Specifications for Coherent Validation Schemes for Sentinel Satellite Products, (Processing Lines and Services)

Deliverable D4.1

Issue 0.1

Due date of deliverable: 01 June 2015
Actual submission date: 24 May 2015

SEN3APP

Processing Lines And Operational Services Combining Sentinel And In-Situ Data For Terrestrial Cryosphere And Boreal Forest Zone

FP7 Grant agreement No 607052

Document History

Date	Update	Contributor	Organization
12/05/2015	Document initiation	Olli-Pekka Mattila	SYKE
13/05/2015	Revision, input and edits	Gabrielle Bippus	ENVEO
22/05/2015	Input	Timo Ryyppö	FMI
22/05/2015	Final editing	Olli-Pekka Mattila	SYKE

Document Release Sheet

Book captain:	Olli-Pekka Mattila (SYKE)	Sign Date
Approval (lead contractor for WP4)	Olli-Pekka Mattila (SYKE)	Sign Date
Endorsement:	Ali Nadir Arslan (FMI)	Sign Date

Dissemination level	
Public	**x**
Restricted to a group (including the Commission Services) specified by the consortium	
Confidential, only for members of the consortium and for the Commission Services	

Table of Contents

Acronyms

CEOS   Committee on Earth Observation Satellites

EO   Earth observation

GEO   Group on Earth Observation

GEOSS   Global Earth Observation System of Systems

QA   Quality assurance

QA4EO   Quality Assurance for Earth Observation

SEN3APP   Acronym for EU FP7 project "Processing lines and operational services combining Sentinel and in-situ data for terrestrial cryosphere and boreal forest zone"

1.   Abstract

Users of satellite data derived products for environmental applications need variety of information to evaluate the suitability of the product for their purposes. This requires quantitative information about the performance of the product, but also documentation of processing steps taken from the satellite raw data to final product. The user is also interested on the authority that has produced the information; the original purpose of the product; detailed description of processing of the data etc. This document describes the procedures taken in SEN3APP to produce this information. The document sets the basic requirements for the information to be provided about the final products. Due to the variety of SEN3APP products and their uses, it is left up to the product providers, i.e. project partner, to select the most suitable statistical measures and figures to describe the performance of the product. The Quality Assurance for Earth Observation (QA4EO) Framework (http://qa4eo.org), developed by the participants in the Committee on Earth Observation (CEOS) and endorsed by the Group on Earth Observation (GEO) is adopted as a guiding documentation, but SEN3APP will not limit to this, as there can be case sensitive needs. The basic information reported by the QA4EO-framework is considered as the minimum quality assessment to be provided for each SEN3APP product.

2.   Scope

This document describes the validation procedures for processing lines, products and services and the concept of NRT-validation applied in SEN3APP project. It also provides the recommended content (Annex 1) for documenting the pre-processing lines for satellite data, product processing lines and for product datasets. The content is adopted from the QA4EO framework, but is also subjected to alteration according to the needs of data, processing line or service provider. The QA4EO is envisioned to include also ground based earth observations (e.g. automated in-situ measurements), but the topics discussed in this document are connected only to assessment and documentation of satellite data products and processing lines in SEN3APP.

This document also follows the QA4EO framework documentation in its structure.

3.   Terminology

Here some key concepts are determined for their meaning in this document:

Data products: If not otherwise indicated data products refers to environmental datasets derived from satellite image data

Earth observation (EO): Here EO is concentrated on satellite based observations

Metadata: Descriptive information about the processing lines, data products and services

Processing line: Collection of software modules and logical data information resources needed to derive a product (satellite image product or satellite data derived product for describing information about environment)

Service: Collection of software modules and logical information resources needed to create access to products by SEN3APP users.

