CCI+ PHASE 1 – NEW ECVS
PERMAFROST

D3.2 SYSTEM SPECIFICATION DOCUMENT (SSD)

VERSION 2.0

29 FEBRUARY 2020

PREPARED BY

Document Status Sheet

Issue	Date	Details	Authors
1.0	31.05.2019	V1	A. Wiesmann, T. Strozzi (GAMMA), S. Westermann (GUIO), A. Bartsch (b.geos)
2.0	29.02.2020	Update of reference documents and section 2.7 Hardware infrastructure	A. Wiesmann (GAMMA), S. Westermann (GUIO), A. Bartsch (b.geos)

Author team

Andreas Wiesmann, GAMMA

Annett Bartsch, B.GEOS

Sebastian Westermann, UIO

Tazio Strozzi, GAMMA

ESA Technical Officer:

Frank Martin Seifert

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Within the European Space Agency (ESA), the Climate Change Initiative (CCI) is a global monitoring program which aims to provide long-term satellite-based products to serve the climate modelling and climate user community. Permafrost has been selected as one of the Essential Climate Variables (ECVs) which are elaborated during Phase 1 of CCI+ (2018-2021).

This novel ECV permafrost product should benefit a wide range of applications and users.

This document outlines the system specifications for the Permafrost_cci processing capable of producing the Permafrost_cci ECVs. This document provides details on the Permafrost_cci processing system. It describes its purpose and intended use as well as the main requirements, functions and components and discusses the main operational scenarios, the necessary infrastructure and highlights the functional design from different perspectives, the users, system operators and developers view. Finally, it summarises the specifications concerning system life cycle design, implementation and maintenance costs and performance.

Three modules are part of the processing system (PS): (1) Pre-processing of satellite and reanalysis data and generation of input files for CryoGrid CCI, (2) Main processing with CryoGrid CCI, and (3) Generation of Permafrost_cci products form raw CryoGrid CCI output and verification/control of data quality. The processing within the Permafrost_cci project is done within the existing supercomputing infrastructure in Norway. Users access the Permafrost_cci data products using the ESA CCI Open Data Portal.

The input data of the PS are EO data, reanalyses data, and auxiliary data such as DEM and masks. The verification is done separately for the different PS modules. It covers system (module) validation. The PS provides tools that facilitate verification tasks such as benchmark test data, test tools and verification reporting.

1 INTRODUCTION

The European Space Agency (ESA) Climate Change Initiative aims to generate high quality Essential Climate Variables (ECVs) derived from long-term satellite data records to meet the needs of climate research and monitoring activities, including the detection of variability and trends, climate modelling, and aspects of hydrology and meteorology.

1.1 Purpose of the document

This document is deliverable D3.2 System Specification Document (SSD) of the Permafrost_cci project requested in the Statement of Work (SoW) [R-1]. The SSD incorporates the requirements described in the System Requirements Document (SRD) [R-6] and specifies the characteristics of an operational ECV production system from a System Engineering point of view.
The system design is based on experience with prototype processors developed prior to this project.
The prototypes are referenced in this document. The processing system (PS) is further developed in the years to come. Consequently, this SSD is a living document and will be complemented when necessary.

1.2 Document status

This document is based on issue 1.0 of the Data Access Requirements Document (DARD), Issue 1.0 of the Product Specification Document (PSD), and Issue 1.0 of the User Requirements Document (URD); refinement of this document will be necessary of catchment of future issues of these documents.

1.3 Applicable documents

[AD-1] ESA 2017: Climate Change Initiative Extension (CCI+) Phase 1 – New Essential Climate Variables - Statement of Work. ESA-CCI-PRGM-EOPS-PF-17-0032. Issue 1.4 r2 EOP-SEP/SOW/0031-1.4 r2

[AD-2] Requirements for monitoring of permafrost in polar regions - A community white paper in response to the WMO Polar Space Task Group (PSTG), Version 4, 2014-10-09. Austrian Polar Research Institute, Vienna, Austria, 20 pp

[AD-3] ECV 9 Permafrost: assessment report on available methodological standards and guides, 1 Nov 2009, GTOS-62

[AD-4] GCOS-200, the Global Observing System for Climate: Implementation Needs (2016 GCOS Implementation Plan, 2015.

1.4 Reference Documents

Table 1: Reference Documents.

Ref	Title	Version
[RD-1]	User Requirements Document (URD)	1.1
[RD-2]	Product Specification Document (PSD)	2.0
[RD-3]	Data Access Requirements Document (DARD)	1.0
[RD-4]	Data Standards Guidelines (DSWG). CCI-PRGM-EOPS-TN-13-0009	2.0
[RD-5]	System Requirements Document (SRD)	2.0
[RD-6]	Algorithm Development Plan (ADP)	2.0
[RD-7]	Product Validation Plan (PVP)	2.0

1.6 Acronyms

AD	Applicable document
ATBD	Algorithm Theoretical Basis Document
B.GEOS	b.geos GmbH
CCI	Climate Change Initiative
CR	Cardinal Requirement (as defined in [AD-1])
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
FOSS	Free Open Source Software
GAMMA	Gamma Remote Sensing AG
GCOS	Global Climate Observing System
GUIO	Department of Geosciences University of Oslo
IPCC	Intergovernmental Panel on Climate Change
L4	Level 4
PE	Permafrost Extent
PS	Processing System
PUG	Product User Guide
RD	Reference Document
RS	Remote Sensing

2 PERMAFROST_CCI PROCESSING SYSTEM OVERVIEW

This section gives an overview of the processing system (PS) with its main modules, functions and components. It also summarises its designated use and the system requirements.

