Quality Assessment Guidelines

Author(s): ***Samuel Hunt, Javier Gorroño, Sarah Douglas, Emma Woolliams***
NPL, Task Experts

Approval: ***Nigel Fox***
NPL, Task Lead

Accepted: ***Clement Albinet***
EOP-GMQ EDAP Technical Officer

EDAP.REP.001 *Issue: 1.1* *23 May 2019*

National Physical Laboratory

AMENDMENT RECORD SHEET

The Amendment Record Sheet below records the history and issue status of this document.

ISSUE	DATE	REASON
1.0	23 April 2019	Initial Issue
1.1	23 May 2019	Colour scheme for matrix revised

TABLE OF CONTENTS

1. INTRODUCTION 3
1.1 Scope 3
1.2 Acronyms & Abbreviations 3
1.3 Reference Documents 4

2. PRINCIPLES OF THE EDAP QUALITY ASSESSMENTS 6
2.1 Considerations for Commercial Missions 6
2.2 Scope of Assessments 6

3. EDAP QUALITY ASSESSMENT GUIDELINES 7
3.1 Product Information 10
3.1.1 Product Details 10
3.1.2 Availability & Accessibility 10
3.1.3 Product Format 12
3.1.4 User Documentation 12
3.1.5 Metrological Traceability Documentation 13
3.2 Product Generation 13
3.2.1 Sensor Calibration & Characterisation Pre-Flight 14
3.2.2 Sensor Calibration & Characterisation Post-Launch 14
3.2.3 Retrieval Algorithm Method – Level 2 Only 15
3.2.4 Retrieval Algorithm Tuning – Level 2 Only 15
3.2.5 Additional Processing 16
3.3 Ancillary Information 17
3.3.1 Product Flags 17
3.3.2 Ancillary Data 17
3.4 Uncertainty Characterisation 18
3.4.1 Uncertainty Characterisation Method 18
3.4.2 Uncertainty Sources Included 19
3.4.3 Uncertainty Values Provided 19
3.4.4 Geolocation Uncertainty 20
3.5 Validation 21
3.5.1 Reference Data Representativeness 21
3.5.2 Reference Data Quality 21
3.5.3 Validation Method 22
3.5.4 Validation Results 22

4. REPORTING DATA PRODUCT QUALITY ASSESSMENTS 23

1. INTRODUCTION

In recent years, the increasing range of applications of Earth Observation (EO) data products and availability of low-cost satellites has resulted in an increasing number of commercial satellite systems. These services may provide complementary capabilities to those of Space Agencies. Due to these advances there is potential for some commercial missions to be considered as candidate European Space Agency (ESA) Third Party Missions (TPMs). TPMs are non-ESA missions, for which ESA assumes some formal responsibility towards the Mission / Data Provider.

Before they can be adopted, these missions require an assessment from ESA of mission suitability and output data quality. For the most efficient exploitation of EO data, assessment of data quality, calibration and validation are indispensable tasks and also form the basis for reliable scientific conclusions. The ESA Earthnet Data Assessment Pilot (EDAP) project is intended to perform this assessment for various missions to ensure the delivered data is fit for purpose.

1.1 Scope

This document provides a set of high-level guidelines by which the EDAP data product quality assessments are to be undertaken. The goal of these guidelines is to ensure both a consistency of approach across the EDAP project and that international community best-practices, such as QA4EO [RD-2], are followed where applicable. Section 2 provides some background information on community best practices. Next, Section 3 describes the EDAP framework to data product quality assessments. Finally, Section 4 describes how to report the data product quality assessment results.

1.2 Acronyms & Abbreviations

ATBD	Algorithm Theoretical Basis Document
CF	Climate & Forecast (Metadata Convention)
CEOS	Committee on Earth Observation Satellites
ECV	Essential Climate Variable
EDAP	Earthnet Data Assessment Pilot
EO	Earth Observation
ESA	European Space Agency
FRM	Fiducial Reference Measurement
FRM4GHG	Fiducial Reference Measurements for Ground-Based FTIR Greenhouse Gas Observations
FRM4SOC	Fiducial Reference Measurement for Satellite Ocean Colour
FRM4STS	Fiducial Reference Measurements for validation of Surface Temperatures from Satellites
L1	Level 1
L2	Level 2

L3	Level 3
NPL	National Physical Laboratory
QA4EO	Quality Assurance Framework for Earth Observation
QA4ECV	Quality Assurance Framework for Essential Climate Variables
SI	International System of Units
TPM	Third Party Mission
WGCV	Working Group on Calibration and Validation

1.3 Reference Documents

RD-1 JGCM, Guide to the Expression of Uncertainty in Measurement (GUM), JCGM 100:2008.

