QA4ECV

Quality Assurance for Essential Climate Variables

Project Number 607405

Deliverable:	D2.3 Outline of QA4ECV Quality Assurance Service
Responsible Partner:	National Physical Laboratory
Delivery date:	August 2017

Outline of QA4ECV Quality Assurance Service

(Version 2.0)

Authors:
Tracy Scanlon [1]
Joanne Nightingale [1]

Other Contributors:
Folkert Boersma [2]
Jan-Peter Muller [3]
Clive Farquhar [4]
Steven Compernolle[5]
Jean-Christopher Lambert[5]

[1] National Physical Laboratory, Teddington, UK
[2] Koninklijk Nederlands Meteorologisch Instituut, Netherlands
[3] University College London, UK
[4] CGI, Leatherhead, UK
[5] BIRA-IASB, Brussels, Belgium

CONTENTS

TERMS & ACRONYMS

APM	Applications Performance Matrix
ATBD	Algorithm Theoretical Basis Document
BIRA-IASB	Belgian Institute for Space Aeronomy
C3S	Copernicus Climate Change Service
CCI	ESA's Climate Change Initiative
CDR	Climate Data Record
CORE-CLIMAX	Coordinating earth observation data validation for re-analysis for climate services (EU FP7 Project)
DOI	Digital Object Identifier
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
EU	European Union
EU-FP7	EU Seventh Framework Programme
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
LPV	Land Product Validation
PUM	Product User Manual
QA	Quality Assurance
QA4ECV	Quality Assurance for Essential Climate Variables (EU FP7 Project)
QI	Quality Indicator
QF	Quality Flag
SMM	System Maturity Matrix
SSD	Service Specification Document

EXECUTIVE SUMMARY

In support of the European Union's Earth Observation Programme's Copernicus Climate Change Service (C3S), the Quality Assurance for Essential Climate Variables (QA4ECV) project aims to fulfil a current gap in the delivery of climate quality satellite derived climate data products. The project will prototype a system for the implementation and evaluation of QA measures for 6 satellite-derived ECV and ECV precursor datasets, thus providing confidence in their application for climate monitoring studies and climate change assessments.

The purpose of developing and implementing a QA4ECV system is two-fold:

1. To provide **ECV data product producers / science teams** with the necessary resources (internationally accepted tools, standards, methodologies) to develop products with embedded QA information that is presented in a clear and common format throughout the Earth Observation (EO) community, and,
2. To provide **ECV data users (scientists – policy-makers)** with robust QA information as a means to quantitatively assess uncertainty and fitness-for-purpose of the data and derived products.

This report provides an outline of the QA service which is to be provided under the QA4ECV project. The ***QA system*** has been implemented is an interactive web-service and the ***documentary framework*** is a series of documentation including procedures, good practice guidance and training which support the QA system.

This document describes the implemented QA system including the reasoning for some of the decisions made in the development of the system. The document refers to other underpinning studies described in other documents. This is an update to the "Outline of Documentary Framework" issued in June 2015.

1 Introduction

1.1 Quality Assurance for Essential Climate Variables

Climate change mitigation and adaptation has risen to the top of the agenda for many governments and international organisations. This has led to the establishment of projects and programmes dedicated to the development of long-term global records of Essential Climate Variables (ECVs) using space-borne assets.

In support of the European Union's Earth Observation Programme's Copernicus Climate Change Service (C3S), the Quality Assurance for Essential Climate Variables (QA4ECV) project aims to fulfil a current gap in the delivery of climate quality satellite derived climate data products. The project will prototype a system for the implementation and evaluation of QA measures for 6 satellite-derived ECV and ECV precursor datasets, thus providing confidence in their application for climate monitoring studies and climate change assessments.

The purpose of developing and implementing a QA4ECV system is two-fold:

1. To provide **ECV data product producers / science teams** with the necessary resources (internationally accepted tools, standards, methodologies) to develop products with embedded QA information that is presented in a clear and common format throughout the Earth Observation (EO) community, and,
2. To provide **ECV data users (scientists – policy-makers)** with robust QA information as a means to quantitatively assess uncertainty and fitness-for-purpose of the data and derived products.