2.1 Context

As depicted in the general overview of Figure 1, the Permafrost_cci PS generates products and supports the process of algorithm improvement, reprocessing and validation. It provides products and services to the permafrost community supporting their climate change impact assessment over a wide range of scales. The PS will be used by the Permafrost_cci consortium but can also be applied by others as the overall workflow is very generic. The key difference to data production in other science projects is their often-missing dissemination, i.e. the work ends with a publication and generated data products are not shared. The PS is specified to provide permafrost-related products such as ground temperature, depth of active layer and permafrost zones based on state-of-the-art technology using the best suited and available EO data and algorithms. The products are produced in a transparent and documented way, with accompanying meta-data, documentation, uncertainty and validation reports (project outreach).

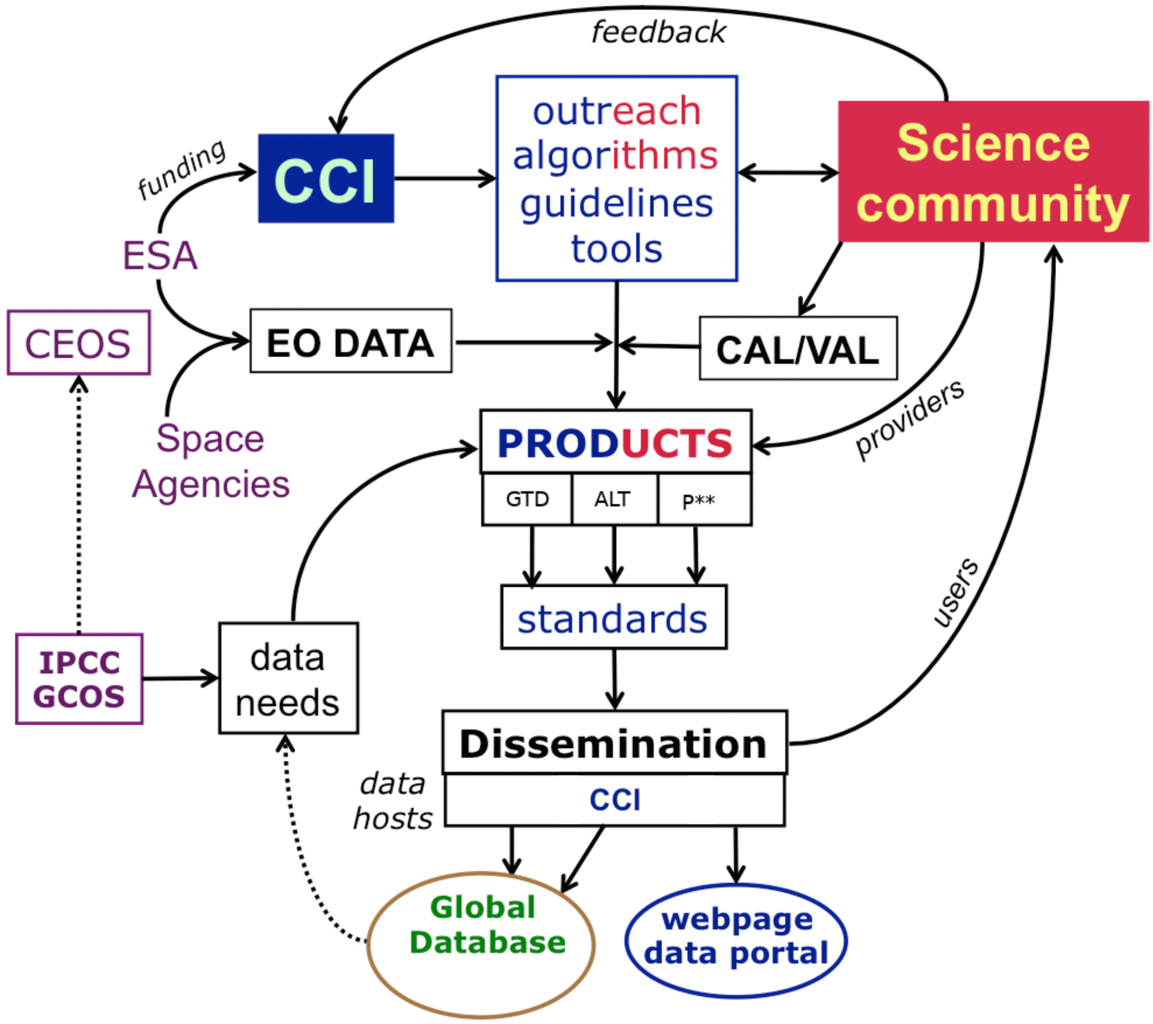

Figure 1: Permafrost_cci PS environment.

A schematic overview of the Permafrost_cci processing system and its relation to developments by the science community for the individual products and their dissemination is shown in .

