority to use the label in a controlled manner we have here added the additional prefix CEOS-FRM. This is not to say that observations not classified as CEOS-FRM are of inferior quality for their intended use, only that their data quality has not been demonstrated to meet CEOS-FRM requirements and in particular their suitability for a particular class of satellite observations. Henceforth we now specifically intend to encourage the use of CEOS-FRM as the label in an analogous manner to CARD4L (CEOS ARD 4 Land) distinguishes it from other potential uses of the ARD label. It should be noted that an FRM acquisition system can be a single measurement system or the combined output of a set of instruments which may be geographically distributed and formally coordinated as a network. If the latter, individual measurements can and should be classed as FRMs in their own right and the output of the network or indeed group of FRM's can also be classified as an FRM network or an FRM dataset.

In Europe, several projects funded by European organisations, EU, ESA, EUMETSAT have evaluated how to establish and evidence FRM quality observations for a wide range of applications spanning surface temperature and reflectance, atmospheric composition, sea and ice level height etc. In many cases this has also led to networks of stations around the globe adhering to the FRM principles. The FRM label is thus now becoming recognised internationally. Many data providers are seeking guidance and formal recognition of their FRM compliance and satellite operators are requesting the validation of data from an FRM compliant source.

The CEOS Working Group on Calibration and Validation (WGCV) has recognised the need to establish a means to enable Cal/Val data providers to robustly evidence that they are CEOS-FRM compliant or at least their progress towards being fully CEOS-FRM compliant. This will allow satellite operators and data users to assess and weigh the Cal/Val data used to validate the performance of a particular satellite product. Recognising that the number of potential CEOS-FRM data providers may become very large, it is necessary that any "endorsement" process be efficient low cost and maintain integrity. The process should also be flexible to address different levels of maturity in the methodologies and products that are being validated as well as technology domains.

3. CEOS-FRM Definition and Principles

CEOS- Fiducial Reference Measurements (FRMs) [1] are independent, fully characterised, and traceable (to a community agreed reference, ideally SI) measurements, tailored specifically to address the calibration/validation needs of satellite borne instruments making measurements of a particular measurand, that follow the guidelines outlined by the GEO/CEOS Quality Assurance framework for Earth Observation (QA4EO) [2]. In many cases, these may be a subset of "in-situ" measurements, at a specific geo-location, both individually and/or as part of a network or campaign, including those from airborne or seaborne platforms or even from another satellite, providing it meets the CEOS-FRM criteria. It should be noted here that the required characteristics of an FRM may also depend on the nature of the satellite instruments sampling method as well as the measurand itself for example a Limb sounding atmospheric composition sensor and one measuring at Nadir will have differing requirements even though the satellite instrument itself may be very similar. The newly emerging SITSats (SI Traceable Satellites) [3] are a good example of a potential satellite-based CEOS-FRM. These CEOS-FRM provide the maximum Return On Investment (ROI) for a satellite mission by providing the required confidence in data products, in the form of independent validation results and satellite measurement uncertainty estimation, over the entire duration of a satellite mission.

The mandatory defining characteristics for CEOS-FRM are:

1. Traceability - FRM should have documented evidence of their traceability (bias and associated uncertainty) to a community agreed reference, ideally tied to the International System of units, SI, e.g. (via a comparison “round robin” or other) with peers and/or a metrology institute together with regular pre and post deployment calibration of instruments. This should be carried out using SI-traceable “metrology” standards and/or community recognised best practices, for both instrumentation and observations.
2. Independence of satellite under test - FRM are independent from the satellite (under-comparison) geophysical retrieval process.
3. Uncertainty budget - A comprehensive uncertainty budget for all instruments used in deriving FRM of a particular measurand, including any transformation of the measurand to match that of the satellite product, is available and maintained.
4. Documented protocols - FRM protocols, procedures and community-wide quality management practices (measurement, processing, archive, documents, etc.) are defined, published and adhered to by FRM instrument deployments and usage.
5. Accessibility - FRM data, including metadata and documentation of processing, are accessible to other researchers allowing independent verification of processing systems. All data and information should be made available in a timely manner and in a form that is readily utilisable by a satellite operator.
6. Representativeness of the measurement - FRM data allow the determination of the on-orbit uncertainty characteristics of satellite geophysical measurements via independent validation activities. It thus requires that the degree of representativeness of the FRM to that of the satellite observation and/or associated retrieval (how well the FRM looks like the satellite measurement?) as well as the satellite to FRM comparison process needs to be documented and the uncertainty assessed. Note for any individual satellite instrument the exact sampling and elements of the comparison process may differ, even within a generic instrument/satellite class, but the documentation and evidence to support the uncertainty analysis must be presented in a manner that can be readily interpreted by a user.
7. Adequacy of uncertainty - The uncertainty of the FRM measurements, including the comparison process, must be commensurate with the requirements of the class of satellite/instrument/measurand they are specified to support.
8. Utility (return on investment) - FRM data are designed to apply to a class of satellite missions (several). They should not be mission specific.
9. It should be noted that in using any CEOS-FRM data it is expected that the user provides a clear acknowledgement of the contribution of the FRM owner/provider in any reporting of validation results, verbal and written as well as to CEOS for the FRM QA framework. Each FRM provider may have their own data policy and required wording for acknowledgement and should be consulted individually if this is not on their documentation.

4. CEOS-FRM Endorsement Process

It is critical that the characteristics and performance of any FRM can be unequivocally considered consistent and trustworthy (within their self-declared metrics) and readily interpretable so that any information derived from it can be reliably utilised by a user. This is not to say that FRM for any particular measurand/application must all meet the same level of performance or indeed necessarily fully meet all the FRM criteria (at least whilst developing a capability) but that any potential user can readily identify and assess suitability for their application. For the FRM classification to have meaning, requires not only a set of common criteria but also obliges some form of independent assessment against them, or at least the potential for such assessment. This could be achieved through the creation of a formal process implemented through a body such as ISO where independent assessment/audit against a set of criteria is undertaken by an approved body funded by the entity seeking accreditation.

Whilst aspects of what is entailed within meeting FRM compliance could benefit from and indeed may be most easily achieved through the use of services adhering to ISO standards, e.g. ISO 17025 for testing and calibration laboratories, it is not anticipated that the formality and inherent cost associated with such an endorsement process will be necessary or even practical at this time.

This framework takes a more pragmatic approach relying on self-assessment and transparency/accessibility of evidence against a set of criteria which are subject to peer review through a board of experts led by CEOS WGCV. The self-assessment matrix templates and associated documentation will be stored and made available through a searchable on-line catalogue accessible through the CEOS cal/val portal [4]. Following submission an FRM quality matrix will remain invisible to the public until the opportunity for a "peer review" validation has been completed. The latter, obliged to be completed within a limited timescale.

In order to be flexible, maximise inclusivity and encourage the development and evolution of CEOS-FRM from new or existing teams, compliance with criteria will be based on a gradation scaling rather than a simple pass/fail. The degree of compliance and associated gradation can then be presented in a Maturity Matrix model - EDAP like [5] to allow intended users of the CEOS-FRM to assess suitability for their particular application and indeed funders to decide on where and what aspects to focus any investment. The matrix model provides a visual 'simple' assessment of the state of any CEOS-FRM for all given criteria making visible where it is mature and where evolution and effort needs to be expended.

In addition to this broad-based summary, an overall classification of the degree of compliance will be provided based on meeting specific gradations for particular criteria. At present, there are four classifications:

Class A – Where the FRM fully meets all the criteria necessary to be considered an FRM for a particular class of sensor.

Class B – Where the FRM meets many of the key criteria and has a path towards meeting the Class A status in the near term.