RD-2 QA4EO, A guide to content of a documentary procedure to meet the Quality Assurance requirements of GEO, QA4EO-QAEO-GEN-DQK-002

RD-3 QA4EO, A guide to "reference standards" in support of Quality Assurance requirements of GEO, QA4EO-QAEO-GEN-DQK-003

RD-4 Evaluation and Quality Control for Observations, https://climate.copernicus.eu/node/244

RD-5 J. Nightingale et al., "Quality Assurance Framework Development Based on Six New ECV Data Products to Enhance User Confidence for Climate Applications," Remote Sens., vol. 10, no. 8, p. 1254, 2018.

RD-6 CARD4L, Product Family Specification, Surface Reflectance, Working Draft (2017)

RD-7 Eaton et al., "NetCDF Climate and Forecast (CF) Metadata Conventions", Version 1.7 (2017). See: http://cfconventions.org

RD-8 Infrastructure for spatial information in Europe (INSPIRE), https://inspire.ec.europa.eu

RD-9 QA4ECV, QA4ECV Product Documentation Guidance: Algorithm Theoretical Basis Document, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20ATBD%20Guidance.pdf

RD-10 QA4ECV, QA4ECV Documentation Guidance: Product User Manual, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20PUM%20Guidance.pdf

RD-11	JGCM, International Vocabulary of Metrology (VIM 3rd Edition), JGCM 200:2012
RD-12	J. Gorroño et al., “A radiometric uncertainty tool for the sentinel 2 mission,” Remote Sens., vol. 9, no. 2, p. 178, Feb. 2017.
RD-13	E. R. Woolliams et al., “Applying Metrological Techniques to Satellite Fundamental Climate Data Records,” J. Phys. Conf. Ser., vol. 972, no. 1, p. 012003, Feb. 2018.
RD-14	C. Merchant et al., “Radiance Uncertainty Characterisation to Facilitate Climate Data Record Creation”, to appear in Journal of Remote Sensing 2019
RD-15	QA4ECV, QA4ECV Guidance: Provenance Traceability Chains, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20Traceability%20 Chains%20Guidance.pdf
RD-16	QA4EO Task Team, A Quality Assurance Framework for Earth Observation: Concept, Version 4.0 (2010). See: http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf
RD-17	I. Barker Snook, Fiducial Reference Measurements for validation of Surface Temperature from Satellites (FRM4STS) Project Brochure (2016). See: http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/D30-FRM4STS-project-brochure-with-cover-16Nov16-signed.pdf
RD-19	QA4ECV, QA4ECV Documentation Guidance: Product Validation and Intercomparison Report, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20Validation%20Guidance.pdf
RD-20	Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., et al. 2016 The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018. (doi:10.1038/sdata.2016.18).

2. PRINCIPLES OF THE EDAP QUALITY ASSESSMENTS

The Quality Assurance framework for Earth Observation (QA4EO), established and endorsed by Committee on Earth Observation Satellites (CEOS), defines the following principle regarding Earth Observation data quality [RD-16]:

"It is critical data and derived products are easily accessible in an open manner and have associated with them an indicator of their quality traceable to reference standards (preferably SI) to enable users to assess its suitability for their application i.e. its "fitness for purpose"."

QA4EO defines high level processes to achieve this, such as well-documented procedures, participation in comparisons and rigorous uncertainty assessments. These processes apply to all EO missions of any kind and it is therefore within this framework that EDAP quality assessments will be performed.

In recent years many initiatives have developed methods to assess satellite mission data quality in the QA4EO context, for example, the Quality Assurance for Essential Climate Variables (QA4ECV) [RD-5] or Evaluation and Quality Control for Observations (ECQO) [RD-4] projects. These guidelines are intended to build on the experience of these projects, this is discussed further in Section 3.

The guidelines are intended to provide a description of the high-level principles and activities that are required in quality assessments for all types of Earth Observation missions, since these are largely common between different domains. More detailed descriptions of domain-specific activities that should be undertaken to fulfil these high-level requirements are presented in further guidelines documents developed by the project task teams.

2.1 Considerations for Commercial Missions

At the top level these guidelines have been drafted to reflect a near "ideal" scenario, which is hoped serve as an aspiration to new space providers as well as space agencies. It is understood that many of the missions the project will assess will only partly comply with these requirements (to different degrees). This is acknowledged in the assessment grading system which is intended to primarily test whether aspects of a given mission are "fit for purpose" within the context of the mission's stated performance and application. This is described in detail in the following section.