Provision of such QA information will demonstrate traceability of products and enable objective comparisons between the same ECVs produced by independent science teams. It will also provide data users with evidence-based confidence in the products and enable evaluation on the fitness-for-purpose of various ECV Climate Data Records (CDRs) for their specific applications.

1.2 Objective

The focus of this document is explaining the QA4ECV Quality Service including how it has been developed and how certain decisions have been arrived at, i.e. through dedicated studies, communication with the community etc.

The Quality Service is partially based on the service outlined in the Service Specification Document (SSD) [1], however it also takes into account comments from the community up to August 2017, the output of the first annual review of the QA4ECV project [2] and various aspects identified through a detailed gap analysis of one land product [3], as well as various studies. It is noted that the QA service will be developed taking into consideration all three ECV domains (land, atmosphere and ocean), not just those considered within the QA4ECV project.

The QA service being implemented within the QA4ECV project is a **prototype only**. Details of further refinements / improvements required to make the QA service routinely operational will be documented at the end of the QA4ECV project.

This document is an update to the documentary framework outline [4] issued in June 2015[1]. Since June 2015, the QA system has been developed; the new version of this document describes the implemented QA system.

[1] Note the change in the title of the document since Version 1 is to minimise confusion of terminology.

2 Basis of the Quality Service

One of the main aims of the QA system being developed under QA4ECV is to bridge the gap between data users and data producers, i.e. allow the transfer of information between the two in easy-to-use and consistent formats (as far as practicable)[2]. Taking this into account, the QA system is split into two main "paths":

- Data providers will follow a series of pages through the QA system to provide evidence relating to their ECV data product. This evidence will be included in a central repository.
- Data users will be able to search and download quality reports about a range of ECV data products from the central repository for comparison.

The ***QA service*** includes:

- The ***QA system*** is a physical system implemented is an interactive web-service through which data products will be assessed, and,
- The ***documentary framework*** is a series of documentation including procedures, good practice guidance and training which support the QA system (and are linked to throughout the QA system).

The structure of the ***QA service*** is summarised in Figure 1[3] (in green) within the context of ECV production (yellow) and dissemination (blue and purple). The ***documentary framework*** which is the associated guidance etc. is within the "Tools & Guidance to establish and evaluate QA" box (in green).

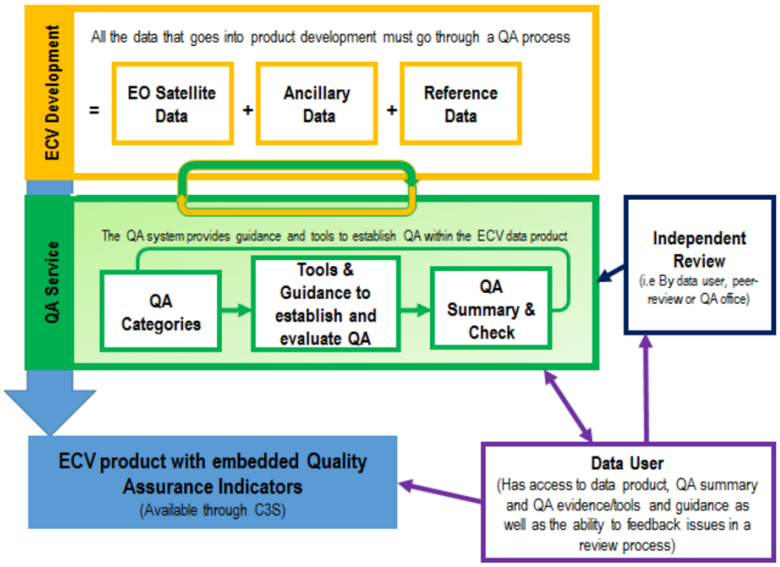

Figure 1: Diagram of system.

[2] Note that several other EU funded projects including EUPORIAS and Core Climax have identified this as a key factor which could improve the overall use of climate data sets. See [6].
[3] Note this diagram is also provided at www.qa4ecv.eu.

3 Data Producer Experience

3.1 Introduction

Data producers will provide information about their products to the QA system which will become their quality records. Once a data producer has entered the system and provided log on details, they will be taken to a page which describes the Quality Service including a diagram showing how different parts of the Service interact (see Figure 1).