Class C – Meets or has some clear path towards achieving the criteria needed to reach a higher class and provides some clear value to the validation of a class of satellite sensors.

Class D - Is a relatively basic adherence to the FRM criteria but where the owners/developers have a strategy and aspiration to progress towards a higher class. This can be considered an entry level class for those starting out on developing an FRM.

The detailed specifications for these classes can be found in section 6.

5. CEOS-FRM Maturity Matrix

5.1 Introduction

The CEOS-FRM Maturity Matrix (MM) provides a high-level colour coded summary of the characteristics of the FRM measurement under analysis against specific criteria. Although the criteria are intended to be generic, how they are interpreted and assessed in detail may vary with the nature of the FRM and sensor class and as a result may require some level of domain expertise to peer review the self-assessed criteria. It is intended that this same concept can be used for an individual measurement/method, the implementation of a method at a geo-located site e.g., a test site as well as a network of such similar sites. The matrix contains a column for each category/criterion and cells for each aspect (subsection of category). Grades are indicated by the colour of the respective grid cell, which are defined in the Key (the grade criteria). The FRM Maturity Matrix can be further subdivided into two parts as follows:

- *Self-assessment CEOS-FRM Maturity Matrix*
- *Verification CEOS-FRM Maturity Matrix (completed by an independent review process not the self-assessment team)*

These matrices are described below:

5.2 CEOS-FRM Description

As a precursor to any CEOS-FRM assessment there has to be a clear description of what the FRM is intended for, what class of satellite instrument and associated measurand is being validated, and who is responsible for the site(s)/instruments(s)/measurements and associated claims. The maturity matrix itself is only a means to facilitate the provision and interpretation of the evidence needed to enable endorsement of the FRM (site, network, method) in relation to the criteria defined in section 3 and the capabilities claimed in the initial declaration.

The following template will form part of an online submission process, with drop down menus and will ultimately allow a searchable catalogue accessible through the CEOS Cal/Val portal. It is intended that the same template be used for a single site/method or that of a coordinated network. It may even be that in some cases individual contributors to a network may additionally provide their own independent FRM. However, it should be noted that in some cases where it is considered essential to validate a wide dynamic range an individual FRM may score differently than when it is considered as part of a network. To reflect some differences between networks or multiple measurement locations we have a dedicated column for specific characteristics. This column should not be completed by individual FRM providers even if they also form part of a network and its content and associated criteria are not included in the overall FRM classification process.

1. **FRM measurand (FRM4?):** What is the FRM measurand? e.g. surface reflectance, Total Column CO2, Land surface Temperature etc. (an initial set of options will be provided plus freeform additions). This should include some information on the range of values for the measurand that the FRM will cover.
2. **FRM for what “class” or classes of Instrument:** Very-high resolution imager, medium resolution imager, Lidar, Atmospheric spectrometer etc. (an initial set of classes will be defined) and where appropriate the characteristics of the observation e.g., at nadir, limb sounding etc. as a sub-class with options for free-flow input if needed.

3. **Temporal and spatial nature of FRM data collection:**
    a. near continuous or regular sampling from a fixed location, a network of near continuous sampling "sites", instrument/method "campaign" based;
    b. from surface-based sensor, airborne, space, autonomous, ground operator;
    c. localised 'point-like' sample-based, integrated/averaged e.g. '' 'line-of-sight' volume for some atmospheric composition remote sensing.
4. **Best achievable uncertainties:** What uncertainty can be achieved for the measurand for the defined class of sensor (including spatial, temporal and vertical representativeness of a measurement for the class of instrument/satellite but not satellite-specific uncertainties)?
5. **FRM owner/operator contact details:** Means to communicate with those responsible for all the information relating to the FRM.
6. **Access to FRM data**: URL (or other) means to obtain FRM data and documentary evidence of FRM characteristics, ideally following FAIR principles.
7. **Approximate start of FRM "like" operations:** When did measurements of this type begin, how long has site existed, has the team been collecting measurements etc., even if not fully FRM compliant?

5.3 Maturity Matrix: Assessment framework

This subsection describes the assessment framework and evidence needed to achieve specific classifications for the various characteristics. It should be noted that in completing the self-assessment for many of the categories there will always be a need for some domain and application specificity in terms of the grade selection and the relative importance in making a grade selection when there may be more than one simple criterion to be considered. It is intended that the FRM provider makes their judgement based on their understanding and interpretation of the category in relation to their specific domain/application and make comments accordingly to support the decision with associated evidence. The independent review will consider the reasonableness of this self-assessment and indeed any comments will provide transparency to any subsequent user of the FRM on the choice of balance and relative importance.

Self-Assessment						Independent Assessor
Nature of FRM	FRM Instrumentation	Operations/ Sampling (single instrument)	Data	Metrology	Completeness, coverage and distribution	Verification
Descriptor	Instrument documentation	Automation level	Data completeness	Uncertainty characterisation	Validation capacity	Guidelines adherence
Location/ availability of FRM	Evidence of traceable calibration	Measurand sampling/ representativeness	Availability and Usability	Traceability documentation	Geographical coverage Temporal sampling	Utilisation/ Feedback
Range of sensors	Maintenance plan	ATBDs on processing/software	Data format	Comparison/ calibration of FRM	Centralized data, processing, quality assessment and adherence to community standards	Metrology verification
Complementary observations	Operator expertise	Guidelines on transformation to satellite pixel	Ancillary data	Adequacy for intended class of instrument/ measurand	Timeliness	Independent verification
					FRM CLASSIFICATION	**A B C D** (to be selected)

Table 1. Maturity Matrix for assessing as CEOS-FRM one measurand for one site or network (and its color-coded- grades)

Grade
Not Assessed
Not Assessable
Basic
Good
Excellent
Ideal
Not Public

The self-assessment matrix is provided by the FRM "owner" and is derived from preparing and/or collating the documentation and evidence needed to support the nature of the declared FRM and its adherence to the CEOS-specified characteristics, Section 2. Note the matrix provides a visualisation of the FRM owners self-assessment and interpretation of what is required which can additionally be clarified through discussion with CEOS experts. Whilst demonstrably achieving green status for all criteria would likely guarantee Class A overall status, it is not necessarily a prerequisite, and indeed evidence of progress towards a higher grade in some categories may well still be sufficient. It should of course be assumed that the classification is progressive and that in meeting a higher overall classification all the elements of the previous classifications are met.

The matrix is intended to help guide an FRM owner/developer towards understanding what is desired and indeed help provide clarity on interpretation. It is also intended to provide a "quick look" "fitness for purpose" assessment for a potential user. It includes the following categories:

- *Nature of FRM*
- *FRM Instrumentation*
- *Operations/sampling*
- *Data*
- *Metrology*
- Completeness, coverage and distribution

Note classifications such as "not public" means that the information may be available but the FRM owner is not able or willing to make it public, but may do so in a controlled way as part of a CEOS "peer review". Similarly, "not assessable" means there is either insufficient information to allow the category to be assessed adequately to meet any of the technical classifications, generally, it might come with some comment such as "in progress" or "under-development". "Not assessed" indicates that this has not been done for some other reason and may be that for the nature of the FRMs intended purpose is outside of the scope of study or not considered necessary or essential.

The independent assessor column provides for a "peer review" of these self-assessments and indeed ultimately a confidence check for users. It is anticipated that if the independent assessment does not comply with that of the self-assessment an updated submission will be undertaken following dialogue between the peer review team and the FRM provider.