The PS can be understood as a value-adding layer between the data provider and the users. A high-level relation diagram of the PS is given in (). There are interfaces to the different user communities, which receive products and can provide feedback. Another interface is with the EO data providers. Depending on the module, EO data are obtained from the providers at CEOS level 3 or 4 and are ingested into the PS. Feedback is given to the providers about issues found with the data, processing improvements and requirements for the continuity of the service. Another interface is towards third-party sources to receive ancillary and validation datasets.

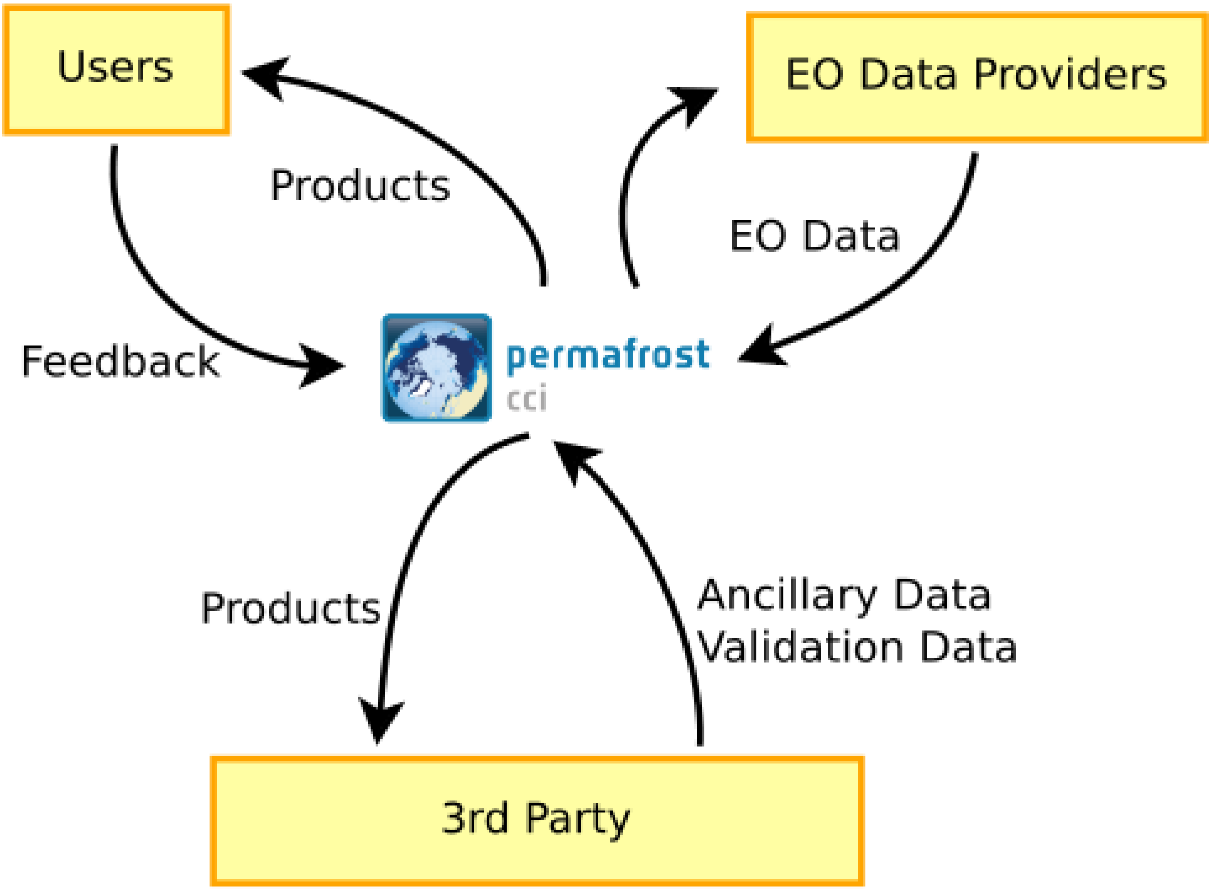

Figure 2: High-level relation diagram of the Permafrost_cci PS.

2.2 User Requirements

User requirements are documented in the URD [RD-1]. It outlines the requirements for Permafrost_cci ECVs (ground temperature - GRD; active layer thickness – ALT and permafrost extent - PFT) obtained through engagement with users from across climate applications, including the detection of variability and trends as well as climate modelling. Involvement of users into the system development process is one of the main points (URq_06).

2.3 Main System Requirements

System requirements are compiled in the SRD [RD-5]. The document lists system requirements grouped into functional, operational and performance requirements many with impact on the system design. Section 7 provides the complete matrix of forward tracing from requirements to sections and also indicates evolution of the requirements with time.

High-level requirements are to generate the Permafrost_cci products (PF-FUN-0010, PF-FUN-0020, PF-FUN-0030). The processing line shall be well defined and flexible for future updates and adaptations (better algorithms, new input data) (PF-SIZ-0080). The available data shall be frequently reported and properly disseminated to the interested user communities (PF-INT-0090).

The main use scenarios include validation and improvement of Earth System Models. That requires the data to be of high value for more general climate change studies (e.g. permafrost interactions with vegetation, water bodies, carbon cycle), as well as for informing field scientists about the potential past evolution of the permafrost environment. The functional scope of the system is not restricted to the processing, reprocessing, validation and improvement cycle, although this is its main purpose. Also functions to make output products and documentation available to users are included in the scope of the system.