Validation: Validation is used here in three meaning: 1) Statistical measures of performance of the products; 2) Measures of performance for the processing lines; 3) A minimum set of information for a user to assess the suitability of the product for the purpose

4.   Background and Context

In GEO's envisioning of creating the Global Earth Observation System of Systems (GEOSS) and for satellite based products to be used (as part of GEOSS network) effectively in public and private sector applications, as well as in research, the data needs to fulfil two key principles:

1. Accessibility/Availability and

2. Suitability/Reliability

QA4EO framework aims at supporting to accomplish the second objective. In order to assure the benefit from EO system derived information for users, there is a need for information about the suitability of the product for the end user's purpose. The user should have information about uncertainties, related to the datasets and products, e.g. uncertainties related to the underlying data, uncertainties occurring in the processing steps and uncertainties used in the final data products. Additionally, the user is likely to be interested in background information about the producer, original purpose of the product, standards associated with the information; performance against similar data sources, models used in the processes etc. Therefore, not only limiting to the quantitative statistical information usually associated with the concept of validation.

5.   Outcomes

The template, provided in Annex 1 and adopted from QA4EO [2], will serve as the template for deliverables D4.2-4.8. The deliverables will be kept separate for clarity and to allow existing documenting systems of the participants (e.g. metadata services) to be used as the main source of quality assurance information.

The template (Annex 1) contains the minimum set of information for the satellite based products and can also be used for documenting the pre-processing lines (deliverable 4.8). If a partner has their own system for collecting the minimum documentation requirements for the quality assurance documentation (Annex 1), the deliverables D4.2-D4.8 can also be a brief descriptions of the system used and have references (links or attached documents or xml- files) to the descriptions of SEN3APP –products in the extent described in Annex 1.

6.   Inputs

In-situ observations and high resolution satellite data will be used for producing qualitative measures for assessing the algorithm performance applied on Sentinel data and the quality of products generated in the SEN3APP project.

The QA4EO framework (see 7. Standards and Traceability).

7.   Standards and Traceability

The documentation and validation of SEN3APP products, processing lines and services will follow the QA4EO- framework to the extent that is applicable, but will also extend the scope if needed. Following guidelines are identified to connect with the SEN3APP- project:

QA4EO-QAEO-GEN-DQK-001 *"A guide to establish a Quality Indicator on a satellite sensor derived data product"* [1]: This document refers to general key components that should be readily available from the satellite sensor operator for e.g. quality measures for digital data available from satellite sensors.

QA4EO-QAEO-GEN-DQK-002 *"A guide to content of a documentary procedure to meet the Quality Assurance requirements of GEO"* [2]: This is the key-guideline for recommended content in the written documentation for quality assurance of earth observation system, process or dataset. The document is also written in the recommended format of a quality assurance document in the context of QA4EO. The same structure is also recommended in SEN3APP for those data producers who do not have standardised practices already (e.g. ISO-19115 metadata model).

QA4EO-QAEO-GEN-DQK-003 *"A guide to "reference standards" in support of Quality Assurance requirements of QA4EO"* [3]: This document gives guidance on identification, establishment and use of reference standards for building solid base for quality indicators.

QA4EO-QAEO-GEN-DQK-004 *"A guide to comparisons – organisation, operation and analysis to establish measurement equivalence to underpin the Quality Assurance requirements of QA4EO"* [4]: Provides guidelines on inter-comparison between similar data products.

QA4EO-QAEO-GEN-DQK-005 *"A guide to establishing validated models, algorithms and software to underpin the Quality Assurance requirements of QA4EO"* [5]: The document suggests general guidelines on validating the models and algorithms used in producing information from satellite data (or other EO observations).

QA4EO-QAEO-GEN-DQK-006 *"A guide to expression of uncertainty of measurements"* [6]: Gives introduction to the ISO/IEC Guide 98-3:2008, also named as "Guide to the Expression of Uncertainty Measurements".

QA4EO-QAEO-GEN-DQK-007 *"A guide to establishing quantitative evidence of traceability to underpin the Quality Assurance requirements of QA4EO"* [7]: This guide summarizes the

traceability of the quality assessment and gives general outline how the final aim of setting up an unequivocal quality indicator for the satellite data derived product.

ISO/IEC Guide 98-3:2008, Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) [8]

8. Validation and quality assurance procedures in SEN3APP

8.1. Quality assurance documentation following QA4EO-framework

QA4EO provides a set of guidelines for building solid quality indicator for the EO (satellite, airborne or in-situ measurement system/procedure) data derived information. The guidelines are very general (at least for now), but provide a good framework to build on more extensive quality assurance for environmental information products.

