GAIA-CLIM Report / Deliverable D1.3

Gap Analysis for Integrated Atmospheric ECV Climate Monitoring:

Report on system of systems approach adopted and rationale

A Horizon 2020 project;
Grant agreement: 640276
Date: 27th November 2015
Lead Beneficiary: NUIM
Nature: R
Dissemination level: PU

Work-package	WP1 (Geographical capabilities mapping)
Deliverable	D1.3
Nature	R
Dissemination	PU
Lead Beneficiary	National University of Ireland Maynooth
Date	27/11/15
Status	Final
Authors	Peter Thorne (NUIM), Joerg Schulz (EUMETSAT), David Tan (ECMWF), Bruce Ingleby (ECMWF), Fabio Madonna (CNR), Gelsomina Pappalardo (CNR), Tim Oakley (MO / GCOS)
Editors	Anna Mikalsen (NERSC)
Reviewers	Karin Kreher, Arnoud Apituley, Greg Bodeker, Barry Goodison, Mark Bourassa, Matthias Buschmann and Ge Peng
Contacts	Peter.thorne@nuim.ie
URL	http://www.gaia-clim.eu

This document has been produced in the context of the GAIA-CLIM project. The research leading to these results has received funding from the European Union's Horizon 2020 Programme under grant agreement n° 640276. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view

Document history

Version	Author(s) / Reviewers	Date	Changes
0	Peter Thorne	05/04/15	
0.1	Peter Thorne with input from Fabio Madonna, David Tan and Joerg Schulz	21/05/15	Substantial additions following first teleconference of task participants
0.1.1	Peter Thorne based upon input from Fabio Madonna	12/06/15	Minor changes to reflect Fabio's input
0.2	Peter Thorne, Karin Kreher, Fabio Madonna, Tim Oakley, Joerg Schulz, Gelsomina Pappalardo, Arnoud Apituley	15/07/15	Substantial additions and edits based upon feedback received.
0.3	Peter Thorne	24/7/15	Edits made based upon outcomes of group call on July 17th
0.31	Peter Thorne	19/8/15	Minor changes to account for resynchronising with the spreadsheet and edits on further read through
0.4	Peter Thorne, Bruce Ingleby, Fabio Madonna	2/9/15	Changes to account for Bruce Ingleby and Fabio Madonna input
1.0	Peter Thorne, Joerg Schulz, Fabio Madonna	10/9/15	Final changes before external review
1.1	Peter Thorne, Mark Bourassa, Greg Bodeker, Matthias Buschmann, Barry Goodison, Bruce Ingleby, Arnoud Atipuley, Stephan	29/10/15	Changes in response to internal and external review comments received

	Bojinski, Karin Kreher, Ge Peng		
1.2	Peter Thorne, Karin Kreher	12/11/15	Tidying changes
2	Peter Thorne	30/11/15	Final submitted version

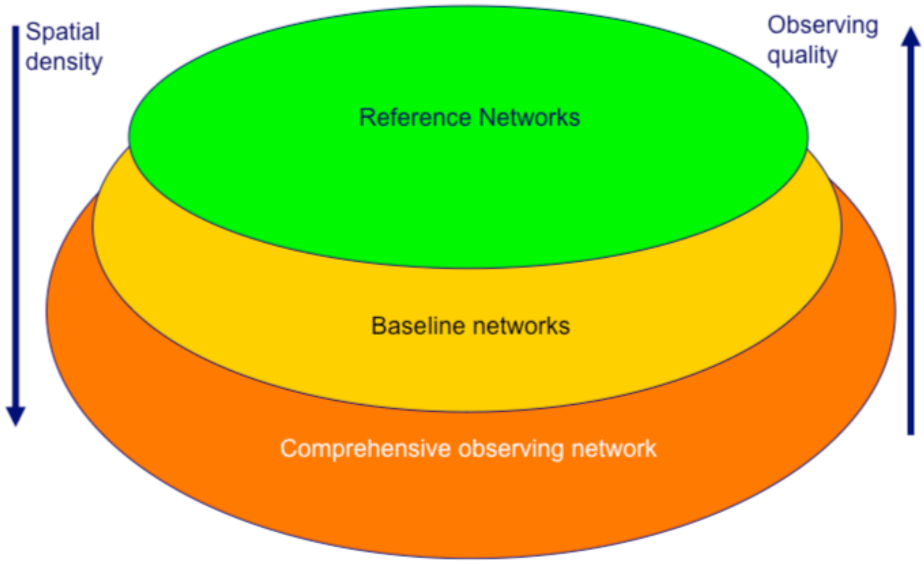
Table of Contents

Executive Summary

In the first instance, this guidance is intended specifically for assigning suitability of candidate non-satellite measurements for satellite calibration and validation, under the Horizon 2020 funded GAIA-CLIM project. However, it is envisaged that it may be adopted more broadly. The guidance builds upon a similar effort to assess climate data record maturity under the Framework Program 7 funded CORE-CLIMAX project.

This guidance exists to support the designation of non-satellite observational capabilities into a structured system of systems architecture consisting of:

- Reference quality networks that have amongst others: strict traceability and comparability, rich metadata, known data origin and quality, and long-term infrastructure support
- Baseline networks that are well characterised and have a long-term monitoring commitment
- Comprehensive networks that consist of a broad range of observational capabilities managed for myriad purposes.



Such a designation has many potential scientific and societal benefits, relating to the appropriate use of the data collected for many applications. The designation is achieved through applying a set of semi-quantitative assessment criteria against the following seven thematic areas, which may reasonably differentiate the observational capability maturity:

1. Metadata
2. Documentation
3. Uncertainty characterisation
4. Public access, feedback, and update
5. Usage
6. Sustainability
7. Software (optional)

1. Document rationale and broader context

The purpose of this document is to provide a framework to semi-objectively classify measurement capabilities, and hence to ensure scientifically rigorous and robust usage. It relates primarily to specific non-satellite observing networks and / or capabilities such as observationally-based research infrastructures. It could also potentially be used on individual instruments / sites, although such a specific assessment would be a substantial undertaking. So, where possible, a consideration of networks / infrastructures that operate to common standards is encouraged.

The aim is to assign observational capabilities into a set of tiers, to ensure optimal usage in subsequent applications, such as satellite calibration and validation or limited area forecasting. It builds upon the concepts of climate dataset maturity developed under the FP-7 CORE-CLIMAX project [Schulz et al., 2015]. These in turn were built upon earlier work undertaken at NOAA [Bates and Privette, 2012]. This document assumes that basic metadata on the measurements to be assessed such as the measurement geo-location(s) and the instrument types are available. Within this current process, a deeper assessment of the data and metadata properties is undertaken, to allow a more rigorous assessment of suitability. The assessment is based upon a number of thematic areas such as documentation, uncertainty quantification and sustainability, which can be used to characterise the critical aspects of measurement system maturity.

The assessment of observational measurement capabilities (this guidance) and derived datasets and products from these measurements (the CORE-CLIMAX based guidance) is somewhat distinct. The taking of measurements is the collection of original data and metadata that is directly or, more commonly, indirectly an estimate of the target measurand. Measurement series result from continuous or periodically repeating observations, using the same or similar measurement techniques, that are processed from the raw measurement to an estimate of the target measurand(s). Derived datasets and analyses use sets of such measurements and apply substantial post-processing steps to aggregate, analyse and, perhaps, filter and / or interpolate. They do not include the collection of primary data. This distinction in what is done in creating a measurement and a dataset and, therefore, what is being assessed, matters. Hence it likely requires separate, but similar, sets of guidance. Consideration was given in the first instance to simply reusing the existing CORE-CLIMAX maturity assessment tables, while writing new measurement-system-specific interpretation guidance. However, it was felt, after considerable discussion, that there were sufficiently unique aspects to assessing the measurements rather than datasets, reanalyses and similarly derived products, to warrant a separate set of tables. In doing so some categories have been removed or made optional, others have been modified, and several entirely new categories and sub-categories have been added. In the longer term it may be possible and desirable to remerge these guidelines, but that would require a new project to be initiated to this end. To enable such a future reconciliation, wherever possible, the CORE-CLIMAX tables have been unchanged to allow traceability and transferability.

Users wishing to assess maturity and suitability of datasets, reanalyses or other approaches that aggregate and analyse large sets of measurements, to create climate or environmental data records for given applications, should refer to and use the CORE-CLIMAX User Guide on the System Maturity Matrix [Schulz et al, 2015]. Users wishing to assess the maturity of a given set of measurements should use the guidance and tables provided in this document. The dividing line between a set of measurements and a climate data record is recognised as meaning distinct things to different users. To attempt to clarify which set of guidance should be used, Table 1 lists some salient features and the likely distinctions, to support the use of the most appropriate set of guidance and tables in any given case.

The guidance in this document should be used for a non-satellite measurement series if …	**CORE-CLIMAX guidance should be used for a climate data record if …**
Data being considered is an (set of) individual time series arising from one or more defined instruments, either at fixed locations or using mobile platforms.	Data being considered have global or at least continental scale coverage arising from satellite data or a substantively aggregated set of non-satellite data.
Available documentation addresses the instrument and / or arises from technical documentation describing the measurement process.	Available documentation addresses the construction, usage and validation aspects of a data product (CDR) in the peer-reviewed literature and/or technical documentation.

Table 1. Decision guidance as to whether the current set of maturity matrices, or those developed under CORE-CLIMAX, are likely most appropriate for a given use case based upon criteria that should permit easy determination.

Like the CORE-CLIMAX User Guide on System Maturity Matrix, there is an explicit limit to how far this guidance can take the user. If applied rigorously, the user can gain an appreciation of the relative maturity of key relevant facets of a set of measurements undertaken, for example, by a network or measurement infrastructure. However, there is not and cannot be, a single threshold that can be used to uniquely decide whether a given set of measurements has reached a given maturity level (Section 2). Rather, the assessment provides the basis for a user to decide upon a defensible level of maturity, and provides a chain of semi-quantitative evidence that can be used to support their assignment. The assessment is intended to define the measurement system maturity, and not the suitability for a given application which will, by its very definition, be application specific.

In reality there are two principal sets of potential users of this guidance and its outcomes. The first set of users consists of people undertaking the assessment or undertaking the measurements to be assessed. For this group of users, it is key that they understand how to implement the assessment outlined in Section 3, and how they can utilise the results to point to ways to improve the maturity and, hence, scientific utility of existing measurement systems. The second set of users consists of scientists, data analysts, policy makers etc. who may use the outcomes of the assessments to guide either their use of data, or decision-making, or both. Subsequent tasks within GAIA-CLIM shall undertake an initial assessment of maturity of many existing measurements and develop, and provide a range of tools to support the second identified set of users.

Given the heterogeneity of surface, ground-based remote sensing, balloon-borne and aircraft measurements (non-satellite measurements) and their funding and governance, this guidance concentrates upon such measurements. In theory, a similar assessment could be made for the satellite-based fundamental measurements (Level 0 and potentially up to level 1A/1B). However, given the GAIA-CLIM remit this guidance does not at this time extend to the satellite domain. Section 4.4 briefly discusses future potential extension in this direction.

The remainder of this guidance is structured as follows. In Section 2 the tiered approach to network measurement capabilities concept is outlined. This includes discussion of the potential scientific and measurement technology and practices benefits that could accrue from an explicit consideration of a tiered network of networks design to non-satellite measurement capabilities. Section 3 contains the substantive assessment criteria, along with the guidance necessary to complete the assessment. Therein each assessment area (or strand) is discussed and guidance on its appropriate completion is given. It is complemented by an excel spreadsheet which can be used to collate the assessment. Finally, Section 4 outlines a number of likely challenges to broader adoption by the scientific community of the concepts detailed herein.

2. Tiered approach to assigning measurement capabilities

Currently, little to no effort has been made to define and broadly agree amongst global stakeholders the measurement and network characteristics underlying a proposed system of systems approach to non-satellite Earth Observation capabilities. This is despite the existence of groups such as Global Earth Observation System of Systems (GEOSS), with the System of Systems implicit in its name. Within the peer reviewed literature, explicit reference to a tiered network of networks approach is, to our knowledge, limited to Seidel et al., 2009. Such a system of systems concept is also present in several recent GCOS documents and NAS, 2009. A tiered set of networks approach is arguably necessary to make sense of the mosaic of observational capabilities at our disposal, and hence use the right measurements for the correct application.

Specifically, for GAIA-CLIM, it is necessary to have a working model from which to:

- Define tiers of capabilities that may define fitness-for-purpose, for different candidate non-satellite measurement programs, to be used to understand and ultimately constrain satellite measurements;
- Assess and map these non-satellite measurement capabilities; and
- Select those measurements that have the necessary metrological (the science of measurement) characteristics to be used in those project work packages concerned with co-location uncertainty quantification, data assimilation and the virtual observatory.

It is hoped that the tier designations and underlying assessment criteria proposed herein can gain broader traction within the Earth Observation community as a whole. But, initially, it is solely necessary to define a working model that is acceptable across GAIA-CLIM, to enable subsequent tasks within the project to be undertaken.

2.1 Requirements for a tiered approach

A perfect measurement is not a metrological possibility, because any measurement will always to some extent differ from the true value of the measurand. In an ideal world, all measurements undertaken to monitor the climate system would be sustained, metrologically traceable and comparable, and have a robustly determined and comprehensive total uncertainty budget. These uncertainties would be commensurate with the best practices in the Guide to Uncertainty in Measurements [JGCM, 2008]. In the real-world, the heterogeneity of different instruments and the complexity of requirements for observations (including process studies, long-term monitoring, real-time applications etc.) require, instead, a tiered system of systems architecture. Such an approach combines the advantages of high-quality achieved by a few selected reference-quality sites, with the ability of baseline networks to both provide a representative sampling and benefit from reference-network innovations, and then with denser coverage achieved by comprehensive observing networks.