2.4 Main Functions

Requirements in this section are:

→ PF-INT-3020 Long-term storage
→ PF-INT-3030 Self-standing documentation
→ PF-FUN-1050 Reprocess also parts of the products
→ PF-FUN-1060 The PS shall be able to do partial processing

The PS hosts input data, performs pre-processing, classification processing, supports validation, and serves the output to users (‘Dissemination’ in), the part that is often missing in other scientific studies. It supports the interaction between the development team and users by information services. Processor interfaces, configurable data management, and version control with easy transfer to operations supports testing and development of new algorithms and continuous improvement.

To fulfil its purpose in such a context the PS provides three high level functions:

1. Production
2. Dissemination
3. Life Cycle Management

In the following we will discuss the fundamental operations of the PS with regard to these three functions.

For production the focus is on processing and repeated reprocessing of complete products. Necessary functions are:

1. Storage to gather and store inputs, intermediate products, output products and auxiliary data;
2. Processors to produce output products from the input data;
3. Processing Control;
4. Quality Control of the intermediate and output products;

In general we distinguish between the pre-processing, the main processing and the post-processing functionality covering the preparation of the input data, the processor itself, and the product generation steps, respectively.

For dissemination, the focus is on the service for the permafrost and climate community. The products are distributed through existing and in development platforms of the ESA CCI data Portal and the Permafrost CCI Website. Functions are:

1. Project Information/Introduction

2. Processing Information
3. Product Description (incl. meta-data)
4. Online Product Access
5. Validation Support
6. Feedback Handling
7. Long-term Preservation

Good Life Cycle Management helps improve service quality and reliability, crucial elements for the attractiveness of the provided service. A small effort should be necessary to implement an improved processor handling, improved algorithms and data of new sensors, given that basic characteristics of the data and the processing do not change in a fundamental way. Consequently, fundamental necessary operations are:

- Test environment (for new processors)
- Access to test or benchmark input data (for tests and comparison)
- Version Management (→ this is linked to the point "Documentation of the Processing in Meta-data")

2.5 The modules of the PS

Like some of the other CCI projects, Permafrost_cci produces products that rely on completely different processing chains in terms of input and ancillary data and the processor. The processing chains for the Permafrost_cci products are organized in modules that are part of the PS. Several products might be produced by the same module. In [RD-5] the following modules are part of the PS:

1. Pre-processing of satellite and reanalysis data and generation of input files for CryoGrid CCI
2. Main processing with CryoGrid CCI
3. Generation of Permafrost_cci products from raw CryoGrid CCI output and verification/control of data quality

Step 2 requires by far the largest share of the total processing time (>95%), with step 1 occupying most of the rest, while step 3 is negligible with regards to processing time. Steps 1 and 2 are realized as independent modules in the PS and could in principle run on independent computing infrastructure. As such, they are connected by data files which are written by step 1 and become read by step 2. In principle, the same could be done for steps 2 and 3, but writing the entire output of step 2 to physical files would create a large amount of data (tens of TB), which would significantly increase runtime due to the writing operations and internal data traffic within the HPC processing environment. For this reason, Steps 2 and 3 have (for year 1) been implemented as a single module, with the Permafrost_cci products being processed "on the fly" from the raw output of Cryogrid CCI.

Further modules can potentially be developed within the Permafrost_cci options and would be added here in the next releases.

Within the framework of the Permafrost_cci project we aim at developing the most efficient, consistent and sustainable system addressing the needs of the corresponding community. We also aim at investigating synergies among the modules especially for the interfaces but potentially also with other CCI-projects if the system benefits. Hence, in the following we address both levels in the SSD, the module level and the system level. The distributed approach of the PS requires a community of producers and consequently a coordination mechanism to ensure continuity and management of multiple outputs.

2.6 High Level Decomposition

The functions and modules listed in the previous section are implemented as functional components. In this section we outline the high-level architecture of Permafrost_cci on this subsystem level. Data-uptake is either through the mostly web-based services of satellite level 3 data providers (e.g. LPDAAC, CCI data portal) and those listed in [RD-3]. A key requirement is that all EO data are open and freely available. Data are then processed in the product generation modules and distributed through the CCI Common Basic Services (CCI data portal). Within this document we focus on the product generation (see Section 3.2). Furthermore, we provide an overview of the components of the PS as a starting point for the operational scenarios in Section 3 and design in Section 4. The specification with all components, functions and interfaces follows in Section 5.

On a high-level we can distinguish the three subsystems: production, user services, and data stewardship based on the way EO data is encapsulated. This can also be viewed from a functional perspective as: production, dissemination, and improvement. Processing storage of the production system is accessed by the module processors and needs not to be openly accessible (i.e. with write access to the storage medium). The user services and archive are taken care by the CCI data portal and is not discussed here.

Production Subsystem

The production subsystem contains the production and development. Production control, processing storage and the processors provide the basic infrastructure for processing. A test environment with read access on input data serves the development needs. Where applicable, main functions of the processing system and processing storage shall be available for development. Due to the distributed processing system, development of new versions will always be performed in parallel to running an older version.

