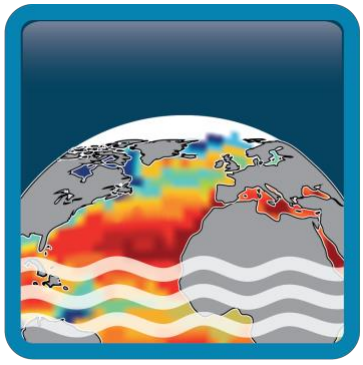
Climate Change Initiative+ (CCI+) Phase 1

Sea Surface Salinity



System Requirement Document (SRD)

Customer: ESA

Ref.: ESA-CCI-PRGM-EOPS-SW-17-0032

Version: v1.1
Ref. internal: AO/1-9041/17/I-NB
Revision Date: 06/12/2019
Filename: SSS_cci-D3.1-SRD-v1.0.docx

Signatures

Author	Stephen Emsley		06/12/2019
reviewed by	Frederic Rouffi		
	Giovanni Corato		
	Rafael Catany (Project Manager)		
APROVED by	Jacqueline Boutin (Science Leader)		06/12/2019
	Nicolas Reul (Science Leader)	<signature>	<date>
	Rafael Catany (Project Manager)		06/12/2019
accepted by	Craig Donlon (Technical Officer)	 Digitally signed by Craig Donlon Date: 2020.01.13 09:52:18 +01'00'	

Diffusion List
Sea Surface Salinity Team Members
ESA (Craig Donlon, Paolo Cipollini)

Amendment Record Sheet

Date / Issue	Description	Section / Page
JAN19/ v1.0	Delivery to ESA	New document
DEC19 / v1.1	Implementation ESA feedback to whole document (i.e. minor revisions)	NA

Table of Contents

Table of Contents

List of figures

List of tables

1 Introduction

1.1 Executive Summary

This document holds the System Requirement Document (SRD), which provides a set of system requirements for the CCI+ Salinity operational system, as requested in the Statement of Work (SOW (Task 3 SOW ref. ESA-CCI-PRGM-EOPS-SW-17-0032), geared to the Salinity Essential Climate Variable (ECV). The SRD is the primary input into the design of the operational system, to be described by the System Specification Document (SSD) and is thus the roadmap that will help the system Engineers to maintain, operate and enhance the system; to develop the software and perform their testing and installation.

1.2 Purpose and Scope

The purpose of the SRD is to specify the system requirements needed to achieve the operational and production goals for the European Space Agency (ESA) Climate Change Initiative Plus (CCI+) Salinity project.

As specified in the SOW (p.39) the SRD shall specify the requirements of a Processing System capable of generating Salinity Essential Climate Variable (ECV) data products as specified in Annex B of the SOW. The SRD shall include verifiable requirements on the following:

- Software requirements for data processing function of each step of its processing chain, including data volumes storage requirements
- Hardware requirements including platform specification of the Processing System for CCI+ Salinity
- Assess the requirements in terms of computing times and expected performance, considering data flow logic and technical limitations.
- Compliance to all processing needs defined by the Task 2 outputs i.e. constrains for the systems specifications based on the requirements appearing in the:
    - User Requirement Document (URD),
    - Product Specification Document (PSD),
    - Data Access Requirement Document (DARD) and

1.3 Intended Audience

The readership of this document is comprised of the CCI+ SSS consortium partners and ESA. There may also be scope, following further investigation, as to the use of this document for

the Software Engineering Working Group (SEWG), towards finding and forming common ground with other ECV projects as is encouraged in the Statement of Work (SOW).

1.4 Assumptions

In addition to requirements specified in the Statement of Work (SOW) and issue 2.0 of the CCI Data Standards Document (DSTD) this document is based on issue 1.4 of the User Requirement Document (URD), issue 1.6 of the Product Specification Document (PSD) and issue 1.4 of the Data Access Requirement Document (DARD); refinement of this document will be necessary in the event of future issues of those documents.

1.5 References

1.5.1 Applicable Documents

ID	DOCUMENT	REFERENCE
SOW	CCI+ Phase 1 – New ECV – Statement of Work	ESA-CCI-PRGM-EOPS-SW-17-0032
DSTD	CCI Data Standards	CCI-PRGM-EOPS-TN-13-0009
URD	CCI+ SSS User Requirements Document	SSS_cci-D1.1-URD-v1r4
PSD	CCI+ SSS Product Specification Document	SSS_cci-D1.2-PSD-v1r6
DARD	CCI+ SSS Data Access Requirement Document	SSS_cci-D1.3-DARD-v1r4

1.5.2 Reference Documents

ID	DOCUMENT	REFERENCE
PROP	Technical Proposal in response to CCI+ Phase 1 – New ECVS - Salinity	ARG-003-039(3) 27th October 2017
RD01	GCOS-200 (also GOOS-214), the Global Observing System for Climate: Implementation Needs (2016 GCOS Implementation Plan, 2015)	WEB LINK

ID	DOCUMENT	REFERENCE
RD02	European Cooperation for Space Standardization: Space Engineering - Software	ECSS-E-ST-40C 6th March 2009
RD03	Hwang, P. A., D. M. Burrage, D. W. Wang, and J. C. Wesson (2011), An Advanced Roughness Spectrum for Computing Microwave L-Band Emissivity in Sea Surface Salinity Retrieval	IEEE Geoscience and Remote Sensing Letters, 8, 547.
RD04	Reul, N., S. Saux-Picart, B. Chapron, D. Vandemark, J. Tournadre, and J. Salisbury, (2009), Demonstration of ocean surface salinity microwave measurements from space using AMSR-E data over the Amazon plume	JGR, 36, DOI: 10.1029/2009GL038860
RD05	GOOS requirements: OOPC (2017), EOV Spec Sheet: Sea Surface Salinity v5.2	WEB LINK
RD06	Escorihuela et al. (2018), Low Frequency Passive Microwave User Requirement Consolidation Study: D-02 White paper on L-band radiometry for earth observation: status and achievements	ESA ITT AO/1-8731/16/NL/IA. Reference: ISARD_ESA_LBAND_TN_565, Issue: 3.0, Date: 5 September 2018
RD07	CF Conventions and Metadata	WEB LINK
RD08	CF Standard Names	WEB LINK
RD09	The Climate Change Initiative Ontology	WEB LINK
RD10	Attribute Convention for Data Discovery (ACCD)	WEB LINK
RD11	UNIDATA Program Center of the University Corporation for Atmospheric Research UDUNITS Software	WEB LINK

1.6 Acronyms

AD	Applicable Document
ACCD	Attribute Convention for Data Discovery
ADP	Algorithm Development Plan
AOPC	Atmospheric Observation Panel for Climate
AR	Assessment Report (of the IPCC)
AR6	IPCC Scientific Assessment Report 6
ATBD	Algorithm Theoretical Basis Document
C3S	Copernicus Climate Change Service
CAR	Climate Assessment Report
CCI	The ESA Climate Change Initiative (CCI) is formally known as the Global Monitoring for Essential Climate Variables (GMECV) element of the European Earth Watch Programme
CCI+	Climate Change Initiative Extension (CCI+), is an extension of the CCI over the period 2017–2024
CDR	Climate Data Record
CEOS	Committee on Earth Observation Satellites
CFOSAT	Chinese French Oceanography Satellite
CGMS	Coordination Group for Meteorological Satellites
CliC	World Climate Research Programme - Climate and Cryosphere Project
CLIVAR	WCRP Climate Variability and Predictability project
CMEMS	Copernicus Marine Environmental Monitoring Service
CMIP	Coupled Model Intercomparison Project
CMUG	Climate Modelling User Group
COP	Conference of the Parties
COWCLIP	Coordinated Ocean Wave Climate Project (of JCOMM)
CR	Cardinal Requirement
CRDP	Climate Research Data Package
CRG	Climate Research Group
CSCDA	Copernicus Space Component Data Access System
CSWG	Climate Science Working Group
DARD	Data Access Requirements Document

DEWG	Data Engineering Working Group
DOI	Digital Object Identifier
DPM	Detailed Processing Model
DTBT3	Database for Task 3
DUE	Data User Element
E3UB	End-to-End ECV Uncertainty Budget
EC	European Commission
ECMWF	European Centre for Medium Range Weather Forecasts
ECSAT	European Centre for Space Applications and Telecommunications
ECSS	European Cooperation for Space Standardization
ECV	Essential Climate Variable
EOV	Essential Ocean Variable (of the OOPC)
ESGF	Earth System Grid Federation
ESM	Earth System Model
EU	European Union
FCDR	Fundamental Climate Data Record
FIDUCEO	Fidelity and uncertainty in climate data records from Earth Observations
FP7	EU Framework Programme 7
FRM	Fiducial Reference Measurements
GAIA-CLIM	Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring
GEO	Group on Earth Observations
GCW	Global Cryosphere Watch
GMECV	Global Monitoring of Essential Climate Variables - element of the European Earth Watch programme.
GNSS	Global Navigation Satellite System
GOOS	Global Ocean Observing System
H2020	Horizon 2020 programme
Hs	Significant Wave Height (see also SWH)
H-SAF	EUMETSAT's Hydrology Satellite Applications Facility
HDD	Hard disk
IOC	Intergovernmental Oceanographic commission (of UNESCO)
IODD	Input Output Data Definition

IP	Implementation Plan
IPCC	Intergovernmental Panel on Climate Change
ISDB	in situ database (of Fiducial Reference Measurements and satellite measurements)
JAXA	Japan Aerospace Exploration Agency
JCOMM	Joint Commission on Oceanography and Marine Meteorology
MOOC	Massive Open Online Course
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NOP	Numerical Ocean Prediction
NWP	Numerical Weather Prediction
Obs4MIPs	Observations for Model Intercomparison Projects
ODP	Open Data Portal
OOPC	Ocean Observation Panel for Climate
PMP	Project Management Plan
PSD	Product Specification Document
PUG	Product User Guide
PVASR	Product Validation and Algorithm Selection Report
PVIR	Product Validation and Intercomparison Report
PVP	Product Validation Plan
QA4EO	Quality Assurance Framework for Earth Observation
QSR	Quarterly Status Report
R&D	Research and Development
RCP	Representative Concentration Pathways
RD	Reference Document
SAF	Satellite Applications Facility
SAR	Synthetic aperture Radar
SISS	Satellite and In situ [Working Group]
SLP	Sea Level Pressure
SMAP	Soil Moisture Active Passive [mission of NASA)
SMOS	Soil Moisture and Ocean Salinity [satellite of ESA]
SoW	Statement of Work

SRAL	SAR Radar Altimeter (of Sentinel-3)
SRD	System Requirements Document
SSD	System Specification Document
SSS	Sea Surface Salinity
SVR	System Verification Report
SWIM	Surface Waves Investigation and Monitoring (instrument of CFOSAT)
SWH	Significant Wave Height (see also Hs)
TOPC	Terrestrial Observation Panel for Climate
TR	Technical Requirement
UCR/CECR	Uncertainty Characterisation Report (formerly known as the Comprehensive Error Characterisation Report)
UNFCCC	United Nations Framework Convention on Climate Change
URD	User Requirements Document
USB	Universal Serial Bus
USGS	United States Geological Survey
VOS	Volunteer Observing ships
WCRP	World Climate Research Programme
WGClimate	Joint CEOS/CGMS Working Group on Climate
WMO	World Meteorological Programme
WWA	World Wave Atlas (of FUGRO)
ADF	Auxiliary Data File
API	Application Program Interface
AUX	Auxiliary
CCI	The ESA Climate Change Initiative (CCI) is formally known as the Global Monitoring for Essential Climate Variables (GMECV) element of the European Earth Watch Programme
CCI+	Climate Change Initiative Extension (CCI+), is an extension of the CCI over the period 2017–2024
CCMM	Configuration Control Monitoring Module
CFOSAT	Climate Forecasting
CMIP	Coupled Model Intercomparison Project
DARD	Data Access Requirements Document
DOI	Digital Object Identifier
EC	European Commission
ECSS	European Cooperation for Space Standardization

ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
FOSS	Free Off-the-Shelf Software
FRM	Fiducial Reference Measurements
GCOS	Global Climate Observing System
GNSS	Global Navigation Satellite System
GOOS	Global Ocean Observing System
GDPR	General Data Protection Regulations
GUI	Graphical User Interface
ISDB	in situ database (of Fiducial Reference Measurements and satellite measurements)
L1 / L2 / L3 / L4	Level 1, 2, 3, 4 Products
L2OS	Level 2 Ocean Salinity
LUT	Look Up Table
NASA	National Aeronautics and Space Administration
Obs4MIPs	Observations for Model Intercomparison Projects
OPeNDAP	Open-source Project for a Network Data Access Protocol
OS	Ocean Salinity / Operating System
OTT	Ocean Target Transfer
Pi-MEP	SMOS Pilot Mission Exploitation Platform
PSD	Product Specification Document
QC	Quality Control
RAM	Random Access Memory
RD	Reference Document
RHEL	Red Hat Enterprise Linux
SAR	Synthetic aperture Radar
SMAP	Soil Moisture Active Passive [mission of NASA)
SMOS	Soil Moisture and Ocean Salinity [satellite of ESA]
SoW	Statement of Work
SRD	System Requirements Document
SSS	Sea Surface Salinity
TR	Technical Requirement

URD	User Requirements Document
UUID	Universal Unique Identifier
VM	Virtual Machine
WCS	World Co-ordinate System
WMS	Web Map Service

1.7 Structure of the document

The is SRD structured as follows:

Section 1, this section, provides an executive summary, and introduction outlining the purpose and scope of this document, reference documents, abbreviations etc.

Section 2 provides a context to the CCI+ Salinity ECV production system.

Section 3 provides an overview of the requirements elicitation and analysis procedure.

Section 4 provides the listing of identified requirements

Annex A provides example product netCDF file headers

2 System Overview

2.1 Purpose and Scope

In the context of the CCI programme, a processing system is a science-driven system that produces the required data products (as seen in URD) to satisfy the GCOS and evolved data requirements [SOW]. This system is built on the heritage of the prototype systems, system specifications and requirements generated in precursor projects, is capable of processing long time series, and rapidly reprocessing them, is under configuration control and maintenance (bug tracking, reprocessing, traceability) and is technically capable of being sustainable in the long term beyond funding from the CCI programme.