2.2 Scope of Assessments

As well as providing the "ideal" scenario in terms of mission quality performance, the guidelines also present the ideal case in terms of the assessments themselves – describing a fully comprehensive mission quality assessment. Within the scope of EDAP such a full study may not be feasible for all of the project's missions. This again is acknowledged in the assessment grading system, described in the following section.

3. EDAP QUALITY ASSESSMENT GUIDELINES

This section outlines the overall framework for the data product quality assessments to be performed for the selected EDAP missions. As mentioned in the previous section, at the top level this has been drafted to reflect a near 'ideal' scenario that ESA may adopt in the future. However, within that framework the specific requirements and degree of compliance that will be applied for each EDAP assessed mission is likely to vary considerably and will depend on the intended applications of the mission and the accessibility of the necessary information. High-level considerations for the different aspects of product quality and a basis for scaling their relative criticality to enable users to readily assess 'fitness for purpose', are then discussed. Following this, Section 4 describes how to practically report the findings of these assessments.

The data product quality assessment approach taken is based on the QA4EO principle and specifically builds from the practical embodiment in the concepts developed by the Evaluation and Quality Control for Observations (ECQO) project [RD-4] for data products entering Copernicus Climate Change Service (C3S) Climate Data Store (CDS). That project itself is in part built on the methods and good practices developed by the EU FP7, Quality Assurance for Essential Climate Variables (QA4ECV) project [RD-5], amongst other similar European projects. The product evaluation activity they undertake is divided into six sections, these are:

- Details
- Generation
- Quality Flags
- Uncertainty Characterisation
- Validation
- Inter-comparison

These sections are themselves divided into sub-sections, which constitute each of the different aspects of the data product that should be assessed and graded, either as Basic, Intermediate, Good or Excellent. The information is gathered in a Quality Assurance Report (QAR) and summarised in a product quality evaluation matrix.

While the quality assessment described here is based on the EQCO framework, we recognise that that framework is primarily aimed at well-established data products of Level 2 (L2), Level 3 (L3) and higher, and therefore we have made several adjustments to the EQCO sections of analysis and grading levels so they better reflect the data products to be assessed within EDAP. These EDAP assessments will be at an earlier processing level, typically Level 1 (L1) or L2, and are less mature (as discussed in Section 2). The resulting product quality evaluation matrix is shown in Figure 1, which contains the following sections of analysis:

- Product Information
- Product Generation
- Ancillary Data
- Uncertainty Characterisation
- Validation

As shown in Figure 1, as well as the EQCO grading scales of Basic, Intermediate, Good and Excellent, our grading scale has the additional levels of Not Assessable and Not Assessed. This covers the expected cases that for some EDAP missions certain aspects of product quality will not be assessed – either because the mission is not yet mature enough to allow the assessment, or, because the assessment is currently outside of the scope of this pilot project.

In the remainder of this section, we look in turn at what each of the sections of product quality (columns in the product quality evaluation matrix) mean and how they may, at a high-level, be assessed. The approach is that we first describe the ideal case, which corresponds to the international community best practices, as introduced in Section 2 and described further below. If this is met, the sub-section would be graded Excellent. Due to the expected relative immaturity of some of the EDAP missions it is unlikely that this level of quality will be met, so whilst remembering that this is the benchmark EO data providers should aim for, a set of less-rigorous but acceptable standards are described corresponding to the lower grades in the assessment, Basic, Intermediate or Good.

Product Information	Product Generation	Ancillary Information	Uncertainty Characterisation	Validation
Product Details	Sensor Calibration & Characterisation Pre-Flight	Product Flags	Uncertainty Characterisation Method	Reference Data Representativeness
Availability & Accessibility	Sensor Calibration & Characterisation Post-Launch	Ancillary Data	Uncertainty Sources Included	Reference Data Quality
Product Format	Retrieval Algorithm Method	} If target mission data product is Level 2	Uncertainty Values Provided	Validation Method
User Documentation	Retrieval Algorithm Tuning		Geolocation Uncertainty	Validation Results
Metrological Traceability Documentation	Additional Processing			

Key
Not Assessed
Not Assessable
Basic
Intermediate
Good
Excellent

Figure 1 – The EDAP Product Quality Evaluation Matrix – uncompleted

3.1 Product Information

The *Product Information* section covers the top-level product descriptive information, product format, and the supporting documentation.

3.1.1 Product Details

Certain basic descriptive information should be provided with any EO data product, some of which is absolutely necessary for the data to be at all meaningful. Additionally, some of this information, such as claimed measurement quality or resolution, can provide both the data product quality assessor and the reader of their report with a frame of reference from which to set expectations for a given product.