2.7 Hardware Infrastructure

The processing within the Permafrost_cci project is done within the existing supercomputing infrastructure in Norway, which is managed nationally by the company UNINETT Sigma2 AS (https://www.sigma2.no/). Processing time is granted on demand on a biannual basis, using the available supercomputing clusters within Norway, which is continuously updated and extended. In year 1 and partially year 2, the Abel cluster at the University of Oslo was used for Permafrost_cci processing (https://www.uio.no/english/services/it/research/hpc/abel/more/index.html). Starting in fall 2019, Abel was gradually discontinued with operations terminating 31 January 2020. As this time frame overlapped with processing of year 2 products, the PS was migrated to the new SAGA cluster (https://documentation.sigma2.no/quick/saga.html) situated in Trondheim, Norway, which is the scheduled replacement for Abel. The migration was successfully finished in December 2019, followed by system verification and subsequent implementation of the year 2 PS.

Storage of input satellite data sets is currently realized on the Norwegian NIRD system (https://documentation.sigma2.no/storage/nird.html) which is also administered through UNINETT Sigma2.
On Abel, year 1 processing was accomplished on Intel E5-2670 (Sandy Bridge) processors at 2.6 GHz, yielding 16 physical compute cores per node. Each node has 64 GB of Samsung DDR3 memory operating at 1600 MHz, giving 4 GiB memory per physical core at about 58 GiB/s aggregated bandwidth using all physical cores. During processing, up to ten nodes (160 cores) were employed simultaneously, depending on availability within the SLURM scheduling system, yielding a total processing time of about two months (ca. 100k CPU hours used in total).
On Saga, year 2 processing is done on Intel Xeon-Gold 6138 2.0 GHz with 40 physical cores per node and 4GB memory per core. Other than on Abel, it is possible to distribute a single processing job over several nodes, using up to 100 cores (instead of 16 as on Abel). During processing, several 100 cores can be employed in parallel, distributed over several processing jobs. In total 200k CPU hours are foreseen to be used in year 2, with a processing time of one to two months, depending on the total computation load on Saga.

From year 1 and 2 processing, the following general constraints for hardware infrastructure for the Permafrost_cci PS can be formulated:

- HPC environment with CPUs with at least 4 GB memory per core
- for time-efficient processing, at least 100 cores should be available
- at least 10TB storage available

These requirements will be continuously updated during the project period.

3. PERMAFROST_CCI PROCESSING SYSTEM WORKFLOW AND OPERATIONAL SCENARIOS

3.1 Roles

The development team consists of scientists and operators that manage the production and continuous development. Actors in the operational scenarios are users with different roles depending on how they use the system:

- Permafrost_cci Users (client)
  - are interested in best existing permafrost products
  - are skilled in permafrost applications
  - provide feedback and proposals
  - request data format compatible with their communities
- Development team
  - mandated to run Permafrost_cci\
  - in dialogue with users
  - develops the PS further
  - issues product versions
- Validation experts

    - are part of the international community
    - support development team
    - provide local expertise
    - feedback on the products
- Auditor/Project Manager
    - project supervision

3.2 User Information and Data Access

Users access the Permafrost_cci data products using the ESA CCI Open Data Portal. Updates to the system are disseminated via the Permafrost_cci website.

Typical functionality performed in the context of user information and data access:

- The development team submits new products to the ESA CCI Open Data Portal. New versions or major updates will be announced on the Permafrost_cci website.
- A user accesses the Permafrost_cci web site as an entry point and gets general information on Permafrost_cci. It also provides links to software, services and resources.
- The user uses the CCI Search catalogue to find Permafrost_cci products. The catalogue also provides metadata and quicklooks.
- The user downloads product documentation (e.g. the Product User Guide).

3.3 Processing System Workflow

The basic processing system production workflow is given in Figure 3. It has the main parts preprocessing, retrieval, product generation and verification/validation. The input data are EO data, reanalyses data, and auxiliary data such as the digital elevation model and masks. The output of the production are the Permafrost_cci products. In the following the two Permafrost_cci modules will be presented in further detail.

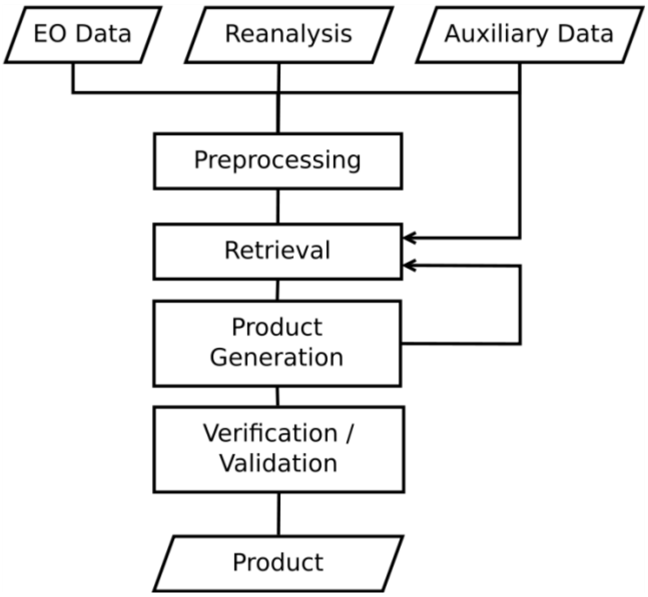

Figure 3: Permafrost_cci module high level diagram

.

Pre-processing

The pre-processing includes all steps necessary that the retrieval can be applied. This covers input and auxiliary data selection and input data preparation (conversion, reprojection, sampling, conversion).