The following subsections detail each category of the matrix in turn with indications of what is required to achieve each grade in the schema. It is assumed that for each category in the matrix, the FRM owner provides a short text descriptor to justify their claim within the box together with any appropriate documentary evidence (or link to it) which would then be stored within the FRM database/catalogue to support any verification review process. Any commercially sensitive or proprietary information that is considered by the FRM owner to remain non-public must be labelled as such clearly and the reasoning justified.

5.3.1 Nature of FRM

Completeness of the general information relating to the nature of the FRM and its basic suitability for the class of instruments/measurand it is intended to be supporting.

Descriptor

Relates to the completeness and adequacy of the information provided by or on behalf of the person, PoC (Point of Contact), who is responsible for the FRM: name, affiliation and email address and all the associated information in the template, Section 5.2, relating to the FRM characteristics. Table 2 shows the assessment criteria for the "Descriptor".

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	All critical information available but incomplete or inaccessible evidence.
Good	Information provided but some evidence would need to be requested.
Excellent	As Ideal but without a comprehensive dedicated website.
Ideal	A complete comprehensive template and an FRM website where all information is clearly and readily available.

Table 2 Nature of FRM - Descriptor - Assessment Criteria

Location/Availability of the FRM

Adequacy of location/availability and evidence made available by the FRM owner of any FRM observation for the specified instrument class/measurand and associated required uncertainty:

- Is the FRM data readily and regularly observable by satellite sensors (e.g., does the FRM location come from generally clear skies? complicated rapidly changing atmospheric conditions, environment etc.).
- Are there any observational complications associated with the location of the FRM e.g., adjacency effects Area/volume of FRM observation to that of the satellite sensor.
- If a network, is the distribution of locations comprehensive (e.g., full range of measurand values, northern and southern hemisphere, full range of major influence quantities like the surface albedo and the aerosol load…) (note this is also specifically addressed in the network column)

Table 3 collects the grades to assess "Location/Availability of the FRM" and should consider the above bullets in its assessment.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Providing occasional but necessarily readily plannable opportunities.
Good	Providing regular FRM data but only for a limited period within a year perhaps only a single season. For a network some aspects of the dynamic range and performance influences are covered for a number of seasons
Excellent	Temporal coverage would typically be at least bi-monthly on average for any one satellite instrument. For a network this should ensure that the most critical aspects of dynamic range and performance influences are assessable. The exact limitations of this should be indicated in the “discovery” descriptions (section 5.2) associated with this FRM.
Ideal	FRM could provide all year coverage of a number of different satellite instruments needs on at least a monthly basis for the same instrument. For a network this should ensure that full dynamic range of the instruments measurements of a measurand are validated and that this is also done in a manner to assess other satellite performance influences e.g. latitude, environmental influences e.g. albedo, atmospheric composition, temperature etc.

Table 3 Nature of FRM - Location/Availability of the FRM - Assessment Criteria

Range of Sensors

How broad a range of different instruments/satellites in a particular generic class e.g. high resolution land imager and ideally number of classes can be served by the FRM, accounting for the characteristics of the FRM as declared in the discovery section (section 5.2) , and the degree/complexity of tailoring needed to meet any specific individual satellite instrument. For example, an FRM providing level 1 radiometric gain in the Vis/SWIR domain may be able to provide services to several classes of instrument: high and low resolution land imagers, atmospheric composition spectrometers, GHG instruments within a given uncertainty criteria and this is a category that is more flag of interest to potential users and not really a scoring criterium.

Table 4 shows how the assessment framework grades the “Range of sensors” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	It would likely only have evidenced suitability for one of two instruments/satellites and relatively limited guidance on how to utilise the FRM.
Good	There would be a large number of instruments in a particular class and some effort to provide information and guidance needed to undertake comparison.
Excellent	There would be most instruments in a particular instrument class and at least some in as a minimum a related class differing perhaps by spatial or spectral range and all information on how to carry out comparison.
Ideal	There would be at least two classes of instrument type with different intended applications e.g. land imaging, atmospheric and most if not all instruments within a class including all information on how they can compare to the FRM.

Table 4 Nature of FRM - Range of Sensors - Assessment Criteria

Complementary Observations

In addition to the specified FRM there are other complimentary observations made at the same time/location for example another FRM or other dataset that may be of value to the satellite observation. For example, a land site that measures Atmospheric parameters, where the atmospheric parameters e.g., AOD or water vapour or temperature may be of value to other missions. The following Table 5 describes how the assessment framework grades the "Complementary Observation".

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Some ancillary data being collected.
Good	Full range of ancillary data needed to support the FRM measurement being made at location.
Excellent	one or more additional observations being made but which are not yet of FRM quality
Ideal	one or more additional FRM quality observations related to the satellite instruments intended measurand for example it could be several atmospheric species, or different surface properties (temp & reflectance)

Table 5 Nature of FRM - Complementary Observations - Assessment Criteria

5.3.2 FRM Instrumentation

This column relates to instrumentation used to establish an FRM, both directly and indirectly for ancillary type observations.

Instrument Documentation

This concerns the availability of identification serial numbers/description of operation and the corresponding documentation in terms of technical manuals for hardware and software nominal measurement operations including:

- **Technical Manuals (Hardware)-** Technical documentation provided by the manufacturer regarding the physical components (mechanical. electronic, etc) of the instrument in all its parts. Usually includes conditions for first measurement settings or calibration, conditions and ranges for nominal operation, as well as basic principles for maintenance.
- **Technical Manuals (Software) -**Technical documentation provided by the manufacturer regarding the software for operating the instrument. Usually includes modules for instrument initialization (e.g., parameter settings, calibration), for nominal operation and routines for basic quality check and warning/alarm for operational malfunction.

In the FRM instrument category, Table 6 describes the assessment criteria for the “Instrument Documentation” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Type of instrument used indicated but limited details/evidence of performance etc
Good	Outline descriptions of instrumentation used and performances but without detailed manuals or ATBDs of software processing steps etc
Excellent	Detailed descriptions of instrumentation, indicative performance (potentially based on manufacturer info)
Ideal	Full range of documentation available describing all instrumentation being used to make the FRM including all ancillary data, their actual performance (not simply manufacturer specified) etc

Table 6 FRM Instrumentation - Instrument documentation - Assessment Criteria

Evidence of Traceable Calibration of FRM Instrumentation

Documentation demonstrating traceable calibration of all appropriate FRM instrumentation, indicating achieved performances and detailed uncertainty budgets. This needs to be evidenced for the location of use of the instrumentation and so may involve pre-deployment laboratory calibrations to an SI standard and some means to show this is valid when taking measurements e.g., comparison to another instrument or transfer standard etc ideally under operational conditions. Traceability requires that there is some independently achieved evidence that the performance to SI or appropriate community reference standard/procedure, is as specified and not simply that a "traceable" calibration artifact has been used or a calibration performed without showing that its use and/or validity is justified for the specific FRM observation. Ideally, an uncertainty budget for the full measurement process of the instrument should be made available together with clear operational ranges over where it is valid. For guidance on how to assess uncertainty including tools to aid its calculation see: QA4EO Learners [6], and QA4EO Tools [7].

Table 7 shows how the assessment framework grades the "Evidence of traceable calibration of FRM instrumentation".

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Evidence of traceability and performance limited potentially to a pre-deployment calibration or manufacturers specification.
Good	Evidence of traceability available together with uncertainty budget but not necessarily independently reviewed or compared.
Excellent	Adequate documentation to make clear the degree of traceability and associated uncertainty although comparison of peers under operational conditions not necessarily undertaken.
Ideal	Fully documented evidence of route of traceability and associated uncertainties (full breakdown including correlations) from the use of the instrument to make a measurement in support of FRM at location of operational use, back to its link to an SI or community agreed reference. This should be presented following the practises indicated by FIDUCEO, and available from the QA4EO website (or similar). This should be evidenced by an independent comparison of performance against as a minimum peers under full range of operational conditions of the instrument. Ideally this would all be carried out following equivalent to ISO 17025.