It is therefore required by the EDAP assessment that any EO product provides the following:

- Product name
- Sensor Name
- Sensor Type
- Product version number
- Product ID
- Processing level of product
- Measured quantity name
- Measured quantity units
- Stated measurement quality
- Spatial Resolution
- Spatial Coverage
- Temporal Resolution
- Temporal Coverage
- Mission coverage

Also recommended is the following (based on INSPIRE metadata):

- Point of contact (Responsible organisation, including email address)
- Product locator (e.g. URL, DOI if applicable)
- Conditions for access and use
- Limitation on public access
- Product abstract (summary of resource)

Table 3-1 shows how a data product's provision of the above information relates to the grade it achieves for this sub-section of the quality assessment.

Table 3-1 – Product Information > Product Details – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Any required information missing.
Intermediate	
Good	All required information available, any recommended information missing.
Excellent	All required and recommended information available.

3.1.2 Availability & Accessibility

This is about how readily the data are available to those who wish to use them. It does not necessarily require cost-free access but is more about following the FAIR Data Principles for scientific data management and stewardship [RD-20], which provide valuable principles for all data applications. These state that:

Data should be **findable**
- Metadata and data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include the identifier of the data it describes
- Metadata and data are registered or indexed in a searchable resource

Data should be **accessible**
- Metadata and data are retrievable by their identifier using a standardised communications protocol
- The protocol is open, free and universally implementable
- The protocol allows for an authentication and authorisation procedure where necessary

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

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

Table 3-2 shows how a data product's provision of the above information relates to the grade it achieves for this sub-section of the quality assessment.

Table 3-2 – Product Information > Availability and Accessibility – Assessment Criteria

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

3.1.3 Product Format

An important aspect of EO data products that ensures they are most easily accessible to the widest variety of users is their file format. CEOS, through initiatives like CARD4L (CEOS Analysis Ready Data for Land) [RD-6], is promoting the concept of Analysis Ready Data (ARD), which attempts to ensure that data are processed to minimum set of requirements to allow immediate analysis of interoperable datasets.

In the ideal case, an assessed mission product format should meet any appropriate CEOS ARD guidelines, for example CARD4L requirements in the case of SAR and high-resolution optical sensors. In the case where these requirements are not met, product formats are graded based on the following:

- the extent to which they are documented;
- whether standard file formats are used (e.g. NetCDF);
- If they comply with standard variable and metadata naming conventions, such as CF Conventions [RD-7], or, for data from the European Union, the INSPIRE directive [RD-8].

Table 3-3 shows how a given EO data product should be graded for its format.

Table 3-3 – Product Information > Product Format – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Non-standard, undocumented data format.
Basic	Non-standard or proprietary data format, or, poorly-documented standard file format.
Intermediate	Data in documented standard file format. Non-standard naming conventions used.
Good	Data in well-documented standard file format, meeting community naming convention standards.
Excellent	Analysis Ready Data standard if applicable, else as *Good*.

3.1.4 User Documentation

Data products should be accompanied with the following minimum set of documentation for users, which should be regularly updated as required:

- Product User Guide/Manual (PUG/PUM)
- Algorithm Theoretical Basis Document (ATBD)

It may be for a given mission that in place of these documents some combination of articles, publications, webpages and presentations provide a similar set of information. For the highest grades however they should be presented as a formal document, since users should not be expected to search the information out. The QA4ECV project provides guidance for the expected contents of these documents [RD-9, RD-10], which they can be evaluated against.

Table 3-4 describes how EDAP grades a products user documentation.

Table 3-4 – Product Information > User Documentation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No user documentation provided, or, documentation out-of-date.
Basic	Limited PUG available, no ATBD. Documentation up-to-date.
Intermediate	Some PUG and ATBD-type information available. May be as formal documents or made up of e.g. articles. Documentation up-to-date.
Good	PUG meeting QA4ECV standard, reasonable ATBD. Documentation up-to-date.
Excellent	PUG ATBD available meeting QA4ECV standard**.** Documentation up-to-date.

3.1.5 Metrological Traceability Documentation

Traceability is defined in the vocabulary of metrology (VIM) [RD-11] as a,

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

and reinforced in the QA4EO procedures. Traceability is therefore a key aspect of achieving reliable, defensible measurements. In this definition an important part of measurement traceability is highlighted – that it is well documented. This of course must be the case for EO data products too.

Various diagrammatic approaches have been developed to present the traceability chains for EO data products (e.g. the QA4ECV guidance, which includes a traceability chain drawing tool [RD-15]). Such a diagram should be included in the documentation for every EO mission. The FIDUCEO project [RD-13] has provided guidance for a more detailed measurement function centred "uncertainty tree diagram" which is ultimately more suitable for Level 1 (and some Level 2) processing and should be the aspiration for missions in the future.