CCI does not build operational processing systems. CCI develops software systems for pre-operational ECV production in a research context. Within CCI the system development work comprises products specification, algorithm development and improvement, uncertainty characterisation, prototype product generation, system definition, sizing and demonstration. The aim is that the resulting specifications will subsequently be handed over to be further developed into operational systems in a non-ESA context.

The SOW restricts the scope to theme (i) of the CCI extension (CCI+) i.e. to maximise scientific return from Europe's investments in space infrastructure, CCI+ will perform the development and qualification of ECVs that were not included in CCI so far. As well as the development and qualification of new ECV retrieval algorithms, an associated task is to build the data processing systems (software) required to produce the ECV products pre-operationally and produce prototype products.

The research context of the CCI+ Salinity project drives the rationale of the system design as:

- Being a help for the CCI+ Salinity Science Team to perform regular and performing computation in view to support them during the different steps of the project. In that respect, in addition to the production system, specific computing capacity through virtual machines (VM) on which the processors will be installed and configured will be made available to the Science researchers.
- Fully addressing the CCI+ expected production volume by running a complete end-to-end ECV processing system. The production will take place at each of the three Years project duration.
- The specification of the production system is a living documentation. Based on the lesson learnt on the production each Year, the system will be improved.

The system described in the present document addresses the generation of satellite-based Sea Surface Salinity (SSS) time series within the frame of the CCI project. The main objective of such production is to provide the best support to the users and more globally to the science community in better understanding the Climate change and in particular its effect on the

salinity of the Oceans and further the impact of such a change on the other geophysics' indicators.

The processing system deployed in the frame of the project is driven by a dedicated system that includes all requested functionalities, among others: a processing orchestrator compatible with the Satellite data processor and their interfaces, handling of multi-cores processing (cluster of processors), management of multiple processing configurations (processor version and associated ADF), on-the-fly resources (re)-allocation, error management, processing historical records management.

The proposed processing system implements a configuration control system allowing simultaneous executions of several configurations of the processor and ADFs and strict traceability of the products generated with various configurations.

2.2 Sea Surface Salinity Data Products

The CCI+ Sea Surface Salinity (SSS) project (in this document also referred to as CCI+ Salinity) aims to develop a method, a system and data products for a long-term global SSS record.

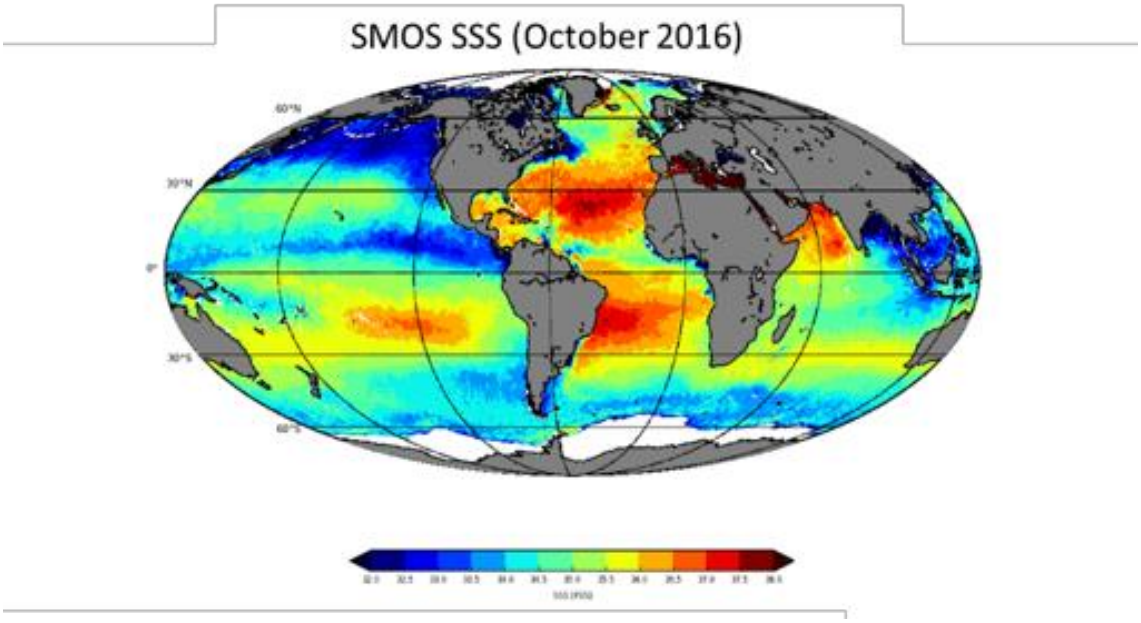


Figure 1 Example of SMOS Level 3 Sea Surface Salinity Global Map

The CCI+ Salinity project aims at deriving global SSS data products from on-going SMOS (ESA/Earth Explorer Soil Moisture and Ocean Salinity) mission, combined with Third Party Missions : the previous NASA/CONAE AQUARIUS-SAC-D and ongoing NASA SMAP (Soil Moisture Active Passive) missions operating with L-band (~ 1.4 GHz) passive microwave

radiometers, combined with US and Japanese AMSR-E, WinSAT and AMSR2 C-band missions for climate salinity related purposes.

2.3 Sea Surface Salinity System Context

Note that there is a distinction between the "CCI+SSS System" and the "CCI+SSS Processor". The CCI+SSS system is the more general term covering the end-to-end capabilities developed within the CCI+SSS project. The CCI+SSS processor specifically refers to the chain by which products specified in the Product Specification Document (PSD) are created.

The end-to-end system ultimately required for CCI+SSS to deliver its users' needs includes a range of interfaces, as illustrated in Figure 2.

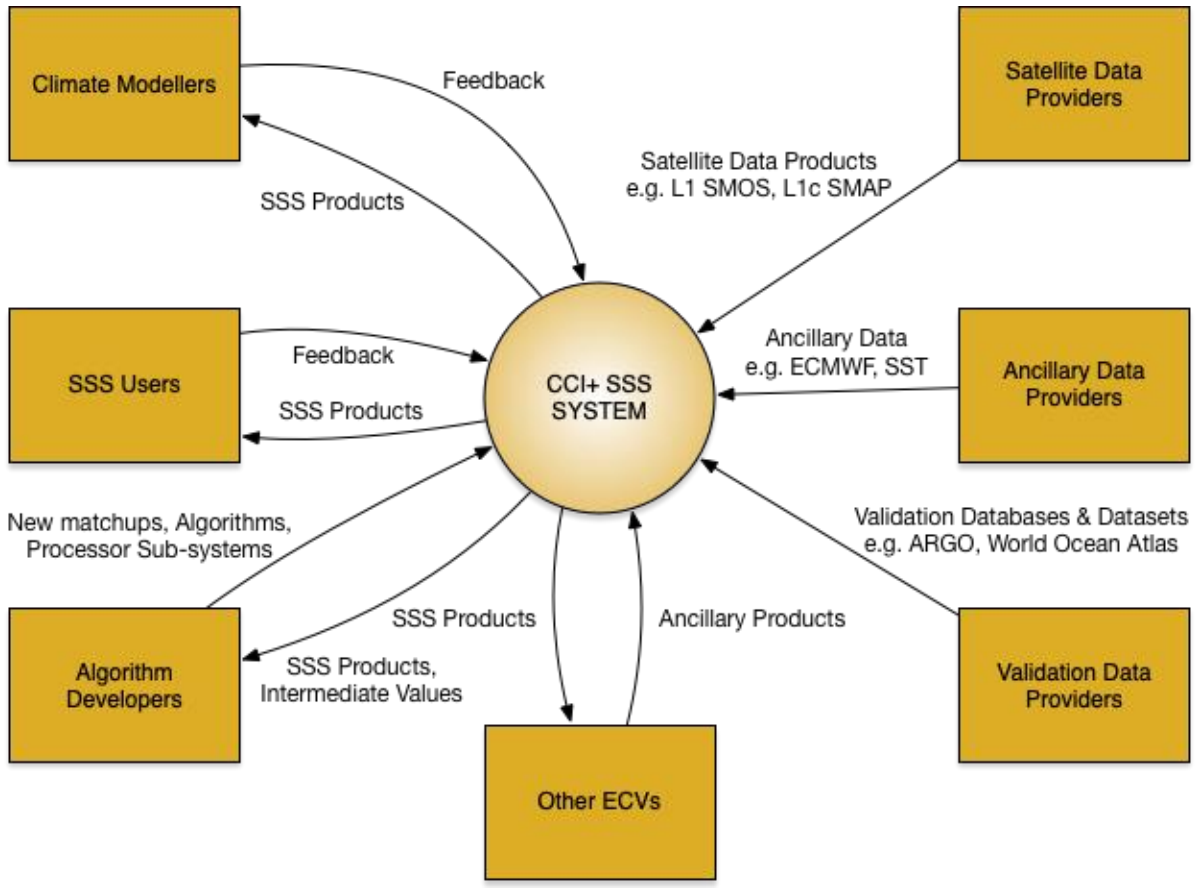


Figure 2 CCI+SSS system in context of its interfaces

According to the existing requirements collected from the SOW and also from the knowledge of the existing data processors (e.g. SMOS L2 OS Data Processor) the following data processing structure is expected.

The SOW establishes the need of introducing a processing chain covering from L1 products up to L3 products, extending it to L4 products whenever they are identified and introduced within the list of deliverable data sets.

The CCI+ Salinity project has chosen to discard the production of L1 data sets parting from L0 datasets, due to the considerable effort spent on this activity in the past, due to the considerable technical effort that should be addressed to achieve that, and particularly, the substantial amount of time that processing such data can take. Instead, L1 products are proposed to be acquired from the agencies/entities that produce them. This covers both L-band and C-band data sources proposed by the Algorithm retrieval team for its processing into L2.

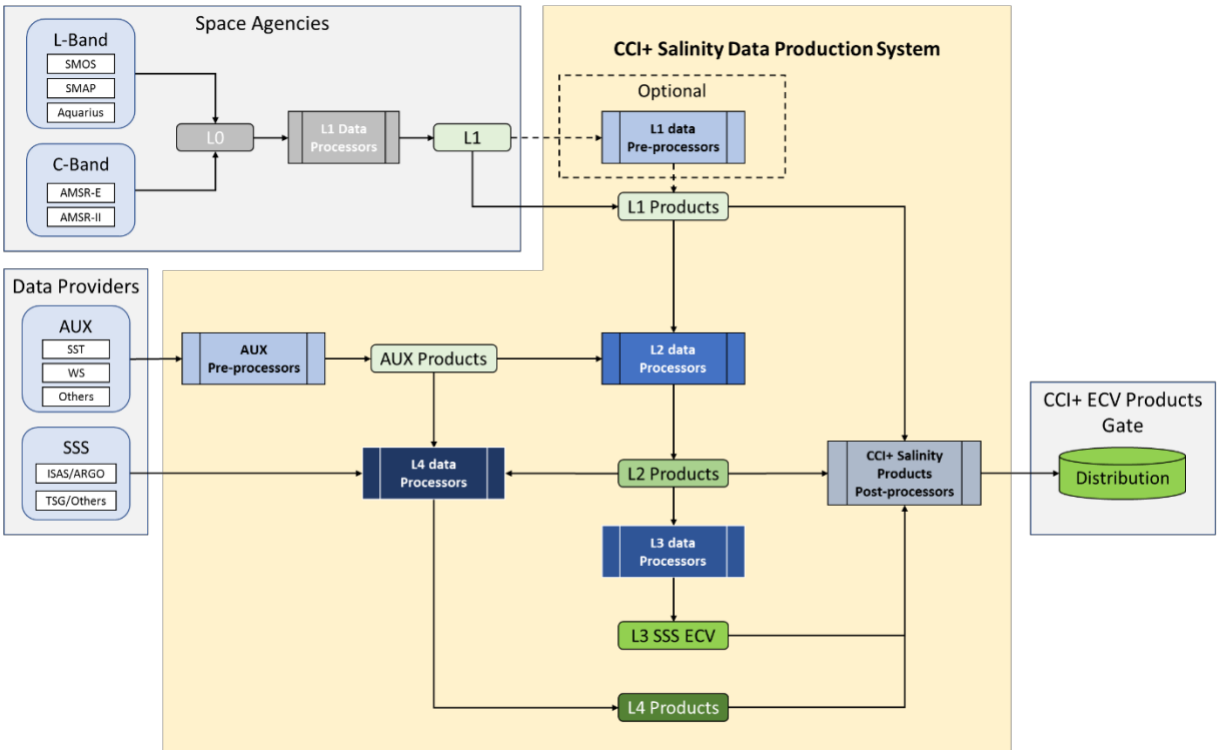


Figure 3 Basic data production structure requirements

It is expected that the various sources of data will have to suffer of some pre-processing prior being ingested by the L2 data processors. The software in charge of such task is what falls within the category of L1 data pre-processors. Their aim -if finally deemed necessary- would be to homogenize the datasets in a way they can be used by the various L2 processors. Note that the Retrieval Algorithm team will propose the testing and evaluation of the current SSS

retrieval methodologies and the investigation of the best retrieval options. This means that data from the sources shall be handled by algorithms that are not currently adjusted to them. Simplest approach to solve this need consists on developing those intermediate pre-processors that prepare L1 data as per specifications of the corresponding algorithms. In this way, modifications over these algorithms, which could be complex and time-consuming, are minimized.

The generation of L2 products requires further information beyond the L1 data sets. Current retrieval algorithms employed for SMAP/Aquarius and SMOS rely in similar forward models, which involve the modelling of the L-band emissivity of the surface of the ocean under the measurement geophysical conditions. The process involves the application of a dielectric constant model and the computation of the Fresnel equation to determine the contribution at each polarization. The process involves the estimation of the roughness contribution, plus the data correction to discard extra-terrestrial signals, like the atmospheric emission and attenuation, the galactic noise and the sun glint. Both the retrieval itself and the corrections rely in a set of auxiliary files that are employed by the nominal L2 data processors.