Table 7 FRM Instrumentation - Evidence of traceable calibration of FRM instrumentation -Assessment Criteria

QA/Maintenance

In the FRM instrument category, Table 8 describes the assessment criteria for the “QA/Maintenance” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	A maintenance plan and operation guidance documents but dependent on training rather than necessarily following documented procedures.
Good	Some level of documentation on operational procedures and maintenance but not necessarily following community best practices.
Excellent	Documented procedures on measurements and instrument characterisation etc together with maintenance plan but not necessarily independently audited but following community best practises.
Ideal	Full ISO 9001 equivalent operations e.g., full documented operational procedures/ maintenance schedule, spare instrumentation etc.

Table 8 FRM Instrumentation - QA/Maintenance - Assessment Criteria

Operator Expertise

In the FRM instrument category, the following Table 9 describes the assessment criteria for the “Operator expertise” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Documented evidence of training on making the measurements by all personnel but supervisor may themselves have relatively limited experience.
Good	At least one member of the team with at least 6 months experience and able to supervise/train short-term personnel in the measurements
Excellent	At least one member of the team has at least one year experience and has time to supervise and review activities of others who have undergone training but may only have up to 6 months experience
Ideal	Team of at least 2 operators, who have both been fully trained on all aspects of the FRM process (or at least a set of people who can cover all aspects) that have been producing this FRM for at least one year. All training records are documented and available.

Table 9 FRM Instrumentation - Operator expertise - Assessment Criteria

5.3.3 Operations/Sampling

Information concerning activities in terms of the level of automatization and documentation available for functional operation/sampling and processing to be representative of a satellite observation.

Automatisation Level

An important indicator of operational robustness and reliability is the degree of automation of the FRM, from data collection through to user access and utilisation:

- Automatic Calibration
- Automatic Measurement
- Automatic QA/QC of instrument parameters
- Automatic Data Transfer
- Automatic processing to satellite sensor observation

In the Operations/Sampling category, Table 10 shows how the assessment framework grades the “Automatisation Level” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The potential to upgrade to some level of automation
Good	At least use of instrumentation capable of automated collection of data
Excellent	A limited set of the above bulleted criteria but including as a minimum automated data collection and delivery to an off-location portal
Ideal	A fully automated process addressing all the above bulleted criteria.

Table 10 Operations/Sampling - Automatisation level - Assessment Criteria

Measurand Sampling

In the Operations/Sampling category, Table 11 shows how the assessment framework grades the "Measurand Sampling" aspect. It should be noted that for some applications e.g., atmospheric composition there needs to be interpretation by the assessor as to the adequacy of meeting volume nature of measurand (including vertical and horizontal smoothing) and any associated protocols defined by the community.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Typically, a single point sample (not at satellite pixels scale) but with an estimate of the impact on the comparison to the satellite measurand due to inadequate representativeness.
Good	A set of samples that seek to scope the observation characteristics of the satellite instruments observation of the measurand for a given pixel but still relatively sparse resulting in an uncertainty that typically needs multiple additional observations to be adequately minimised.
Excellent	A broadly sampled target (in all necessary dimensions (spatial and temporal) with observational based analysis of the effect of non-representativeness and a resultant uncertainty not dominating that of other contributions to the use of the FRM.
Ideal	The sampling characteristics of the FRM allow full representation of the satellite observation for all observation conditions of the sensor (both the instantaneous observation and/or global) such that the resultant uncertainty contribution is small compared to the process as a whole. Note: to achieve full representativeness this is likely to require some form of network or "set" of FRM observations and potentially very dense sampling if localised "point like" observations are being made by the FRM. The latter aspect is not addressed here but in the column **Completeness, coverage and distribution** described later.

Table 11. Operations/Sampling - Measurand Sampling - Assessment Criteria

ATBDs on Processing/Software

In the Operations/Sampling category, the following Table 12 shows how the assessment framework grades the "ATBDs on processing/software" aspect. This relates to any algorithm needed to combine or process data, including any ancillary data, into a form that can be compared with the satellite observation.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Some basic description of ATBD and its performance
Good	Documented but not necessarily fully accessible or peer reviewed but with some evidence of validation results
Excellent	Fully documented and tested ATBDs with at least evidence of comparison of results against defined/documented test datasets with some uncertainty assessment.
Ideal	Fully documented and independently validated ATBD for the FRM and any associated algorithms to derive a suitable measurand comparable with that of the satellite under test. This should evidence performance including full uncertainty assessment using community agreed data sets or formal intercomparison with peers

Table 12 Operations/Sampling - ATBDs on processing/software - Assessment Criteria

Guidance on Transformation of FRM to Satellite Sensor

In the Operations/Sampling category, the following Table 13 shows how the assessment framework grades the “Guidance on transformation of FRM to satellite sensor” aspect. This relates to the process and description of the process that any user would need to undertake to enable the FRM to be meaningfully compared with that of a satellite instrument.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	A basic description of how the FRM can be transformed to a satellite instrument measurement of a measurand
Good	Documented methodology on how to transform FRM into generic satellite instrument observational space with the potential to provide additional support on a case by case basis.
Excellent	Documented procedure of methodology with uncertainty estimates to enable users to transform FRM into satellite instrument observational space for most instruments of a particular class with examples. Notes- as per ideal
Ideal	Fully automated procedure with documented description of methodology to enable FRM to be transformed into satellite instrument observational space for the specific measurand with associated uncertainties for most sensors of a particular class. Note the transformation could be to match a direct level 1 measurand or indeed a level 2 product for example an atmospheric profile where vertical averaging and covariances need to be employed and accounted for.

Table 13 Operations/Sampling - Guidance on transformation of FRM to satellite instrument - Assessment Criteria

5.3.4 Data

Descriptive information concerning the data (result of instrument measurement) shall be provided in relation to the specifics of the data details itself, availability and usability, format and ancillary products (when part of operation or processing)

Data Completeness

The list of the required information is listed below:

- Data filename (according to filename convention if any)
- Instrument ID and Name
- Instrument Type
- Processing Level
- Measured quantity name and units.
- Stated Measurement quality
- Spatial Coverage and Resolution
- Vertical Coverage, Registration and Resolution
- Temporal Coverage and Resolution
- Spectral Coverage and Resolution
    - Uncertainty

    - Ancillary/auxiliary data and its origin
    - ID of any processing algorithms used
    - Time stamp
- PoC (Responsible organisation, including email address)
- Data Access (e.g. URL, DOI if applicable)
- Restriction for access and use, if any.

In the Data category, the following Table 14 shows how the assessment framework grades the "Data Completeness" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	FRM data available but not necessarily formally tagged with meta data although should be available from FRM owner.
Good	Essential information to allow usability of the FRM data to be delivered as part of metadata all other data available from a URL not necessarily in an automated manner
Excellent	all critical characteristics to be fully machine readable with the remainder to be readily accessible from a URL
Ideal	all FRM data characteristics to be fully machine readable and integrated with the delivered data through an open access portal with automated delivery capability.

Table 14 Data - Data Completeness - Assessment Criteria

Availability and Usability

This is about how readily the FRM data are available to those who wish to use them. It does not necessarily require cost-free access but is more about following the FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles for scientific data management and stewardship [8] which provide valuable principles for all data applications and which are:

Data should be findable

- Metadata and data are assigned a globally unique and persistent identifier.
- Data is described with rich metadata.
- Metadata clearly and explicitly includes the identifier of the data it describes.
- Metadata and data are registered or indexed in a searchable resource.