Table 3-5 shows how the EDAP grades the metrological traceability documentation, based on its completeness.

Table 3-5 – Product Information > Metrological Traceability Documentation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No traceability chain documented.
Basic	Traceability chain diagram and/or uncertainty tree diagram included, missing some important steps.
Intermediate	Traceability chain and/or uncertainty tree diagram documented identifying most important steps and sources of uncertainty.
Good	Rigorous uncertainty tree diagram, with, where appropriate a traceability chain documented, identifying all reasonable steps of and accompanying sources of uncertainty.
Excellent	Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes traceability to SI.

3.2 Product Generation

The Product Generation section covers the processing steps undertaken to produce the data product itself. This starts with an assessment of the calibration of the instrument measurements to L1. If the mission under assessment produces a L2 data product, then additional steps of assessment must be undertaken.

3.2.1 Sensor Calibration & Characterisation Pre-Flight

The pre-flight calibration and characterisation campaign should encompass a given sensor's behaviour to an extent and sufficient quality that is "fit for purpose" within the context of the mission's stated performance. What this requires is specific to given instrument types, which will be discussed in the sensor specific assessment implementation guidelines and will require a degree expert judgement.

Table 3-6 shows how EDAP grades pre-flight sensor calibration and characterisation.

Table 3-6 – Product Generation > Sensor Calibration & Characterisation Pre-Flight – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Pre-flight calibration & characterisation not documented or information not available.
Basic	Pre-flight calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose.
Intermediate	Pre-flight calibration & characterisation covers most important aspects of instrument behaviour at a level of quality to be judged fit for purpose.
Good	Pre-flight calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Calibration traceable to SI or community reference, characterisation meets good practice.
Excellent	As *Good*, additionally calibration and characterisation includes the measurements needed to assess uncertainties at component level and their impact on the final product.

3.2.2 Sensor Calibration & Characterisation Post-Launch

As in the pre-flight case, the post-launch calibration and characterisation activity should encompass a given sensor's behaviour to an extent and sufficient quality that is "fit for purpose" within the context of the mission's stated performance. What this requires is specific to given instrument types and calibration methods, which will be discussed in the sensor specific assessment implementation guidelines and will require a degree expert judgement. However, in general where a CEOS or FRM method/test-site is available this should as a minimum be used

Table 3-7 shows how EDAP grades post-launch sensor calibration and characterisation.

Table 3-7 – Product Generation > Sensor Calibration & Characterisation – Post-Launch – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Post-launch calibration & characterisation not documented or not available.
Basic	Post-launch calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose.
Intermediate	Post-launch calibration & characterisation covers most important aspects of instrument behaviour at a level of quality to be judged fit for purpose and uses appropriate community infrastructure/methods (CEOS/FRMs).
Good	Post-launch calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance and uses appropriate community infrastructure/methods (CEOS/FRMs).
Excellent	Post-launch calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Measurements fully traceable to SI or community reference at an uncertainty commensurate with the product specification and carried out regularly across the full range of observational conditions of the product and dynamic range.

3.2.3 Retrieval Algorithm Method – Level 2 Only

For many types of L2 products there are typically a variety of potential retrieval methods that may be used to derive them. These may vary in ways such as model complexity and computational efficiency – resulting in higher or lower quality final products.

As with the L1 sensor calibration, the L2 retrieval method should be of a sufficient quality that is "fit for purpose" within the context of the mission's stated performance across all stated use cases (e.g. scene types). What this requires is specific to a given variable's retrieval methods and will require a degree of expert judgement.

Table 3-8 shows how EDAP grades the algorithm retrieval method used to generate L2 products.

Table 3-8 – Product Generation > Retrieval Algorithm Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Retrieval method not documented.
Basic	Retrieval method too simple to be judged "fit for purpose" in terms of the mission's stated performance.
Intermediate	Reasonable retrieval method used, judged "fit for purpose" in terms of the mission's stated performance for most expected use cases, with at least a sensitivity analysis carried out.
Good	Sophisticated retrieval method used, "fit for purpose" in terms of the mission's stated performance all expected use cases and validated performance against similar algorithms or with empirical evidence.
Excellent	State-of-the-art retrieval, easily "fit for purpose" in terms of the mission's stated performance, full uncertainty budget derived and validated.

3.2.4 Retrieval Algorithm Tuning – Level 2 Only

Level 2 retrieval algorithms often require some initial "tuning" or calibration against reference data, such as in-situ measurements. The reference datasets used in this process must be of a sufficient quality, size and representativeness (in terms of factors like scene type or dynamic range) in order to achieve the mission's stated performance across all stated use cases. What this requires is specific to the retrieval method used and may require some expert judgement.