Within CCI+ Salinity, there is a requirement to obtain information about these auxiliary variables via different data sources from the current approaches and explore their advantages and limitations. While the official L2 processing chains have already in place components to provide such data into given spatial grid and times (e.g. ISEA grid at ESA), they are unlikely to be capable of perform the same task using other data sources from the pre-configured ones without having to implement modifications into them. Therefore, there is a requirement of the systems to include a component for the production of the auxiliary files under the specifications of each retrieval algorithm, as well as to homogenize and integrate the various sources that can be, in a given moment, being used to generate them.

In the next step, the required L2 products are generated by the L2 data processors, which will produce SSS values at orbit level whenever is required. The L2 products will be the inputs of the other two levels requested by the SOW. On one hand, L3 data processors will construct the synoptic maps at space and time scales as specified by the Retrieval Algorithm team, what usually conveys the performance of post-retrieval data corrections (systematic biases removal, spurious data filtering, etc.) and the re-projection and/or interpolation of the L2 orbits, plus the time integration of the information at the pre-defined scales. On the other hand, in the case of L4 products being generated, L2 data files will be also used jointly with other data sources. These external datasets can consist on other existing SSS products, to be merged with L2 SSS ECV products from several instruments and to generate in this way L4 SS ECV data sets, or they can imply the derivation of other variables (e.g. density) requiring SSS data to be extracted, jointly with other auxiliary information (SST).

The four specified levels of data (L1 to L4) are required to be delivered in specific formats and standards, namely netCDF CF conventions, Data Standards CCI requirements [RD-29] and

obs4MIPs formats. Aiming to optimize the resources and reduce the number of possible points of failure, the propositions is to set a post-processing component able to convert the various products into those standards for their delivery and distribution via the CCI+ Data Gate.

Note that, while the Systems Engineering team discards to perform L1 data processing, datasets can be refactored by the proposed post-processors to reshape them into the standards the [SOW] stands by. Nonetheless, the activity requires a previous assessment before acquiring final compromise for delivering such products.

2.4 Top-Level System Components

The aim is to develop an automated high-performance processing chain capable of delivering Level 2/3/4 sea surface salinity for each contributing satellite mission and merged products.

For this SRD the following assumptions have been made, that the CCI Salinity System will consist of five major components:

- The CCI Salinity Data Ingestion Subsystem (CCI-SSS-IS) that will retrieve input products, ingesting them into the catalogue, a database, and deriving metadata.
- The CCI Salinity Processing Subsystem (CCI-SSS-PS) that will generate the Level 2 products from the Level 1 products (and other input data sets), and Level 3 and Level 4 products.
- The CCI Salinity Archiving Subsystem (CCI-SSS-AR) that will archive all generated data products.

A top-level dataflow through the system components is represented in

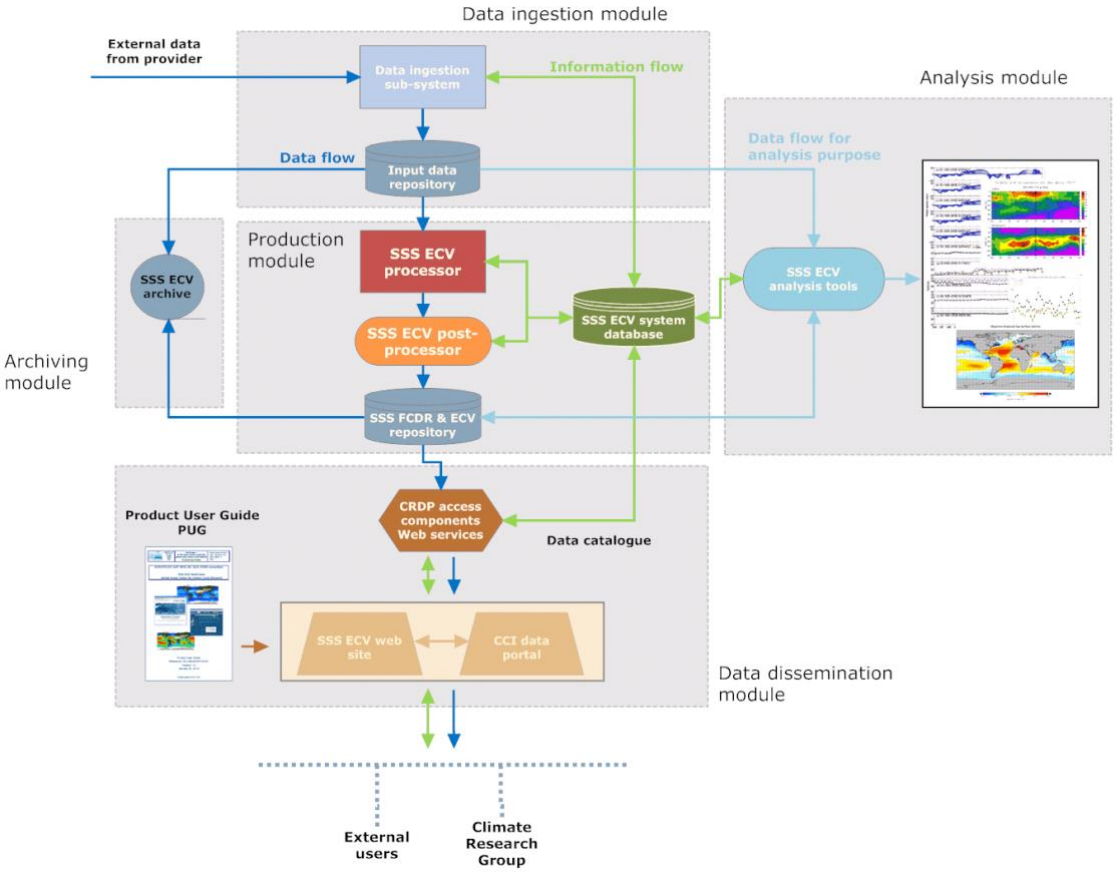


Figure 4 CCI+ Salinity production dataflow. Notice that Analysis module is not longer required due to the existence of CCI+SSS toolbox

2.5 System Software Modules

The assessment of the requirements in terms of software includes the following elements, which will be revisited during the project and adapted as necessary. Note that these are top level modules and that the processing chain itself inherits from the existing software that are (or will be) re-used and adapted to fit the specific needs of the project. This is in particular applicable to the SMOS, SMAP and Aquarius processing chains which will be modified in turn in order to align part of the chains to fit with the best selected algorithms as per outcome of the Science Team work.

Apart from the SMOS L2 processing chain which is well known by the Science and Engineering Teams through their work on L2 OS maintenance, SMAP and AQUARIUS will require some more involvement to be integrated in the system and modified in order to test and further apply some of the algorithms tuned by the Science Team.

2.5.1 Data Acquisition Module

The implemented system shall count with capability to retrieve the external files from their sources at the corresponding state of maturity. More specifically, L1 data files and auxiliary data sources. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.2 L1 Pre-processing Module

According with the requirements that will be provided by the Retrieval Algorithm team, the system shall include a pre-processing module to allow for the employment of L1 data files into existing L2 data processors for all the sources under consideration. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.3 Auxiliary Data Pre-processing Module

The platform shall be able of ingesting information about auxiliary parameters required for the L2 retrieval from data sources not necessarily matching the nominal ones used by the official algorithms. To enable this capability, a set of pre-processors shall be put in place able to produce AUX data files under the format and specifications associated to these L2 algorithms.

This module shall also cover the potential need of pre-processing any data set necessary for L4 applications, whenever they are related to L4 SSS ECV datasets or derived variables.

The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.4 L2 Data Processing Module

The components under this module will be in charge of the production of SSS ECV at L2, by ingesting L1 data previously pre-processed. The module shall be coordinated with the Auxiliary Data Pre-processing Module to ensure the availability of any auxiliary data file required by the L2 components for the performance of their activities. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.5 L3 Data Processing Module

The various L3 data sets shall be generated within this module, which shall cover the production of this higher-level dataset under a configurable environment. Because most L3 data processing is separated at various stages, design shall be as modular as possible, as well as being able to delimit specific times and areas. The aim of this is being able to produce both global and regional products. Retrieval Algorithm team envisages the possibility of generating

some regional products of interest. Its generation shall be covered following this strategy. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.6 L4 Data Processing Module

This part of the system shall cover the generation of synoptic SSS ECV products, based in SSS from various data sources and/or the derivation of L4 variables, which shall be defined by the Retrieval Algorithm team. The module shall ensure the ingestion of the required auxiliary information and the incorporation of any data level produced by the system and necessary for the L4 data production. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.7 Data Post-processing Module

This component shall be in charge of converting the various data levels (L2 to L4, and potentially L1) into the formats and standards required by the [SOW]. To do so, it shall be able to read all the outputs of the project and reshape them as needed. This module shall have connectivity with the internal archiving system of the platform, plus supporting any data distribution needs identified in Task 4. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes.

2.5.8 Look-up Table Generator and Tools

In addition to all the above, the system shall be able to produce any Look-Up Table (LUT) that is not externally provided and that is required to be generated within the activities of the Systems Engineering component of the project. LUTs are a useful way to optimize computation times, what increases the performance of the data production. Due to that, they are commonly put in place in EO systems.

The system shall also support the integration of additional tools to support the activities of the teams working in Task 2, Task 4 and Task 5.

2.5.9 Configuration Control and Monitoring Module

Based on the learning of the team, configuration control and monitoring of the data production are key to resolve some potential points of failure. In the case of platforms supporting components performing activities with their own configuration files, version of the configurations and their matching with the Data Processing Plan specifications are important. To do so, the system shall contain a Configuration Control and Monitoring module able to check the configuration of the system in a given moment. It shall also address the coordination between the different pieces of the data production chain, ensuring that data generation at a given sub-level is only triggered when all the data and information necessary for it is available. In this sense, the CCMM shall be in charge of the orchestration of the data production. In addition, CCMM shall host monitoring tools to detect any potential bottle-neck

or problems during the data processing. To do so, CCMM shall be able to ingest the automatic logging reports generated by each of the modules embedded into the system and serve as a diagnostic tool.

2.6 Design & Implementation Constraints

2.6.1 Operating System & Programming Languages

Compatibility shall be assured between the various software modules integrated into the platform. Due to the better performance and control, high performance platforms are usually deployed under Linux solutions. The check of the current software indicates some restrictions that shall be taken into account. For instance, SMOS L2 OS data processor nominally operates over RHEL5, which nowadays is out of date, what can trigger some security concerns into the platform. Interactions of the Engineering Systems team with the SMOS L2OS algorithm development team have allowed to learn that a new version of L2OS (v670) is coming in the following months. Such version of the software has made compatible with RHEL7 and CentOS 7. Considering that these operative systems are more recent and maintained, it is recommended the installation of CentOS7 as Linux hosting system.

It is expected no special constraints with the other algorithms, but in order to prevent potential conflicts, the platform shall be able to host virtual machines (VMs) that can be configured to comply with the individual limitations (including OS) imposed for each integrated piece of software.

The specifications about additional libraries and auxiliary software shall be acquired from the software documentation associated with each processor.

In addition to the OS, the platform shall be able to run tools and auxiliary processors, which can appear in different languages. The platform shall include Python 2.7.5 or above to allow for these tools to be run. 54% of users stated a programming language preference for a software library to be written in MATLAB, 28% for Python and 11% Fortran (URD).

2.6.2 Processing Hardware & Performance

The data processing constraints are hard to resolve. From the analysis of the existing pieces of software to be integrated within the systems, the most demanding in terms of data computation is the retrieval of L2 SSS from SMOS data (i.e. the L2OS SMOS data processor). Hence, this software is used as a reference to escalate the size of the platform in terms of RAM and cores.

The platform shall be able to host multi-threading and multi-processing, plus additionally run multiple instances of the same processing chain. The strategy for the data processing shall optimize the data processing attending to the specific data level. For instance, L2 processing takes place at orbit level, L3 and L4 products will require post-processing, including spatial

and temporal merging plus likely re-gridding of the products and interpolations. Therefore, the platform shall have capability to assess when enough L2 data has been processed to generate a given L3 or L4 product.

Because of L2 data processing is the most demanding and complex, it is recommended to split it from L3 and L4 product generation, which could have more specific requirements (e.g. requiring having the entire L2 time series in advance for, for example, determining post-retrieval land/sea contamination corrections). In the same ways, while L2 is stricter in terms of specifications, it will make use of far larger amounts of RAM and processing power. L3 and L4 data products are, in comparison, much less demanding, and hence, easier to fit within the systems specifications.

In the case of L2, as said, SMOS L2 OS data processor is taken as a reference. Current nominal processing requirements for SMOS L2 are currently set on a maximum of 9GB RAM and 4 threads as optimal hardware configuration. Under such configuration, each SMOS L1c data file takes a maximum of 110 minutes to be processed. The escalation of the processing to multiple instances is, however, limited by the OTT computation done by the post-processor of L2OS data processor. OTT is only computing after completing a day worth of SMOS L1c data, and it requires the presence of the results from the previous 10 days from the one being under computation. Considering the maximum number of orbits per day of SMOS (29), it is possible to estimate the maximum scalability of the system, having into account the limit of the OTT computation. This yields a maximum result of 120 cores and 270GB RAM, to produce 1 day of SMOS L2 data in a maximum of 110 minutes.

This yields a performance of 13:1 with a single processing chain attached to the system. However, this situates the performance far from the target (11 years of SMOS L1c data in 4 months**). The platform shall be able to process SMOS L1c data at a rate of 35:1**. To obtain the required performance, the platform shall integrate at least 3 processing chains as the above indicated. The approach consists on splitting the time series into 3 even parts with an overlapping of at least 11 days between them to compensate the data of low quality obtained until OTT values are stable (what takes 10 days' worth of data). This escalation would meet the specifications for SMOS L2 OS data processing (40:1 approximately), situation the specifications of the platform in 360 cores and 810 GB RAM. A preventive approach, and to take into consideration additional processing requirements or increment of needs as per evolution of the processor, would set these goals at 400 cores and 1TB RAM, just for SMOS L2 data generation.