The very best measurements that we can ever hope to make would have full metrological traceability to SI units or accepted standards, and have the smallest possible technological achievable associated total uncertainty budgets. These measurements have exacting requirements. Thus for both technical and financial reasons, their widespread and sustained deployment across the globe, at the required density to be the sole source of observations, is not feasible. This is particularly so when considering the myriad possible applications for measurements of the atmospheric, oceanic and terrestrial ECVs. There will always be a need for additional measurements, of lower absolute quality, to provide geographical and temporal detail. Such measurements are still useful for a broad range of applications, assuming that they are used appropriately. Some of these measurements will need to be sustained, to enable characterisation of regional variability and change for longer-term climate monitoring.

From the perspective of network operators, there are distinct advantages to a system-of-systems architecture. It provides an aspirational trajectory for sites, such that sites in a given tier can work towards promotion to a higher tier. It also provides a potential mechanism by which innovations in instrumentation and techniques can 'trickle-down', aiding all measurements and application areas.

Firstly, however, to maximize the return-on-investment of the currently available and future non-satellite observational capabilities portfolio, it is necessary to clearly define measurement capability tiers, which individual non-satellite observational programs can be placed into. In that way, users can employ the measurements appropriately and with confidence. It is, therefore, necessary to create criteria which are as objective as possible by which to designate a given candidate measurement series or measurement program into the most appropriate tier. Finally, mapping these capabilities in various ways can aid end users to make informed and appropriate decisions and analyses.

2.2 Proposed tiers for non-satellite measurement capabilities and possible system of systems benefits

It is proposed that GAIA-CLIM uses the tier designations defined in Seidel et al., 2009, and discussed further in GCOS, 2014 (Figure 1). The tier designation should be a function of demonstrable measurement qualities such as: traceability, metadata, comparability, data completeness, documentation, record longevity, measurement program stability and sustainability, etc.. Following the example of CORE-CLIMAX, it is intended that these aspects be assessed semi-quantitatively, through a combination of self-assessment and external-assessment of capabilities, against a consistently defined set of assessment criteria. Solely self-assessment may be possible for certain aspects, whereby only the network or site staff have the knowledge necessary to undertake the assessment. The assessment process has a range of benefits to both the institutions / individuals undertaking the measurements and to end-users, as will become apparent later. Sites or networks may both transition up and (hopefully less frequently) down between tiers and, as such, periodic reassessment is encouraged.

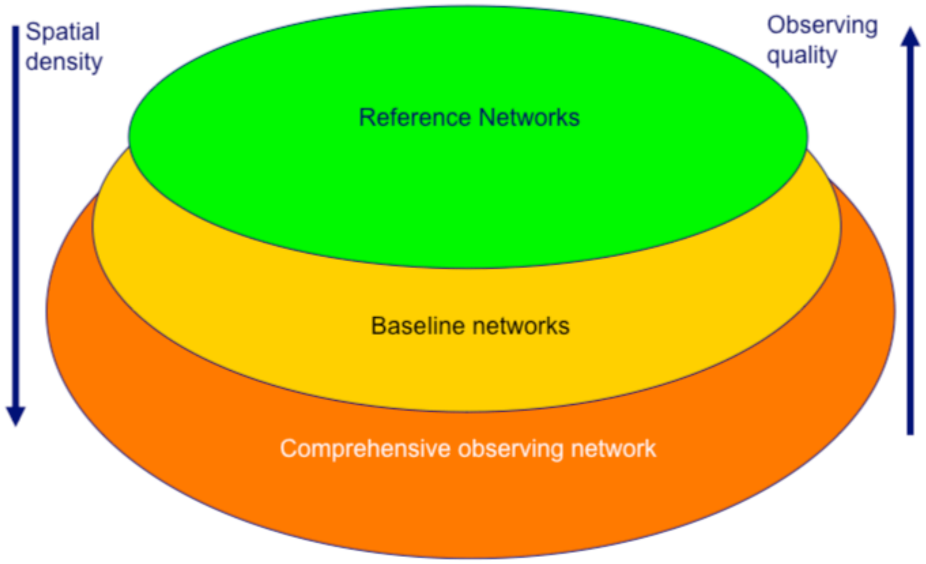


Figure 1. Proposed tiers in a system of systems approach to be adopted within GAIA-CLIM.

The starting point is a schematic view of measurements as an inherently interlinked "system of systems". In general terms, measurements typically involve a trade-off between properties such as fidelity & traceability (i.e. the degree to which the values reproduce the real-world state and have fully-characterised uncertainties), and properties such as representativeness (both in terms of sampling and resolution). The proposed system of systems recognises that resulting datasets and analyses / reanalyses are generated via a combination of measurements and subsequent analysis and computational protocols. Presently, there is a distinct trade-off between spatio-temporal data completeness and data fidelity. In part this arises because the synergies and benefits of a coordinated system of systems approach are not being realised.

GAIA-CLIM envisages a possible future in which fidelity and geographic completeness are improved for all components within the system of systems through robust, sustained and co-ordinated engagement, both between and within the different observing tiers. For many of the non-satellite systems, we still consider/manage them operationally as entirely independent networks. Take for example radiosondes, we have GRUAN, GUAN and the total network, which fits well into the proposed tiers. But very few of the National networks consider their locations and operational schedules as a component of an upper-air network incorporating radiosondes, profiling radars, aircraft, lidars etc.. There are exceptions for some subsets of observational capabilities. For example, EUMETNET tries to coordinate observations undertaken by European National Meteorological Services, and this and similar efforts may prove a model going forwards. Such sustained engagement would encompass aspects such as:

- Pro-active network design (including rationalisation of programs) and co-location of existing observing capabilities to maximise scientific return on investment;
- incremental improvements in instrument technology;

- step-change introduction of new measurement techniques;
- continued development, and greater adoption of, "best-practice" in all component systems;
- improved metrological characterisation and uncertainty quantification;
- iterative life-cycles of dataset generation, validation/evaluation and reprocessing; and
- better observationally constrained data assimilation systems through use of additional data streams and traceable observational uncertainty estimates.

Three essential elements for realising these improvements are:

- Sustained communication and coordination amongst the various tiers and the networks, both national and international, which contribute to them, with clear procedural protocols to ensure effective integration;
- robust operational frameworks capable of delivering iterative reassessments and reprocessing; and
- targeted research that will identify, and address, key obstacles and limitations.

Such an approach is beyond the remit of GAIA-CLIM funded activities and charter to achieve. Rather, it is more appropriately achieved through relevant global governance activities, such as the WMO Integrated Global Observing System (WIGOS), which was officially endorsed by the World Meteorological Organization at its 2015 Congress. The WIGOS concept explicitly envisages an integrated approach to the use of observing systems. The designation and adoption of the tiered approach and assessment criteria are a pre-requisite to realising this vision, to which GAIA-CLIM can contribute.

2.3 Tier defining characteristics

It is proposed that GAIA-CLIM defines the measurement capabilities in the following way (modified from GCOS, 2014).

2.3.1 Global reference observing networks

These networks provide metrologically traceable observations, with quantified uncertainty, at a limited number of locations, or for a limited number of observing platforms, for which traceability has been attained.

- The measurements are traceable through an unbroken processing chain (in which the uncertainty arising in each step has been rigorously quantified) to SI units, Common Reference Points defined by BIPM, or community recognised standards (ideally recognised by National Measurement Institutes), using best practices documented in the accessible literature.
- Uncertainties arising from each step in the processing chain are fully quantified and included in the resulting data. Combined expanded coverage factors (2 standard deviations of traceable uncertainty estimates which are referred to as expanded

coverage factors in the GUM), are reported for each data point. Individual components of the uncertainty budget are available. Where uncertainties are correlated, these are appropriately handled.

- Full metadata concerning the measurements is captured and retained, along with the original raw data, to allow subsequent reprocessing of entire data streams as necessary.
- The measurement and its uncertainty are verified through complementary, redundant, observations of the same measurand on a routine basis.
- The observations program is actively managed and has a commitment to long-term operation, to the extent possible.
- Change management is robust including a sufficient program of parallel and/or redundant measurements to fully understand any changes that do occur. Unnecessary changes are minimised.
- Measurement technology innovation is pursued. New measurement capabilities through new measurement techniques, or innovations to existing techniques, which demonstrably improve the ability to characterize the measurand, are encouraged. These innovations must be managed in such a way as to understand their impacts on the measurement series before they are deployed.

2.3.2 Global baseline observing networks

These networks provide long-term records that are capable of characterising regional, hemispheric and global-scale features.

- The baseline network is a globally and regionally representative set of observations capable of capturing, at a minimum: global, hemispheric and continental scale changes and variability. As such, a baseline network may be considered a minimum and highest priority subset of the Comprehensive networks, which should be actively curated and retained.
- The measurements are periodically assessed, either against other instruments measuring the same geophysical parameters at the same site or, alternatively / in addition, through intercomparison campaigns held under international or national auspices. These activities provide understanding of the relative performance of different techniques in use. Ideally, such intercomparisons should include reference quality measurements / networks, to realise scientific benefits.
- Representative uncertainties, that are based upon understanding of instrument performance or peer reviewed lines of evidence, are available.
- Metadata about changes in observing practices and instrumentation are retained.
- The observations have a long-term commitment.
- Changes to the measurement program are minimized and managed (by overlapping measurements, or measurements with complementary instruments over the change), with efforts made to quantify the effects of changes in an appropriate manner.
- The measurements aim to meet stakeholder stated requirements.

2.3.3 Comprehensive observing networks

These networks provide high spatio-temporal density data information necessary for characterising local and regional features.

- The comprehensive networks provide observations at the detailed space and time scales required to fully describe the nature, variability and change of a specific climate variable, if analysed appropriately. They include regional and national operational observing networks.
- Representative uncertainties based upon e.g. instrument manufacturer specification and knowledge of operations should be provided. In their absence gross uncertainties based upon e.g. expert or operator judgement should be provided.
- Metadata should be retained.
- Although encouraged, long-term operation is not required.

3. Objectively assessing measurement capabilities

The measurement system maturity matrix (SMM), like its counterpart for Climate Data Records (CDRs) developed under CORE-CLIMAX, is a tool to assess various facets of the maturity of a measurement. The matrices assess to what extent current (at time of production of the Guidance) measurement best practices have been met and, hence, the maturity of the candidate measurement system.

The measurement maturity is distinct from its applicability to a given problem, where additional concerns such as measurement location, frequency etc. pertain. Such aspects are end-user specific, and cannot be captured within the matrices detailed herein. However, the assessment results herein, in combination with such additional information, can be used to help inform users to decide upon the appropriate measurements for their use case.

The assessment can be performed either on individual instruments / sites, or for entire networks. A network will typically constitute a federated collection of sites, under the umbrella of an organisation that is generally recognised by the community. Examples are the GCOS Reference Upper Air Network, Network for Detection of Atmospheric Composition Change, and Total Carbon Column Observing Network. For sites, instruments and networks, the assessed measurement program may consider multiple measurement techniques and/or Essential Climate Variables. In some cases, it may be preferable to consider aspects of a network on a disaggregated level, either site-wise or instrument-technique-wise. Such an assessment is encouraged where it adds interpretative value, and should be agreed in the rules of the round phase (Section 3.2).

Finally, a note of caution: measurement best practices may well change in future, necessitating new versions of this Guidance. Please ensure you are using the most up to date version of this guidance, and ensure the specific guidance version used is retained as metadata alongside the assessment.

3.1 Maturity assessment concept

There are 6 mandatory major categories and one optional major category, where assessments are made, which overlap with, but are not identical to, those used to assess CDRs under the CORE-CLIMAX System Maturity Matrix approach. Where they overlap, in many cases the guidance differs substantially, to reflect the frequently substantial distinction between the measurements and derived CDRs. The strands for assessing measurement maturity herein are as follows:

- Metadata
- Documentation
- Uncertainty characterisation
- Public access, feedback, and update
- Usage
- Sustainability
- Software (optional)

The software option should be completed only for those measurements where substantive routine post-processing is undertaken, to convert the basic measurement to the finally presented geophysical time series. For example:

- the conversion of digital count data returned from a radiosonde to the ground segment to temperature and humidity profiles; or
- from backscattered photons collected and counted by a lidar to a geophysical profile of an atmospheric parameter, like aerosol extinction coefficient or water vapor mixing ratio.

Although this requirement to assess software maturity will often apply, there are many instances where it is not the case, such as standard meteorological surface station networks. In cases where anything more than very basic automated processing (such as resistance to temperature for a platinum resistance thermometer) of the measurements, from the measured parameter to derived parameters is being undertaken, the software strand should be completed. Otherwise, this strand should be noted as not relevant, with necessary justification being given instead in the assessment. Where a combination of external and internal assessments is being performed, assessors should agree on whether the software category strand is to be assessed ahead of time (Section 3.2).

Within each category are a number of sub-categories. For each of these sub-categories, the assessment will assign a score from 1 to 6, that reflects the maturity of the measurement with respect to that facet of the measurement system. The scores may help to inform a decision upon maturity of a given candidate measurement system. All aspects of the assessment are important. Weakness in any one strand will, inevitably, impact on the quality or usability of the measurements. For example, if the metadata and user documentation are assessed as weak, but uncertainty characterisation strong, there is reduced value in the observations, as the necessary context for end-users to use the measurements appropriately is missing.

3.1.1 Maturity scores and tiered networks concept

The maturity can, alternatively, be considered in three broad categories that give information on the scientific grade and sustainability of the measurements being assessed. This is similar to the CDR assessment in CORE-CLIMAX, which in turn, builds upon the earlier NOAA assessment process. However, the category definitions are fundamentally distinct from those for a CDR reflecting the real distinctions between CDR and measurement maturity considerations.

- Maturity scores 1 and 2 establish **Comprehensive Measurement Capability** (CMC, Comprehensive network type measurements): The instruments are placed in the field and recording data, but may not be well curated or metrologically understood and calibrated.
- Maturity scores 3 and 4 establish a **Baseline Measurement Capability** (BMC, Baseline network type measurements): At this stage the measurements are better characterised and understood, and intended to be run for the long-term. These may be considered a substantial, sustained contribution to the system of systems. However, they lack strict traceability and comparability.