Table 3-9 shows how EDAP grades a mission's retrieval algorithm tuning.

Table 3-9 – Product Generation > Retrieval Algorithm Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Retrieval algorithm tuning not documented.
Basic	Algorithm tuned, but to data that is not of a sufficient quality or sufficiently representative to be judged "fit for purpose" in terms of the mission's stated performance.
Intermediate	Algorithm tuned to data that is of a quality or representativeness that is "fit for purpose" in terms of the mission's stated performance for most use cases.
Good	Algorithm tuned to data traceable to SI, potentially through an FRM.
Excellent	Algorithm tuned to data traceable to SI, potentially through an FRM. Representative of all stated use cases and all input parameters fully traceable with robust uncertainties.

3.2.5 Additional Processing

Additional processing steps are separate to the main sensor calibration or retrieval processing. These may include processes like resampling or the generation of classification masks. Additional processing steps must themselves be assessed for quality based on their "fitness for purpose" in the context of the mission.

Each additional processing step should be separately assessed and based on the criteria described in Table 3-10, and then a combined score determined.

Table 3-10 – Product Generation > Additional Processing – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Additional processing steps not documented.
Basic	Additional processing steps documented. Some important additional processing steps may not be fit for stated purpose.
Intermediate	Additional processing steps documented. All significant additional processing steps are fit for stated purpose.
Good	Additional processing steps documented. All additional processes steps fit for stated purpose.
Excellent	All additional processing steps fully documented and state-of-the-art.

3.3 Ancillary Information

In addition to its core measured variables, the assessment of which is covered by the *Product Generation* section, EO data products typically contain a variety of additional ancillary information to facilitate interpretation and further analysis of the data. This section of the mission quality assessment evaluates this ancillary information both in terms of its quality and completeness (i.e. do users have access to all the relevant information they need).

3.3.1 Product Flags

Product Flags offer users important extra layers of useful descriptive information on top of the measurements themselves. They can include information on the performance of the instrument, such as indicating periods of unusual instrumental behaviour where the data should not be used, or classification information as to the type of pixel.

For the user is it important that flags are clearly named and documented and that they cover an appropriate breadth of information. What is exactly required will depend on the instrument type and the intended use case. For the EDAP criteria for grading a product's flags see Table 3-11.

Table 3-11 – Ancillary Information > Product Flags – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Product flags not available or not documented.
Basic	A limited set of product flags of poorly documented product flags available.
Intermediate	A limited set of well documented product flags available, but mostly binary in nature e.g. relative to a threshold.
Good	A reasonable set of well documented product flags available, including meaningful gradation i.e. % of clouds
Excellent	A comprehensive set of well document product flags full gradation where appropriate and many provided or calculable at pixel level.

3.3.2 Ancillary Data

Ancillary data provides users with vital additional data layers to properly define, interpret and analyse a product's measurement data.

As a minimum all information required to properly define the primary measured data should be included in the data product. For example, in the case of optical sensors this includes information such as the sensor spectral response function or the viewing and illumination geometries. Where appropriate this data should be uncertainty quantified.

Other information, though not strictly required to define the measurement, may be useful to interpret the measurements or for further analysis. This may include information such as meteorological data. Inclusion of this kind of data, though it may be available or derivable from other sources, is convenient for users and considered advantageous. For the EDAP criteria for grading ancillary data provision see Table 3-12.

Table 3-12 – Ancillary Information > Ancillary Data – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Ancillary data not available.
Basic	Key ancillary data provided that is required to define measurement.
Intermediate	All ancillary data provided that is required to define measurement.
Good	All ancillary data provided that is required to define measurement, uncertainty quantified where appropriate. Some additional ancillary data provided required to interpret measurements.
Excellent	All ancillary data provided that is required to define measurement, uncertainty quantified where appropriate. All key additional ancillary data provided required to interpret measurements.

3.4 Uncertainty Characterisation

To ensure measurements are both meaningful and defensible it is crucial that they come with rigorously evaluated uncertainty estimates. This section of the mission quality assessment evaluates the methodology used to estimate uncertainty values for a given mission, the extent of the mission's analysis and how the values are provided.

3.4.1 Uncertainty Characterisation Method

A comprehensive description of how to evaluate sources of uncertainty in a measurement and propagate them to a total uncertainty of the final measurand is provided by the metrological community in the Guide to the Expression of Uncertainty in Measurement (GUM) [RD-1]. This is the approach that should be taken by all EO missions.