Aquarius and SMAP L1 data processing to L2 requires significantly less hardware, in comparison. The main reason is the shorter time series for each of them, and the significantly fewer number of measurements associated to each orbit. Processing of both datasets with their nominal processors is estimated in about 15% of the SMOS requirements, i.e. 60 cores and about 120 GB RAM.

L2 to L3 and L4 is also expected to require far less data processing than SMOS L1c to L2. To take that into account, and also to cover pre-processors and post-processors plus tools, additional 60 cores and 120GB RAM shall be considered.

Nevertheless, this estimation is based on the hypothesis of processing L1C to L2 SMOS data using the original SMOS grid. This scenario is the most demanding in terms of computational resources and changes in the computational grid (i.e. subsampling pixel) can dramatical reduce the computational cost.

The total specifications for the data processing are, therefore, set up to 520 cores and 1.24TB RAM and it will be reassessed once the computational GRID is defined.

2.6.3 Data Storage

The system shall need to handle considerably large datasets. SOW establishes the requirement of producing full time series of SSS ECV and derived data sets from L2 to L4 (and potentially L1). This production involves the storage of high-sized files, especially at L1 and L2, plus the addition of all the auxiliary information required for the production chain.

Table 1 Estimation of storage requirements based on satellite data products

SATELLITE DATA				
PRODUCT	COVERAGE	YEARS	GB / YR	TB DATA
SMOS L1C	2010 to present	10	5800	58
SMOS L2	2010 to present	10	85	0.85
SMAP L1C	2015 to present	5	63	0.315
SMAP L2C	2015 to present	5	85	0.425
SMAP L2C (70km)	2015 to present	5	210	1.05
Aquarius L1A	2011 to 2015	4	27000	108
Aquarius L2 CAP	2011 to 2015	4	6.6	0.0264
Aquarius L2 OR	2011 to 2015	4	34	0.136

AMSR-E L2A	2002 to 2011	9	912.5	8.2125
AMSR2 L1R	2012 to present	8	511	4.088
WindSat L1	2003 to present	17	1000	17
WindSat L2	2003 to present	17		0
AMSR-E L2B	2002 to 2011	9	1.6	0.0144
AMSR2 L2	2012 to present	8		0
			TOTAL	**198TB**

The table above defines the storage requirement for processed satellite data based on the data provided in DARD §3.1.

DARD §3.2 estimates 0.15TB storage of in-situ test data. DARD §3.3 estimates 1.39TB of inter-comparison data. DARD §3.4 estimates 1.12TB of ancillary data. However, that did not include ECWMF data requirements, not NCEP. This assessment does not add substantially to the data storage requirement although the proposal included a 15TB storage requirement for auxiliary / ancillary data files, which is higher than that estimated in the DARD.

These estimations don't take into consideration any ancillary data that could be needed for L4 data processing of derived variables or sub-variables. This is not, either, contemplating the possibilities of having to store various reprocessing versions of the products, which could be the case (to be determined by Retrieval Algorithm team). Just with the nominal data here indicated, the storage volume that is envisaged is in excess of 200TB. In order to host any other aspects of the project that could take place (e.g. other data for L4 ECVs or ECVs sub-variables) this amount shall be extended up to 250TB as nominal storage space to be considered within the system as minimum storage requirement, for each reprocessing data version to be stored.

Taking these aspects into consideration, a comfortable storage requirement can be set in up to **250 TB per cycle**, what will allow for additional data sources not contemplated in the above estimations, plus the dissemination requirements. Note that, each intermediate data product shall be preserved, in case corrections have to be applied to later stages of data processing, not having to reprocess again from L1.

2.6.3.1 Storage used in the first year

During the first year, 51 TB of SMOS, Aquarius and SMAP data have been downloaded and archived in the adwäisEO storage to be available for the processing system. Table 2 resumes the input data available in the CCI+ Salinity storage at the 1 June 2019.

Additional 176GB of data produced in this year are archived in the same storage and made available via a dedicated ftp server.

Table 2 Data available in the CCI+ Salinity Storage at 1 June 2019

PRODUCT	COVERAGE	Volume [TB]
SMOS L1C	From 2010	41
SMOS AUX	From 2010	9.1
SMAP L2C (40Km)	From 2015	0.9
Aquarius L3	2011 to 2015	0.09
	Total	**51**

2.6.4 Processing Platform / Infrastructure

The considerable processing and storage requirements require a High-Performance Computing (HPC) infrastructure, provided by adwäisEO and hosted in the European Reliance Centre East (EBRC) a Tier IV datacentre situated in Luxembourg. The current computational capability of AdwäisEO relays on more than 1000 Intel E5 physical cores supported by 28 PB nearline storage and more than 11 PB online (on disk). The infrastructure hosts a cloud for a flexible allocation of resources connected to a HPC system for an efficient usage of the computational power.

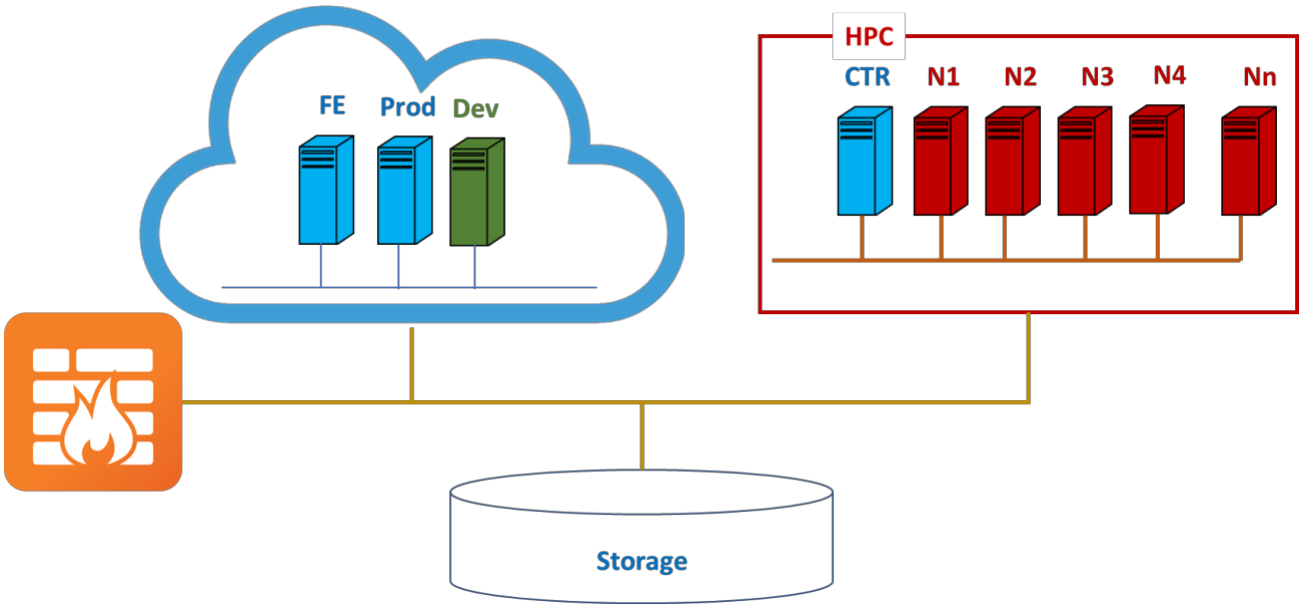


Figure 5 Example of CCI+ Salinity processing system implementation

The CCI+ Salinity processing system is installed on the adwäisEO cloud and interconnected with the its HPC system and storage (see Figure 5). The components mentioned in the above schema are:

- **Prod:** production server hosting a database and used to manage the processing;
- **Dev:** development server hosting the same services of Prod but used for testing and development purposes;
- **FE**: front-end server, in this moment consisting of a ftp server
- **CTRL:** cluster orchestrator of the HPC
- **N1…Ni:** computation nodes. The number of nodes is a function of the targeted system performance and can be varied according to processing needs.
- **Storage:** storage areas. All files identified in the database are stored on these volumes. Once identified into dedicated database tables and visible from the computation nodes, the volumes can be used to store information. The storage is connected with both cloud and HPC environments in order that the produced products can be easily made available to the FTP server for the dissemination.

The CCI+ Salinity processing system is based on a virtual application (vApp) dedicated to project consisting of 3 virtual machines (VM) CentOS based and running in the adwäisEO Cloud. The vApp is connected to the adwäisEO HPC cluster where at the first June 2019 16 Physical Nodes nodes are available for running the processing. Each node is equipped with:

- 2 XEON E5, 28 cores with frequency between 2.4 and 2.6 Ghz
- 256 DDR4 RAM
- 2 TB SSD

for a total of 448 Cores and 4 TB of RAM. In terms of storage, an elastic policy has been implemented in order to dynamically allocate up to 0.35 PB of disk space.

Figure 6 One of the adwäisEO server room hosting the CCI+ SSS processing system

2.7 Future Contingencies

This section describes any contingencies that might arise in the design of the system that may change the development direction. Possibilities include lack of interface agreements with outside agencies or unstable architectures at the time this document is produced. Address any possible workarounds or alternative plans.

3 Requirements Overview

3.1 Requirements Elicitation

The Requirements Elicitation process essentially consisted of fathering system requirements from source documents to build the stock of initial stakeholder requirements, from the Statement of Work with particular reference to the CCI+ Salinity Annex; user requirements from the project deliverable User Requirement Document (URD); product specification requirements from the Product Specification Document (PSD); data access requirements from the Data Access Requirement Document (DARD); and CCI data standard requirements from the CCI Data Standards document (DSTD).

Whereas the SOW and DSTD are static documents the URD, PSD and DARD are living documents, which will likely change throughout the course of the project and consequently this document, the System Requirement Document (SRD) is also to be considered as a living document that will require to be updated in the course of the development process to consider any additional requirements and/or changes to existing requirements.

3.2 Requirements Analysis

ECSS-E-ST-40C Space Engineering – Software Standard (RD02) lists the expected outputs of software requirements analysis as:

- Functional and performance specifications, including hardware characteristics, and environmental conditions under which the software item executes, including budgets requirements
- Operational, reliability, safety, maintainability, portability, configuration, delivery, adaptation and installation requirements, design constraints
- Software product quality requirements
- Security specifications, including those related to factors which can compromise sensitive information
- Human factors engineering (ergonomics including HMI usability) specifications
- Data definition and database requirements
- Validation requirements
- Interfaces external to the software item
- Reuse requirements

ANNEX D of ECSS-E-ST-40C states that requirements must be uniquely identified, include traceability of each requirement derived from higher level documentation, should be characterised by priority level, and by means of verification.

3.2.1 Organization of System Requirements

ANNEX D of ECSS-E-ST-40C considers the following categorization schema of system requirements, which will be adopted, omitting those not applicable. They are:

- FUN Functional Requirements
- PRF Performance Requirements
- INF Interface Requirements
- OPL Operational Requirements
- RES Resource Requirements
- CON Design and Implementation Constraints
- VRF Verification Requirements
- DOC Documentation Requirements
- PRY Portability Requirements
- QTY Quality Requirements
- RLY Reliability Requirements
- MTY Maintainability Requirements
- ORG Organizational Requirements

A second level of classification is specific to the CCI+ Salinity domain and will complement that categories above by designating the subject area related to the requirement:

- GEN General
- ACQU Data Acquisition
- PRE Data Pre-processing
- PROC Data Processing
- POST Data Post-processing
- PROD Products
- SOFT Software
- DIST Data Distribution
- ACCS External Data Access and Analysis
- ARCH Data Archival and Purging
- MON Monitoring and Control
- RCVR Anomaly Recovery
- FRMT Data Formats
- DATA Data Provider Interfaces
- ORG Organizational Interfaces
- USER User Management
- FCLY Facility Resources
- HW Hardware Resources
- OPS Operations & Maintenance

- CM Configuration Management
- QC Quality Control

3.2.2 Description of System Requirements

System requirements will be recorded with a unique identifier in a table of the following format:

Table 3 Template for System Requirement Description

Requirement ID	**Requirement Title**	**Priority**
Requirement Statement		
Sources: *Source References*		**Verification**
Comments: *List of comments (optional)*		
Issues: *List of Issues (optional)*		

A requirement description includes the following main fields:

- **Requirement ID**: A unique identifier of the format SR-NNN-CAT-SUBJ where NNN is a number unique within the whole set of requirements recorded herein. CAT and SUBJ indicate the assignment to a Requirements Category and a Requirements Subject, respectively (see above). CAT and SUBJ are useful for locating the requirement as they are reflected as sections within this document.
- **Requirement Title**: A short noun form indicating the topic of the requirement.
- **Requirement Statement**: The actual requirement, given in the typical textual form "The System shall ...."

Additionally, for each record the following requirement attributes are given:

- **Source References**:
- **Priority**: A priority value that indicates the importance of the requirement based on the MoSCoW prioritization schema.
  - MUST-HAVE Essential for operational capability
  - SHOULD-HAVE Important, but may be fully met in later releases
  - COULD-HAVE Non-essential, nice to have if resources allow
  - WONT-HAVE Not planned, possibly considered in the future
- **Verification Method**: A verification method to be applied in the course of the verification process to confirm that the requirement is fulfilled by the system. One of:

- INSPECT — Verify by observation or examination
- ANALYZE — Verify by showing theoretical compliance
- DEMONSTRATE — Verify by qualitative means
- TEST — Verify by quantitative means

Optionally, for each record the following supplementary information may be given:

- **Comments**: Information useful for the understanding of the requirement, or related ongoing or planned activities.
- **Issues**: A list of issues determined for resolution in subsequent stages of the development process. The description indicates who is responsible for resolution and when this is planned.