The data producer is then asked to provide information about their products over several pages, with each page covering a Quality Indicator (QI). Each page can be accessed (and the progress on each page observed) using the bars shown in Figure 2 on the first page. The following sections describe the contents of each page and associated guidance etc. which is available.

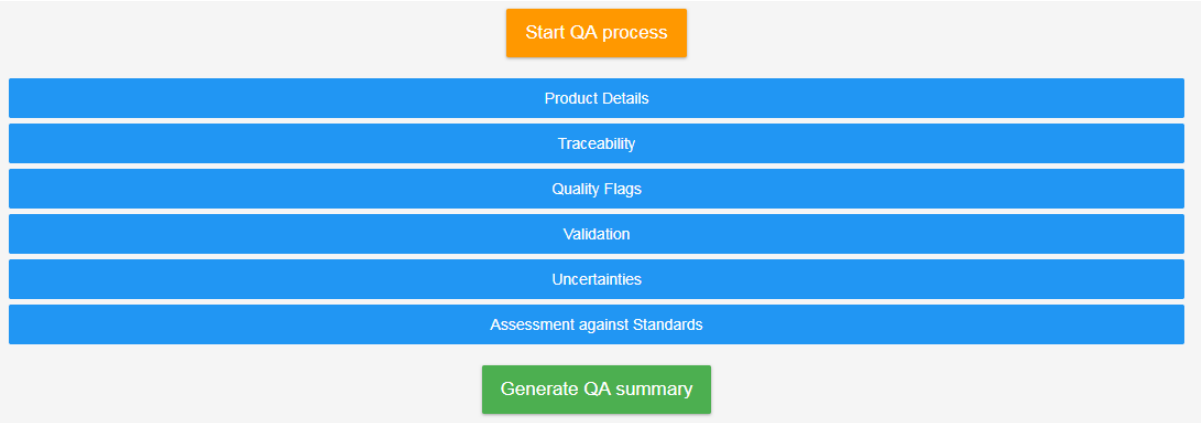

Figure 2: QA indicators within the QA system.

3.2 Product Details

The product details page allows the data producer to enter information about the data product. They may select the appropriate ECV and ECV product from the GCOS list of ECVs [5]. Generic information about the product is requested including: spatial resolution, temporal resolution, spatial coverage, temporal coverage, in-situ datasets used, satellite or airborne datasets used, etc.

The information requested has been determined from consideration of ECV data products as well as previous lists drawn up by other EU funded projects (for example in Core Climax (See [6] for further details). Certain fields, such as that relating to the CHARMe tool and the question on the granularity of a Digital Object Identifier (DOI) have been derived from consideration of other projects (See [6] for further details).

In addition, the page asks for details of the documentation which describes the data product including the Algorithm Theoretical Basis Document (ATBD) and Product User Manual (PUM). Guidance and templates are provided for both and are linked to on this page. These documents have been derived from consideration of existing similar documents, for example those produced on behalf of ESA (for the Climate Change Initiative (CCI) and Glob projects). For each document, a one-page guidance document is provided along with a template in word.

3.3 Traceability Chain

The traceability chain page provides a tool which allows users to draw their traceability chains. The current version of the tool is shown in Figure 3. Guidance on the use of the tool and conventions for the development of traceability chains are also provided on this page.