- Maturity scores 5 and 6 establish a **Reference Measurement Capability** (RMC, Reference network type measurements): These measurements are extremely well characterised, with strict traceability and comparability, and robustly quantified uncertainties. The measurements are actively managed and curated, and envisaged as a sustained contribution to the observational system.

3.1.2 Interpreting the maturity assessment results

The major categories of the SMM are subdivided into several sub-categories, and assessment scores are assigned based on scores in these sub-categories. It should be noted that the numbers require interpretation for each assessed measurement series, because the circumstances under which the measurements were taken may affect what maturity level can be reasonably expected to be attained. A degree of expert judgment will, therefore, always be required to finally assign a measurement system into a given category, that reflects the totality of the assessment, including all relevant sub-category scores. All relevant sub-category scores should be considered to aid both data providers and users. In particular, data providers should consider low-scoring sub-categories as target areas for further work to improve the overall usefulness, accessibility, useability, and utility of their measurement program. Figure 2 provides a visual summary of the typical output that may accrue, and can be used to make a final assessment on measurement system maturity.

Metadata	**Documentation**	**Uncertainty charaterisation**	**Public access, feedback and update**	**Usage**	**Sustainability**	**Software (optional)**
Standards	Formal Description of Measurement Methodology	Traceability	Access	Research	Siting environment	Coding standards
Collection level	Formal Validation Report	Comparability	User feedback mechanism	Public and commercial exploitation	Scientific and expert support	Software documentation
File level	Formal Measurement Series User Guidance	Uncertainty Quantification	Updates to record		Programmatic support	Portability and numerical reproducibility
		Routine Quality Management	Version control			Security
			Long-term data preservation			

Legend

1	2	3	4	5	6	Not applicable

Figure 2. Hypothetical example assessment. For this example assessment it was agreed that the software strand was not applicable but that the two additional optional sub-categories were. Blacked out entries arise because not all major strands have the same number of minor categories.

Within Figure 2 it is possible to ascertain areas of both strength and weakness. In the hypothetical example given there is a clear lack of usage for non-research purposes, for example, which highlights a potential avenue to improve return-on-investment. Similarly, version control is assessed as lacking, and this points to an area that could be improved in future. Conversely, access, updates and preservation are rated highly, as is scientific use and support. From the data provider's perspective, such an assessment may inform strategic developments to the measurement program. From the data user's perspective, the assessment should provide an indication of applicability to their intended use case.

When considering an assessment of a network, rather than an individual site or instrument, in certain categories or sub-categories it shall be appropriate to perform the assessment on a per-asset (instrument or site) basis, rather than a network-wide basis. This is particularly the case for the Sustainability strand, but may also be applicable elsewhere if there are intra-network heterogeneities in protocols pertaining to e.g. metadata, uncertainty quantification or documentation. In such cases, and where practical, the assessment should be performed individually on each unique subset and stored in the assessment report metadata. Both the network wide mean score and range of scores should then be reported in the summary. Such a refined assessment helps ensure both appropriate network sub-selection for certain applications, and a fair assessment, that may help network operators and coordinators identify and address intra-network issues.

In the following subsections we provide instructions on how to assign scores to each of the sub-categories. The sub-categories sometimes include criteria that cannot easily be assessed by an external assessor without asking the provider of the data, a step that could be done in a formal audit type assessment.

3.1.3 Practical application considerations

The SMM is provided as a multi-level Excel file where the scores are input in the pages associated with the sub-categories. These scores are then automatically used to mark the range of scores for the major category. If a sub-category is not filled a maturity of 1 will be set. There are two exceptions: one in the category Usage and one in the category Sustainability.

1. In the Usage category, usage of a measurement is considered for applications in research and decision-making. Which columns are taken into account depends on the intention of the measurement system. For instance, if the description is only pointing to intended use in research, then that category alone shall be used to compute the overall usage maturity.
2. Within Sustainability, the siting environment is only applicable to fixed measurement assets that are always made from the same fixed location. This particular sub-category assessment should not be completed for mobile non-repeating observing assets such as aircraft measurements or field campaigns. However, observational assets that take repeated profiles, along a consistent transect, may be suitable to be assessed in this category.

Where either of these categories are not applicable, the entry in the equivalent plot to Figure 2 should be grey shaded to indicate its non-relevance rather than left blank.

It is very important to use a unique measurement system name and identification number (version) when the SMM is filled. This shall match the name and identification information on the measurement description form (Appendix A). Also, documentation of the assessment date, to follow the evolution in maturity of a particular measurement system, is very important if changes in measurement maturity are to be tracked through time. Sufficient assessment metadata should be appended to enable the tracking of multiple assessments of a candidate measurement system over time. This should include the version of this Guidance document that was used.

3.2 How to perform an assessment

Assessments should be repeated and refined on a multi-year cycle to capture both improvements and degradations in performance of the observing networks, and new insights. Thus ensuring that at any time the appropriate data are being employed to the appropriate scientific tasks. An assessment using the maturity concept should be conducted by an assessment leader that organises the assessment, provides needed guidance to the participants, and collects and analyses the results. It is likely to be useful to have a specific meeting to agree on the analysis results before publication.

It is intended that this guidance be updated relatively infrequently. The over-arching assessment framework in this document should remain stable for a considerable period of time, and not get substantively dated. This has required in many cases generic rather than specific guidance where details may reasonably be expected to change with evolving instrumental, metrological and community best-practices developments. For example, in the next section the guidance refers to "*appropriate high-quality metadata standards, which permit inter-operability of metadata*.", rather than referring to a current standard that may reasonably quickly become superseded. This is one of several examples where this guidance requires additional interpretation in the context of the state-of-the-art at the time of any assessment.

Where a substantive assessment of the state of multiple networks, instruments or sites is being organised it is therefore recommended to create an additional supplement of specific assessment criteria details or '*rules of the round'*, which provides additional guidance on such aspects. This guidance should be agreed by all participants, and should be retained alongside the completed assessments in such cases, to permit full interpretation of the assessment round results.

3.3 Metadata

Metadata is 'data' about data. Metadata should be standardised, as complete as possible, and adequately document how the measurement was attained. This involves aspects such as instrumentation, siting, observing practices etc. The measurement system should use appropriate high-quality metadata standards, which permit inter-operability of metadata. If an ISO standard is defined, then the assessment in future would be against such a standard.

However, at the present time no such universally agreed standard exists that pertains across all aspects of EO science. There are emerging efforts under WIGOS [WIGOS, 2015a,b] to create universal metadata standards[1], and there are several de facto working standards such as CF-compliant file headers. Unless and until an ISO standard is developed and applied, the assessors' judgement will be required as to the appropriateness of the standards being adhered to (see rules of the round sub-section above).

In this category the maturity is assessed using three sub-categories that consider the standards used, the metadata at the collection level, i.e., valid for the complete data record; and at file level, i.e., valid for the data at a specific granularity.

3.3.1 Standards

Standards – It is considered to be good practice to follow recognized metadata standards. These may differ depending upon the instrument or measurement program under consideration, and may be determined on a network / infrastructure-wide basis. As discussed previously currently no ISO-standard for metadata exists.

1	No standard considered
2	No standard considered
3	Metadata standards identified and/or defined and partially but not yet systematically applied
4	Score 3 + standards systematically applied at file level and collection level by data provider. Meets international standards
5	Score 4 + meta data standard compliance systematically checked by the data provider
6	Score 5 + extended metadata that could be useful but is not considered mandatory is also retained.

Table 2: The 6 maturity scores in metadata sub-category **Standards**.

Note: It is likely that this sub-category can only be fully assessed by the measurement initiator. An external assessment can be made by asking the data provider directly, or if the metadata and data are freely available from a portal (which would tend to indicate a mature measurement system). However, signs for used standards can be found by looking at the data record documentation and/or at a sample data file.

The assessment can be made as follows:

Score 1 and 2: No standard is considered. Data are made available solely as is with at most the geographical measurement location, time of observation and instrument type metadata provided that enables use, but prohibits measurement understanding.

[1] https://www.wmo.int/pages/prog/www/wigos/documents/Cg-17/Cg-17-d04-2-2(3)-add1-MANUAL-ON-WIGOS-approved_en.docx

Score 3: Standard identified/defined means that the measurement originator has identified or defined the standard to be used, but has not yet systematically applied it. The information about this most often can be found in Format description documents available from web pages, or from statements on web pages.

Score 4: A systematic application requires that you can find the relevant metadata protocol identifier and details in every file of the measurement product and descriptions.

Score 5: This means that the measurement provider has implemented procedures to check the metadata contents. This could be ascertained by a check on consistency of metadata header information in individual data files.

Score 6: This score will be attained if, in addition to mandatory metadata, additional optional metadata is collected, retained and transmitted. This score may not apply to some data streams where all metadata is considered mandatory but may help differentiate truly well performing measurement series in other cases, where metadata is differentiated into mandatory and optional classes such as the WIGOS metadata standards [WIGOS, 2105a,b] for example.

3.3.2 Collection level metadata (including change records)

Collection Level metadata – these are attributes that apply across the whole of a measurement series, such as processing methods (e.g., same algorithm versions), general space and time extents, creator and custodian, references, processing history etc. Discovery metadata through e.g. use of digital object identifiers, can form part of this and ensure long-term discoverability. Collection level metadata allows other people to find out what the measurement series contains, where it was collected, where and how the series is provided, and what usage restrictions apply.

1	None
2	Limited
3	Sufficient to use and understand the data independent of external assistance; Sufficient for data user to extract discovery metadata from metadata repositories
4	Score 3 + Enhanced discovery metadata
5	Score 4 + Complete discovery metadata meets appropriate (at the time of assessment) international standards
6	Score 5 + Regularly updated

Table 3*:* The 6 maturity scores in sub-category ***Collection Level***

The assessment can be made as below:

Score 1: Data files have no global attributes.

Score 2: Only attributes like location, space and time coverage, custodian of data, are provided, but no information on measurement/processing methods or history are available.

Score 3: All relevant information on processing (for example: software used, recording platform, raw data type) and for general understanding the data (such as references and comments) is provided. Also, contains information on how to extract discovery metadata from repositories.

Score 4: Score 3 + more information on discovery metadata (for example, how to obtain raw data and the necessary information to enable a user to reprocess those data). This may include relevant information such as instrument batch, set-up, time averaging period etc. and the availability of a data doi.

Score 5: Score 4 + all the available information on the data are provided with the data using an internationally recognized and agreed defined standard, that is appropriate to the measurement system in question at the time of the assessment. There may exist several such standards, and an appropriately agreed standard should be used if defined for the 'rules of the round'.

Score 6: Score 5 + Updates are provided whenever new metadata become available. For example: information on events impacting the quality of the measurement series, or the addition of commentary metadata such as publications written about the data record.

3.3.3 File Level

File level attributes are those specific to the granuality of the data (on a per measurement basis) and vary with each measurement entity. The file level metadata includes such elements as time of observation, location, measurement units, measurement specific metadata such as ground check data, measurement batch number, ambient conditions at time of observation etc.. Such metadata are necessary to understand and properly use the individual measurements.

1	None
2	Limited
3	Sufficient to use and understand the data independent of external assistance
4	Score 3 + Limited location (station, grid-point, etc.) level metadata along with unique measurement set metadata (e.g. batch, set-up, time, averaging period)
5	Score 4 + Complete location (station, grid-point, etc.) level and measurement specific metadata
6	

Table 4: The 6 maturity scores in sub-category ***File Level***

The assessment can be made as follows:

Score 1: Data files contain no variable attributes.

Score 2: Data geographical coordinates are described and data units are provided.

Score 3: The data files are provided with measurement geographical coordinates, units, valid range, and missing and/or fill values.

Score 4: Score 3 + measurement footprint details are provided. There is some location level (i.e., station level for an in situ data set, pixel level for a swath level satellite data) information available in the data files. An example for location level metadata is surface type. In addition there is information on the instrument batch, the instrumental set-up, measurement time and averaging period.

Score 5: Score 4+ additional location level metadata such as level of confidence in the retrieval for each data location is provided for a balloon ascent. Includes vicarious metadata, where necessary, to interpret the measurement, such as precipitation or cloud fraction for those measurement techniques potentially impacted.

Score 6: Not used. There is no innovation possible beyond Score 5.

3.4 Documentation

Documentation is essential for the effective use and understanding of a measurement record. There are three sub-categories to assess the completeness of user documentation. Note that the description of operations category used in the CORE-CLIMAX CDR maturity assessment model was not deemed applicable to measurements, and so is not utilised herein. Although the category has 3 sub-categories, it is possible that two or more of these categories may be covered by a single document for a given candidate measurement. For example, the formal description of measurement methodology may be written in such a way as to also constitute / contain a user guide.

3.4.1 Formal description of measurement methodology

Formal description of measurement methodology refers to a description of the physical and methodological basis of the measurements, network status (if applicable), processing of the raw data and dissemination. It shall often be used as a manual by the site technicians for how to take the measurements. For non-satellite measurement capabilities this can cover such aspects as descriptions of measurement principles, methods of observation, calibration procedures, data filtering, data processing, corrections, aggregation procedures, data distribution etc.. As such documents are most often grey literature, it is required to also have a peer-reviewed publication(s) on the methodology to increase the maturity. Where software is involved in the processing of the data, its availability should be assured. For measurements that involve substantial post-processing to get from the raw measurement to the processed measurement series, the optional software elements strand (Section 3.7) should be completed.