4 Requirements Listing

Note that the requirements listing in this document applies to the system under development and do not include either the cardinal requirements or the generic requirements enumerated in the main section of the Statement of Work, but only the technical requirements derived from ANNEX B: Salinity ECV of the SOW and requirements identified in the URD, PSD and DARD of the CCI+ Salinity project

4.1 Functional Requirements

Functional requirements specify 'what' the system has to do. They define the purpose of the system.

4.1.1 General (FUN-GEN)

SR-0010-FUN-GEN	**SSS Observing System**	**MUST-HAVE**
CCI+ Salinity **shall** directly address GCOS Action O11 to maintain and grow a global ocean salinity observing system for the assessment of ocean salinity and freshwater content change and its contribution to the global hydrological cycle.		
Sources: *SOW ANNEX B TR-1*		**INSPECT**

SR-0020-FUN-GEN	**Space-based SSS Continuity**	**MUST-HAVE**
CCI+ Salinity **shall** directly address GCOS Action O32 to ensure continuity of space-based SSS measurements.		
Sources: *SOW ANNEX B TR-1*		**INSPECT**

SR-0030-FUN-GEN	Develop Retrieval Algorithms	MUST-HAVE
In Phase 1 the CCI+ Salinity project **shall** develop an initial data set in the modern satellite era to capitalise on the rich satellite in situ and other data holdings available during that period. The focus **shall** be on the development, testing and improvement of dedicated sea surface salinity retrieval algorithms with respect to climate science user requirements, as compiled at high level by GCOS.		
Sources: *SOW ANNEX B TR-2*		**INSPECT**
Comments: The emphasis of this requirement is to ensure that the CCI+ Salinity project focuses on quality and understanding. The 2009-2020 period is a "golden era" in which SMOS, SMAP, Aquarius, AMSR-E, AMSR-II, are available and suitable satellite data for surface roughness estimation (e.g. ENVISAT ASAR, altimeters, Sentinel-1A&B, CFOSAT). Furthermore, in situ data from Argo and other sources offer improved data for verification/validation work.		

SR-0040-FUN-GEN	Deliver SSS ECV Products	MUST-HAVE
The CCI+ Salinity project **shall** develop and deliver Sea Surface Salinity ECV products primarily derived from microwave satellite measurements.		
Sources: *SOW ANNEX B TR-6*		**INSPECT**
Comments: It is clearly necessary to ensure that there is homogeneity between the in situ Fiducial Reference Measurements (FRM) and satellite records. Furthermore, a variety of FRM data must be used for validation activities.		

SR-0050-FUN-GEN	Deliver Validated Prototype Products	MUST-HAVE

The CCI+ Salinity project **shall** deliver validated prototype products using agreed validation methods and metrics developed within a research environment to climate users for assessment and feedback.	
Sources: *SOW ANNEX B TR-7*	**INSPECT**

4.1.2 Data Acquisition (FUN-ACQU)

SR-0060-FUN-ACQU	**Input Data Acquisition**	**MUST-HAVE**
The CCI+ Salinity project **shall** acquire the L1 Data specified in DARD §3.1 as input to the ECV Data Products production process		
Sources: *PROP §2.2.3.2.1.2, DARD §3.1*		**INSPECT**

SR-0070-FUN-ACQU	**Ancillary Data Acquisition**	**MUST-HAVE**
The CCI+ Salinity project **shall** acquire the Ancillary data (In-situ & Inter-comparison) specified in DARD §3.2 & 3.3 as input to the ECV Data Products production process		
Sources: *PROP §2.2.3.2.1.2, DARD §3.2, DARD §3.3*		**INSPECT**

SR-0080-FUN-ACQU	**Auxiliary Data Acquisition**	**MUST-HAVE**
The CCI+ Salinity project **shall** acquire the Ancillary specified in DARD §3.4 as input to the ECV Data Products validation and inter-comparison processes.		
Sources: *PROP §2.2.3.2.1.2, DARD §3.4*		**INSPECT**

4.1.3 Data Pre-processing (FUN-PRE)

SR-0090-FUN-PRE	**L1 Data Pre-processing**	**MUST-HAVE**
According with the requirements that will be provided by the Retrieval Algorithm team, the system **shall** include a pre-processing module to allow for the employment of L1 data files into existing L2 data processors for all the sources under consideration.		
Sources: *PROP §2.2.3.2.1.2*		**DEMONSTRATE**

SR-0100-FUN-PRE	**Auxiliary Data Pre-processing**	**MUST-HAVE**
The platform **shall** be capable of ingesting information about auxiliary parameters required for the L2 retrieval from data sources not necessarily matching the nominal ones used by the official algorithms. To enable this capability, a set of pre-processors **shall** be put in place able to produce AUX data files under the format and specifications associated to these L2 algorithms. This module **shall** also cover the potential need of pre-processing any data set necessary for L4 applications, whenever they are related to L4 SSS ECV datasets or derived variables.		
Sources: *PROP §2.2.3.2.1.2*		**DEMONSTRATE**

SR-0110-FUN-PRE	**Input Data QC - Logging**	**MUST-HAVE**
The pre-processing modules **shall** generate logging reports of the activities, to be used for monitoring and problem analysis purposes. A data QC report **shall** be generated.		
Sources: *PROP §2.2.3.2.1.2*		**INSPECT**

SR-0112-FUN-PRE	Look-up Table Generator & Tools	MUST-HAVE
T he system **shall** be able to produce any Look-Up Table (LUT) that is not externally provided and that is required to be generated within the activities of the Systems Engineering component of the project. LUTs are a useful way to optimize computation times, what increases the performance of the data production. Due to that, they are commonly put in place in EO systems. The Systems Engineering team contemplates the possibility of those LUTs and the potential to support their development and compilation. It is worth to mention that LUTs shall be defined by the Retrieval Algorithm team, which shall address the requirement to Task 3 by means of an associated Table Generation Requirement Document, in which the procedure and requirements for the LUTs are indicated. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes. A data QC report **shall** be generated.		
Sources: *PROP §2.2.3.2.1.2*		**INSPECT/TEST**

4.1.4 Data Processing (FUN-PROC)

SR-0120-FUN-PROC	Level 2 Data Processing	MUST-HAVE
The System **shall** have Processor Module(s) implementing the Processing Algorithms for the production of SSS ECV at L2, by ingesting L1 data previously pre-processed. The module **shall** be coordinated with the Auxiliary Data Pre-processing Module to ensure the availability of any auxiliary data file required by the L2 components for the performance of their activities. The module **shall** generate logging reports of the activities, to be used for monitoring and problem analysis purposes. A data QC report **shall** be generated.		
Sources: *PROP §2.2.3.2.1.2*		**INSPECT/TEST**

SR-0130-FUN-PROC	Level 3 Data Processing	MUST-HAVE

The System **shall** have Processor Module(s) implementing the Processing Algorithms for the production of SSS ECV at L3, by ingesting L2 data previously processed. Because most L3 data processing is separated at various stages, design **shall** be as modular as possible, as well as being able to delimit specific times and areas. The aim of this is being able to produce both global and regional products. Retrieval Algorithm team envisages the possibility of generating some regional products of interest. Its generation **shall** be covered following this strategy. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes. A data QC report **shall** be generated.
Sources: *PROP §2.2.3.2.1.2* — **INSPECT/TEST**

SR-0140-FUN-PROC	**Level 4 Data Processing**	**MUST-HAVE**
The System **shall** have Processor Module(s) implementing the Processing Algorithms for the production of synoptic SSS ECV products, based in SSS from various data sources and/or the derivation of L4 variables, which **shall** be defined by the Retrieval Algorithm team. The module **shall** ensure the ingestion of the required auxiliary information and the incorporation of any data level produced by the system and necessary for the L4 data production. The module **shall** generate logging reports of the activities, to be used for monitoring and problem analysis purposes. A data QC report **shall** be generated.		
Sources: *PROP §2.2.3.2.1.2*		**INSPECT/TEST**

4.1.5 Data Post Processing (FUN-POST)

SR-0145-FUN-PROC	**ECV Data Post-processing**	**MUST-HAVE**
The System **shall** have Post-processor Module(s) that **shall** convert the various data levels (L2 to L4, and potentially L1) into the formats and standards required by the [SOW]. To do so, it shall be able to read all the outputs of the project and reshape them as needed. This module **shall** have connectivity with the internal archiving system of the platform, plus		

supporting any data distribution needs identified in Task 4. The module shall generate logging reports of the activities, to be used for monitoring and problem analysis purposes. A data QC report **shall** be generated.	
Sources: *PROP §2.2.3.2.1.2*	**INSPECT**

4.1.6 Product Distribution (FUN-DIST)

SR-0150-FUN-DIST	**FTP Data Access Protocol**	**MUST-HAVE**
The data **shall** be distributed using the FTP protocol.		
Sources: *URD §3.4*		**INSPECT**
Comment: 81% of users asked require FTP data access		

SR-0160-FUN-DIST	**Obs4MIPS Delivery**	**COULD-HAVE**
The Contractor **shall** submit the final ECV products produced in Year 3 to the obs4MIPs initiative, including the preparation of data sets in obs4MIPs format, and of the required associated obs4MIPs documentation (proposal, technical note, etc.)		
Sources: *SOW ANNEX B TR-17*		**INSPECT**
Comment: Only 2% of users surveyed requested Obs4MIPS delivery (*URB §3.4*)		

SR-0170-FUN-DIST	**CCI Open Data Portal Delivery**	**MUST-HAVE**
The CCI Open Data Portal project provides a mechanism to access multiple ECV data sets via standard climate science protocols. CCI+ Salinity **shall** submit sea surface salinity ECV products to the CCI Open Data Portal for further dissemination to the user community.		

Sources: *SOW ANNEX B TR-18*	**INSPECT**
Comment: 6% of users surveyed requested OPeNDAP delivery (*URB §3.4*) Note that OPeNDAP is one of several CCI Data Portal access methods; which include WMS and WCS, and using the climate toolobox Python API, GUI and command line.	

SR-0190-FUN-DIST	**Public Availability of ECV Products**	**MUST-HAVE**
The suite of CCI+ Salinity ECV products produced **shall** be made publicly available together with the validation results immediately following the completed validation.		
Sources: *SOW ANNEX B TR-42*		**INSPECT**

4.2 Product Requirements

4.2.1 Functional Considerations (FUN-PROD)

SR-0200-FUN-PROD	**Global Ocean Coverage**	**MUST-HAVE**
The CCI+ Salinity project **shall** cover the global ocean, including full coverage of both northern and southern hemispheres as far as possible.		
Sources: *SOW ANNEX B TR-8*		**DEMONSTRATE**

SR-0210-FUN-PROD	**Mission Lifetime Coverage**	**MUST-HAVE**
All CCI+ Salinity products **shall** cover the full mission lifetimes of the satellite missions selected in SOW [TR-15].		
Sources: *SOW ANNEX B TR-9*		**DEMONSTRATE**

Comments: TR-15 identifies the following missions: SMOS 2009- ; SMAP 2015- ; AMSRE 2002- ; AMSR-11 2012-

SR-0220-FUN-PROD	**Available Product Processing Levels**	**MUST-HAVE**
CCI+ Salinity products **shall** be available to users as Level-1 (where appropriate), Level-2 and Level-3 product versions, and potentially as higher-level derived products if required by the users.		
Sources: *SOW ANNEX B TR-10; URD §3.4*		**INSPECT**
Comments: the intent of L1 and L2 instrument products is to develop an archive of lower-level data from which to initiate reprocessing activities. 20% of users surveyed would prefer L2 original swath data. 37% of surveyed users would prefer L3 gridded data.		

SR-0230-FUN-PROD	**Available Higher-Level Products**	**MUST-HAVE**
CCI+ Salinity products **shall** include daily, monthly, seasonally and annually aggregated and potentially as higher-level derived products if required by the users.		
Sources: *SOW ANNEX B TR-11; URD §3.4*		**INSPECT**
Comments: 43% of users surveyed would prefer L4 analysed data; including several satellite missions.		

4.2.2 Operational Product Requirements (OPL-PROD)

SR-0240-OPL-PROD	**Product Variables**	**MUST-HAVE**

In Phase-I, CCI+ Salinity shall provide, products and higher-level merged product time-series that shall include the following variables:

Sea surface salinity
Appropriate [RD-3] derived-variables;
Appropriate [RD-3] supporting variables;
Other information relevant to the processing and use of SSS data from space.

NOTE: SOW ANNEX B §3.1 include the following GCOS requirements:

ECV Product	Sea Surface Salinity (SSS)
Sub-variables	Bulk surface salinity[1] Skin surface salinity Near surface salinity at stated depth
Derived variables	Evaporation & precipitation estimates River runoff & glacial/land ice melting rates Sea surface density Sea surface alkalinity Sea surface pO2
Supporting variables	Sea surface temperature Sea surface winds Precipitation [additionally other fields used in data processing e.g. RFI maps, galactic

[1] the Sea Surface Temperature community is moving away from using the term 'bulk' as it is misleading term (i.e. it refers to the SST at or over some arbitrary depth). The Sea Surface Salinity community should consider the same issue.

	contributions, surface ocean roughness etc.