There is strong interest in the scientific research as well as in technical development to create comprehensive metadata for services, data processing systems, computer models and datasets to make the information search and processing easier with machine readable information, to make systems communicate better with each other and to open opportunities for data semantic data processing. The general outline of QA4EO framework complies well with this background. The QA4EO-QAEO-GEN-DQK-002 [2] guideline sets up minimum requirements for the documentation of dataset and will act as the general template (Annex 1) for the information that should be provided for each data product and as template for Deliverables 4.2-4.8. The set of minimum requirements here is somewhat modified and can also be further developed (in the document, by including sub-sections) according to needs by the product providers, i.e. project partners, or from requests of the users.

The minimum requirements can also be documented by using standard metadata formats, such as ISO-19115 Geographic information – Metadata, if they are used by the participating institute. The systems, where the information is stored (typically in xml-formats) should then be described in D4.X instead of the information itself. The xml-realisation of the product documentation can be attached to the deliverable.

8.2. Numerical validation of products

The characteristics of products derived from satellite data in SEN3APP differ from each other, and therefore it's left to the product providers to select the methodology for numerical evaluation of the product performance.

The project partners can choose the most suitable methodology and indicators for reporting the performance of the satellite based products. QA4EO-QAEO-GEN-DQK-006 provides an introduction to ISO/ IEC "Guide to the Expression of Uncertainty Measurements", which can be a recommended source for establishing performance indicators.

Each partner will justify the selected methods for producing performance indicators in Deliverable 4.X for the particular products.

8.3. NRT- validation

The NRT- validation is a new concept that will be tried in SEN3APP for selected products. The key-concept is to derive up-to-date comparison of EO based products with available in-situ datasets. This information can be displayed as figures, numerical values or as additional data layer together with the satellite based product. The timelines (annual, monthly, daily, or NRT) and extent of the validation information depends on the availability of in-situ data. Following in-situ datasets and products have been identified for testing the concept:

- SYKE snow course network: The network consists of over 100 snow courses over Finland. The data is suitable for analysis over large areas, but due to manual procedures in in-situ data storage and quality control, the dataset is closer to traditional algorithm validation. → Data can be utilized for annual regional evaluation of snow products.

- Weather station networks: Both national and international weather station networks offer data on snow depth and snow covered area. This data is more readily available in daily basis. → The data can be used for larger areas on NRT-bases, although data policies may restrict the use.

- Web-camera network: A network of web-cameras is under development in Finland for monitoring vegetation and snow phenology. This data, when available, could be used as NRT- validation data source. The coverage of the network is currently fairly sparse, ~10 stations in Finland and data processing is still under development. → Dataset could be used for quick point-wise evaluation of snow products.

One or two of these data sources will be tested for validation purposes of mainly snow products in SEN3APP.

8.4. Validation of SEN3APP processing lines and services

Processing lines and services must be built in an efficient way that they won't waste available resources. This comes more and more important when using virtualization where the resources are shared between several virtual machines. If one of the machines is using too much, for example, memory from the resource pool due to a bad coding or memory leak it can slow down the whole virtual group even other machines were properly configured.

Many of the virtual environment software have their own monitoring tools but there are also common tools which can be used for both virtualized and physical servers regardless of the system provider. The best results are got by using the combination of different monitoring tools for different purposes. For example, FMI uses iLO (Integrated Lights-Out) to manage Hewlett-Packard servers and both Nagios and Check_MK to monitor network, software, disk usage and memory to name a few.