1	Limited scientific description of methodology available from data collector or instrument manufacturer
2	Comprehensive scientific description available from data collector or instrument manufacturer

3	Score 2 + Journal paper on measurement methodology published
4	Score 3 + Comprehensive scientific description available from Data Provider
5	Score 4 + Comprehensive scientific description maintained by Data Provider
6	Score 5 + Journal papers on measurement system updates published

Table 5: The 6 maturity scores in sub-category ***Formal description of measurement methodology***

The assessment can be made as follows:

Score 1: Documentation of the measurement technique principles and processing chain is available and discoverable, e.g. on the Internet using a recognized search term.

Score 2: Complete documentation of the measurement technique and processing steps is available, which includes all the steps that were used to process from the raw measurement basis, such as digital counts, to the final product, such as a temperature profile.

Score 3: In addition to Score 2 a journal paper in a recognised appropriate scholary journal, outlining the measurement principles and processing is available. This can be checked using tools such as Web of Science.

Score 4: Measurement technique information sufficient for a third party to reproduce the measurement at another location is available from the measurement provider, e.g., an instrument manual describing how to take the measurements, and any necessary processing software package is available.

Score 5: This score is related to updates of the documentation, following updates of the measurement techniques or metadata (see Public Access, Feedback and Update). A sign for maintenance is if the instrument manual has proper document version numbering and is referring to a specific version of the measurement series record.

Score 6: Each substantive update to the measurement technique is published in the peer reviewed literature.

3.4.2 Formal validation report

A **Formal validation report** contains details on the validation activities that have been undertaken to assess the fidelity / reliability of the measurement record. It describes uncertainty characteristics of the measurement record found through the application of uncertainty analysis (see section on Uncertainty Characterisation), and provides all relevant references.

1	None
2	Informal validation work undertaken.
3	Instrument has participated in certified intercomparison campaign and results available in grey

	literature
4	Report on intercomparison to other instruments, etc.; Journal paper on product validation published
5	Score 4 + Sustained validation undertaken via redundant periodic measurements
6	Score 5+ Journal papers describing more comprehensive validation, e.g., error covariance, validation of qualitative uncertainty estimates published

Table 6: The 6 maturity scores in sub-category ***Formal validation report***

Score 1: No validation is done, and hence no report;

Score 2: Report on limited validation activities, undertaken using other measurement techniques, or by comparison to vicarious measurements or relevant model-based analyses / reanalyses is available, but no formal published validation / characterisation of the measurement series exists.

Score 3: The measurement technique has been evaluated in a formally recognized national or international intercomparison or validation campaign. For example for a radiosonde the model has participated in either a CIMO (Commission for Instruments and Methods of Observations) intercomparison, or a regional comparison that includes instruments that participated in one or more such CIMO campaigns. The results of the comparison or validation are available in a suitable report, but are not peer reviewed, and the comparison data is available for analysis.

Score 4: The measurement technique has been evaluated and validated using appropriate techniques, and compared to other independent techniques that measure the same measurand and have similar maturity. Analyses verifying the performance of the measurement technique are available in the peer-reviewed literature.

Score 5: The measurement technique is regularly validated using appropriate techniques, and regularly contributes to internationally recognised intercomparison activities. These validation reports are publicly available although may not be peer-reviewed.

Score 6: More papers on instrument characterisation are published and measurement developer/provider maintains up-to-date information on the validation activities and resulting uncertainty estimates in their data series.

3.4.3 Formal measurement series user guidance

Formal measurement series user guidance – This document contains details necessary for measurement users to discover and use the data in an appropriate manner. It includes aspects such as the technical definition of the measurement series, overview of instrumentation and methods, general quality remarks, validation methods and estimated uncertainty in the data, strength and weakness of the data, format and content description, references, and processing details. It may be that this same documentation also constitutes the formal description of measurement technique.

1	None

2	Sufficient information on the measurements available to allow user to ascertain minimum set of information required for appropriate use
3	Comprehensive documentation on how the measurement is made available from data collector or instrument manufacturer including basic data characteristics description
4	Score 3 + including documentation of manufacturer independent characterisation and validation
5	Score 4 + regularly updated by data provider with instrument / method of measurement updates and/or new validation results
6	Score 5 + measurement description and examples of usage available in peer-reviewed literature

Table 7: The 6 maturity scores in sub-category **Formal Measurement Series User Guidance**

The assessment can be made as follows:

Score 1: Data collector / instrument manufacturer has not provided any documentation on the measurements and how they were taken.

Score 2: There is sufficient information regarding the measurements and how they were taken to enable informed use of the data, for at least some applications. However, the information is not complete.

Score 3: A reviewed (for example by the data provider) set of documentation is available from data collector's, network's or instrument manufacturer's webpages. The documentation is complete.

Score 4: Score 3 + the documentation includes steps that have been undertaken to independently characterise the instrument performance. For example, the use of an ice bath to calibrate a thermometer, or a well-characterised lamp check for a lidar.

Score 5: Score 4+ Updated guidance is available from data provider's web page. A sign of updating is increasing version numbering and date. This is related to both updates in the measurement technique itself and its understanding. This may include new validation techniques, or results or new methods of observation and their impact.

Score 6: Score 5+ the measurement technique description is published in the peer reviewed literature, and there are one or more example usage applications documented either in the description paper or subsequent application papers.

3.5 Uncertainty characterisation

The category Uncertainty Characterisation assesses the practises used to characterise and represent uncertainty in a measurement series. Four sub-categories are considered with the aim to encompass traceability, the validation process, how uncertainty is quantified, and if an automated quality monitoring process is implemented that increases the efficiency of production and validation. <u>Note that uncertainty nomenclature and practices must follow established definitions [JGCM, 2008 or any subsequent updates to this] to attain a score of 5 or 6 in any of the sub-categories</u>.

3.5.1 Traceability

Traceability is the property of the result of a measurement whereby it can be related to stated references, usually national or international standards such as SI units, through an unbroken chain of comparisons, and these processing procedures all have stated / quantified uncertainties. To support a claim of traceability, the provider of a measurement must document the measurement process or system used to establish the claim, and provide a description of the chain of comparisons that were used to establish a connection to a particular stated reference. Any measurement claiming SI traceable means that any unit used shall be traceable back to the seven well-defined base units of the SI system: the metre, the kilogram, the second, the ampere, the Kelvin, the mole, and the candela. Alternatively, traceability can be attained to recognized community standards, where SI traceability is not possible. Full traceability on a sustained basis requires in-depth instrument understanding and regular comparisons to standards, and will typically involve and be certified by National Measurement Institutes. A fully traceable measure shall always have an associated total uncertainty budget that accounts for the uncertainty arising in all of the processing steps.

1	None
2	Comparison to independent stable measurement or local secondary standard undertaken irregularly
3	Score 2 + independent measurement / local secondary standard is itself regularly calibrated against a recognized primary standard
4	Score 3 + processing steps in the chain of traceability are documented but not yet fully quantified.
5	Score 4 + traceability in the processing chain partly established
6	Score 5 + traceability in the processing chain fully established

Table 8: *The 6 maturity scores in sub-category **Traceability***

The assessment can be made as follows:

Score 1: No attempt has been made to ascertain the absolute or relative performance of the measurements.

Score 2: Periodic comparisons are made against secondary standards to ascertain drift or gross biases. For example, a temperature sensor is compared to the reading from a thermometer shelter, or a lidar is calibrated against a stable lamp or radiosonde profile. This permits traceability to a secondary standard, which is stable but of unknown absolute quality.

Score 3: Score 2 + the independent comparison measurement is itself periodically calibrated against a primary standard from a National Measurement Institute, or other holder of certified primary measurement standards. Continuing the first example under Score 2, the shelter thermometer is periodically calibrated against an NMI certified calibration thermometer.

Score 4: Score 3 + the processing steps in the traceability chain from the fundamental measurement to SI or community recognized standards have been identified, and at least gross estimates for the uncertainties in some of these steps have been estimated.

Score 5: Score 4 + many of the processing steps in the measurement are understood and quantified in a rigorous manner.

Score 6: Score 5 + the traceability is fully established and verified, and a peer reviewed paper describing the measurement series and its uncertainty is published.

3.5.2 Comparability

Comparability - This category evaluates the extent to which the product has been validated to provide realistic uncertainty estimates and stable operations through in-the-field comparisons. Such validation is substantively distinct from traceability in that it relates to a sustained program of comparison both in the measured environment, and using lab-based experiments to ascertain potential biases, drifts and artefacts between two measurements. Unlike for traceability, the comparison need not be to a measure that itself is traceable directly or indirectly to SI or community standards. However, for the highest quality measurements such comparisons should be against measurements that are themselves traceable. This could be through intercomparison campaigns, with fixed or mobile standards available in the network, or through complementary traceable measurements using distinct techniques on a sustained basis.

1	None
2	Validation using external comparator measurements done only periodically and these comparator measurements lack traceability
3	Score 2 + Validation is done sufficiently regularly to ascertain gross systematic drift effects
4	Score 3 + (Inter)comparison against corresponding measurements in large-scale instrument intercomparison campaigns
5	Score 4 + compared regularly to at least one measurement that has a traceability score >=5
6	Score 5 + compared periodically to additional measurements including some with traceability assessment >5

Table 9: The 6 maturity scores in sub-category **Comparability**

The assessment can be made as follows:

Score 1: No validation activity has been performed on the measurements.

Score 2: The measurement is validated only periodically. For example, there are annual comparisons to a similar instrument that does not have SI traceability as part of routine maintenance.

Score 3: Regular comparisons to a similar measurement, or appropriate characterisation technique, to ascertain measurement relative performance in a sustained manner. For example, ground-checks for radiosondes using manufacturer standard ground-check recalibrations, or regular comparisons of a lidar system to radiosondes launched contemperaneously.

Score 4: Score 3 + instrument is characterised against other similar instruments or instruments measuring the same measurand in intercomparison campaigns such as e.g. the CIMO intercomparison for radiosondes, the screen temperature / humidity comparisons carried out in Algeria, or radiometer intercomparisons at Davos. Ideally such comparisons shall be carried out in a range of environments (tropical, sub-tropical, temperate, polar), to ascertain environmental effects.

Score 5: Score 4 + compared to well characterised measurements from an independent technique or instrument on a regular basis.

Score 6: Score 5 + compared to fully traceable measurements on a periodic basis to provide robust quantification of absolute biases and drifts.

3.5.3 Uncertainty quantification

Uncertainty quantification - This sub-category evaluates the extent to which uncertainties have been fully quantified and their ease of use.

1	None
2	Limited information on uncertainty arising from systematic and random effects in the measurement
3	Comprehensive information on uncertainty arising from systematic and random effects in the measurement
4	Score 3 + quantitative estimates of uncertainty provided within the measurement products characterising more or less uncertain data points
5	Score 4 + systematic effects removed and uncertainty estimates are partially traceable
6	Score 5 + comprehensive validation of the quantitative uncertainty estimates

Table 10: The 6 maturity scores in sub-category **Uncertainty quantification**

The assessment can be made as follows:

Score 1: No validation, and therefore no uncertainty quantification.

Score 2: Only limited information on uncertainty is available because of limited validation, but it is possible to partition random and systematic effects.

Score 3: Comprehensive information is available, so that the nature of the uncertainty is well understood. For example, whether the uncertainty varies depending upon: geographic region, atmospheric state, and instrument geometry. Uncertainties are estimated for each step of the measurement production.

Score 4: Score 3 + quantitative comprehensive information described in Score 3 is available for each data point of the measurement profile or series.

Score 5: Score 4 + the systematic effects are removed and uncertainty estimates are partially traceable to SI or community accepted standards. In addition, where applicable, the correlated and uncorrelated uncertainty terms in the measurement series or profile are quantified. For example, the calibration of an instrument may be an uncertainty that is absolutely correlated, whereas the effects of fluctuating cloud cover may be uncorrelated or partially correlated in the series.

Score 6: Score 5 + the uncertainty estimates are fully traceable and validated, using other high quality traceable data, on a sustained basis.

3.5.4 Routine quality monitoring

Routine quality monitoring is the monitoring of data quality while processing the data. Quality monitoring is a robust and quantitative measure of how closely an individual measurement conforms to an expectation against which the observations can be compared and assessed. Such quality monitoring helps to assess, in near real time, major issues with the measurements, and permits proactive management. It may lead to a stop and restart of processing activities or measurement series if any type of error is detected. In that sense it can save significant resources whilst minimizing bad data volumes, and is a clear sign for a mature observing system with active management.

Routine data quality monitoring may require an integrated approach that includes several steps, depending on the level of complexity of quality assurance procedures. This is directly linked to the complexity of the calibration procedures required for each measurement technique, and on the level of complexity of the processing chain (see optional assessment area Software). Moreover, robust data quality monitoring also depends on the availability of co-located redundant measurements, or high quality estimates based upon e.g. data assimilation based short-term forecasts. Such data facilitate the assessment of the data quality through inter-comparison of different time series, and through the development of higher-level synergistic products.

Monitoring of data quality control can be manually applied by site operators and scientists, or performed automatically, or both. Quality checks are typically realized through a flagging system applied to the data. Such a system shall typically include several or all of the following steps.

1. Data file format checks: catch files with missing metadata or data, incorrect data formatting, or any other type of gross errors.

2. Consistency checks: identify unreliable values based upon our understanding of the physics of the considered ECV. For example, negative relative humidity values or values that exceed substantially 100% for a sustained period cannot be correct.

3. Calibration: verify that calibration procedures have been applied and recorded for each measurement technique following traceable procedures and, when possible, performed using different calibration approaches and reference tools. This step may also include the provision of maintenance information, and reports on the expected and the actual Instrument performance by the site operators and scientists.

4. Uncertainty: identify those data whose uncertainty is beyond thresholds considered useful for most intended applications. Such thresholds may be application specific, and depend upon to what extent the uncertainties can be segmented into systematic and random components.

5. Retrieval chain: ensure that all the processing steps from the basic data to processed products have successfully completed; this also includes the number of corrections typically applied to the data, as required by each measurement technique (e.g. multiple scattering, gas absorption, multipath corrections, radiation bias corrections etc.). If an automatic data processing is used, checks are implemented in the calculus chain.