Sources: *SOW ANNEX B TR-14; SOW ANNEX B §3.1*	**INSPECT**
Comments: This requirement is incomplete because URD/PSD has not identified any variables that are required	

SR-0250-POL-PROD	**Mission Datasets**	**MUST-HAVE**

Sea Surface Salinity ECV products **shall** be delivered consisting of consolidated time series constructed from a multiple instrument data set that includes:

Source	**Type**	**Dates**
SMOS	L-band interferometer	2009-
Aquarius	L-band radiometer	2011-2015
SMAP	L-band radiometer	2015-
AMSR-E	C-band radiometer	2002-
AMSR-2	C-band radiometer	2012-

Sources: *SOW ANNEX B TR-15*	**INSPECT**
Comments: As stated this requirement is a duplicate of **SR-0210-FUN-PROD** but other satellite mission data may be used where appropriate within the CCI+ Salinity project such as novel satellite roughness measurements (e.g. SAR imager, altimeters, GNSS, CFOSAT, sun-glitter, scatterometers etc). None have been identified in the URD/PSD	

SR-0260-OPL-PROD	Product Temporal Frequency	SHOULD-HAVE
As part of data merging methods, time-dependent and sampling biases in products from different instruments **shall** be investigated, and strategies **shall** be developed and implemented to correct for these effects. GCOS [RD01] state hourly to monthly frequency for sea surface salinity ECV GOOS [RD05] state requirements depending on phenomena under investigation: • Air-Sea Flux (including Ice/Sea Flux): daily • Coastal Shelf exchange processes: daily • Fronts & Eddies: weekly • Riverine: monthly		
Sources: *SOW ANNEX B TR-16; URB §2.1 Table 1; URD §2.2 Table 2; USB §2.4*		**INSPECT**

SR-0279-OPL-PROD	Product Spatial Resolution	COULD-HAVE
As part of data merging methods, time-dependent and sampling biases in products from different instruments **shall** be investigated, and strategies **shall** be developed and implemented to correct for these effects. GCOS [RD01] state horizontal resolution between 1-100km. GOOS [RD05] state requirements depending on phenomena under investigation: • Air-Sea Flux (including Ice/Sea Flux): 100km • Coastal Shelf exchange processes: 1km • Fronts & Eddies: 10km • Riverine: 50km		
Sources: *SOW ANNEX B TR-16; URB §2.1 Table 1; URD §2.2 Table 2; USB §2.4*		**DEMONSTRATE**

SR-0280-OPL-PROD	**Threshold & Goal Values**	**MUST-HAVE**
Recognising the broad range of requirements from different users and applications, quantitative user requirements **shall** be specified as *Threshold* and *Goal* values, where: • *Threshold* is the minimum requirement to be met to ensure that data are useful; • *Goal* is an ideal requirement above which further improvements are not necessary; GOOS [RD05] state requirements depending on phenomena under investigation: • Air-Sea Flux (including Ice/Sea Flux): 0.01 • Coastal Shelf exchange processes: 0.1 • Fronts & Eddies: 0.1 • Riverine: 0.1		
Sources: *SOW ANNEX B TR-24; URD §2.2 Table 2*		**TEST**

Table 4 Requirement settings for satellite sea surface salinity from ESA L-band study [RD06]

	PHENOMENA	SPATIAL SCALES	TEMPORAL SCALES	ACCURACY
Air-Sea Interaction	Climate change/long-term changes (> 10 years)	10°	1 year	0.01
	Barrier layer	1.5°	1 month	0.1
Air-Sea Interaction/Climate Variability	ENSO/IOD/SSS anomalies	10°	1 month	0.2
Ocean Circulation	mesoscale/eddy propagation	100 km	1 month	0.2
	density compensation	0.5°	1 month	0.3

	PHENOMENA	SPATIAL SCALES	TEMPORAL SCALES	ACCURACY
	Tropical Instability Waves	100 km	8 days	0.2
	Thermohaline Circulation	100 km	2 weeks	0.2
Freshwater Fluxes	River plumes	1°	1 month	1
	Rain	0.5°	< 1hour	1
	Ice melting	50 km	8 days	1
Carbon cycle and biochemistry	Air-sea CO2 fluxes	150 km	1 month	0.2
	Alkalinity	150 km	1 month	0.2

SR-0290-OPL-PROD	**User Resolution, Coverage & Accuracy**	**SHOULD-HAVE**
To satisfy user requirements the CCI+ Salinity project **shall** be global with a frequency of at least weekly and resolution at least 0.25˚ with an accuracy at least 0.3		
Sources: *URD §3.4.1*		**INSPECT**
Comments: User's surveyed suggest the requirement for spatial resolution shall be at least 0.25˚ (39%), that coverage should be global (83%), at least 9 years duration (94%), 3 to 7-day frequency (61%) and most (74%) of surveyed users would be satisfied with accuracy between 0.1 and 0.3.		

SR-0300-OPL-PROD	**Dataset Production Schedule**	**MUST-HAVE**
The Contractor **shall** process and deliver the following products: Project Year 1:		

• Round Robin algorithm inter-comparison data sets for all satellite data sets to be used within the CCI+ Salinity project. • Initial Version-1 of individual mission test data sets (L1 and above) • Initial Version-1 L2/L3/L4 merged test time series up to the present day • Initial Version-1 of the ISDB Project Year 2: • Version-2 of individual mission L1 and L2 data sets up to the present day • Version-2 L3/L4 merged time series up to the present day • Version-2 of the ISDB Project Year 3: • Version-3 of the sea surface salinity ECV individual mission L2 dataset up to the present day (Final Phase-I data) • Version-3 of the sea surface salinity ECV L3/L4 data set up to the present day (Final Phase-I data) • Version-3 of the sea surface salinity ECV ISDB up to the present day (Final Phase-I data)

Sources: *SOW ANNEX B TR-45; PSD*	**DEMONSTRATE**
Comment: URD §3.4 Figure 9 suggests that only 28% of users were satisfied with products being updated only annually; 22% expected them to be updated continually i.e. as soon as improvements to alleviate error estimates and provide better accuracy were developed.	
Issue: *§6 Table 5 states that irrespective of this requirement in Year 1 the data to be delivered will be:* *L4 weekly, 50km smoothing, 25km grid size, global coverage, 01/2010-10/2018* *L4 30 days, 50km smoothing, centred 1st and 15th of the month, 25km, global 01/2010-10/2018*	

4.2.3 Product Quality (QTY-PROD)

SR-0310-QTY-PROD	**Level 4 Error Specification**	**MUST-HAVE**

The CCI+ Salinity Level 4 product **shall** contain the random error, systematic error, standard deviation of the bias, as well as good/bad flags computed from different indicators (chi-squared, number of outliers).	
Sources: *PSD §1.2(2)*	**INSPECT**

SR-0320-QTY-PROD	**Correction of Intra-mission Biases**	**MUST-HAVE**
As part of data merging methods, time-dependent and sampling biases in products from different instruments **shall** be investigated, and strategies **shall** be developed and implemented to correct for these effects.		
Sources: *SOW ANNEX B TR-16*		**DEMONSTRATE**

SR-0330-QTY-PROD	**Quality Indicators / Flags**	**SHOULD-HAVE**
In addition to ECV product uncertainties the URD survey required data to be quality flagged as good/bad, for each quality check and for selected quality checks. In the case of L3 and L4 products, it is not possible to specify each quality control flag (data are already binned). Flagging data as good/bad, or if they exceed a certain threshold would be a possibility.		
Sources: *URD §3.5; PSD §5.1*		**DEMONSTRATE**
Comment: In the case of the L4 product released during the first phase of the CCI+SSS project it is not possible to specify each quality flag since the data sets are created from lower level data that result in gridded, higher quality products. For production, a detailed documentation is needed including all steps in the data and product delivery chain.		

SR-0340-QTY-PROD	**Assessment Long-term Stability**	**MUST-HAVE**

The long-term stability the CCI+ Salinity time series delivered **shall** be assessed. GCOS [RD01] require long-term stability 0.001 / decade	
Sources: *SOW ANNEX B TR-39; URD §2.1 Table 1*	**TEST**

4.2.4 Product Uncertainty (RLY-PROD)

SR-0350-RLY-PROD	**Product Uncertainty**	**SHOULD-HAVE**
The CCI+ Salinity project **shall** provide a validated estimate of uncertainty at product grid/pixel level for each data product. GCOS [RD01] require measurement uncertainty of 0.01		
Sources: *SOW ANNEX B TR-20; URD §2.1 Table 1*		**TEST**
Comment: it may not be possible to validate uncertainty using in situ data due to the lack of measurements available in the time series		

SR-0360-RLY-PROD	**ECV Product include Uncertainty**	**MUST-HAVE**
Uncertainties **shall** be reported within the ECV products for every geophysical measurement.		
Sources: *SOW ANNEX B TR-33*		**DEMONSTRATE**

SR-0370-RLY-PROD	**Define Uncertainty Methodology**	**MUST-HAVE**
The Contractor **shall** develop a practical method to provide uncertainty estimates for each geophysical data product produced at the pixel/grid level.		

The end-to-end uncertainty budget will estimate the uncertainties that arise in each step of the retrieval process, including from: • instrument noise characteristics, • atmospheric correction, • geolocation, • pixel classification, • geophysical product retrieval. Include all potential sources of uncertainty, for example those introduced from the use of: • external ancillary data (*e.g.* meteorological fields), • forward radiative transfer models, • incorrect retrieval model assumptions, *etc.* Analyse how these different sources of uncertainty combine into the total product uncertainty and make an estimate of this. Identify the major sources of uncertainty, and also investigate sources of uncertainty that are difficult to quantify (*e.g.* unknown optical properties). Uncertainties in validation data sources shall be characterised. Where independent measurement of each component of ECV product uncertainty is impossible, the uncertainty budget must be derived using a theoretical/simulation approach. Document this analysis in the **End-to-End ECV Uncertainty Budget (E3UB)**.

Sources: *SOW §3.3.4*	**ANALYZE**
Comment: During the first phase of the CCI+SSS project, data sets will include information about random noise and systematic errors, number of outliers and standard deviation of the bias. Error estimation for the L4 CCI+SSS product is described in E3UB	

SR-0380-RLY-PROD	**Report Uncertainty to PUG**	**SHOULD-HAVE**
The method used to derive and validate uncertainties, the characteristics of those uncertainty estimates and advice on how uncertainty estimates are to be used for each product **shall** be fully reported in the PUG.		
Sources: *SOW ANNEX B TR-21*		**INSPECT**

SR-0390-RLY-PROD	User-defined Uncertainty Needs	COULD-HAVE
User requirements for ECV product uncertainties **shall** be included in the user requirements analysis, including how the uncertainties should be expressed and used in the CCI+ Salinity ECV products (*e.g.* how should the uncertainties be broken down into their random and systematic components). Users requirements for the specification of spatial and temporal error-correlation characteristics of the products **shall** be analysed. Consideration **shall** be given to uncertainties that are difficult or impossible to quantify numerically, and how these should best be communicated to users. Often these are related to limitations of sampling, or to retrieval model assumptions.		
Sources: *SOW ANNEX B TR-25; URD §3.5*		**INSPECT**
Comment: The URD survey indicates that no single means of communicating uncertainty satisfied all users. This requirement has been satisfied.		

4.2.5 Product Validation (VRF-PROD)

SR-0400-VRF-PROD	Pre-defined Validation Metrics	MUST-HAVE
Validation **shall** be performed against metrics pre-defined defined by the contractor and endorsed by the user community. NOTE: The Product Validation Plan (PVP) produced by the science team identifies validation metrics and methods.		
Sources: *SOW ANNEX B TR-35; URD §3.5 Figure 17*		**INSPECT**
Comment: The USD survey indicates that ARGO Buoy measurements were the preferred in situ dataset for product validation.		

SR-0410-VRF-PROD	ECV Product Validation	MUST-HAVE
A full validation of all sea surface salinity ECV products produced **shall** be performed.		
Sources: *SOW ANNEX B TR-36*		**TEST**

SR-0420-VRF-PROD	Uncertainty Validation	MUST-HAVE
Validation **shall** quantify the uncertainty of the sea surface salinity ECV products as well as the quality of the product uncertainty estimates themselves. The long-term stability of all ECV time series delivered **shall** be assessed.		
Sources: *SOW ANNEX B TR-37*		**TEST**

SR-0430-VRF-PROD	In situ FRM Database	MUST-HAVE
A database of relevant and ideally independent in situ Fiducial Reference Measurements and satellite measurements (ISDB) **shall** be developed to serve the CCI_ Salinity project validation, research and development needs.		
Sources: *SOW ANNEX B TR-38*		**INSPECT**
Comment: Note that CCI+ Salinity will work together with the ESA Pi-MEP activity. Validation data may include unconventional verification/validation measurements and alternative satellite capabilities.		

SR-0440-VRF-PROD	ISDB include Uncertainties	MUST-HAVE
All measurements in the ISDB **shall** include uncertainty estimates.		

Sources: *SOW ANNEX B TR-40*	**INSPECT**

SR-0450-VRF-PROD	**ISDB Documentation**	**MUST-HAVE**
The structure, functionality and operation of the ISDB and its interfaces **shall** be documented in a dedicated Technical Report.		
Sources: *SOW ANNEX B TR-41*		**INSPECT**

SR-0460-VRF-PROD	**ISDB Uncertainties Reported to PUG**	**MUST-HAVE**
The methods used to derive and validate ISDB uncertainties and the characteristics of those uncertainty estimates for each product **shall** be fully reported in the PUG.		
Sources: *SOW ANNEX B TR-42*		**INSPECT**

4.2.6 Product Format (INF-FRMT)

SR-0470-INF-FRMT	**CCI Data Standards**	**MUST-HAVE**
CCI Data Standards [DSTD] **shall** be fully implemented. NOTE: This requirement is not specific enough i.e. full implementation of the DSTD will be satisfied by fulfilling the subsequent requirements. in this section.		
Sources: *SOW §3.5.1*		**INSPECT**

SR-0480-INF-FRMT	**Use netCDF-4 (Classic) format**	**MUST-HAVE**

CCI projects **shall** produce data according to the CCI Data Standards Requirements and use the netCDF-4 (classic) format, the mode set to NC_CLASSIC_MODEL, but compression may be used.	
Sources: *DSTD R-1; PSD §2.1*	**INSPECT**
Comment: 94% of surveyed users requested netCDF format (*URB §3.4*)	
Issue: *If netCDF-4 cannot be used, data may be produced in netCDF-3*	

SR-0490-INF-FRMT	**Use CF Convention**	**MUST-HAVE**
CCI projects **shall** produce data according to the CCI Data Standards Requirements and conform to the CF (Climate and Forecasting) convention [RD07] The following CF global variables **shall** be included in product files: • title - short description of the dataset • institution – where data produced (use CCI common vocabulary [RD09]) • source – original data source(s), including ADFs, with DOIs as comma-separated list • history – processing audit trail of dataset • Conventions – the CF version		
Sources: *DSTD R-1; DSTD § 2.5.1; ; PSD §2.1*		**DEMONSTRATE**

SR-0500-INF-FRMT	**Use ACCD Convention**	**MUST-HAVE**
Since CF and ACCD conventions are aligning CCI projects **should** produce data according to the Attribute Convention for Data Discovery [RD08]. The following discovery metadata shall be included in additional to the CF global variables: • summary – paragraph describing the dataset • keywords – comma separated list of words • id – separate from tracking_id, unique, no white space, URL, URN, DOI etc.		