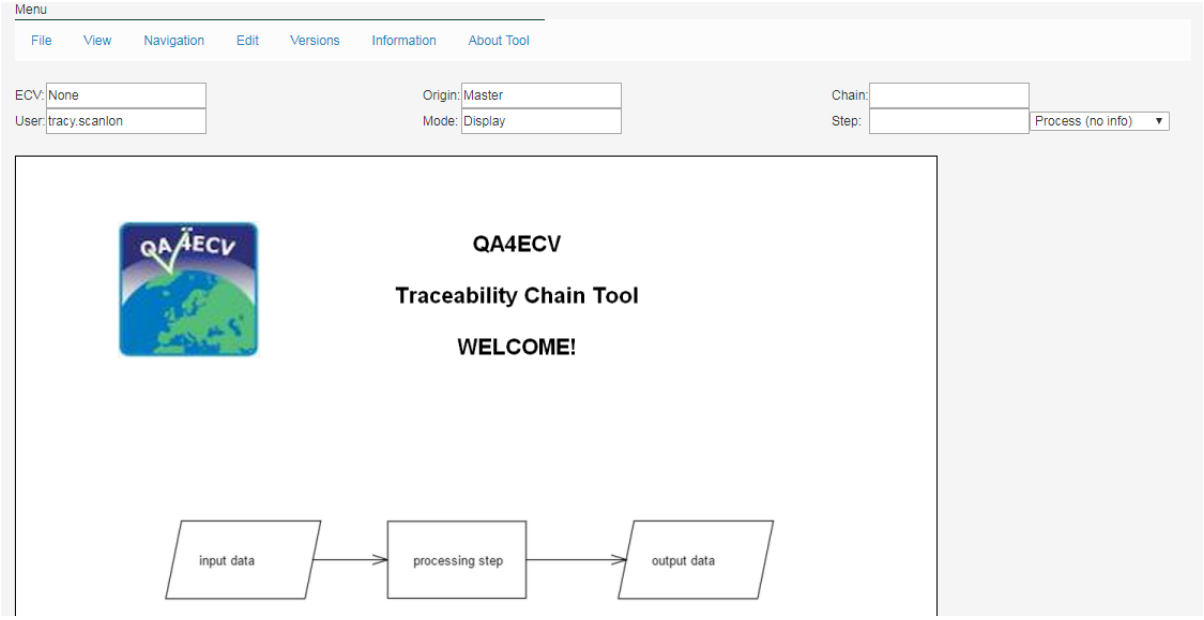

Figure 3: Traceability chain tool.

3.4 Quality Flags

The Quality Flags (QFs) section requests information from the data producer about the pixel level quality information that is provided with the data product. This includes questions on the meanings of each field and how the QFs have been derived.

A study has been undertaken to determine what information should be requested on this QI page [7]. The study has sought to determine what type of QFs should be provided with data products based on those commonly provided in a reference sample of data products.

The content of the QF page is dependent on the ECV domain chosen on the product details page.

3.5 Validation

The validation page requests information about the validation activities which the product has been subjected to. The content of the validation page is dependent on the ECV domain chosen on the product details page.

For each domain, data producers are asked to provide the validation report(s), comparisons with independent reference data and comparisons with other ECV datasets (i.e. intercomparisons). For the validation report, guidance and a template are provided.

For the terrestrial ECVs validation page, the data producer is asked to provide information about how their product compares to the Land Product Validation (LPV) validation hierarchy and for the atmospheric ECVs validation page, the data producer is asked to provide information about the validation protocol steps which the product has been subjected to (those developed under QA4ECV).

3.6 Metrological Traceability

The metrological traceability category requests evidence that their product contains uncertainty information and details on how the uncertainty estimates have been produced. This includes information on how the uncertainties from each dataset have been included into the final uncertainty estimate and how uncertainties introduced at each stage of the processing have been accounted for.

3.7 Assessment against Standards

This page asks product producers for evidence that a data product has been reviewed or assessed against the System Maturity Matrix (SMM) from the Core Climax project and GCOS requirements.

System Maturity Matrix

The aim of the SMM (described at [8]) is to evaluate the production of the ECV CDRs to ensure they follow best practices for science, engineering and utilisation; it is not to assess the quality of the data itself.

The SMM is incorporated into the QA System through an interactive table in which the assessed level of maturity of a data product can be selected and this will provide a colour to the table. Some of the fields may be pre-populated based on information provided under other QIs, for example, where a validation report has been provided, this would result in a pre-populated assessment of "2" against "Formal Validation Report"; this assessment may be changed by the data provider if required.

An example of a randomly filled in table from the Quality System is shown at Figure 4. The instructions on how to use the SMM [9] will be provided as part of the Quality System.