6. Redundancy checks: measurement intercomparisons and cross-checking with other techniques measuring the same ECV, if physically co-located. In addition, the calculation of site atmospheric state best estimates, that combine information from several synergistic measurement platforms, can help to learn more about measurements health status. Such activities can augment the routine checking by providing an estimate of the utility of data streams. These higher-level checks can also point out deficiencies that are not necessarily detectable within individual data stream checks.

7. Time series analysis: routine near real time analysis of the collected time series may help identifying inconsistencies and mistakes in the applied procedure, or non-physical anomalies in the measurement series. Intercomparisons of co-located redundant measurements may also help in investigating time series.

8. Collection of feedback, through the implementation of a website with a combination of: diagnostic plots browser with thumbnail views, an interactive plotting capability, a data quality documentation, a problem reporting system and instrument and maintenance logs.

Data quality flags should be applied without rejecting data as subsequent innovations in instrument understanding may permit reprocessing and recovery of good values.

1	None
2	None
3	Methods for routine quality monitoring defined
4	Score 3 + routine monitoring partially implemented
5	Score 4 + monitoring fully implemented (all production levels)
6	Score 5 + routine monitoring in place with results fed back to other accessible information, e.g. meta data or documentation

Table 11: The 6 maturity scores in sub-category **Routine quality monitoring**

The assessment can be made as follows:

Score 1: No automated quality monitoring in place.

Score 2: No automated quality monitoring in place.

Score 3: A metric (e.g., radiometric noise of one or more channels of the instrument used is significantly above specification, number of good measurements is below a threshold value, agreement between duplicate measurements, measurements fail to attain stated height requirements, procedures, data used in comparisons, setting of thresholds for deviations, etc.) for routine quality monitoring has been defined.

Score 4: Score 3 + the proposed monitoring is partially implemented, e.g., for a subset of the measurements that contribute to a global collection but not to the remainder.

Score 5: Score 3 + quality monitoring is implemented for all the measurements. Variants in performance are reported to the technicians undertaking the measurements and resolved in a timely manner.

Score 6: Score 5 + Results of routine quality monitoring are reflected in metadata and documentation. For example, the quality monitoring procedures and results are described in the peer reviewed or grey literature.

3.6 Public access, feedback and update

This category contains five sub-categories related to archiving and accessibility of the measurement record, how feedbacks from user communities are established, and whether these feedbacks are used to update the measurement record. It also concerns version control and archival and retrieval of present and previous versions. A mature measurement system would be available routinely to allow operational use, with formal version control, and mature archival procedures. Furthermore, a mature measurement data stream would have an established mechanism to collect, and act upon, user feedback.

3.6.1 Access

Access evaluates the ease of distributing the raw and processed data, documentation, and any necessary source code used to process the data from the raw measurement to geophysical or radiance parameter space, to users. Public access means that the data are available without restrictions for at least academic use, but such access may still be subject to a reasonable fee. The raw data may only be provided upon request, but a mechanism for requesting should be readily apparent in such cases. The highest scores in this category can only be attained for data provided free of charge without restrictions on use and re-use.

Data provider here means either the data collector or organisations such as space agencies, national meteorological centres or research institutes. An institutionalised data provision is considered to be more robust (and hence mature), compared to the provision by an individual investigator or group.

1	Data may be available through request to trusted users
2	Data available for use through originator
3	Data and documentation available through originator

4	Score 3 + available through recognized data portal
5	Score 4 + source data, code and metadata available upon request
6	Score 5 + no access restrictions apply

Table 12: 6 maturity scores in sub-category **Access**

The assessment can be made as follows:

Score 1: Measurement record is not ready yet to be given to users; it may be available to beta-users for testing. Data originator is still conducting initial validation of the observed product.

Score 2: Measurement record is now ready to be given to users without any restrictions on academic usage. Users can get the measurement data either by requesting it from the data originator, or from a publicly accessible site.

Score 3: Measurement series and appropriate documentation to understand the measurements is publicly available for academic use through either the data provider or a publicly accessible site. Academic re-use is permitted.

Score 4: As Score 3 + measurement series are available through a recognised and measurement-appropriate data portal such as the Copernicus Climate Change Services Data Portal, NDACC portal, or NOAA's National Centers for Environmental Information.

Score 5: As Score 4 + the source data, metadata and any processing code is also archived by the data provider, allowing subsequent reprocessing of the full measurement series if required by a third party.

Score 6: As Score 5 but there are no restrictions on use or re-use of the data, metadata, or code, and all aspects are made available free of charge.

3.6.2 User feedback mechanism

User feedback is important for developers and providers of measurement records to improve quality, accessibility, etc. of a given measurement series. This category is to evaluate whether mechanisms are established to receive, analyse, and use user feedback. Feedback can reach a measurement provider in many ways, but needs to be organised in such a way that it can be used to improve a measurement record and/or the service around it. In the scientific community, measurement records are presented and discussed at workshops and conferences. A scientist may take messages back to his/her lab and start to think and realise improvements, if resources are available. A higher maturity for gathering feedback is obviously reached when a measurement record has been institutionalised and the responsible institute has established regular feedback processes.

1	None
2	Ad hoc feedback

3	Programmatic feedback collated
4	Score 3+ consideration of published analyses
5	Established feedback mechanism and international data quality assessment results are considered
6	Score 5 + Established feedback mechanism and international data quality assessment results are considered in continuous data provisions

Table 13: 6 maturity scores in sub-category **User feedback mechanism**

The assessment can be made as follows:

Score 1: Measurement record is intended as what you see is what you get, and so no feedback mechanism is constituted.

Score 2: Ad hoc feedback received and may be acted upon.

Score 3: A programmatic collection of user feedback is instigated that may relate to a broad network of measurements, and lessons learnt are disseminated either formally or informally periodically.

Score 4: Score 3 + the measurement program takes into account findings documented in the peer reviewed literature.

Score 5: The measurement program has a well-established and recognized system for the collection of metadata, which allows users to provide and track feedback. The results of international comparisons and campaigns are considered.

Score 6: An international review panel (such as a network task team or management group) that meets regularly would indicate a mature system, that took account of innovations and feedback. A further sign of this is to check whether interim data records are provided (operational continuation of a measurement record employing the same procedures), and if feedback is also considered for this.

3.6.3 Updates to record

Updates to record evaluates if data records are systematically updated when new observations or insights become available, or if this is done in ad hoc fashion if at all. A more ad hoc update cycle is indicative that the update very much depends on irregular funding, and is not done by a bigger institution that provides the update as part of an operationally oriented service. More mature measurement series will tend to be updated in an operational manner that assures both their sustainability and their suitability for applications requiring reliable data updates. The most mature measurement systems distribute data in near real-time so that it can be used in forecasting applications.

1	None
2	None

3	Irregularly following accrual of a number of new measurements or new insights
4	Regularly updated with new observations and utilising input from established feedback mechanism
5	Regularly operationally by stable data provider as dictated by availability of new input data or new innovations
6	Score 5 + initial version of measurement series shared in near real time

Table 14: 6 maturity scores in sub-category **Updates to record**

The assessment can be made as follows:

Score 1: No update is made to the measurement series after initial release.

Score 2: No update is made to the measurement series after initial release.

Score 3: There are irregular updates to the measurement series record available to the public. Such updates may result from user feedback, innovations in understanding, or simply constitute a string of new measurements. Such updates are made in an ad hoc (un-timetabled) manner.

Score 4: This can be seen by regular updates for the measurement records, accompanied by documentation of updates at reasonable frequency. For example, a regular daily, monthly or annual update occurs to append new observations. Updates periodically include innovations to account for user feedback. In cases where no feedback has been received, despite a facility for feedback being made available, this should be stated.

Score 5: The updates to append data are made on a stated regularity, allowing the operational usage of the measurement series in applications. Updates periodically take into account methodological innovations that improve the utility of the measurement series. Such updates are clearly differentiated from straight data updates.

Score 6: Score 5+ a version (which may not be the final processed version) is made available in near real time (typically defined as within 2-3 hours) for applications that can make use of this information for forecasting purposes.

3.6.4 Version control

Version control allows a user to trace back the different versions of algorithms, software, format, input and ancillary data, and documentation used to generate the measurement record under consideration. It allows clear statements about when and why changes have been introduced, and allows users to document the precise version of the data they used, thus enabling replication of users' analyses. Typically, a mature version control will have a documented version control protocol that is openly documented and may include in addition to version number a date stamp on each version. The most mature version control should allow users to retrieve previous versions if required.

1	None
2	None
3	Versioning by data collector
4	Version control institutionalised and procedure documented
5	Fully established version control considering all aspects
6	Score 5 + all versions retained and accessible upon request

Table 15: Six maturity scores in sub-category - **Version control**

The assessment can be made as follows:

Score 1: No versioning system in apparent use for the measurement series.

Score 2: No versioning system in apparent use for the measurement series.

Score 3: The measurement series has an informal version control undertaken by, and documented by, the data collector that is used internally to document versions;

Score 4: Data version control is transferred from the data collector to an institutionally maintained archive, and formalised. The version control protocol shall be documented. For example a versioning N.x.y.z might be instituted, and the reasons for incrementing any of N x, y, or z will be clearly articulated.

Score 5: Data provider has established full version control for the measurement record including versions of algorithms, software, format, input and ancillary data, and documentation.

Score 6: Score 5 + all historical versions, since instigation of version control, can be made available to interested users upon request.

3.6.5 Long-term data preservation

Long-term data preservation relates to the preservation of measurement series records. According to Long Term Data Preservation (http://earth.esa.int/gscb/ltdp/) guidelines an archive should keep more than one copy, use different media/technologies, and different locations. Most important is to retain the raw data (e.g. the solar spectral measurements of an FTIR) and necessary metadata, which may allow subsequent reprocessing.

1	None
2	None
3	Local archive retained by measurement collector
4	Each version archived at an institutional level on at least two media

5	Data, raw data and metadata is archived at a recognised data repository such as a National Meteorological Service, national archive or international repository.
6	Score 5 + all versions of measurement series, metadata, software etc. retained, indexed and accessible upon request

Table 16. Six maturity scores in sub-category – **Long-term data preservation**

The assessment can be made as follows:

Score 1: No archiving system in apparent use for the measurement series.

Score 2: No archiving system in apparent use for the measurement series.

Score 3: The measurement series has a local archive, maintained by the instrument data collector, which may be used to retrieve data on an ad hoc request basis, but is dependent upon the data collector or a single small group.

Score 4: Data archival is transferred from the data collector to an institutionally maintained archive and formalised. The data is preserved on at least two media, in two distinct locations.

Score 5: Data archival is undertaken by a recognised institution with expertise in data preservation. The preservation extends to raw data, metadata, software, and data versions.

Score 6: Score 5 + all historical versions since instigation of archival can be uniquely identified, and made available to interested users upon request.

3.7 Usage

This category contains two sub-categories related to the usage of measurement series in research applications and for decision support systems. Public and commercial exploitation means the use in applications that directly support economic or public decisions, e.g., a radiosonde measurement may be used in an NWP model or forecast assessment, or an ozone measurement may be used to monitor stratospheric ozone conditions, and hence the effectiveness of the Montreal Protocol and its amendments. In addition all usages in creating climate data records, and citations in reports, such as the Intergovernmental Panel for Climate Change (IPCC) reports, that support decisions and policy making on mitigation and adaptation are countable for the public and commercial exploitation sub-category.

The two sub-categories allow for a separate assessment of the usage of measurement records, i.e., the assessment result can state a high maturity for usage in research, and a lower or no maturity for public and commercial exploitation. For the overall score, it is important to know for which application area(s) the measurement was intended. This information shall come from Section 1 of the GAIA-CLIM Measurement Record Description Form (see Appendix A). If this description is only pointing to use in academic research, then only that category shall be used to display the overall maturity for this category.

3.7.1 Research

Research applications of a measurement series can be evaluated by its appearance in publications and citations of such publications.

1	None
2	Benefits for research applications identified
3	Benefits for research applications demonstrated by publication
4	Score 3 + Citations on product usage occurring
5	Score 4 + product becomes reference for certain applications
6	Score 5 + Product and its applications become references in multiple research fields

Table 17: 6 maturity scores in sub-category **Research**

The assessment can be made as follows:

Score 1: Measurement series is not used yet.

Score 2: An available research plan, or similar document, outlines actual or intended usage of the measurement series in research applications.

Score 3: A peer reviewed publication exists, that describes the usage of the measurement series in a research application.

Score 4: The peer reviewed publication under score 3 is cited by peer reviewed publications of other applications.

Score 5: The measurement series is used as a reference / contributing series in almost all peer reviewed publication for a specific application.

Score 6: The measurement series is used as reference in almost all peer reviewed publication for applications in different research fields, e.g., climate modelling and climate system analysis.

3.7.2 Public and commercial exploitation

As described above under usage for **Public and Commercial Exploitation** covers any direct use in real-time monitoring, forecasts, infrastructure planning, support to agencies or other business areas such as insurance and indirect support, e.g., through citations in IPCC reports, to decision and policy making in socio-political contexts.

1	None
2	Potential benefits identified
3	Use occurring and benefits emerging

4	Score 3 + societal and economical benefits discussed, data being distributed via appropriate data portals.
5	Score 4 + societal and economical benefits demonstrated
6	Score 5 + influence on decision (including policy) making demonstrated

Table 18: 6 maturity scores in sub-category **Public and Commercial Exploitation**

The assessment can be made in the following manner:

Score 1: Product is not used yet for any public or commercial application.

Score 2: An available report suggesting that the measurement series can be used for certain public or commercial applications exists, and can be found online or in a recognised repository.

Score 3: Product has been used in public and / or commercial applications, and a report(s) is available through appropriate data portals for use. For example, the data is available via the Climate Data Store of the Copernicus Climate Change Service, or is used in NWP or reanalyses.