SEN3APP processing lines and services are validated using the currently available and used software by FMI ICT. CPU loads and memory usage are monitored to see how the system is performing and is there a need to optimize the code and after that to optimize the resources that will be provided to the processing line or service. Once the processing itself fulfils the requirements, data storage and transfer rates are validated. For example, monitoring the data rates and transferring times it is possible to see if there are some bottle necks and is there a need to reroute the data through routers and firewalls that have less traffic and higher band width.
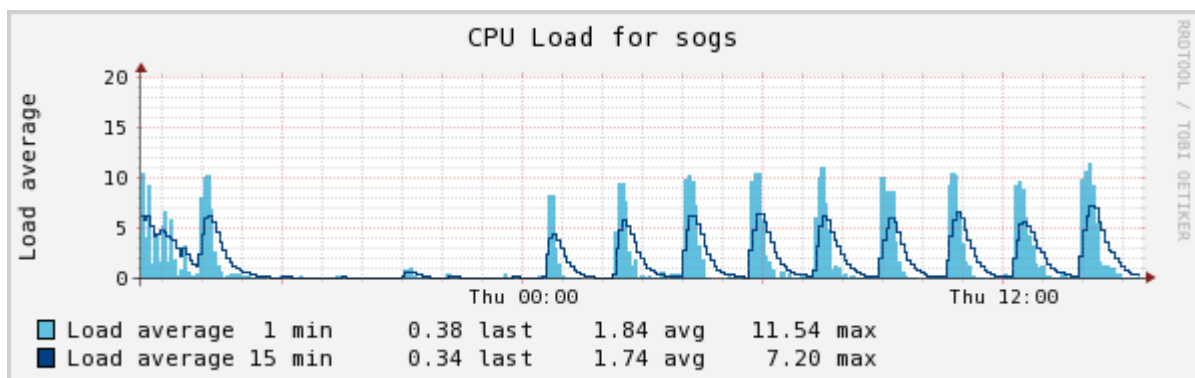

Figure 1. Server CPU load monitoring result from Check_MK software.

References

[1] QA4EO-QAEO-GEN-DQK-001 *"A guide to establish a Quality Indicator on a satellite sensor derived data product"*

[2] QA4EO-QAEO-GEN-DQK-002 *"A guide to content of a documentary procedure to meet the Quality Assurance requirements of GEO"*

[3] QA4EO-QAEO-GEN-DQK-003 *"A guide to "reference standards" in support of Quality Assurance requirements of QA4EO"*

[4] QA4EO-QAEO-GEN-DQK-004 *"A guide to comparisons – organisation, operation and analysis to establish measurement equivalence to underpin the Quality Assurance requirements of QA4EO"*

[5] QA4EO-QAEO-GEN-DQK-005 *"A guide to establishing validated models, algorithms and software to underpin the Quality Assurance requirements of QA4EO"*

[6] QA4EO-QAEO-GEN-DQK-006 *"A guide to expression of uncertainty of measurements"*

[7] QA4EO-QAEO-GEN-DQK-007 *"A guide to establishing quantitative evidence of traceability to underpin the Quality Assurance requirements of QA4EO"*

[8] ISO/IEC Guide 98-3:2008, Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

ANNEX 1

XXX Validation Report – QA4EO Documentation

Deliverable Dex.xx

Issue x.x

Due date of deliverable: dd Month YYYY
Actual submission date: dd Month YYYY

SEN3APP

Processing Lines And Operational Services Combining Sentinel And In-Situ Data For Terrestrial Cryosphere And Boreal Forest Zone

FP7 Grant agreement No 607052

Document History

Date	Update	Contributor	Organization
dd/mm/yyyy	Update	Names	Org.

Document Release Sheet

Book captain:	Name (ORG)	Sign Date
Approval (lead contractor for WPX)	Name (ORG)	Sign Date
Endorsement:	Name, Title, (ORG)	Sign Date

Dissemination level	
Public	x
Restricted to a group (including the Commission Services) specified by the consortium	
Confidential, only for members of the consortium and for the Commission Services	

Table of Contents

Acronyms

Introduction

1. **Identifier**

2. **Title**

3. **Author**

4. **Authority**

5. **Issue, Version Number, Date**

6. **Abstract**

7. **Scope, Purpose**

8. **Key terminology, Definitions**

9. **Background, Context, Requirements**

10. **Inputs**

11. **Outcomes**

12. **Methodology, Processing**

13. **Standards and Traceability**

14. **Evaluation of Performance**

15. **(Evidence to Support Performance Indicator)**

References