Software readiness	Metadata	User documentation	Uncertainty characterisation	Public access, feedback and update	Usage
Coding standards	Standards	Formal description of scientific methodology	Standards	Public Access/Archive	Research
Software Documentation	Collection level	Formal Validation Report	Validation	Version	Decision Support System
Numerical Reproducibility and Portability	File level	Formal Product User Guide	Uncertainty quantification	User Feedback Mechanism	
Security		Formal description of operations concept	Automated Quality Monitoring	Updates to Record	

Legend

0	1	2	3	4	5	6

Figure 4: Randomly filled in System Maturity Matrix table from the Quality System.

3.8 Quality Summary and Checking Records

Once information has been provided to the Quality System by the data producer, there will be an option to provide a product "Quality Summary" which is a report summarising the available Quality Information for a data product. At this time, the information provided into the Quality System will be made available on the user's area of the Quality System (see further details in Section 3).

Once all QA records have been provided to the QA system, these will need to be checked to ensure that all relevant information has been provided by the data producer and all content

is suitable for wider dissemination[4].

A checklist is provided against which the records can be assessed as well as a procedure [10]. The checklist provides three levels for each aspect considered important to the quality of the ECV data records.

3.9 Summary of Supporting Studies and Available Guidance / Templates

Decisions made in the development of the QA system are underpinned by a series of supporting documents. These are listed in Table 1.

Table 1: Underpinning Studies Contributing to the Design of the Quality System

Document Name	Purpose	Ref.
Linkages between the QA Service of QA4ECV and Other Projects and Initiatives	The QA system and documentary framework will take into consideration QA procedures and methods, as well as guidance and ideas, developed in other similar EU-funded projects and by established organisations. These will be incorporated into QA4ECV as far as possible, to ensure a coherent approach to quality within the ECV community.	[6]
Review of Pixel Level Quality Flags Provided with ECV Data Products	The QI study will consider the QIs which are provided in ECV products similar to those being produced under the scope of the QA4ECV project. The ECVs selected will come from varied sources, i.e. NASA, ESA etc. to determine if there are commonalities between ECVs or within organisations. The study will determine the QIs provided (i.e. data fields) and how they are provided, i.e. as separate data layers, as binary fields etc. The report will also provide generic recommendations for the provision of ECV QIs.	[7]
Review of Validation Standards and Practices for Terrestrial ECV Products	The validation study reviews current standards and practices implemented in the land validation community which is one the of the QA system QIs. To achieve this, this report studies the current standards / good practices available within the land validation community including information available from other similar projects. This is followed by the study of two validation reports to determine which validation studies are commonly undertaken and how the results are presented.	[11]

The documentary framework provides the underpinning documentation to support the Quality System. The documents which are to form the documentary framework (listed in Table 2) will be provided on relevant QI pages of the Quality System (see Section 3 for further details).

[4] Note, this does not include assessment of the quality of the data, i.e. the stability, accuracy etc.

Table 2: Documents forming the Documentary Framework

Document Name	Purpose	Ref.	Page of Quality System
ATBD one page guidance	To provide data producers with an easy-to-use guidance document on the contents of an ATBD and how this should be presented.	[12]	Product Details
ATBD template	To provide a template which data producers can use to create an ATBD if they do not already have one. Meets the requirements of the ATBD guidance.	[13]	Product Details
PUM one page guidance	See “ATBD one page guidance”.	[14]	Product Details
PUM template	See “ATBD template”.	[15]	Product Details
Traceability chains one page guidance	To provide guidance on what the format of a traceability chain should be and the type of information that the supporting text boxes should contain.	[16]	Traceability Chain
Traceability chain tool user manual	A manual on how to use the tool including details of how to construct a chain and how to add supporting information.	[17]	Traceability Chain
Validation one page guidance	See “ATBD one page guidance”.	[18]	Validation
Validation template	See “ATBD template”.	[19]	Validation
QA Records Checklist	To provide a list of requirements against which Quality Records will be checked. This will be available to the data producers so that they are aware of what they should achieve.	[20]	Introduction to QA system

4 Data User Experience

4.1 Introduction

It is expected that users of ECV data will use the system primarily for the following two functions (for the available ECV records):

1. To identify suitable datasets which meet their needs, and,
2. To discover information available about datasets which they utilise.

To allow this, the system provides a static label which can be applied to a dataset, as well as a search functionality and the ability to download summary and assessment reports for individual products. The following describes these aspects of the data user experience.