Retrieval

In the retrieval process the algorithms from the ADP [RD-6] are applied.

Product Generation

In the product generation step the resulting maps are put in the delivery product format as requested and specified in the PSD [RD-2].

Verification/Validation

The products undergo defined procedures to ensure product integrity and quality. This step must not be confused with the product validation process that is done in WP42x following the Product Validation Plan (PVP) [RD-7].
The system operators are in charge of system (module) verification. The verification is done separately for the different PS modules. It covers:

- System (module) validation (after an upgrade or new installation)

The PS provides tools that facilitate these tasks such as

- Benchmark test data (PF-OPE-4050)
- Test tools (PF-OPE-4050)
- Verification Report (PF-RAM-5040)

3.4 Algorithm Improvement

The development team decides about features or processes to be improved in order to meet user requirements. The development team implements the improvements as new versions of processors and if necessary, as a modified workflow definition. The development team tests and validates the new products. The development team also decides about new versions to be released.
An improvement cycle is defined as:

1. New requirements are identified and analysed
2. Modified processor implementation
3. Test production
4. Validation
5. Decision to a) go to 1. and iterate again, b) implement go on, c) stop here
6. Release a new version (code freeze) while retaining older ones for cross check
7. Start production

After each validation a decision is taken if the improved algorithm is accepted, further developed or the development is stopped. Only in the case of acceptance, the development leads to a new version of the processor and full reprocessing of the archive or implementation for new data. This cycle of innovation and improvement typically takes 5-10 years. With an effective control mechanism in place, the update cycle can be reduced to a few years.

4. FUNCTIONAL DESIGN

Here we will discuss and present the major functional blocks of the Permafrost_cci PS.

4.1 Services

Most of the necessary services are provided through the CCI environment such as the data exchange storage, the document management system and user interaction. For the PS development a software repository and an issue tracker are hosted at GUIO as part of the PS.

Processor software repository

An important element of the modern software development process is source control (or version control). Cooperating developers commit their changes incrementally to a common source repository, which allows them to collaborate on code without resorting to crude file-sharing techniques (shared drives, email). Source control tools track all prior versions of all files, allowing developers to "time travel" backward and forward in their software to determine when and where bugs are introduced. These tools also identify conflicting simultaneous modifications made by two (poorly-communicating) team members, forcing them to work out the correct solution (rather than blindly overwriting one or the other original submission). The software repository contains the actual processing code and all prior versions. The write access to the processor repository is restricted to the development team. As all software changes are updated directly in the repository, the software changes are published almost immediately and are made available for review.

Issue Tracker

During software development a Redmine issue tracker is used (http://www.redmine.org/). Redmine is a flexible project management web application written using Ruby on Rails framework. Redmine integrates the version control system into its user interface and manages the access control to the version control system resulting in a state-of-the-art FOSS software development environment.

4.2 Processors

The processors cover the necessary tools to produce the different Permafrost_cci products. The Permafrost_cci is organized in modules covering the production of the different products. In general, a distributed processing approach is followed. Consequently, the modules are portable.

Requirements:
→ PF-SIZ-0040 Flexible production
→ PF-FUN-5010 Data overwrite

Table 2: Permafrost_cci modules

COMPONENT	PURPOSE	CONTENT	IMPLEMENTATION
Preprocessor	Generate input data for CryoGrid	input data	new processor, building on Globpermafrost preprocessor
CryoGrid processor	generate depth- and time-resolved ground temperature data	raw model output	New processor
Postprocessor	Generate L4 Products from raw output, including projection to target coordinate system	L4 products	new processor, building on GlobPermafrost postprocessor

4.3 Concept for continuous improvement

Continuous improvement is an important aspect in the CCI projects. To ensure a transparent process Software Modularity, Software Version Concept, Version Control, and Version Numbering are important issues. This section defines the structures and functions that extend the production environment for continuous improvement. Focus is on flexibility, rapid testing and prototyping. The concepts described are processors, versioning, and a test environment. The concept of processors and versioning contribute to the modularity of the system.

Requirements:
→ PF-OPE-4010 Development under version control
→ PF-OPE-4020 Decoupled from own research

The software of the PS and the processing algorithms code are under version control. The software repository contains the actual processing code and all prior versions. All software changes are documented in the repository. Version numbering of the processor is reflected in the repository by revisions and tags. Revisions are usually linked to commits and indicate the sequential order of documented changes. Tags are set to indicate software releases of frozen software states. Subversion is a good candidate for version control. Together with Redmine it is a complete FOSS version control and issue tracking system.

Data processors help to organize the data processing in modules. Due to the differing input and output datasets/formats, the modules are normally not shared among products (even if the functionality is the same). A processor is a software component that can be parametrised and that generates a (higher level) output product of a certain type from one or several input products of certain types. A PS module consists of the sequential call of processors. Each processor has its own version information. Parameters and environment variables are provided in dedicated parameter files within the code. Feedback is received by a return code, messages on stdout/stderr and in log files.