- naming_authority – organisation providing the id, URI or reverse-DNS naming
- keywords_vocabulary – if using a controlled vocabulary
- cdm_data_type – selected from enumeration: Grid, Image, Point, Radial, Station, Swatch, Trajectory
- comment – miscellaneous information not captured elsewhere
- date_created – date this version created in ISO8601:2004 extended format
- creator_name – name of the person (or type if creator_type specified)
- creator_url – URL of person (or type if creator_type specified)
- creator_email – email of person (or type if creator_type specified)
- project – Climate Change Initiative Plus – European Space Agency
- geospatial_lat_min – decimal degrees north, range -90 to +90
- geospatial_lat_max – decimal degrees north, range -90 to +90
- geospatial_lon_min – decimal degrees east, range -180 - +180
- geospatial_lon_max – decimal degrees east, range -180 to +180
- time_coverage_start – format yyymmddThhmmssZ
- time coverage_end – format yyymmddThhmmssZ
- time_coverage_duration – ISO8601 duration string e.g. P1Y
- time_duration_resolution – ISO8601 duration string e.g. P1D
- standard_name_vocabulary – name of controlled vocabulary for standard name
- license – restrictions on data access and distribution

Sources: *DSTD R-1*	**DEMONSTRATE**

Comment: creator_type is one of person, group, institute or position

SR-0510-INF-FRMT	**CCI Specific Metadata**	**MUST-HAVE**

The following CCI specific metadata **shall** be included in product files:

- platform (use CCI common vocabulary [RD09])
- sensor (use CCI common vocabulary [RD09])
- spatial_resolution
- key_variables
- tracking_id – UUID (Universal Unique Identifier) version 4
- product_version (see DSTD §2.8)

SR-0510-INF-FRMT	CCI Specific Metadata	MUST-HAVE
for gridded (level 3/4) data on a regular lat/lon grid the following **shall** be included: • geospatial_lat_units • geospation_lon_units • geospatial_lon_resolution • geospatial_lat_resolution		
Sources: *DSTD §2.5.1*		**DEMONSTRATE**
Comment: Software to generate UUID available from: http://www.ossp.org/pkg/lib/uuid/ and CCI **shall** use version 4 (random number based) for consistency with CMIP5		

SR-0520-INF-FRMT	Main Variables use CF Standard Names	MUST-HAVE
CCI projects **shall** produce data according to the CCI Data Standards Requirements and CF Standard Names **shall** be used for the main variables [RD08]		
Sources: *DSTD R-1*		**DEMONSTRATE**
Comment: The CF Standard Name sea_surface_salinity is the salt content of sea water close to the sea surface. The unqualified term 'salinity' is generic and does not necessarily imply any particular method of calculation. The units of salinity are dimensionless, and the units attribute should normally be given as 1e-3 or 0.001 i.e. parts per thousand.		

SR-0530-INF-FRMT	Key Primary Variables	MUST-HAVE
The key primary variables in the file and their related ancillary variables (e.g. uncertainty) shall be identified, and the range of their expected values shall be indicated. • Use `key_variable` netCDF global variable to identify primary variable • Ancillary variables of key variable, such as uncertainty or quality flags, should be identified by the `ancillary_variable` attribute of the primary variable		

- Coordinate variables are identified with the same name as the corresponding dimension e.g. `time, lat, lon`
- For each variable `valid_range` should be defined e.g. values outside this range considered as missing data. Alternatively use `valid_min` and `valid_max`.
- For each variable `actual_range` should be defined, i.e. of data in file
- Where quality flags included, zero or no bit set should be taken as good data

In addition to the fields recommended in 2.5.2 of DSTD PSD §2.4.2 proposes additional fields:

- `_FillValue` – a value used to represent missing or undefined data.
- `units` – a string recognised by UNIDATA UDUNITS package [RD11]
- `scale_factor` – if present for a variable, the data are to be multiplied by this factor after the data are read by an application.
- `add_offset` – if present for a variable, this number is to be added to the data after it is read by an application.
- `long_name` – descriptive name for the variable
- `comment` – miscellaneous information about the data or methods used to produce it.
- `source` – The method of production of the original data. If it was model-generated, source should name the model and its version; If it is observational, source should characterize it
- `grid_mapping` – identifies a variable that defines a grid mapping

Sources: *DSTD R-6; PSD §2.4.2*	**DEMONSTRATE**

Comment: If both `scale_factor` and `add_offset` attributes are present; the data are first scaled before the offset is added.

Annex 1 provides an example of correctly formatted netCDF variable definition.

Issue: *PSD §2.4.2 includes a field `auxiliary_data_N` as a string describing additional data used to produce this variable. This is not considered useful as the CF field `comment` is perfectly suitable for this.*

SR-0540-INF-FRMT	**Gridded Data**	**MUST-HAVE**

For consistency across CCI gridded products, variables shall have, as a minimum, the following dimensions: time, latitude, longitude (or alternative horizontal grid)
Sources: *DSTD R-7* — **DEMONSTRATE**
Comment: The first CCI+ Salinity product files will be delivered on the global (Cylindrical) Equal-Area Scalable Earth Grid 2.0 (EASE 2) Annex 1 provides an example of correctly formatted netCDF grid definition.

SR-0550-INF-FRMT	**Additional Data Formats**	**COULD-HAVE**
Projects who have commitments to produce data in other formats (e.g. GeoTiff), shall do this **in addition** to the standardized products, and shall ensure these products **comply as much as possible** to the CCI Data Standards (e.g. filenames, metadata)		
Sources: *DSTD R-2*		**INSPECT**

SR-0560-INF-FRMT	**INSPIRE Metadata**	**MUST-HAVE**
The CCI projects shall ensure that INSPIRE compliant metadata records are created for each dataset. For datasets held in the CCI Open Data Portal, this requires the CCI projects to provide information to the Data Portal Team.		
Sources: *DSTD R-4*		**DEMONSTRATE**

SR-0570-INF-FRMT	**CCI Ontology Terms**	**MUST-HAVE**
CCI Data Producers shall use terms from the CCI vocabulary tables [RD09] in the netCDF global attributes, or if terms are missing, they shall request that they are added to the tables.		

SR-0570-INF-FRMT	**CCI Ontology Terms**	**MUST-HAVE**
Sources: *DSTD R-5*		**DEMONSTRATE**
Issue: *Several terms required for CCI+ Salinity are missing from the CCI Ontology version 0.0.1 (2018-11-27) and consequently additions will need to be requested by the CCI+ Salinity team e.g. the platforms SMAP and Aquarius are missing.*		

SR-0580-INF-FRMT	**DOI Product Identifiers**	**SHOULD-HAVE**
Digital Object Identifiers (DOI) **shall** be assigned to all ECV data sets made publicly available.		
Sources: *SOW ANNEX B TR-12*		**INSPECT**

SR-0590-INF-FRMT	**Directory Structure**	**SHOULD-HAVE**
CCI Data Producers **shall** use the common directory structure for all output data made available to users. The common CCI directory structure is arranged as follows: `/<archive root>/<type>/<version>/<time>`		
Sources: *DSTD R-8*		**INSPECT**

SR-0600-INF-FRMT	**File naming Convention**	**MUST-HAVE**
CCI Data Producers **shall** use the CCI file naming convention for all output data made available to users. The filename will follow the format: `ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<AdditionalSegregator>]-<IndicativeDate>[<IndicativeTime>]-fv<File version>.nc`		

Sources: *DSTD R-9; PSD §2.2*	**INSPECT**

Comment: Several field values need to be added to the CCI Ontology (RD09) e.g. <CCI Project> and <Data Type> for CCI+ Salinity and Sea Surface Salinity are not currently in the Ontology.

<Processing Level> is defined in RD09

L0	Unprocessed and payload data at full resolution. No CCI recommendations regarding formats or content for data at this processing level
L1A	Reconstructed unprocessed instrument data at full resolution, time referenced and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters, computed and appended, but not applied, to L0 data
L1B	Level 1A data processed to sensor units
L1C	Further processed Level 1B data (e.g. correcting radiances, mapping onto a spatial grid)
L2	Retrieved environmental variables at the same resolution and location as the level 1 source
L2P	Geophysical variables derived from Level 1 source data at the same resolution and location as Level 1 data, typically in a satellite projection with geographic information. These data are the fundamental basis for higher level CCI products.
L3	L2 variables mapped on a defined grid with reduced ancillary data requirements:
L3U	Uncollated L3U: L2 data granules remapped to a space grid without combining any observations from overlapping orbits.
L3C	Collated L3C: Observations combined from a single instrument into a space-time grid
L3S	Super-collated L3S: Observations combined from multiple instruments into a space-time grid
L4	Data set created from an analysis of lower level data that result in gridded, gab-free products

<Product String> is defined by the CCI+ Salinity team but in the first instance, for L4 products, will be MERGED i.e. data from more than one sensor.

<Indicative Date> in format YYY[MM[DD]] should best represent the observation date of the dataset.

<Indicative Time> in format [HH[MM[SS]]] is optional <Additional Segregator> is optional and may be used to ensure different datasets do not generate the same filename, or to provide additional information. fv<File Version> is in the form n{1,}[.n{1,}
Issue: *Several terms required for CCI+ Salinity are missing from the CCI Ontology version 0.0.1 (2018-11-27) and consequently additions will need to be requested by the CCI+ Salinity team e.g. <CCI Project> and <Data Type> for CCI+ Salinity.*

4.3 Algorithm Development (FUN-PROC)

SR-0610-FUN-PROC	**Development of Improved Algorithms**	**MUST-HAVE**
The Contractor **shall** devote significant effort to developing improved algorithms specifically for use in CCI+ Salinity NOTE: Note a S.M.A.R.T requirement and implied by following requirements.		
Sources: *SOW ANNEX B TR-26*		**IGNORE**

SR-0620-FUN-PROC	**Improve Retrieval Algorithms**	**MUST-HAVE**
The Contractor **shall** conduct research and development and explore new algorithms that could address known weaknesses in SSS retrievals from satellite data sets.		
Sources: *SOW ANNEX B TR-27*		**DEMONSTRATE**

SR-0630-FUN-PROC	**Calibration & Aging Biases**	**MUST-HAVE**

The Contractor **shall** investigate and account for satellite instrument biases taking account of changes in calibration with instrument aging.	
Sources: *SOW ANNEX B TR-28*	**ANALYZE**

SR-0640-FUN-PROC	**Algorithm Performance Metrics**	**MUST-HAVE**
The contractor **shall** define a set of metrics to be used when assessing the performance of algorithms.		
Sources: *SOW ANNEX B TR-29*		**ANALYZE**

SR-0650-FUN-PROC	**Perform Round-Robin Inter-comparison**	**MUST-HAVE**
The contractor **shall** conduct a “round robin” inter-comparison of SSS retrieval algorithms and make a detailed assessment of performance using metrics defined in [TR-29]. The pre-processing, retrieval algorithm approach, as well as the ancillary data used as part of the retrieval, **shall** all be as consistent as possible.		
Sources: *SOW ANNEX B TR-30*		**ANALYZE**

SR-0660-FUN-PROC	**Select Definitive Retrieval Algorithms**	**MUST-HAVE**
Based on the outcome of [TR-30], The contractor **shall** select a set of definitive retrieval algorithms to be applied to data from different instruments. The selection **shall** be based on defined metrics but must also strive to deliver a multi-mission dataset that is as consistent as possible in order to avoid inter-instrument biases within the ECV.		

Sources: *SOW ANNEX B TR-31*	**DEMONSTRATE**

SR-0670-FUN-PROC	**Investigate Alternate Measurements**	**COULD-HAVE**
The Contractor **shall** investigate the potential of other measurements (e.g. GNSS, CFOSAT, Sentinels, sun glitter etc.) to better meet GCOS requirements for the sea surface salinity ECV.		
Sources: *SOW ANNEX B TR-34*		**DEMONSTRATE**

SR-0680-FUN-PROC	**Use of C-band Radiometers**	**COULD-HAVE**
There is some evidence that C-band radiometers can provide estimates of SSS [RD04] and this shall be explored to extend the time series prior to L-band measurements		
Sources: *SOW ANNEX B TR-19*		**DEMONSTRATE**

4.4 Software Design & Implementation

SR-0690-DOC-GEN	**System Specification Document**	**MUST-HAVE**
The System Specification Document (SSD) **shall** include details on the following: • *Trade-off* criteria and trade-off analysis; • *Engineering methodologies* adopted; • A quantitative justification for *cost-effectiveness* of the system platform, particularly in relation to Cloud facilities; • Security measures preventing malicious access to the system; • A *design walkthrough* describing fully usage of the system; • Conformance to *EU General Data Protection Regulations* (GDPR).		