4.2 Quality Label

The quality label, shown in Figure **5**, allows the user to see a quick assessment of the quality of a product[5]. This label is provided on the assessment summary report which is produced from the checking of quality records for a data product. The colours relate to the degree to which each level has been met on the quality checklist.

Figure 5: Quality Label utilised to graphically show the assessment of a data product as part of the QA4ECV system.

[5] The idea of the quality label is based on the GEO Label utilised by GEOSS in their datasets. For more information see [6].

4.3 Search Functionality

There are two facilities available for searching the database of records which have been subject to the QA system:

1. Basic search – which allows simple conditions in the search, i.e. products which have provided a traceability diagram.
2. Applications Performance Matrix (APM)[6] – which allows a more complex search allowing the user to define a set of criteria which they want an ECV product to match. The search will return a matrix showing how different products meet these criteria.

[6] The APM was originally suggested within the Core Climax project. For more information see [6].

5 QA Service Training

To ensure the QA service is used as intended, training will be provided for data providers. The training on the QA system will be undertaken / made available as:

- Several interactive training sessions held by NPL for all of the project partners,
- A video available online, and,
- A user manual which is available as both a long and short version.

A document detailing training undertaken is available [21].

6 References

[1] J. Nightingale, N. P. Fox, T. Scanlon, F. Boersma, J.-C. Lambert, A. De Rudder, J. Schulz, and J. P. Muller, “QA4ECV: Service Specification Document (D2.1),” 2016.

[2] S. Bojinski, “QA4ECV First Technical Review Report,” 2015.

[3] T. Scanlon, J. Nightingale, A. De Rudder, S. Compernolle, J. P. Muller, and F. Boersma, “QA4ECV: Gap Analysis of QA4ECV ECV Products (D2.2),” 2015.

[4] T. Scanlon, J. Nightingale, F. Boersma, J. P. Muller, and C. Farquhar, “QA4ECV: Outline of Documentary Framework (D2.3) - Version 1.0,” no. Version 1.0, pp. 1–23, 2015.

[5] I. P. Subject, “The Global Observing System for Climate : Implementation Needs,” 2016.

[6] T. Scanlon, S. Douglas, J. Nightingale, S. Compernolle, and J.-C. Lambert, “Review of Quality Aspects in EU ESA and US funded ECV-related projects,” 2017.

[7] T. Scanlon, J. Nightingale, S. Compernolle, and J.-C. Lambert, “Review of Pixel Level Quality Flags Provided with ECV Data Products,” 2017.

[8] J. J. Bates and J. L. Privette, “A Maturity Model for Assessing the Completeness of Climate Data Records,” *Eos --- Trans. Am. Geophys. Union*, vol. 93, no. 44, p. 441, 2012.

[9] Core Climax, “Core Climax: Climate Data Record Assessment Instruction Manual,” 2013.

[10] T. Scanlon and J. Nightingale, “QA4ECV: Quality Checklist Procedure,” 2017.

[11] T. Scanlon and J. Nightingale, “Review of Validation Standards and Practices for Terrestrial ECV Products,” 2017.

[12] T. Scanlon, “QA4ECV: Documentation Guidance: Algorithm Theoretical Basis Document,” 2017.

[13] T. Scanlon, “QA4ECV: Product Documentation Template: Algorithm Theoretical Basis Document,” 2017.

[14] T. Scanlon, “QA4ECV: Product Documentation Guidance: Product User Guide,” 2017.

[15] T. Scanlon, “QA4ECV: Product Documentation Template: Product User Guide,” 2017.

[16] T. Scanlon, “QA4ECV Guidance: Provenance Traceability Chains,” 2017.

[17] Martin Milnes, “Traceability Tool User Guide,” 2016.

[18] T. Scanlon, “QA4ECV: Product Documentation Guidance: Product Validation and Intercomparison Report,” 2017.

[19] T. Scanlon, “QA4ECV: Product Documentation Template: Product Validation and Intercomparison Report,” 2017.

[20] T. Scanlon, “QA4ECV Guidance: Quality Checklist,” 2017.

[21] T. Scanlon and J. Nightingale, “QA4ECV Overview of Training on the QA System,” 2017.