Table 3: Permafrost_cci repository details

COMPONENT	PURPOSE	CONTENT	IMPLEMENTATION
Software Repository	Stores all versions of the processor code in a transparent way, with branches, authorship	Code	Subversion Redmine Tools

4.4 System Documentation

The documentation contains the PS documentation consisting of manuals, specifications and reports, as well as the product documentation consisting of product specifications, manual and validation reports. At this stage no advanced functionality such as collaborative editing etc. seems to be necessary so that the basic functionality of any FOSS CMS might be sufficient for this task.
Since the Permafrost_cci processing chain was to a large extent compiled new in year 1, documentation will gradually be improved in years 2 and 3, following the consolidation of the code.
Requirements addressed by this section are:
→ PF-FUN-0040 Product and PS Description
→ PF-RAM-5020 Self-standing documentation

Table 4: Permafrost_cci document management details

COMPONENT	PURPOSE	CONTENT	IMPLEMENTATION
Documentation Management	Stores documentation in a structured and transparent way	Documentation	Project Website

The PS documentation includes requirement documents, design and interface control documents, test documents, manuals, and maintenance information. Permafrost_cci deliverables to name here are ATBD and the PUG. The SRD and SSD define requirements and design of the system.

5. DEVELOPMENT, LIFE CYCLE, COST AND PERFORMANCE

This section discusses the system development in the future, potential development strategies, efforts and costs. The development is driven by several factors such as the availability of new technology, faster algorithms, new scientific findings and improved product algorithms, new available EO data, and user needs.

5.1 Re-use and Development

Development is needed to bring the existing prototypes of the PS modules to a higher operational level satisfying the requirements listed in the previous sections and to add the missing components such as those for user services, data handling, life cycle management, archiving etc.

Requirements addressed by this section are:
→ PF-OPE-4030 Development based on User Requirements
→ PF-OPE-4020 Development decoupled from research
→ PF-OPE-4060 Freeze prototype
→ PF-FUN-5030 Verification of implementation

In the Table 5 we summarise the tools that were used, adapted, configured and integrated during development of the PS within the CCI and beyond.

Table 5: Permafrost_cci development tools

NAME	USAGE	REMARKS
Subversion	Version control	FOSS
Redmine	Issue tracking	FOSS
Python3	Scripting, netcdf4 reader, pyresample, rasterio	FOSS
CryoGrid prototype	Running prototype in retrieval	GUIO

5.2 System Life Cycle Drivers and Considerations

The PS needs to be incrementally adapted to integrate new functional extensions, improved algorithms and input datasets. New EO data make adaptations necessary and most likely also have an impact on the hardware infrastructure. The life cycle plan cannot be static as it is not foreseeable. Currently the following driving factors are identified:

- Availability of the existing processor module prototypes
- Functional extension of the system

- New workflows
- Improved algorithms
- New Sensors
- Hardware improvements
- Dependencies on 3rd parties (other ECVs, data providers, new users)

To answer the first two points the system is initially based on the prototype. Incrementally, additional components and functions are added and interfaces to data providers and users are extended. The third and fourth point of workflow and algorithm development requires the addition of tools for validation and user feedback.
The new sensors and the increased data volume are a qualitative change, too. The existing methods need to be adapted to make use of new sensors. For the longer perspective renewal of hardware and optional change of software layers must be taken into account. The PS design is prepared for this by the modularity of its functional components. The last item is not so relevant for Permafrost_cci at the moment as the dependence on other CCI projects is minor.

Requirements:
→ PF-INT-3020 Long-term storage
→ PF-FUN-1020 Unique identifier
→ PF-FUN-1030 Structured storage
→ PF-FUN-1050 Reprocess Products
→ PF-OPE-4010 PS shall be under version control
→ PF-OPE-4050 Test tools
→ PF-RAM-5040 Verification

5.3 Sizing and Performance Analysis

In the SRD [RD-5], no specific requirements are present concerning the processing time performance. At the moment it is mainly labour hours that drive the processing rather than CPU core hours. Full reprocessing of historical data requires a variable amount of work, depending on the product (see tables below). The data storage budget for inputs and outputs for historical data and for the yearly increase of acquired data is in the low TB range initially.

There exist requirements on disk space that are modest:
→ PF-SIZ-2060 Space for input data
→ PF-SIZ-2050 Space for auxiliary data
→ PF-SIZ-2040 Space for output data
→ PF-SIZ-2030 Run on available hardware
→ PF-SIZ-2010 Time series

The budgets for data storage and processing capabilities are estimated in Table 6. The budget for data storage mainly depends on the amount of input data to be managed. This comprises historical data and data acquired continuously.

Table 6: Permafrost_cci annual data storage budget

DATA	PRODUCT	TIME SPAN	HISTORICAL DATA
MODIS	MOD11A1 and MYD11A1	Annual	2-3 TB
Reanalysis	ERA-5	Annual	100 GB
Total		Annual	2 – 3 TB

5.4 Cost Estimation

Storage and processing hardware have been provided at no costs in year 1, and tentatively will be free of charge in years 2 and 3. Depending on providers, the computation cost (ca. 100k CPU hours) in the cloud is in the order of 15k€, when using a commercial provider.
Development costs are on the order of 400 k€.

Requirements addressed by this section are:
→ PF-OPE-0100 Minimum maintenance and cost

The costs for the PS are composed of costs for storage, processing, network, development and integration, operations, dissemination and labour. The different modules have different needs and some of the efforts are covered through the GIUO infrastructure. The cost estimates per year in kilo Euros are summarized in Table 7. Costs from operational production of on-demand services move from the products to the user services after initial setup.