A rigorous treatment of uncertainty in EO data should consider the error-covariance between product pixels. Pixel-level errors are often highly correlated on scales that are very relevant to the kind of analysis typically performed, for example, the combination data from different spectral bands or spatial binning. Additionally, many scientific applications, such as data assimilation or optimal estimation retrieval algorithms, can exploit data error-covariance information to achieve more accurate results.

The field of Earth Observation metrology has progressed greatly in recent years. Operational missions are developing different approaches to evaluate and distribute metrologically rigorous uncertainties for L1 and L2 product. For example, ESA's Sentinel-2 mission has developed an on-the-fly, pixel-level uncertainty evaluation tool [RD-12]. There have also been some initiatives, like the FIDUCEO project, that attempt to apply metrology to historical sensor data records [RD-13].

That said, Earth Observation metrology is still a developing field and it is still more common for uncertainties to be evaluated in a manner that does not comply with the GUM. It is still typical for values like the specification performance value or single offset from a comparison sensor to be quoted as the uncertainty.

Table 3-13 describes how EDAP grades a mission's uncertainty characterisation methodology.

Table 3-13 – Uncertainty Characterisation > Uncertainty Characterisation Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Uncertainty characterisation not performed, or method not documented.
Basic	Uncertainty established by limited comparison to measurements by other sensor/s Not by independent assessment and then comparison.
Intermediate	Limited use of GUM approach, and/or, an expanded comparison to measurements by other sensors.
Good	GUM approach to estimate measurement uncertainty with full breakdown of components and separated as Type A or B classification.
Excellent	GUM approach to estimate measurement uncertainty, including a treatment of error-covariance.

3.4.2 Uncertainty Sources Included

In addition to the methodology used to determine the uncertainty caused by given error sources the breadth of different error sources analysed must also be assessed. This again is judged on the basis of what is "fit for purpose" in the context of a mission's stated performance. All contributions relevant at the required level of uncertainty should be included in the mission's uncertainty budget. Again, what is required is specific to given instruments and will require a degree expert judgement. This will be discussed further in the sensor specific assessment implementation guidelines.

Table 3-14 describes how EDAP grades the extent of uncertainty sources included in a mission's uncertainty characterisation analysis.

Table 3-14 – Uncertainty Characterisation > Uncertainty Sources Included – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Uncertainty characterisation not performed, or sources analysed not documented.
Basic	Some important sources of uncertainty missing.
Intermediate	Most important sources of uncertainty included.
Good	All important sources of uncertainty included.
Excellent	All reasonable sources of uncertainty included.

3.4.3 Uncertainty Values Provided

As described in Section 3.4.1, uncertainty values should be provided in EO data products per-pixel, in a manner that describes the pixel error-covariance. Since it is not practical to provide a full error-covariance matrix for an EO data product due to their data volume various approaches have been developed to approximate this. For example, the FIDUCEO project FCDRs contain three components of uncertainty – *independent*, *structured* and *common* – to describe the three typical scales of error correlation [RD-14].

It is still typical however for uncertainty values to be provided, if at all, on a per-product or, more often, a per-mission basis – losing a great deal of information significant to users. Table 3-15 shows the EDAP grades missions for the extent of uncertainty information they provide.

Table 3-15 – Uncertainty Characterisation > Uncertainty Values Provided – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No uncertainty information provided.
Basic	Single uncertainty value provided for whole mission.
Intermediate	Single uncertainty value provided for subsets of data, e.g. per product.
Good	Total uncertainty per pixel is provided, with basic breakdown of key components no error-covariance.
Excellent	Uncertainties per pixel provided with error-covariance information for all appropriate components.

3.4.4 Geolocation Uncertainty

Geolocation uncertainty is typically described as a circular error associated to a certain confidence level (e.g. 95%). It is a less common for the geolocation uncertainty to be described in a more detailed manner. For example, the geolocation error might be dependent on the latitude position, time of the year.

Similar to the measurement uncertainty in 3.4.1, the uncertainty associated to the geolocation requires a description of an error-covariance matrix when the product information is processed. Furthermore, the inaccurate geolocation of a pixel can result in an incorrect estimation of the measurement irrespectively of the uncertainty associated to the data product. Thus, the measurement uncertainty in 3.4.1 and the geolocation one are interrelated.

Table 3-16 gives the EDAP grading criteria for geolocation uncertainty.

Table 3-16 – Uncertainty Characterisation > Geolocation Uncertainty – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No uncertainty information provided
Basic	Single uncertainty value provided for whole mission.
Intermediate	Uncertainty value provided includes dependency on several variables.
Good	Uncertainty value provided includes dependency on several variables. Includes error-covariance information between pixels
Excellent	Uncertainty value provided includes dependency on several variables. Includes error-covariance information between pixels and impact on measurement uncertainty.