Score 4: The results of studies in Score 3 are used for a relevant public or commercial system. For example, a state or national government report on the planning is available, which cites the study using the measurements under consideration, or the forecast resulting from their use enables decisions by public and commercial actors.

Score 5: The results of studies in Score 4 are used in an application area, and have resulted in demonstrable societal and economic benefits.

Score 6: Substantive contribution to national and international public decision making, and applications such as climate policy discussions or to economic applications. One can also point to the use of a measurement series in other applications, which have economical benefits, such as use by an insurance company for decision making or use in a climate service, e.g., the major application areas mentioned in the WMO Global Framework of Climate Services (agriculture and food security, disaster risk reduction, health and water).

3.8 Sustainability

This category pertains to aspects of sustainability, and hence suitability, of any given measurement program for scientific, operational, and societal applications. For a measurement program to be used in critical applications, its long-term sustainability must be assured. There are three primary strands to sustainability of a measurement program that relate to: siting environment, scientific and expert support, and programmatic (funding) support.

Where an international measurement network is being assessed, the network shall typically consist of individual measurement sites operated by distinct legal entities, with distinct funding mechanisms, and in a variety of siting environments. In such cases, there are two options. One is to provide a typical score, that is representative of the network as a whole, but this is then not indicative of the maturity of individual contributing sites. The alternative, preferred option, is that this assessment be performed site-wise, with the site-by-site scores retained as metadata associated with the assessment, and the range of scores highlighted appropriately in the assessment summary by providing both a mean value and the range. The latter approach is preferred because it enables, for example, applications that require a representative sampling environment, to use the site-by-site scores metadata provided to retain only the appropriate subset of the network that is sited in regionally representative locales. A site-by-site assessment also avoids conflating contributing entities with long-term commitment with other contributiors which may be less secure. This then helps network coordinators to highlight potential areas for within-network improvement / remediation. The range of individual site scores across the network may also provide a useful indicator of the overall maturity of the network.

3.8.1 Siting environment

Siting environment only applies to fixed measurement assets, for which observations are taken repeatedly from a single location (including weather balloons which originate from a constant location but may drift), or mobile observations using repeating transects. Non-repeating measurements made from aircraft and other mobile platforms should leave this entry blank, and use solely the remaining strands to assign a score under sustainability. Within this category, consideration is limited to the representativeness of the site / transect of its immediate surrounding environment / landscape. Questions of network design are outside the scope of this maturity assessment, although clearly are important in network design and expansion considerations.

1	None
2	Site environment is stable in the short term
3	Score 2 + site ownership is sustainable
4	Score 3 + Site is representative of a broader region around the immediate location
5	Score 4 + site ownership, immediate environment is likely to be unchanged for decades
6	Score 5 + long-term ownership and rights are guaranteed

Table 19. Six maturity scores in category **siting environment**

The assessment can be made in the following manner:

Score 1: No information is available about the siting of the instrument used to make the measurement, or its representativity of the local surroundings and their environmental conditions.

Score 2: The instrument location is known and characterised by photography, satellite imagery or other means and the environment unlikely to be modified, beyond maintaining the environment stability by e.g., grass mowing, tree and bush management etc., by direct human influence in the short-term.

Score 3: As Score 2, plus the ownership of the site is sustainable such that the measurement program is viable at the specific location for the foreseeable future.

Score 4: As Score 3, plus the site is representative of a broader region surrounding its immediate location. Here broader region may be application and ECV dependent. For use in satellite characterisation (the purpose of GAIA-CLIM) this may extend to a typical satellite pixel field of view, for example, where the thermal, albedo and other surface characteristics are sufficiently homogeneous for the measurement to be deemed representative.

Score 5: As Score 4, plus the site ownership and the immediate surrounding environment is likely to be unchanged for decades. Evidence for this may arise from planning documents, government ownership, or other relevant national land designations.

Score 6: As Score 5 but the long-term site ownership and management is assured. For example the measurement is undertaken on managed government property that is protected by statute.

3.8.2 Scientific and expert support

Scientific and expert support evaluates the degree of scientific, technical and measurement science expertise that underpins the measurement programme. Higher quality networks will benefit from sustained curation, development, and exploitation that typically arises from a strong infrastructure support basis, and a continuous recruitment policy, that is able to fill in the personnel and skills gaps that might occur.

1	None
2	Minimal scientific support required to sustain the program is available
3	Relevant instrument expertise is available to support the measurements
4	Score 3 + at least two experts available to support the measurement program operation
5	Active instrumentation research and development being undertaken
6	

Table 20. Six maturity scores in category **Scientific and expert support**

The assessment can be made in the following manner:

Score 1: No scientific or expert support is available to the measurement program.

Score 2: A minimal level of scientific or technical support is available, sufficient to maintain the measurement program in a sustained manner in the absence of major failures or events.

Score 3: There are effectively sufficient resources available to ensure continuation and upkeep of the measurement system, on a sustained basis, which may include calibration / replacement of sensors, effecting repairs and monitoring of instrument performance to identify and correct obvious faults.

Score 4: As Score 3, but the maintenance and upkeep is not dependent upon a single engineer or scientist, such that the support for the measurement series can be sustained.

Score 5: In addition to sustained upkeep, there is active scientific assessment of the measurements and investigation of potential improvements in either the instrument or its performance characterisation, including traceability and uncertainty quantification, being undertaken.

Score 6: Not used as no further support beyond score 5 is envisaged.

3.8.3 Programmatic support

This category assesses the long-term programmatic support that underpins the measurement program. Typically, higher quality measurements will be supported by sustained national or international programs, and infrastructure support that can assure longer-term operation and sustainability.

1	None
2	Project based funding support available
3	Score 2 + with expectation of follow on funding
4	Score 3 + not dependent upon a single investigator or funding line
5	Sustained infrastructure support available to finance continued operations for as far as can be envisaged given national and international funding vagaries
6	Score 5 + support for active research and development of instrumentation or applied analysis of the observations

Table 21. Six maturity scores in category **Programmatic support**

The assessment can be made in the following manner:

Score 1: No dedicated programmatic support is evident for the measurement program.

Score 2: There is dedicated funding support, but it is tied to a project and, therefore, the support is not envisaged to be continuous.

Score 3: As Score 2, but there is a reasonable expectation that funding will be renewed.

Score 4: As Score 3, but the measurement program is supported by multiple investigators and / or funding streams, to ensure long-term sustainability.

Score 5: The measurement program funding arises from a sustainable funding stream, such as national or international infrastructure funds, which are stable and unlikely to be removed in the foreseeable future.

Score 6: As Score 5, but site is also funded to actively analyse and develop the measurement program, ensuring that the highest possible quality observations are always undertaken. This may be ascertained by evidence of peer reviewed papers, book chapters, or membership of committees / working groups / task teams of high quality observational networks such as GRUAN, NDACC, AERONET, EARLINET, and TCCON.

3.9 Software readiness (optional)

As noted at the start of Section 3 this major strand is optional, and shall apply only to those measurements where routine automated and substantive processing occurs from the raw measured data to the provided geophysical parameters of the measurement series.

Cases where this would be appropriate would include measurement series where the directly measured parameter is a digital count, a radiance, a photon count or some other indirect proxy for the reported measurand, where processing exists to convert from the measured quantity to the reported quantity. Conversely, where the measurement constitutes a direct proxy for the measurand, such as a platinum resistance thermometer or anemometer, and the conversion is facile, the software readiness category is not appropriate.

It should be agreed, and documented in the assessment, whether this strand is applicable or not ahead of time, when deciding the rules of the round. Where it is not applicable, the column should be greyed out in the summary (see Figure 2). Note that the software readiness strand is solely related to the software that is used in the production of the primary measurement products. It does not consider software, often created and curated by third parties, used in subsequent applications of the data, including post-processing and dataset construction.

In this major category there are four sub-categories. These are mainly meant to be for self-assessment because the information is rarely publicly available**.** The software readiness category provides information on the availability and maintainability of software used to generate the measurement record. All software used to manipulate the measurement to its distributed product should be assessed. High maturity is indicative of a system that is institutionally well understood, and doesn't depend on specific individuals that have knowledge of the software since its origin. Software becomes more easily understandable if the programming follows standards and the installation and usage is documented. Software is also maintainable if it can be ported to other locations and across operating systems. More mature software may tend to also be open-source, and open-source code should be encouraged where it can be attained. However, for cases where the data are used operationally it may not be possible or practical to share the full processing code.

3.9.1 Coding standards

Coding standards are a set of conventions/rules specific for a coding language, which describes style, practices and methods that greatly reduce the probability of introducing bugs. This is especially important in a team environment, or group collaboration, so that uniform coding standards are used, and helps to reduce oversight errors and save time for code reviews. It is key to assuring the maintainability of the code at reasonable cost. There are ISO standards available for software coding which may be applicable. If such ISO standards are to be used should be agreed in the 'rules of the round'.

1	No coding standard or guidance identified or defined
2	Coding standard or guidance is identified or defined, but not applied
3	Score 2 + standards are partially applied and some compliance results are available
4	Score 3 + compliance is systematically checked in all code, but not yet compliant to the standards
5	Score 4 + Measurement provider has identified departures from the standards and actions are planned to achieve full compliance
6	Code is fully compliant with standards

Table 22: The 6 maturity scores in sub-category **Coding standards**

Coding standards can be evaluated as follows:

Score 1: There is no evidence available that coding standards have been considered.

Score 2: Standard identified/defined means that the measurement record producer has identified or defined the standards to be used, but has not applied it. The information about this most often can be found in software description documents or programming guidelines available from web pages, or by asking the measurement provider.

Score 3: This means that the measurement provider has started to apply the standards, and implemented procedures to check the compliance. This information may be available by asking the measurement provider.

Score 4: Score 3 + procedures are systematically applied to check the compliance, and the results are often available as internal reports.

Score 5: standards are systematically applied in all code and compliance is systematically checked in all code. Code is not fully compliant to the standards. Improvement actions to achieve full compliance are defined.

Score 6: At this stage the software shall be fully compliant with its description and the documented standard. This includes procedures to check the compliance and the results of the unit tests conducted.

3.9.2 Software documentation

Software Documentation is key to ensuring usability, portability and operator understanding. This sub-category is concerned primarily with whether the code is documented with proper headers, change history, and sufficiently complete and understandable comments describing the processes. Further steps are whether the README file is up-to-date, there is documentation available, which describes design rationale and architectural overview of the software, and there is a software installation and user manual available.

1	No documentation
2	Minimal documentation
3	Header and process description (comments) in the code
4	Score 3 + a draft software installation / user manual available
5	Score 4 + enhanced process descriptions throughout the installation / user manual complete
6	Score 5 + code and documentation is publicly available from a webpage

Table 23: The 6 maturity scores in sub-category **Software documentation**

The assessment can be made, for example, as below:

Score 1: No software documentation exists.

Score 2: There are header and limited comments in the code and installation instructions available, but no other documentation is available.

Score 3: README file should at least contain information on "Configuration instructions", "Installation instructions", "Operating instructions", "Copyright and licensing", "Contact information", etc..

Score 4: Score 3 + Software User Manual should at least contain information on software concept and design and providing instructions for installing and using the software.

Score 5: Code is very well documented and installation/user manual is complete and available on data provider's web page.

Score 6: The code and documentation is openly available through a website to allow users full understanding of the processing suite.

3.9.3 Portability and numerical reproducibility

Portability and numerical reproducibility concerns the usability of the software in different environments (different computing platforms such as Linux, Solaris, Mac OS, Windows etc. and different compilers such Intel, IBM, GNU, Portland, etc), and whether the results are numerically reproducible. It is important for migrating software from old to new computer systems and from one place to another.

1	Not evaluated
2	Reproducible under identical conditions
3	Reproducible and portable
4	Third party affirms reproducibility and portability
5	Score 4 + third party can install the code operationally
6	Score 5 + Turnkey system

Table 24: The 6 maturity scores in sub-category **portability and numerical reproducibility**

The assessment can be made, for example, as below:

Score 1: Not evaluated means this has not been considered at all.

Score 2: Measurement series investigator affirms that the software reproduces results when rerun on the same platform with the same input and same compiler. This information can be obtained by asking the investigator.

Score 3: The software produces numerically reproducible results to specified precision on different computing platforms (such as Linux, Solaris, Mac OS, Windows etc.), and/or with different compilers (such Intel, IBM, GNU, Portland, etc).

Score 4: Score 3 + 3rd party can install the code operationally with minimal manual efforts. Runs reveal that the output is numerically reproducible (within machine rounding errors). This information shall typically be found in software description documents available from measurement series investigator's web pages.

Score 5: Score 4 + the code is already used by a 3rd party in operational environment under configuration control. This shall be described in the software installation/user manual.

Score 6: Turnkey is software that is designed, supplied, built or completely installed and ready to operate. The term implies that the end user just has to turn a key and start using the software, e.g., Linux OS. This shall be described in the software user manual.

3.9.4 Security

Security is associated with software contents that either have the potential to destroy files and complete environments or are related to file transfer between compute environments. Both should not be contained in software. The security category also checks whether the file system can be accessed from outside, as this may hamper the integrity of the measurement series generation environment.

1	Not evaluated
2	Data provider affirms no security problems
3	Submitted for data provider's security review

4	Passes data provider's security review
5	Continuously passes periodic data provider's review
6	

Table 25: The 6 maturity scores in sub-category **Security**

The assessment can be made, for example, as below:

Score 1: Not evaluated at this stage means that software security issues have not been considered to date.

Score 2: Data provider has done the testing for security issues in the code and found none. This information can be obtained by asking the data provider.

Score 3: This information can be obtained by asking the data provider. This is a necessary step before porting the software from a research environment to an operational environment.

Score 4: This means the software has passed data provider's quality assurance and security tests. Information on this shall be obtained from software installation/user manual.

Score 5: Data provider does security assessment periodically and also whenever there is a software update, and the results shall be available from updated software installation/user manual.

Score 6: Not used.