Data should be accessible

- Metadata and data are retrievable by their identifier using a standardised communications protocol.

- The protocol is open, free and universally implementable.
- The protocol allows for an authentication and authorisation procedure where necessary.

Data should be interoperable

- Metadata and data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- Metadata and data use vocabularies that themselves follow FAIR principles.
- Metadata and data include qualified references to other (meta)data.

Data should be reusable

- Metadata and data are richly described with a plurality of accurate and relevant attributes.
- Metadata and data are released with a clear and accessible data usage license.
- Metadata and data are associated with detailed provenance.
- Metadata and data meet domain-relevant community standards

In the Data category, the following Table 15 shows how the assessment framework grades the “Availability and Usability” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The data set does not appear to be following the FAIR principles but can be obtained from the provider.
Good	The data set meets some of the FAIR principles and/or there is an associated data management plan that shows progress towards the FAIR principles and there is a clear route on how to get access to the data.
Excellent	The data set meets many of the FAIR principles and has an associated data management plan and is available either free of cost or through an easy-to-access commercial licence.
Ideal	The data set fully meets the FAIR principles and has an associated data management plan and is available either free of cost or through an easy-to-access commercial licence

Table 15 Data - Availability and Usability - Assessment Criteria

Data Format

An important aspect of data that ensures ease of access to the widest variety of users is their format. Metadata and flags offer users important extra layers of useful descriptive information, in addition to the measurements themselves, that can be crucial to their analysis.

In the ideal case, the data format would meet the appropriate Committee on Earth Observation Satellites (CEOS) metadata guidelines.

In the case where such a standard does not exist, product format is graded based on the following:

- the extent to which it is documented.
- whether a widely used "standard" file format is used (e.g., NetCDF) i.e., not proprietary or specific to a single satellite sensor or the FRM site owner.
- whether it complies with standard variable, flag and metadata naming conventions, such as the Climate and Forecast (CF) metadata Conventions [9], or, for data from the European Union, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive [10].
- whether flags and metadata provide an appropriate breadth of information

In the Data category, the following Table 16 shows how the assessment framework grades the "Data Format" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The data set has relatively limited information, related to flags and metadata, and does not meet commonly agreed standards.
Good	The data set satisfies many of the format principles and there is an associated plan showing progress towards the commonly agreed standards. If is proprietary in nature there is support to translate it into a more common standard
Excellent	It would be as Ideal but not fully compliant with agreed common standards and state-of-the-art methodologies. However, the FRM provider will provide or help support development of a reader to allow a user to access the data.
Ideal	The dataset is equipped with an appropriate breadth of information including flags and metadata. The data format complies with standard variable, flag and metadata naming conventions, such as the Climate and Forecast (CF) metadata Conventions, or, for data from the European Union, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive.

Table 16 Data - Data Format - Assessment Criteria

Ancillary Data

Throughout the processing chain there may be a requirement for external input data. The ancillary datasets used during the processing should be identified to the user (where possible accounting for potential commercial sensitivity). Ancillary datasets must be of sufficient quality, including the application of suitably rigorous metrology, for example, in the form of SI traceability where necessary for the FRM's intended use.

The suitability of the ancillary data for its application must also be considered, with respect to the FRMs stated performance requirements. For example, the quality, size and representativeness of algorithm input data. The requirements will be specific to any algorithm used and may require some expert judgement, particularly if required as part of any satellite observation.

In the Data category, the following Table 17 shows how the assessment framework grades the "Ancillary Data" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Ancillary data used in product generation, specified to some extent, though incomplete. Not entirely of a sufficient quality to be judged "fit for purpose" in terms of the FRMs stated performance.
Good	Ancillary data used in product generation, specified, though not necessarily on a per observation basis. Mostly of a sufficient quality to be judged "fit for purpose" in terms of the FRMs stated performance.
Excellent	Ancillary data used in product generation, fully specified per FRM observation, but not necessarily fully traceable. Ancillary data used are of sufficient quality to be judged "fit for purpose" in terms of the FRMs stated performance
Ideal	Ancillary data used in product generation, fully specified per FRM observation, and are fully and traceable to SI. Ancillary data used are of sufficient quality to be judged "fit for purpose" in terms of the FRMs stated performance.

Table 17 Data - Ancillary Data - Assessment Criteria

5.3.5 Metrology

Metrology is the science of measurement. This section (matrix column) covers the aspects of the FRM related to measurement quality in relation to the measurand of the satellite, including calibration, traceability, uncertainty, comparison and overall adequacy. This column relates to the FRM as a whole and accounts inclusively for the aspects already assessed as part of FRM instrumentation. Note when considering a network, it is the consistency and harmonisation of the network that may be relevant in this context particularly if individual members of the network are independently assesses against the FRM criteria.

Uncertainty Characterisation

To ensure measurements are both meaningful and defensible, it is crucial that they include rigorously evaluated uncertainty estimates. A comprehensive description of how to evaluate sources of uncertainty in a measurement and propagate them to a total uncertainty of the final measurand, is provided by the metrological community in the Guide to the Expression of Uncertainty in Measurement (GUM) [11]. The GUM approach, where practical, should be applied to all FRM and EO satellite missions. A tailored version of this together with training material can be found in QA4EO Training [12] and QA4EO Tools [7]. Increasingly, providers of operational and reprocessed data products are applying different approaches to evaluate and distribute metrologically rigorous error-covariance information for L1 and L2 satellite products at the per-pixel level, as required by climate studies. However, it is still not unusual for uncertainties (or performance estimates) to be evaluated in a manner that does not comply with the principles of the GUM, for example, the performance specification value or single offset from a comparison sensor may be quoted as the uncertainty. For an FRM this is not in general considered adequate. For an FRM we consider here the uncertainty of the FRM quantity i.e. that which has been transformed in a manner that allows it to be directly compared to that of a satellite sensor observed pixel. Note that this is an assessment of the how the uncertainty has been assessed and relates to the estimate provided in the discovery section (section 5.2) for the FRM. It is not intended to be judgemental as to whether the resultant uncertainty is useful for the intended application only that it has been adequately assessed.

In the Metrology category, the following Table 18 shows how the assessment framework grades the "Uncertainty Characterisation" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Uncertainty established by limited comparison to measurements by other sensor/s.
Good	Limited use of GUM approach, and/or, an expanded comparison to measurements by other sensors. Most important sources of uncertainty are included.
Excellent	Full GUM approach is used to estimate measurement uncertainty, all important sources of uncertainty are included. Uncertainty per satellite sensor pixel provided
Ideal	Full GUM approach is used to estimate measurement uncertainty, including a treatment of error-covariance. Per satellites sensor pixel uncertainties in components, e.g., random systematic – as appropriate for the error-correlation structure of the data

Table 18 Metrology - Uncertainty Characterisation - Assessment Criteria

Traceability Documentation

Traceability is defined in the vocabulary of international metrology (VIM) as a,

"property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty"

and reinforced in the QA4EO procedures QA4EO Home. It should not be confused with documentary traceability which ensures there is a correct paper trail for QA purposes. Traceability is therefore a key aspect of achieving reliable, defensible measurements. In this definition, an important part of measurement traceability is highlighted – that it is well documented. This of course must be the case for FRM data products too. Various diagrammatic approaches have been developed to present the traceability information, chains from primary realisation through to field observed measurand, for EO data products (e.g. the QA4ECV guidance, which includes a traceability chain drawing tool [13]). Such a diagram should be included in the documentation for every EO mission. The FIDUCEO project [14] has provided guidance for a more detailed measurement function centred on an "uncertainty tree diagram" which is ultimately more suitable for Level 1 (and some Level 2) processing as is typical for the end use of the FRM.