SR-0690-DOC-GEN	System Specification Document	MUST-HAVE
Sources: *SOW §3.4.4 p39*		**INSPECT**

SR-0700-DOC-GEN	Compliance to ECSS	SHOULD-HAVE
The contractor **shall** identify the correspondence between the documentation set and applicable software within the CCI+ Salinity project and those required by the applicable Software Standard, *e.g.* appropriate components of ECSS-E-ST-40C. NOTE: The deliverables identified in the SOW do not have a 1-to-1 correspondance to the artifacts identified in ECSS-E-ST-40C neither does the CCI+SSS project meeting schedule represent those identified in ECSS-E-ST-40C Figure 4-2 Overview of the software life cycle process.		
Sources: *SOW §2.8 p25*		**DEMONSTRATE**

SR-0710-CON-SOFT	FOSS & Component Re-use	SHOULD-HAVE
The Contractor **shall** Fully capitalise on existing European assets through their reuse, particularly Open Source scientific tools and prototype ECV processing systems from prior projects. All investment in tools and technology shall be optimal and justified.		
Sources: *SOW §3.4.4 p39*		**INSPECT**

SR-0720-CON-SOFT	Open Source Software	MUST-HAVE
All code developed **shall** be Open Source (MIT License) as far as is possible by the licensing constraints of existing ECV processing systems reused by the Contractor.		
Sources: *SOW §3.4.4 p39*		**INSPECT**

SR-0730-CM-SOFT	Software Version Control	MUST-HAVE
All Open Source code developed by the Contractor **shall** be maintained in a common CCI GitHub code repository, details of which shall be provided by the Agency at project kick-off. This is to ensure source code version control		
Sources: *SOW §3.4.4 p39*		**INSPECT**

SR-0740-CM-SOFT	System Sustainability	SHOULD-HAVE
The system **shall** be designed to be sustainable. This implies the requirements for configuration control and maintenance (bug tracking, reprocessing, traceability), operability and transferability are priorities.		
Sources: *SOW §2.8 p25*		**INSPECT**

SR-0750-CON-SOFT	Modular & Flexible Design	SHOULD-HAVE
The system **shall** be designed to be regularly updated to include algorithm improvements. The design should also be modular and flexible while at the same time capable of rapid reprocessing, thus the overall design needs to be developed with end-to-end throughput of the ECV production as a design priority.		
Sources: *SOW §2.8 p25*		**INSPECT**

SR-0760-CON-SOFT	Programming Language	SHOULD-HAVE

The system **shall** be implemented in the native language of the standalone processing chains along with Python to provide binding between components. For SMOS the processor is written in C.	
Sources: *SOW §2.8 p25; USB §3.6*	**INSPECT**
Comment: User survey indicated that users would like tools for data extraction / reading and 54% desired any user-centred software library be written in MATLAB, 28% requested Python and 11% Fortran.	

4.5 System Infrastructure

SR-0770-PRF-HW	**Infrastructure Sizing**	**MUST-HAVE**
Each CCI project team (the contractor) **shall** ensure that the system is adequately dimensioned to accommodate the growing volumes of input and output data, and the increasing computational loads needed to process, re- process, quality control, validate, and disseminate multi-decadal, global, ECV data products, of the required climate quality, in a timely manner.		
Sources: *SOW R-17*		**DEMONSTRATE**

SR-0780-PRF-HW	**Processor / RAM Sizing**	**MUST-HAVE**
Based solely on SMOS L2 processing performance metrics in order to process 11 years of SMOS L1c data in 4 months**.** The platform shall be needed to process SMOS L1c data at a rate of 35:1. **The total specifications for the data processing are, therefore, set up to 520 cores and 1.24TB RAM.**		
Sources: *SRD 2.6.2*		**ANALYZE**
Comment: Note that this metric is based only on processing SMOS L2 on native grid it and must be considered as a higher threshold for CPU/RAM requirements.		

SR-0790-PRF-HW	**Storage Sizing**	**MUST-HAVE**
Based DARD and summarised. In section 2.6.3 the lower threshold for storage requirements is 250TB.		
Sources: *SRD 2.6.3; DARD §3*		**ANALYZE**

SR-0800-PRF-HW	**Automated HPC Processing Chain**	**MUST-HAVE**
Given the large amounts of data to be processed, the Contractor **shall** develop an automated high-performance demonstration processing chain.		
Sources: *SOW ANNEX B TR-43*		**INSPECT**

SR-0810-PRF-HW	**Computing Resource Constraints**	**MUST-HAVE**
The CCI+ Salinity processor **shall** be implemented on a sufficiently powerful (possibly distributed) computing infrastructure that is capable of extension and evolution and processing all the required products within the project schedule.		
Sources: *SOW ANNEX B TR-44*		**DEMONSTRATE**

Page Intentionally Blank

Annex A: Example netCDF

netCDF variable example

```
       float variable1(time, lat, lon) ;
                  variable1:long_name = "a_longer_descriptive_name_of variable" ;
                  variable1:standard_name = "CF_standard_name_here" ;
                  variable1:units = "unit from UDUNITS" ;
                  variable1:valid_range = 0f,10f ;
                  variable1:actual_range = 1f,9f ;
                  variable1:anciliary_variables = 'variable1_uncertainty variable1_flag' ;
        float variable1_uncertainty(time, lat, lon) ;
               variable1:long_name = " uncertainty associated with variable1" ;
                  variable1:units = "unit from UDUNITS" ;
           byte variable1_flag(time,lat,lon) ;
                  variable1_flag:long_name = "status flag associated with variable1" ;
                  variable1_flag:standard_name = "status_flag" ;
                  variable1_flag:_FillValue = -128b ;
                  variable1_flag:valid_range = 0b,2b ;
                  variable1_flag:flag_values = 0b,1b,2b ;
                  variable1_flag:flag_meanings = "good_quality problem1 problem2" ;

   //global attributes:
                  :key_variables = "variable1" ;
```

netCDF grid example

```
netcdf CCI_example_grid {
dimensions:
       time = UNLIMITED ;
        lat =  1200 ;
        lon =  1200 ;
         nv = 2 ; //number of vertices (to define the grid boundaries)
variables:
        float lat(lat) ;
                lat:standard_name = "latitude" ;
                lat:units = "degrees_north" ;
                lat:bounds = "lat_bnds" ;
        float lon(lon) ;
                lon:standard_name = "longitude" ;
                lon:units = "degrees_east" ;
                lon:bounds = "lon_bnds" ;
        double time(time) ;
                time:standard_name = "time" ;
                time:units = "seconds since 1970-01-01 00:00:00 0:00" ;
                time:bounds = "time_bnds" ;
        float lat_bnds(lat,nv)
        float lon_bnds(lon,nv)
        double time_bnds(time,nv)
        float ecv_variable(time, lat, lon) ;
                ecv_variable:long_name = "a_longer_descriptive_name_of_the_variable" ;
```

Annex B: Compliance Matrix

SRD ID	TITLE	SOURCE
SR-0010-FUN-GEN	SSS Observing System	SOW ANNEX B TR-1
SR-0020-FUN-GEN	Space-based SSS Continuity	SOW ANNEX B TR-1
SR-0030-FUN-GEN	Develop Retrieval Algorithms	SOW ANNEX B TR-2
SR-0040-FUN-GEN	Deliver SSS ECV Products	SOW ANNEX B TR-6
SR-0050-FUN-GEN	Deliver Validated Prototype Products	SOW ANNEX B TR-7
SR-0060-FUN-ACQU	Input Data Acquisition	PROP §2.2.3.2.1.2, DARD §3.1
SR-0070-FUN-ACQU	Ancillary Data Acquisition	PROP §2.2.3.2.1.2, DARD §3.2, DARD §3.3
SR-0080-FUN-ACQU	Auxiliary Data Acquisition	PROP §2.2.3.2.1.2, DARD §3.4
SR-0090-FUN-PRE	L1 Data Pre-processing	PROP §2.2.3.2.1.2
SR-0100-FUN-PRE	Auxiliary Data Pre-processing	PROP §2.2.3.2.1.2
SR-0110-FUN-PRE	Input Data QC - Logging	PROP §2.2.3.2.1.2
SR-0112-FUN-PRE	Look-up Table Generator & Tools	PROP §2.2.3.2.1.2
SR-0120-FUN-PROC	Level 2 Data Processing	PROP §2.2.3.2.1.2
SR-0130-FUN-PROC	Level 3 Data Processing	PROP §2.2.3.2.1.2
SR-0140-FUN-PROC	Level 4 Data Processing	PROP §2.2.3.2.1.2
SR-0145-FUN-PROC	ECV Data Post-processing	PROP §2.2.3.2.1.2
SR-0150-FUN-DIST	FTP Data Access Protocol	URD §3.4
SR-0160-FUN-DIST	Obs4MIPS Delivery	SOW ANNEX B TR-17
SR-0170-FUN-DIST	CCI Open Data Portal Delivery	SOW ANNEX B TR-18
SR-0190-FUN-DIST	Public Availability of ECV Products	SOW ANNEX B TR-42
SR-0200-FUN-PROD	Global Ocean Coverage	SOW ANNEX B TR-8
SR-0210-FUN-PROD	Mission Lifetime Coverage	SOW ANNEX B TR-9
SR-0220-FUN-PROD	Available Product Processing Levels	SOW ANNEX B TR-10; URD §3.4
SR-0230-FUN-PROD	Available Higher-Level Products	SOW ANNEX B TR-11; URD §3.4
SR-0240-OPL-PROD	Product Variables	SOW ANNEX B TR-14; SOW ANNEX B §3.1
SR-0250-POL-PROD	Mission Datasets	SOW ANNEX B TR-15
SR-0260-OPL-PROD	Product Temporal Frequency	SOW ANNEX B TR-16; URB §2.1 Table 1; URD §2.2 Table 2; USB §2.4

SR-0279-OPL-PROD	Product Spatial Resolution	SOW ANNEX B TR-16; URB §2.1 Table 1; URD §2.2 Table 2; USB §2.4
SR-0280-OPL-PROD	Threshold & Goal Values	SOW ANNEX B TR-24; URD §2.2 Table 2
SR-0290-OPL-PROD	User Resolution, Coverage & Accuracy	URD §3.4.1
SR-0300-OPL-PROD	Dataset Production Schedule	SOW ANNEX B TR-45; PSD
SR-0310-QTY-PROD	Level 4 Error Specification	PSD §1.2(2)
SR-0320-QTY-PROD	Correction of Intra-mission Biases	SOW ANNEX B TR-16
SR-0330-QTY-PROD	Quality Indicators / Flags	URD §3.5; PSD §5.1
SR-0340-QTY-PROD	Assessment Long-term Stability	SOW ANNEX B TR-39; URD §2.1 Table 1
SR-0350-RLY-PROD	Product Uncertainty	SOW ANNEX B TR-20; URD §2.1 Table 1
SR-0360-RLY-PROD	ECV Product include Uncertainty	SOW ANNEX B TR-33
SR-0370-RLY-PROD	Define Uncertainty Methodology	SOW §3.3.4
SR-0380-RLY-PROD	Report Uncertainty to PUG	SOW ANNEX B TR-21
SR-0400-VRF-PROD	Pre-defined Validation Metrics	SOW ANNEX B TR-25; URD §3.5
SR-0410-VRF-PROD	ECV Product Validation	SOW ANNEX B TR-35; URD §3.5 Figure 17
SR-0420-VRF-PROD	Uncertainty Validation	SOW ANNEX B TR-36
SR-0430-VRF-PROD	In situ FRM Database	SOW ANNEX B TR-38
SR-0440-VRF-PROD	ISDB include Uncertainties	SOW ANNEX B TR-40
SR-0450-VRF-PROD	ISDB Documentation	SOW ANNEX B TR-41
SR-0460-VRF-PROD	ISDB Uncertainties Reported to PUG	SOW ANNEX B TR-42
SR-0470-INF-FRMT	CCI Data Standards	SOW §3.5.1
SR-0480-INF-FRMT	Use netCDF-4 (Classic) format	DSTD R-1; PSD §2.1
SR-0490-INF-FRMT	Use CF Convention	DSTD R-1; DSTD § 2.5.1; ; PSD §2.1
SR-0500-INF-FRMT	Use ACCD Convention	DSTD R-1
SR-0510-INF-FRMT	CCI Specific Metadata	DSTD §2.5.1
SR-0520-INF-FRMT	Main Variables use CF Standard Names	DSTD R-1
SR-0530-INF-FRMT	Key Primary Variables	DSTD R-6; PSD §2.4.2
SR-0540-INF-FRMT	Gridded Data	DSTD R-7
SR-0550-INF-FRMT	Additional Data Formats	DSTD R-2
SR-0560-INF-FRMT	INSPIRE Metadata	DSTD R-4
SR-0570-INF-FRMT	CCI Ontology Terms	DSTD R-5
SR-0580-INF-FRMT	DOI Product Identifiers	SOW ANNEX B TR-12
SR-0590-INF-FRMT	Directory Structure	DSTD R-8

SR-0600-INF-FRMT	File naming Convention	DSTD R-9; PSD §2.2
SR-0610-FUN-PROC	Development of Improved Algorithms	SOW ANNEX B TR-26
SR-0620-FUN-PROC	Improve Retrieval Algorithms	SOW ANNEX B TR-27
SR-0630-FUN-PROC	Calibration & Aging Biases	SOW ANNEX B TR-28
SR-0640-FUN-PROC	Algorithm Performance Metrics	SOW ANNEX B TR-29
SR-0650-FUN-PROC	Perform Round-Robin Inter-comparison	SOW ANNEX B TR-30
SR-0660-FUN-PROC	Select Definitive Retrieval Algorithms	SOW ANNEX B TR-31
SR-0670-FUN-PROC	Investigate Alternate Measurements	SOW ANNEX B TR-34
SR-0680-FUN-PROC	Use of C-band Radiometers	SOW ANNEX B TR-19
SR-0690-DOC-GEN	System Specification Document	SOW §3.4.4 p39
SR-0700-DOC-GEN	Compliance to ECSS	SOW §2.8 p25
SR-0710-CON-SOFT	FOSS & Component Re-use	SOW §3.4.4 p39
SR-0720-CON-SOFT	Open Source Software	SOW §3.4.4 p39
SR-0730-CM-SOFT	Software Version Control	SOW §3.4.4 p39
SR-0740-CM-SOFT	System Sustainability	SOW §2.8 p25
SR-0750-CON-SOFT	Modular & Flexible Design	SOW §2.8 p25
SR-0760-CON-SOFT	Programming Language	SOW §2.8 p25; USB §3.6
SR-0770-PRF-HW	Infrastructure Sizing	SOW R-17
SR-0780-PRF-HW	Processor / RAM Sizing	SRD 2.6.2
SR-0790-PRF-HW	Storage Sizing	SRD 2.6.3; DARD §3
SR-0800-PRF-HW	Automated HPC Processing Chain	SOW ANNEX B TR-43
SR-0810-PRF-HW	Computing Resource Constraints	SOW ANNEX B TR-44

End of document