4. Challenges to adoption

The approach introduced herein will be used in the first instance solely for the internal purposes of GAIA-CLIM. During the development of this guidance, a number of internal project partners have attempted to use it to classify a number of non-satellite measurement systems, and that feedback has been used to modify the criteria to ensure that this is fit-for-purpose, at least for the specific needs of GAIA-CLIM. So, we can be reasonably confident that this should be applicable more broadly to aid the consideration of maturity of non-satellite measurement characteristics for various possible purposes.

However, there is also a broader need to articulate and adopt a system of systems approach, which this documentation may help to nurture [GCOS, 2014, 2015]. There are significant challenges to its likely broad adoption which were highlighted in the recent GCOS meeting in Ispra [GCOS, 2014], and which are expanded upon here. The most appropriate mechanism to take this forwards, at least in the atmospheric domain, would be through the WIGOS program, recently adopted by WMO at its 2015 Congress.

4.1 Naming nomenclature for existing networks across and within domains

Perhaps the largest challenge is that currently a broad range of non-satellite measurement networks and infrastructures have been called ‘reference’, ‘baseline’ or ‘comprehensive’ that, when assessed against the criteria detailed in Section 3, would instead fall within a different category. The lack of clarity historically regarding a system of systems architecture, taken together with fractured governance and support structures, has led to a varied use and adoption of network nomenclature and practices both across, and within, Earth Observation science disciplines. This means that what different sub-communities concerned with environmental measurements refer to as ‘baseline’, ‘reference’ or indeed ‘comprehensive’ network measurements is not always the same. Often it is not even remotely similar. If a system of systems approach is to be broadly adopted, significant further work is required to reconcile the disparate approaches to network designations, and to manage the transition to a more trans-disciplinary approach to network assignations. There are several risks / challenges in any such transition:

1. National or international funding support for a measurement program may be tied to its present designation. There is a risk in enforcing any change that the funding support for the program is endangered. Take, for example, the Ocean reference network. This network is not a reference network in the sense advocated here, but rather closer to baseline capability. But, it is still the best set of observations available, and risking its loss would be a significant mistake.
2. Users may use a measurement program because of its current designation, and may get confused if measurement programs are reassigned or renamed without adequate consultation or justification.
3. The observers undertaking the measurement program may not fully understand the implications if updates to protocols and / or practices are required.

4. Ensuring program support sustainability and harmonization of practices across national boundaries.

On the flip side to these concerns is that allowing the status quo to continue means that users referring to e.g., a 'reference' network in the marine, atmospheric and composition communities (just as by way of an example) may be comparing measurement programs that are widely differing concerning their fundamental measurement characteristics and qualities and, therefore, suitability for a given application. The status quo places the responsibility of understanding the measurement systems and networks on a system-by-system and even ECV-by-ECV basis firmly on the end-user. Experience shows that end-users are, understandably, unlikely to have either the time or the necessary knowledge / expertise to fully understand the distinctions that may exist between similarly named programs and assume, incorrectly, that they are equivalent. This is a barrier to the effective usage of existing EO capabilities by scientists, policy makers and other end users, and will continue to be so unless and until a more holistic approach, such as suggested in this guidance, is adopted.

4.2 End-User Adoption

It is clear that alongside adoption and designation of a tiered network capabilities framework, it is necessary to provide material to aid users to understand what the tiers mean, and to show real case examples of how they can be used. GAIA-CLIM will, through its work packages, provide case study examples in the domain area of satellite measurement characterisation. But, further examples in other domain areas and application areas are necessary, that will be beyond the remit of GAIA-CLIM.

4.3 Realising technological and scientific benefits of a tiered set of capabilities

Even if the tier designations and criteria documented herein were adopted, there would remain the challenge of ensuring linkages between the different components to realise benefits. This includes aspects such as infrastructure co-location, intercomparison campaigns, information sharing, training and development etc.. Such inter-linkages will become both more obvious and more realisable if a system-of-systems architecture approach and assessment is adopted. Some subset of these aspects that touch upon satellite calibration / validation are covered within the living Gap Assessments and Impacts Document of GAIA-CLIM, which the interested reader is encouraged to refer to (see www.gaia-clim.eu).

4.4 Potential future applicability to the satellite domain

The tiers and their designations for GAIA-CLIM detailed herein pertain explicitly only to non-satellite measurement capabilities. Their extension to satellite measurements is non-trivial. Thus, the guidance in Section 3 is explicitly solely for application to non-satellite measurements.

In particular, the relation of fidelity and spatio-temporal completeness, that is clearly applicable to the non-satellite measurements domain, does not readily apply to satellite measurements. For satellites the fidelity instead depends on instrument design and its characterisation both prior to the launch, and using onboard calibration. For channels where clouds have an impact, it also depends upon the efficacy of cloud detection techniques. Also, metadata content is less of a concern as the historic evolution of the metadata has led to standards, which are both comprehensive and broadly applied, with very little difference among satellite data suppliers. Finally, satellite systems do not form networks in a geographical sense with the exception of geostationary sensors that always observe the same area.

However, some of the characterisations given for the observational tiers for non-satellite capabilities are broadly applicable to satellite based measurement systems, with some additional interpretation.

- Reference quality measurements in space (often called benchmark measurements within that community) would fulfill criteria on very high accuracy and traceability to the SI standard. Currently, no such system exists but several, such as the Climate Absolute Radiance and Refractivity Observatory (CLARREO) and Traceable Radiometry Underpinning Terrestrial- and Helio- Studies (TRUTHS) missions, have been proposed. Such missions potentially represent a calibration laboratory in orbit for the purpose of accurately measuring climate change. A specific value of the posited CLARREO / TRUTHS style measurements lies in their high value to function as reference for remaining space-based instruments not built specifically to measure climate change. Important for satellite reference systems, in particular in the infrared range, is that measurements are taken with high spectral resolution that enables analysis and characterisation of the performance of instruments in space, for instance with respect to changing spectral response of filter radiometers that measure an integral over a broader spectral range. In addition, higher spectral resolution may also be calibrated more accurately. Comparison of such posited measurements to other satellite measurements would establish an unbroken chain for SI traceable accuracy on orbit. Close to such a system is the GNSS-RO technique where the base unit is a time delay, that is traceable, and may be able to constitute a reference measurement, assuming all steps in the processing chain can be understood and their uncertainty quantified.

- The category baseline, as described for ground-based observations, has little in common with satellite systems as satellites either are in orbit and measuring everywhere or do not exist. There is no effective minimal set of measurements that a satellite takes – it is either operational or it is not. The closest analogy in the satellite domain to the non-satellite baseline network concept, therefore, is the provision of long-term (multi-decadal) measurements in some sub-set of the emissions spectra that can be used to characterise change and variability in a range of ECVs on climate timescales. Many satellite data records start in the late 1970s. The measurements are mostly in the visible, infrared, and microwave spectral ranges, but were built for the purpose to observe weather and not climate. A baseline concept would ensure their continuation into the future to enable multi-decadal continuous monitoring.

  As in the non-satellite domain, new measurements may provide enhanced monitoring in spectral domains with a long measurement heritage. For instance hyper-spectral infrared as delivered by the IASI instrument aboard the Metop satellite have high spectral resolution, and are approximately an order of magnitude more accurate than historic infrared measurements. Such instruments can serve as a comparator for historic instruments, establishing an unbroken chain of inter-satellite calibrations that enable relative calibration to more modern, better characterised, measures even if absolute calibration remains elusive.

Baseline implies the need for sustained missions, which is best achieved for operational weather observations. Such measurements may be also achieved by more operational ocean and land surface oriented missions, such as both launched and planned Sentinel missions.

- Comprehensive capability for satellite missions needs to be interpreted very differently from in situ networks, as little of the characterisation provided for non-satellite measurements fits. However, it might be interpreted in a way that this class is established as a catch-all for all other Earth Observing missions not captured above. These additional missions expand the ability to measure more components of the Earth system, with higher accuracy over shorter periods, fostering process understanding. Or they contribute by proofing measurement concepts for future missions. In many cases they eventually transition to a baseline capability.

Acknowledgements

This work has been based upon the substantial work undertaken by CORE-CLIMAX and a number of precursor studies assessing dataset maturity. Without these preceding efforts this work would not have been possible. Karin Kreher, Arnoud Apituley, Greg Bodeker, Barry Goodison, Mark Bourassa, Matthias Buschmann and Ge Peng provided feedback based upon early drafts that served to improve the Guidance.

References

Bates, J. J. and J. L.Privette, (2012), A maturity model for assessing the completeness of climate data records, Eos Trans. AGU, 93(44), 441.

GCOS, 2014, Workshop on the review of the GCOS Surface Network (GSN), GCOS Upper-Air Network (GUAN), and related atmospheric networks, *Ispra, Italy, April 2014*, GCOS-182 [http://www.wmo.int/pages/prog/gcos/Publications/gcos-182.pdf]

GCOS, 2015, Status of the Global Observing Systemfor Climate, GCOS-195 [http://www.wmo.int/pages/prog/gcos/Publications/GCOS-195_en.pdf]

JGCM, 2008 Evaluation of measurement data – Guide to the expression of uncertainty in measurement [www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf].

National Academy of Sciences. 2009. 'Observing weather and climate from the ground up: A nationwide network of networks.' http://dels.edu/Report/Observing-Weather-Climate-from/12540

J. Schulz, V. John, A. Kaiser-Weiss, C. Merchant, D.Tan, E. Swinnen and R. Roebeling, 2015: European Climate Data Record Capacity Assessment, in prep for Geoscience Data Journal

Seidel, D. J., F. H. Berger, et al. (2009). "REFERENCE UPPER-AIR OBSERVATIONS FOR CLIMATE Rationale, Progress, and Plans." Bulletin of the American Meteorological Society **90**(3): 361-+.

WIGOS, 2015a, WMO TECHNICAL REGULATIONS (WMO-No. 49) - MANUAL ON WIGOS

WIGOS, 2015b, WMO TECHNICAL REGULATIONS (WMO-NO. 49) - MANUAL ON WIGOS Attachment WIGOS METADATA STANDARD

Glossary

AERONET	Aerosol Robotic Network
BIPM	International Bureau of Weights and Measures
CDR	Climate Data Record
CF-compliant	Climate Forecast convention compliant data
CIMO	Commission for Instruments and Methods of Observation
CLARREO	Climate Absolute Radiance and Refractivity Observatory
CORE-CLIMAX	Coordinating earth observation validation for RE-analysis for CLIMAte ServiceS
EARLINET	European Aerosol Research Lidar Network to establish an aerosol climatology
ECV	Essential Climate Variable
EO	Earth Observation
EUMETNET	EU Meteorological Network
GAIA-CLIM	Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring
GCOS	Global Climate Observing System
GEOSS	Global Earth Observation System of Systems
GNSS-RO	Global Navigational Satellite System Radio Occultation
GRUAN	GCOS Reference Upper-Air Network
GUAN	GCOS Upper-Air Network
GUM	Guide to Uncertainties in Measurements
IASI	Infrared Atmospheric Sounding Interferometer
IPCC	Intergovernmental Panel on Climate Change
ISO	International standards Office
NAS	National Academy of Sciences
NDACC	Network for the Detection of Atmospheric Composition Change
NOAA	National Oceanographic and Atmospheric Administration
NWP	Numerical weather Prediction
SI	Systeme International of fundamental measurement units
SMM	System Maturity Matrix
TCCON	Total Carbon Colum Observing Network
TRUTHS	Traceable Radiometry Underpinning Terrestrial- and Helio-Studies
WIGOS	WMO Integrated Global Observing System
WMO	World Meteorological Organisation

Appendix A GAIA-CLIM measurement description

(General Note: This measurement description shall not become longer than 5 pages per measurement system described. Please only state to the most important facts and use tables and bullet lists to provide information where appropriate.**)**

(Type Measurement system Name and if available digital identifier here. The name must be unique and should include instrument type and location and / or network identifier):

Version	Author	Reviewers (if any)	Date

(Please use the above table to note version control on this record)

1 Intent of the document

(Provide information on what measurement system is being described and for what application(s) it was created. Keep in mind that the information is targeted at users of any level who wish to use the measurements for scientific applications. Users may not be expected to be experts for in situ, remote sensing or reanalysis techniques.)

2 Point of contact

(Please provide a point of contact: Organisation and Contact details (at least a contact name, organisation and e-mail address)).

3 Measurements description

(Provide a link to an existing technical product specification or provide the information in a form of a table in this document. The specification shall at least include measured variable names (identifying any that are Essential Climate Variables) and units (including uncertainty estimates indicators if provided), length of record, spatial coverage, spatial and temporal sampling.)

4 Data origin

(Provide a basic description of the methodology used to derive the measurements including a description of data processing methods such as the processing used to convert from a digital count transmission from a radiosonde to a geophysical profile estimate.)

5 Validation of an uncertainty estimation

(Provide a summary of any validation activities performed for the measurement product and provide a summary of uncertainty quantification of the product including whether the measurement is metrologically traceable to SI units or accepted standards (tabulated form appreciated).

6 Considerations for scientific applications

(Provide information on the applicability of the product for the possible scientific application including limitations. This includes aspects such as the ability to measure the full diurnal cycle, geographical representativity, sampling frequency etc.)

7 References

(Provide a complete list of references used in this document and, if applicable, provide additional reading references on measurement principles, retrievals, modelling, validation, uncertainty characterisation, product, and applications.)

Appendix B Measurement maturity assessment spreadsheet

The Guidance is most easily completed using the associated excel spreadsheet to record the maturity of candidate measurement systems. These spreadsheets are based upon the guidance outlined in Section 3. There is a spreadsheet for each major assessment strand. The spreadsheets are given below in the order that they arise.