The criteria here are based on the degree of completeness and availability of the information. In the Metrology category, Table 19 shows how the assessment framework grades how the the "Traceability" aspect is documented.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Traceability chain diagram and/or uncertainty tree diagram included, but missing some important steps.
Good	Traceability chain and/or uncertainty tree diagram documented identifying the most important steps and sources of uncertainty.
Excellent	Rigorous uncertainty tree diagram, with a traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty.
Ideal	Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes and evidences full traceability to SI.

Table 19 Metrology - Traceability Documentation - Assessment Criteria

Note: whilst the above indicates the use of traceability chain/uncertainty tree diagrams, it is acceptable to provide alternative means of representing the same information, but, the important criterion being the ability for the reader to be able to readily assimilate the information and associated evidence.

Comparison/Calibration of FRM Products

A key element of traceability is the ability to evidence the performance of the measurement, in this case FRM, against a defined community agree reference, ideally SI. For this element, it is not only the FRM instrument measurement, which is considered in an earlier category, but the full end to end of the transformation process and the comparison process carried out against the satellite product itself i.e. accounting for any representational differences, subtle product definition differences, environmental effects etc. The following lists the uses a comparison can support:

A. A comparison can be used to validate that observation values are within an expected tolerance.

B. A comparison can be used to evaluate the uncertainty associated with the FRM and/or its ability to compare with a satellite observation.

C. A comparison can be used to validate independently determined uncertainties.

Traditionally, comparisons have been used for approach (A), i.e., to monitor whether satellite and reference measurements agree within the satellite's requirements. As the community progresses in its capabilities, approaches (B) or (C) can be considered viable to attempt.

In comparison, it is necessary to consider uncertainties associated with:

1. The FRM reference / compared observations.
2. The satellite instrument observations.
3. The comparison process itself.

Uncertainties associated with the comparison process itself include uncertainties related to the fact that the reference measurements and the satellite observations may be different, or uncertainties related to processing steps used to make the two measurements more equivalent (for example, by scaling or sampling observations to a common grid). If the uncertainties associated with (1) the FRM observations and (3) the comparison process themselves are much smaller than (2) uncertainties associated with the satellite instrument observations, then the comparison can be used for approach (B), i.e., the comparison can be used to calculate/evaluate the uncertainty associated with the satellite measurements.

A metrological approach to comparisons would follow approach (C). That is, the three types of uncertainty associated with the two measurements (satellite and non-satellite) and the comparison process are independently evaluated and then the comparison is used to validate the uncertainties. It is for this reason that the FRM needs an uncertainty evaluation independently determined and that we need to consider the uncertainty associated with the comparison process itself ideally all independently of each other.

Over the last decade as the FRM process has developed, European funding agencies have established a variety of FRM projects within which there has been organised a community comparison. In some cases there has been the inclusion of a measurement or references that can be considered SI or community defined allowing the concept of a calibration, see for example Home - FRM4STS [15].

For this category, we are looking for evidence on how the FRM can demonstrate its performance preferably through participation in a community organised comparison where possible. CEOS and other international bodies will endeavour to establish opportunities to enable FRM providers to participate in community comparisons to underpin this critical characteristic.

In general, comparison here refers to comparison with peer like activities e.g. other similar FRM providers or instrument users. However, particularly for networks consideration can be made of comparison between members of a network and indeed the network or FRM output by comparison to a satellite, providing the satellite used for the assessment has an uncertainty commensurate with that of the FRM and of course is not intended to be one that relies on the FRM for its validation.

In the Metrology category, the following Table 20 shows how the assessment framework grades the “Comparison/Calibration of FRM products” aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Evidence to show some level of consistency with an independent observation.
Good	Participation with results largely consistent with declared FRM performance capabilities across most of scope and many satellite instruments but in an informal comparison/calibration exercise
Excellent	Participation with results largely consistent with declared FRM performance capabilities across most of the scope and for many types of satellite instrument in a formally organised comparison/calibration exercise.
Ideal	Participation, with results fully consistent with declared FRM performance capabilities, in a formal independently organised comparison/calibration exercise for full scope of FRM and wide range of satellite instruments.

Table 20 Metrology - Comparison/Calibration of FRM products - Assessment Criteria

Adequacy for Intended Class of Satellite instrument/measurand

The majority of this assessment framework is compliance to a self-declared set of performance criteria and for a particular class/es of satellite instruments measurements of a measurand. In the main, it does not judge if the level of uncertainty that is declared is at an appropriate level to assess the performance of an instrument. Indeed, many instruments in the same class can have very different requirements for uncertainty and so in general it is for the user to make this assessment based on their particular application needs.

However, here we seek to make a quantitative assessment considering the requirements of a typical "high performance" instrument within a particular class for a stated measurand. For this self assessment process we ask the FRM provider to indicate in their view the type of application for a particular measurand their FRM is likely to be suited to e.g. climate, single image variability, single sensor relative measurements etc and thus the classification is adapted accordingly. However, whilst labels: ideal, excellent together with the colour coding etc are for practical purposes still used, the classification is not intended to be as subjective in nature, with an FRM of "basic" in this category still being of value for its specific type of application. This category is not considered within the assessment of the FRM's overall assessment.

In the Metrology category, the following Table 21 shows how the assessment framework grades the "Adequacy for intended class of instrument" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Uncertainty is sufficient to provide validation for all satellite instruments , although for those instruments making quantitative observations where temporal, longer term studies and or those considered "risk" critical they are unlikely to be of a sufficient uncertainty level.
Good	Uncertainty comparable to the majority of satellite instruments used for operational/monitoring applications where the specific demands of climate are not required.
Excellent	Uncertainty comparable to that of the highest performing satellite instruments typically designed for climate. Data from FRMs in this class would be able to provide robust validation for any satellite instrument in its specified class and calibration for some.
Ideal	Uncertainty typically smaller than that of any satellite instrument in its class for a given measurand and can be used to provide robust post-launch calibration of all satellite instruments in the class including those for climate applications.

Table 21 Metrology - Adequacy for intended class of instrument and measurand - Assessment Criteria

5.3.6 Completeness, coverage and distribution

In most of the cases a limited set of measurements acquired at a single location are not sufficient to produce FRM datasets sampling adequately and over the entire useful range the measurand values and influencing quantity values over which validation needs to be carried out. For some FRMs, the results of different individual "measurements" from the same or different geographical locations and times (each potentially assessable as an FRM in its own right) may be combined in a network-based approach to validation, ideally through a common processing chain to provide a combined "network"/multi-parameter value as the FRM to compare with the satellite product. This can be a truly combined result or simply a common/harmonized processing chain ideally formalized in a Standard Operating Procedure (SOP) reflecting the concept and rules of a community agreed system. In Beyond the concept of FRM maturity level for a single measurement, this category/column provides information concerning the level of maturity for those characteristics of this combined/common product which are relevant and specific to this category for the FRM assessment. The aspects identified as important to this category are the following:

- Validation Capacity

- Geographical coverage and sampling
- Temporal coverage and sampling
- Centralized data, processing, quality assessment and adherence to community standards
- Timeliness

Validation Capacity

This important aspect concerns criteria assessing the properties of the measuring system (with the exception of the geographical and temporal coverage and sampling handled separately) in terms of completeness and capability to properly get characteristics of the measurand over its full range of possible values and with appropriate sampling of its variations (e.g.: short-term, seasonal, long-term, vs. latitude, etc.) reachable by the measuring system intended as a set of coordinated instruments deployed as network and/or in a context of a well-designed campaign. We mention, as examples, the capability to get range and sampling of the most influencing quantities (e.g.: constituent profile shape, surface albedo, cloudiness, temperature contrast, aerosol load, etc.), range and sampling of the most influencing measurement parameters (e.g.: Solar Zenith Angle, View Zenith Angle, South Atlantic Anomaly, etc.), and atmospheric state when relevant for the measurement.