Table 7: Permafrost_cci processing cost estimates

PRODUCT	COMPONENT	YEAR 1	YEAR 2	YEAR 3
Permafrost	Hardware	NA	NA	NA
	Development	200	100	100
	Operations	15	15	15
	Infrastructure	NA	NA	NA
	Total	**215**	**115**	**115**

6 REQUIREMENTS TRACEABILITY

Table 8: Requirements traceability matrix

ID	TITLE	REF
PF-FUN-0010	Develop and validate algorithms to approach the GCOS ECV and meet the wider requirements of the Climate Community (i.e. long term, consistent, stable, uncertainty-characterized) global satellite data products from multi- sensor data archives. (CR-1)	§2.3
PF-FUN-0020	Produce, validate and deliver consistent time series of multi-sensor global satellite ECV data products for climate science. (CR-2)	§2.3
PF-FUN-0030	The *Permafrost_cci* system shall be able to generate different products: 1. GTD, when the parameter is ground temperature at a certain depth 2. ALT, if the parameter is active layer thickness 3. PFR if the parameter is permafrost extent (fraction) 4. PFF if the parameter is permafrost-free fraction 5. PFT if the parameter is fraction underlain by talik 6. PZO if the parameter is permafrost zone	§2.3
PF-FUN-0040	Generate and fully document a production system capable of processing and reprocessing the data, with the aim of supporting transfer to operational activities outside CCI (such as C3S). [CR-4]	§2.3
PF-OPE-0050	All project documentation shall be made publicly available via the CCI Open Data Portal: http://cci.esa.int.	§2.3
PF-OPE-0060	The PS shall capitalise on existing European assets through their reuse, particularly Open Source scientific tools and prototype ECV processing systems from prior projects. (heritage)	§2.3
PF-INT-0070	The global Permafrost community shall play an active role in its creation according to given guidelines and advice from a strategic operations team. They shall also give feedback from the implementation to the strategic team.	§2.3
PF-SIZ-0080	The system shall implement a data production line that is sufficiently flexible to continuously update and extend the database (e.g. with data from new sensors or better acquisitions).	§2.3

ID	TITLE	REF
PF-INT-0090	The available data shall be frequently reported and properly disseminated to the interested user communities.	§2.3
PF-OPE-0100	Minimum maintenance and cost	§2.3
PF-FUN-1020	The Products shall be uniquely identified.	§5.2
PF-FUN-1030	The PS shall store data in a structured way using type, revision, date.	§5.2
PF-FUN-1040	If input data is retrieved directly from a third party ground segment, the PS has to ensure that links are maintained and functionality is regularly checked.	§4.3
PF_FUN-1050	The PS shall be able to reprocess also parts of the products.	§2.4, §5.2
PF-FUN-1060	The PS shall be able to do partial processing.	§2.4, §5.2
PF-FUN-1070	The PS shall be flexible in form of delivery for maps and data.	§2.4, §5.2
PF-FUN-1080	The PS shall include proper Product Description	§4.4
PF-SIZ-2010	The PS shall be able to do long time series processing in due time.	§5.3
PF-SIZ-2020	The PS shall be able to do reprocessing in due time.	§5.3
PF-SIZ-2030	The PS shall be able to run on the available hardware infrastructure	§2.7
PF-SIZ-2040	The PS shall have sufficient space for output data	§5.3
PF-SIZ-2050	The PS shall have sufficient space for auxiliary data	§5.3
PF-SIZ-2060	The PS shall have sufficient space for input data	§5.3
PF-INT-3010	The PS shall have the capability and interfaces to extend for future adaptations.	§5.3
PF-INT-3020	The PS shall have or use a long-term storage for its products	§2.4
PF-INT-3030	The PS shall have a self-standing documentation	§4.4

ID	TITLE	REF
PF-OPE-4010	Development of the PS shall be under version control.	§5.1
PF-OPE-4020	The system should be decoupled from the research	§5.1
PF-OPE-4030	Development of the system shall be based on the the user requirements, the selected algorithms and the validation protocols used to generate the baseline products for the worldwide glacier inventory.	§2.2
PF-OPE-4040	The PS development shall be overseen by a science team that drives the development process.	§3.1
PF-OPE-4050	Each PS installation includes a set of test tools, data and benchmark data to test PS integrity (end-to-end, interfaces)	§5.1
PF-OPE-4060	If a module is based on a prototype, the prototype state has to be frozen until it is implemented.	§5.1
PF-OPE-6611	The verification is regarded as successful, when all tests agree within TBD limits. Hashes are to be preferred where applicable.	§3.3
PF-FUN-5010	The operational processor shall not overwrite existing data. Versioning shall be used instead.	§5.2
PF-RAM-5020	The system developed shall be detailed as a separate self-standing document providing an overview of the system and its components, functionality of the system and its subsystems, inputs, outputs, resource key interfaces, and resource requirements.	§5.1
PF-FUN-5030	Verification of the correct implementation of the prototype system against the algorithms developed in Task 2 is a fundamental part of the process.	§2.3
PF-RAM-5040	The verification shall be documented in a Verification Report. It shall contain the chosen approach and the justification, the selected verification data set and the verification results.	§2.3
PF-RAM-5050	The PS shall provide means against data loss of its input / output products.	§2.3
PF-FUN-5060	All data stored in the system shall be available for the long-term (at least 15 years).	§2.3