Measurement system / program name here

maturity level as of mm/dd/yyyy

GAIA-CLIM Measurement System Maturity Matrix

Maturity	METADATA	DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, AND UPDATE	USAGE	SUSTAINABILITY	SOFTWARE READINESS
1	No metadata available	Limited scientific description of the measurement methodology available	None	Restricted availability through request	None	None	Conceptual development
2	Very basic metadata available	Comprehensive scientific description of the measurement methodology, report on limited validation, and limited measurement series user guide	Limited steps taken towards assuring traceability and comparability; limited information exists on systematic and random measurement uncertainties	Data avaliable from originator	Benefit for research applications identified; Potential public and commercial opportunities identified	Measurement program is sustainable in the short-term	Research grade code
3	Standards defined or identified; sufficient to use and understand the data and extract basic discovery metadata	Score 2 + paper on methodology published; Validation report available from data collector or in grey literature; comprehensive user guidance is available	Score 2 + limited traceability and comparability assured; comprehensive documentation on measurement uncertainties present and methods for routine quality monitoring defined	Data and documentation publically available from originator, feedback collated, irregular updates, initial versioning and local archival	Benefits for research applications demonstrated; Public and Commercial use occuring and benefits emerging	Measurement program is sustainable and has minimum level of necessary support to assure minimal quality standards are maintained	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability, numerical reproducibility and no security problems
4	Score 3 + standards systematically applied; meets international standards for the measurement metadata collection; enhanced discovery metadata; limited location level metadata	Score 3 + comprehensive scientific description available from data provider; report on inter comparison available; paper on validation published; user guide available from data provider includes details of validation and characterisation	Score 3 + steps required to establish traceability are defined; (inter)comparison against corresponding measurements in organised campaigns available; quantitative estimates of uncertainty available and routine monitoring partially implemented	Data and documentation available through a recognised data portal; feedback mechanism considers published analyses; version control formalized, rocbust archival on multiple media	Score 3 + research citations on product usage occurring; societal and economical benefits discussed	Measurement program has medium-term sustainability and is not liable to a single point of failure	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review
5	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + traceability partly established; measurements regularly compared to a measurement of similar or greater traceability; systematic uncertainties removed and uncertainty estimates are partially traceable; routine quality monitoring fully implemented	Source data, code and metadata archived and available upon request; established feedback mechanism; regular update cycle; fully established version control; data archival at recognized national or international long-term repository	Score 4+ product becomes reference for certain research applications; societal and economic benefits are demonstrated	Measurement program is long-term sustainable and robust to possible sources of failure	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party installs the code operationally
6	Score 5 + regularly updated and using extended metadata where defined	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + traceability established; measurements are regularly compared to other traceable measurements to verify; comprehensive validation of the quantitative uncertainty estimates that are fiully traceable; routine monitoring in place with results noted in meta data or documentation	Score 5 + no data access restrictions; active consideration of user feedback; data available in initial version for near-real time applications; all versions retained, indexed and available through a recognised repository	Score 5 + product and its applications become references in multiple research fields; Influence on decision and policy making demonstrated	Measurement program is sustainable and striving for constant improvement	Score 5 + fully compliant with standards; Turnkey System

1 & 2	Operational measurement capability
3 & 4	Baseline measurement capability
5 & 6	Reference measurement capability

MEASUREMENT SYSTEM MATURITY EVALUATION GUIDELINES

Maturity	METADATA	Standards	Collection level metadata (including change records)	File level
1	No metadata available	No standard considered	None	None
2	Very basic metadata available	No standard considered	Limited	Limited
3	Standards defined or identified; sufficient to use and understand the data and extract basic discovery metadata	Metadata standards identified and/or defined and partially but not yet systematically applied	Sufficient to use and understand the data independent of external assistance; Sufficient for data provider to extract discovery metadata from meta data repositories	Sufficient to use and understand the data independent of external assistance
4	Score 3 + standards systematically applied; meets international standards for the measurement metadata collection; enhanced discovery metadata; limited location level metadata	Score 3 + standards systematically applied at file level and collection level by data provider. Meets international standards	Score 3 + Enhanced discovery metadata	Score 3 + Limited location (station, grid-point, etc.) level metadata along with unique measurement set metadata (e.g. batch, set-up, time, averaging period)
5	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + meta data standard compliance systematically checked by the data provider	Score 4 + Complete discovery metadata meets appropriate (at the time of assessment) international standards	Score 4 + Complete location (station, grid-point, etc.) level and measurement specific metadata
6	Score 5 + regularly updated and using extended metadata where defined	Score 5 + extended metadata that could be useful but is not considered mandatory is also retained.	Score 5 + Regularly updated	

MEASUREMENT SYSTEM MATURITY EVALUATION GUIDELINES

Maturity	DOCUMENTATION	Formal description of measurement methodology	Formal Validation Report	Formal Measurement series User Guidance
1	Limited scientific description of the measurement methodology available	Limited scientific description of methodology available from data collector or instrument manufacturer	None	None
2	Comprehensive scientific description of the measurement methodology, report on limited validation, and limited measurement series user guide	Comprehensive scientific description available from data collector or instrument manufacturer	Informal validation work undertaken.	Sufficient information on the measurements available to allow user to ascertain minimum set of information required for appropriate use
3	Score 2 + paper on methodology published; Validation report available from data collector or in grey literature; comprehensive user guidance is available	Score 2 + Journal paper on measurement methodology published	Instrument has participated in certified intercomparison campaign and results available in grey literature	Comprehensive documentation on how the measurement is made available from data collector or instrument manufacturer including basic data characteristics description
4	Score 3 + comprehensive scientific description available from data provider; report on inter comparison available; paper on validation published; user guide available from data provider includes details of validation and characterisation	Score 3 + Comprehensive scientific description available from Data Provider	Report on intercomparison to other instruments, etc.; Journal paper on product validation published	Score 3 + including documentation of manufacturer independent characterisation and validation
5	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + Comprehensive scientific description maintained by Data Provider	Score 4 + Sustained validation undertaken via redundant periodic measurements	Score 4 + regularly updated by data provider with instrument / method of measurement updates and/or new validation results
6	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + Journal papers on measurement system updates published	Score 5+ Journal papers describing more comprehensive validation, e.g., error covariance, validation of qualitative uncertainty estimates published	Score 5 + measurement description and examples of usage available in peer-reviewed literature

MEASUREMENT SYSTEM MATURITY EVALUATION GUIDELINES

Maturity	UNCERTAINTY CHARACTERISATION	Traceability	Comparability	Uncertainty quantification	Routine Quality Monitoring
1	None	None	None	None	None
2	Limited steps taken towards assuring traceability and comparability; limited information exists on systematic and random measurement uncertainties	Comparison to independent stable measurement or local secondary standard undertaken irregularly	Validation using external comparator measurements done only periodically and these comparator measurements lack traceability	Limited information on uncertainty arising from systematic and random effects in the measurement	None
3	Score 2 + limited traceabaility and comparability assured; comprehensive documentation on measurement uncertainties present and methods for routine quality monitoring defined	Score 2 + independent measurement / local secondary standard is itself periodically calibrated against a recognized primary standard	Score 2 + Validation is done sufficiently regularly to ascertain gross systematic drift effects	Comprehensive information on uncertainty arising from systematic and random effects in the measurement	Methods for routine quality monitoring defined
4	Score 3 + steps required to establish traceability are defined; (inter)comparison against corresponding measurements in organised campaigns available; quantitative estimates of uncertainty available and routine monitoring partially implemented	Score 3 + processing steps in the chain of traceability are documented but not yet fully quantified.	Score 3 + (Inter)comparison against corresponding measurements in large-scale instrument intercomparison campaigns	Score 3 + quantitative estimates of uncertainty provided within the measurement products characterising more or less uncertain data points	Score 3 + routine monitoring partially implemented
5	Score 4 + traceability partly established; measurements regularly compared to a measurement of similar or greater traceability; systematic uncertainties removed and uncertainty estimates are partially traceable; routine quality monitoring fully implemented	Score 4 + traceability in the processing chain partly established	Score 4 + compared regularly to at least one measurement that has a traceability score >=5	Score 4 + systematic effects removed and uncertainty estimates are partially traceable	Score 4 + monitoring fully implemented (all production levels)
6	Score 5 + traceability established; measurements are regularly compared to other traceable measurements to verify; comprehensive validation of the quantitative uncertainty estimates that are fully traceable; routine monitoring in place with results noted in meta data or documentation	Score 5 + SI traceability in the processing chain fully established	Score 5 + compared periodically to additional measurements including some with traceability assessment >5	Score 5 + comprehensive validation of the quantitative uncertainty estimates	Score 5 + routine monitoring in place with results fed back to other accessible information, e.g. meta data or documentation

MEASUREMENT SYSTEM MATURITY EVALUATION GUIDELINES

Maturity	PUBLIC ACCESS, FEEDBACK, AND UPDATE	Public Access/Archive	User Feedback Mechanism	Updates to Record	Version control	Long-term data preservation
1	Restricted availability through request	Data may be available through request to trusted users	None	None	None	None
2	Data avaliable from originator	Data available for use through originator	Ad hoc feedback	None	None	None
3	Data and documentation publically available from originator, feedback collated, irregular updates, initial versioning and local archival	Data and documentation available through originator	Programmatic feedback collated	Irregularly following accrual of a number of new measurements or new insights	Versioning by data collector	Local archive retained by measurement collector
4	Data and documentation available through a recognised data portal; feedback mechanism considers published analyses; version control formalized, rocbust archival on multiple media	Score 3 + available through recognized data portal	Score 3+ consideration of published analyses	Regularly updated with new observations and utilising input from established feedback mechanism	Version control institutionalised and procedure documented	Each version archived at an institutional level on at least two media
5	Source data, code and metadata archived and available upon request; established feedback mechanism; regular update cycle; fully established version control; data archival at recognized national or international long-term repository	Score 4 + source data, code and metadata available upon request	Established feedback mechanism and international data quality assessment results are considered	Regularly operationally by stable data provider as dictated by availability of new input data or new innovations	Fully established version control considering all aspects	Data, raw data and metadata is archived at a recognised data repository such as a National Meteorological Service, national archive or international repository.
6	Score 5 + no data access restrictions; active consideration of user feedback; data available in initial version for near-real time applications; all versions retained, indexed and available through a recognised repository	Score 5 + no access restrictions apply	Score 5 + Established feedback mechanism and international data quality assessment results are considered in continuous data provisions	Score 5 + initial version of measurement series shared in near real time	Score 5 + all versions retained and accessible upon request	Score 5 + all versions of measurement series, metadata, software etc. retained, indexed and accessible upon request

MEASUREMENT SYSTEM MATURITY EVALUATION GUIDELINES

Maturity	USAGE	Research	Public and commercial exploitation
1	None	None	None
2	Benefit for research applications identified; Potential public and commercial opportunities identified	Benefits for research applications identified	Potential benefits identified
3	Benefits for research applications demonstrated; Public and Commercial use occuring and benefits emerging	Benefits for research applications demonstrated by publication	Use occurring and benefits emerging
4	Score 3 + research citations on product usage occurring; societal and economical benefits discussed	Score 3 + Citations on product usage occurring	Score 3 + societal and economical benefits discussed, data being distributed via appropriate data portals.
5	Score 4+ product becomes reference for certain research applications; societal and economic benefits are demonstrated	Score 4 + product becomes reference for certain applications	Score 4 + societal and economical benefits demonstrated
6	Score 5 + product and its applications become references in multiple research fields; Influence on decision and policy making demonstrated	Score 5 + Product and its applications become references in multiple research fields	Score 5 + influence on decision (including policy) making demonstrated

Maturity	SUSTAINABILITY	Siting environment	Scientific / expert support	Programmatic support
1	None	None	None	None
2	Measurement program is sustainable in the short-term	Site environment is stable in the short term	Minimal scientific support required to sustain the program is available	Project based funding support available
3	Measurement program is sustainable and has minimum level of necessary support to assure minimal quality standards are maintained	Score 2 + site ownership is sustainable	Relevant instrument expertize is available to support the measurements	Score 2 + with expectation of follow on funding
4	Measurement program has medium-term sustainability and is not liable to a single point of failure	Score 3 + Site is representative of a broader region around the immediate location	Score 3 + at least two experts available to support the measurement program operation	Score 3 + not dependent upon a single investigator or funding line
5	Measurement program is long-term sustainable and robust to possible sources of failure	Score 4 + site ownership, immediate environment is likely to be unchanged for decades	Active instrumentation research and development being undertaken	Sustained infrastructure support available to finance continued operations for as far as can be envisaged given national and international funding vageries
6	Measurement program is sustainable and striving for constant improvement	Score 5 + long-term ownership and rights are guaranteed		Score 5 + support for active research and development of instrumentation or applied analysis of the observations

MEASUREMENT SYSTEM MATURITY EVALUATION GUIDELINES Note that this set of criteria is optional and should only be applied to relevant measurement systems that make substantive use of softwa

Maturity	SOFTWARE READINESS	Coding standards	Software Documentation	Portability and Numerical Reproducibility	Security
1	Conceptual development	No coding standard or guidance identified or defined	No documentation	Not evaluated	Not evaluated
2	Research grade code	Coding standard or guidance is identified or defined, but not applied	Minimal documentation	Reproducible under identical conditions	Data provider affirms no security problems
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability, numerical reproducibility and no security problems	Score 2 + standards are partially applied and some compliance results are available	Header and process description (comments) in the code	Reproducible and portable	Submitted for data provider's security review
4	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	Score 3 + compliance is systematically checked in all code, but not yet compliant to the standards.	Score 3 + a draft software installation / user manual available	3rd party affirms reproducibility and portability	Passes data provider's security review
5	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party installs the code operationally	Score 4 + Measurement provider has identified departures from the standards and actions are planned to achieve full compliance	Score 4 + enhanced process descriptions throughout the installation / user manual complete	Score 4 + 3rd party can install the code operationally	Continues to pass the data provider's review
6	Score 5 + fully compliant with standards; Turnkey System	Code is fully compliant with standards.	Score 5 + code and documentation is publicly available from a webpage	Score 5 + Turnkey system	