3.5 Validation

CEOS Working Group on Calibration & Validation (WGCV) defines validation as,

"the process of assessing, by independent means, the quality of the data products"

Validation therefore should assess the consistency between both the data values and their uncertainties with those of independent reference data.

3.5.1 Reference Data Representativeness

By the representativeness of the set of reference data we refer to the extent to which the measurements reflect the satellite measurements that they are being used to validate (e.g. point to pixel considerations), over the full extent of measurements the satellite may make (e.g. dynamic range, seasonal variation). This may in general require the use of a variety of different datasets to cover different observation conditions. Table 3-17 describes how EDAP grades the extent of validation reference data representativeness.

Table 3-17 – Validation > Reference Data Representativeness – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Reference measurements assessed to be somewhat representative of the satellite measurements, covering a limited range of satellite measurements. Typically a one-off campaign.
Intermediate	Reference measurements assessed to be mostly representative of the satellite measurements, covering a primary range satellite of measurements and at adhoc opportunities (no formal documented regular timescale).
Good	Reference measurements assessed to be well representative of the satellite measurements, covering a reasonable range of the satellite's measurements and carried out using FRM or community approved methods. Carried out on a regular timescale of approximately annual basis but not necessarily based on need.
Excellent	Reference measurements independently assessed to be fully representative of the satellite measurements, covering the satellite's full range of measurements and with full assessment of uncertainties and carried out on a regular basis determined by product performance.

3.5.2 Reference Data Quality

In the same way these guidelines describe how to assess the quality of satellite mission data, similar considerations must be made of the reference data used to validate them. Primarily, this concerns the following:

- Is uncertainty and error correlation information provided with the data?
- Have the data uncertainties been estimated with the GUM methodology?
- Is the data traceable to SI or a community reference standard?

The highest quality validation reference data therefore comes from activities such as the ESA Fiducial Reference Measurement (FRM) projects (e.g. [RD-17]), which provide uncertainty assessed validation references data traceable to SI.

Table 3-18 describes how EDAP grades validation reference data quality.

Table 3-18 – Validation > Reference Data Quality – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Uncertainty information not available for reference data.
Intermediate	Reference data comes a single uncertainty for the entire dataset.
Good	Reference data comes with full uncertainty information, assessed following the GUM and traceable to community reference or SI (e.g. FRM)
Excellent	Reference data comes with full uncertainty and error-correlation information, assessed following the GUM and traceable to SI (e.g. FRM).

3.5.3 Validation Method

A metrologically-rigorous validation should assess both the satellite measurements and their associated uncertainties. Commonly values such as the statistical spread of the results may be used to estimate the uncertainty, however this often may not provide a realistic estimate of the actual uncertainty.

Validated uncertainties provide evidence of the credibility of the uncertainty estimate given.

Table 3-19 shows how EDAP grades validation methodology.

Table 3-19 – Validation > Reference Data Representativeness – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Methodology is simple comparison, uncertainties not considered.
Intermediate	Methodology assess satellite measurements, simple uncertainty estimated e.g. from statistical spread for results.
Good	Methodology assesses satellite measurements and reference data w.r.t. their uncertainties.
Excellent	Methodology assess satellite measurements and reference data w.r.t. their error-covariance and validates those uncertainties.

3.5.4 Validation Results

This final sub-section of the validation quality assessment deals with the results of the validation activities themselves. In the best case these will show both validated satellite measurement and uncertainties and will have been obtained by a group independent of the satellite mission owner.

The results should be documented in a Validation report, prepared following the QA4ECV guidance for expected content [RD-19].

Table 3-20 how EDAP grades validation results.

Table 3-20 – Validation > Validation Results – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Validation results show some agreement between satellite and reference measurement.
Intermediate	Validation results show good agreement between satellite and reference measurements within uncertainties in most cases.
Good	Validation results show excellent agreement between satellite and reference measurements, within uncertainties. Analysis performed independently of satellite mission owner.
Excellent	Validation results show excellent agreement between satellite and reference measurements, within uncertainties. Uncertainty validated. Analysis performed independently of satellite mission owner.

4. REPORTING DATA PRODUCT QUALITY ASSESSMENTS

The data product quality assessment for a given mission should be reported by populating the template EDAP Data Product Quality Assessment Report, based on the considerations outlined in Section 3. The template of this report is available alongside these guidelines.

The concept of this report is that it is a summarising form that covers each section of analysis required by our approach (see Section 3) from which supporting and justifying documentation is referenced to provide more detailed information. These could include documents such as ATDBs or product validation reports, which may have been developed internally or externally to the project.

At the end of the Data Product Quality assessment report the product evaluation matrix should finally be coloured in based on the results of the analysis.