The classification guidance "criteria" for this aspect are below:

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The measuring system deployment cannot properly get the measurand characteristics. There is, however, a clear aspiration and a strategy to progress towards a better equipped system.
Good	The measuring system is well deployed and well-coordinated. It has the capability to get some of the characteristics of the measurand over a good but not exhaustive range of possible values.
Excellent	The measuring system consists of a coordinated set of instruments deployed as a network and/or in a context of a well-designed campaign. It has the capability to get most of the characteristics of the measurand over most of possible range values and its dependence on the most influencing quantities.
Ideal	The measuring system consists of a coordinated set of instruments deployed as a network and/or in a context of a well-designed campaign. The measuring system is fully equipped to properly get characteristics of the measurand over its full range of possible values and with appropriate sampling of its variations, and its dependence on the most influencing quantities.

Table 22 Completeness, coverage and distribution - Validation Capacity - Assessment Criteria

Geographical Coverage and Sampling

Validation of a satellite's performance requires assessment under a range of geographical, environmental, atmospheric conditions and spanning the dynamic range of the parameter being measured. This means that often a single FRM "site" can only do part of the job and multiple FRM are required to assess the full functionality. However, a coordinated collection of FRM geographically distributed can seek to address this. This aspect assesses the ability of the "network" to meet the coverage and sampling needs to assess the broad functionality of a satellite. A network of measurement locations well distributed on both hemispheres spanning the full dynamic range of the measurand under test and the most important environmental conditions and atmospheric states is clearly of high value. However, this needs to be contrasted with the cost and thus a balance of minimum need and redundancy needs to be considered (i.e., return on investment). A dedicated network design study might be required to obtain an optimal distribution of FRM sites with maximum return on investment.

The classification guidance "criteria" for this category are below:

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The "network" is composed of few measurements concentrated in a localized area, and/or limited range of parameter values.
Good	The "network" offers an adequate density of measurements for some areas of one hemisphere.
Excellent	The "network" is spread over all relevant geographical areas over both hemispheres and covers the most important environmental conditions and atmospheric states, but with a notable difference of density among the areas and broad coverage of parameter values but not necessarily detailed intermediate values.
Ideal	The network is regularly spread over all the geographical regions and covers the most important environmental conditions and atmospheric states with density able to ensure full "redundant" sampling spanning full detailed validation of the satellites performance.

Table 21 Completeness, coverage and distribution - Geographical coverage and sampling - Assessment Criteria

Temporal Coverage and Sampling

As for the Geographical Coverage, a network offers an added value when is able to follow a satellite measurement in its temporal evolution over the relevant period of time and at the relevant frequency, for example to observe a dynamic atmospheric feature or detect a long term drift. This aspect assumes a certain level of coordination among the individual components/measurements of the network.

In the "Completeness, coverage and distribution" category, Table 22 shows how the assessment framework grades the "Temporal Coverage and Sampling" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The network is not coordinated in terms of observation making and offers only a basic temporal coverage and sampling capability to cover some satellite observations.
Good	The network, even if not coordinated, has the temporal coverage and sampling capability to cover most of the satellite observation mode.
Excellent	The elements of the network are coordinated and optimized to have a temporal coverage and sampling to cover most of the satellite observation mode.
Ideal	The elements of the network are coordinated and optimized to have a temporal coverage and sampling that allow a full representation of the satellite instrument observation for all the observation conditions.

Table 22 Completeness, coverage and distribution – Temporal Coverage and Sampling - Assessment Criteria

Centralized data, processing, quality assessment and adherence to community standards

The organization and the level of coordination of a network of FRM instruments or of network-like operation of FRM instruments is, probably, one of the most important aspects that define and characterize its level of maturity as a federated provider of FRM. In particular, the following elements are usually interrelated are part of this aspect:

- Adoption of common standards, operating procedures and measurement protocols
- Instrument intercalibration
- Centralized data acquisition
- Centralized data processing
- Centralized assessment of data quality and of network performance
- Centralized data distribution (common access point, common data and metadata format, data versioning…) and additional services

In the "Completeness, coverage and distribution" category, shows how the assessment framework grades the "Centralized data, processing, quality assessment and adherence to community standards" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The organization of the network is barely sketched, without planning to which all participants adhere. Intercalibration is sporadic if at all. There is no centralized system for standardized collection, processing and quality assessment/quality control of the data.
Good	The network, even without a structured organization, has resources and capability for future development, including procedures for intercalibrations, standardization and a centralized data collection, processing and quality assessment/quality control. It encourages adoption and utilization of community standards.
Excellent	The network has an organization even if not rigorously and formally structured following the latest management standards. It has started with intercalibrations, standardization and with some level of centralization of data and processing or at least following common procedures and a common access point. There is a clear intent to proceed towards a structured organization and a full centralization of the activities.
Ideal	The network has a structured organization including a steering committee, chairs, working groups and periodic meetings. Instruments are regularly intercalibrated following standards procedures. Data are collected and archived with a standard format in a centralized database. Processing (e.g. from raw data to higher levels) follow standards, runs over a centralized system or the processing kit is distributed according to clear SW versioning approach, and the data is quality assessed and controlled at network level. The homogeneity of the network data quality across stations is assessed and documented regularly.

Table 23 Completeness, coverage and distribution - Centralized data, processing, quality assessment and adherence to community standards - Assessment Criteria

Timeliness

The Timeliness aspect is particularly important for operational missions, for which the timeliness of validation and monitoring information can be crucial. Timeliness of FRM availability is also essential to characterize *in extenso* the in-orbit performance of a recently launched satellite instrument during its intense but also short commissioning phase. Some level of timeliness can also be relevant to ensure proper validation of satellite data before its use by other applications and users. Data with too long dissemination time are useless to characterise some events (e.g., satellite data assimilation for meteorology and air quality forecasting services). The ideal dissemination time depending on the user and the application, the grades provided hereafter assess how timeliness is managed rather than its compliance with the requirements of a specific application.

In the "Completeness, coverage and distribution" category, Table 4 shows how the assessment framework grades the "Timeliness" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	The organization of the measuring system is basic. Information about the timeliness in providing data from the network are few, sporadic and discontinuous.
Good	The organization of the measuring system is developing. There is a clear intent to proceed towards a systematic and timeliness provision of data.
Excellent	The measuring system is well organized, data are provided with continuity and with a timeliness that cover most of the operational measurement mode.
Ideal	The measuring system has a structured organization. The data are provided with continuity with a scheduled timeliness which allow validation and monitoring of all the operational measurement mode.

Table 24 Completeness, coverage and distribution - Timeliness - Assessment Criteria

6. Independent Assessor FRM Maturity Matrix

The Independent Assessor column of the FRM Maturity Matrix (MM) should be filled by an independent assessor with respect to the specific FRM. It is intended that this process will be undertaken by experts under the auspices of CEOS WGCV and conceptually complemented by peer review. As part of the verification process an overall classification/grading of the FRM will be assigned based on its degree of compliance and planned progress in relation to the FRM criteria.

In this MM the grading process is largely achieved through the Verification Column.

6.1 Verification

The overall goal is to verify that the FRM is consistent with the criteria in section 3 as self-assessed by the FRM owner and that the evidence provided fully supports the assessment. This verification column is again subdivided into categories to provide some granularity to the verification process. It should be noted that achieving FRM endorsement is only the first step and a continual effort to maintain and/or improve performance and compliance is expected with the possibility that FRM endorsement can be removed, a grade increased/reduced. To this end some form of confirmation of self-assessment is expected on an annual basis as is verification. However, the latter is likely to rely on an exception basis or reports from users, unless some significant change is indicated by the FRM owner.

Guidelines Adherence

This is the overall summary evaluation of the degree of compliance with the FRM guidelines in section 3 and is essentially a review of the self assessed classifications in the rest of the MM and how many, and in which categories, green classifications are reached and those that are blue e.g. good, basic or below. Verification here also considers if the provided evidence is adequate to justify the claim. Note the degree of 'greenness' for the purposes of the overall FRM classification is weighted towards those categories impacting the quality aspects and should not be unduly impacted by categories deemed to be not applicable to the FRMs specified application.

In the context of the Verification, the following Table 22 shows how the assessment framework grades the "Guidelines adherence" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	60% of all applicable categories should be at least "basic" and if not there should be a clear strategy to progress within a short (<3 month) timescale. At least half of those in the metrology and instrument columns should be at least basic. Those categories in basic should have a strategy to progress towards greater compliance.
Good	More than 60% of applicable categories must meet the "good" category and those in "basic" should indicate a strategy to progress. >30 % should be in the green classification. There should be no "basic" classifications in the metrology or Instrument columns and any in these columns only indicating "good" should indicate a strategy to progress
Excellent	All applicable categories are good or above with > than 60% in the green classification and those in the Metrology or instrument columns must meet excellent or above.
Ideal	All applicable categories in the matrix fully meet the good or above classification with at least 80% reaching the green classification i.e. Excellent or Ideal and at least half reaching the ideal category. At least half of those in the metrology and FRM instrument column should meet the ideal classification with of course all others in these categories being at least excellent.

Table 22 Verification - Guidelines Adherence - Assessment Criteria

Utilisation/Peer Review/Feedback

This category is typically one which develops with time and considers the degree of utilisation/impact of the FRM by and upon the user community e.g. citations. It also takes account of any feedback provided on the website for the specific FRM (positive and negative) and/or any other input from colleagues who have the opportunity to independently review and provide comment to the CEOS oversight team on the FRM data and associated evidence provided for self-assessment. Any feedback from the user community will be considered by the CEOS team on a regular basis and may ask for clarifications from the FRM owner before changing any gradation for this category. This category and associated grading is somewhat subjective in nature.

In the context of the Verification, the following Table 23 shows how the assessment framework grades the "Utilisation/peer review/feedback" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Limited evidence of community usage.
Good	Some Citations with mixed positive and negative comments from the community.
Excellent	Well cited FRM, some positive feedback from the user community and limited negative comments.
Ideal	Highly cited FRM, with significant positive feedback from the user community.

Table 23 Verification - Utilisation/peer review/feedback - Assessment Criteria

Metrology Compliance

One of the critical distinguishing characteristics of an FRM compared to similar observations is the evidenced degree of metrological rigour carried out by the FRM providers. This category independently considers and grades the evidence that is described in the self-assessment MM related to this topic (instrument and metrology columns) and provides an overall grading based on an amalgam of the information provided and interpretation from the CEOS experts. In essence this is an independent review of the evidence and claims that the FRM provider has self-declared.

In the context of the Verification, the following Table 24 shows how the assessment framework grades the "Metrology Compliance" aspect. Note "adequacy" is not explicitly assessed as part of this classification as the performance of the FRM is declared explicitly in the discovery section (section 5.2) and it is for the user to determine specific adequacy for their application.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Limited evidence of traceability chain diagrams and/or uncertainty tree diagrams. Uncertainty information based on limited independent assessment for all or some aspects of the FRM.
Good	Traceability chain and/or uncertainty tree diagram documented identifying most important steps and sources of uncertainty with some reasonable level of overall uncertainty evaluation for most of the critical elements of the FRM.
Excellent	Rigorous uncertainty tree diagram, with a traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty for all aspects of the FRM. Full GUM (or equivalent) approach is used to combine and propagate uncertainty throughout the FRM.
Ideal	Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes traceability to SI with low uncertainty. Uncertainty evaluated with full treatment of error-covariance.

Table 24 Verification - Metrology Compliance - Assessment Criteria

Independent Verification

This category allows the prospect of an independent audit of the FRM through comparison against similar FRM and/or satellite sensors organised and carried out under the auspices of the CEOS expert team. This may be the result of a formal community comparison such as those already carried out in some FRM4xx projects or against a satellite sensor with adequately known performance that can be considered a "reference". Although where possible this independent verification step will be carried out at or around the time of evaluation of the FRM, and indeed may have been used as evidence by the FRM provider in the self-assessment section of the MM it will likely be undertaken at times of convenience or where it is considered critical and urgent by a potential user. Grading here will be based on the degree of consistency with any comparison reference and of course dependent on the uncertainty of the reference and the comparison process. It is thus quite possible that a less than ideal grading can be obtained simply because the comparison used for verification was not of sufficient quality, and similarly, a not assessed or not assessable grade because no opportunity has arisen to undertake this verification step. However, if the FRM provider chooses not to participate in a formally arranged activity without good reason this will lead to a lower grade and an appropriate comment.

The following Table 25 shows how the assessment framework grades the "Independent Verification" aspect.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Some comparison evidence but limited ability to confirm or otherwise the declared FRM uncertainty
Good	Full compliance of declared FRM uncertainties through comparison to a reference of good but higher uncertainty than the FRM or near but not full compliance against a reference of comparable or lower uncertainty.
Excellent	Full compliance of declared FRM uncertainties through comparison to a reference with comparable uncertainties.
Ideal	Full compliance of declared FRM uncertainties through independent comparison to a reference of lower overall uncertainty

Table 25 Verification - Independent Verification- Assessment Criteria

6.2 CEOS-FRM Overall Classification

To provide overall summary guidance to a user we have created the following four classes to represent an overall grading for the FRM, with the intention that Class A is highly demanding and an aspiration for many, but should be the target for climate applications. Class B being the minimum desired for most quantitative applications and Class D considered to be the minimum starting point for a new provider starting out in the process.

Class A – Where the FRM fully meets all the criteria necessary to be considered an FRM for a particular class of instrument and measurand.

It should achieve a class of Ideal in the "guidance adherence" criteria in the verification section of the MM and green (at least excellent) for all other verification categories where these have been carried out.

Class B – Where the FRM meets many of the key criteria and has a path towards meeting the Class A status in the near term.

It should achieve at least Excellent in the "guidance adherence" criteria in the verification section of the MM and green (at least excellent) for all other verification categories where these have been carried out. Ideally it should indicate a path towards achieving the high class.

Class C – Meets or has some clear path towards achieving the criteria needed to reach a higher class and provides some clear value to the validation of a class of satellite instruments/measurands.

It should achieve at least Good in the "guidance criteria" in the verification section of the MM and at least good for all other verification categories where these have been carried out. Ideally, it should indicate a path towards achieving the high class.

Class D – Is a relatively basic adherence to the FRM criteria but where this is a strategy and aspiration to progress towards a higher class. This can be considered an entry level class for those starting out on developing an FRM.

It should achieve at least Basic in the guidance criteria in the verification section of the MM and at least Good for all other verification categories where these have been carried out. FRM owners/developers must indicate a path towards achieving the high class.