



climate change initiative

European Space Agency

System Specification Document Version 1 (SSDv1)



glaciers
cci

Prepared by: Glaciers_cci consortium
Contract: 4000101778/10/I-AM
Name: Glaciers_cci-D5.3_SSDv1
Version: 0.8
Date: 17.04. 2014

Contact:
Frank Paul
Department of Geography
University of Zurich
frank.paul@geo.uzh.ch

Technical Officer:
Stephen Plummer
ESA ESRIN



University of
Zurich^{UZH}



UNIVERSITY
OF OSLO



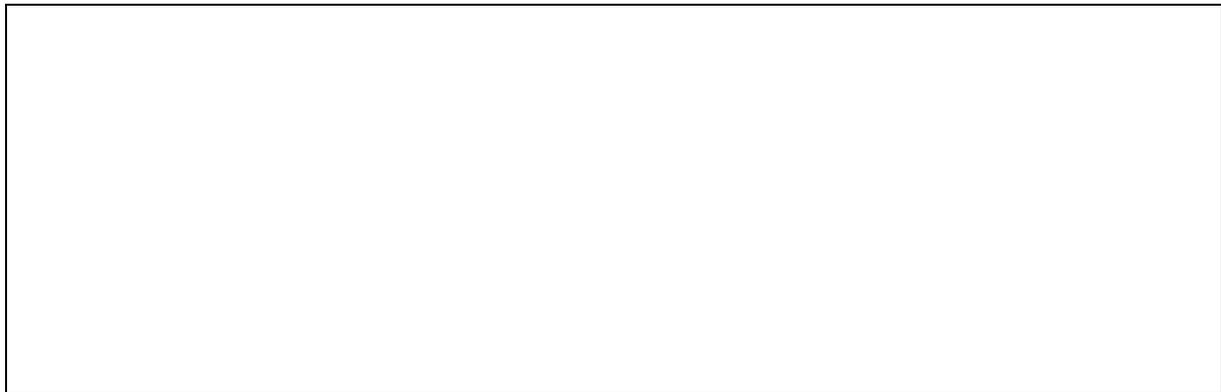
University of
BRISTOL



UNIVERSITY OF LEEDS

Document status sheet

Version	Date	Changes	Approval
0.1	13.02. 2013	First draft	
0.2	03.04. 2013	Extended draft considering all input deliverables	
0.3	06.05. 2013	Feedback from consortium included	
0.4	30.08. 2013	Feedback from TO included	
0.5	19.11. 2013	Missing sections implemented	
0.6	20.12. 2013	Consortium feedback integrated	
0.7	17.01. 2013	Final adjustments	
0.8	17.04. 2014	Feedback from TO included	



The work described in this report was done under ESA contract 4000101778/10/I-AM. Responsibility for the contents resides with the authors that prepared it.

Author team:

Andreas Wiesmann, Othmar Frey, Tazio Strozzi (Gamma), Frank Paul (GIUZ), Andreas Käab, Christopher Nuth (GUIO), Thomas Nagler, Helmut Rott, Kilian Scharrer (Enveo), Francesca Ticconi, Andrew Shepherd (SEEL)

Glaciers_cci Technical Officer at ESA:
Stephen Plummer

Table of Contents

EXECUTIVE SUMMARY	4
1. INTRODUCTION	5
1.1 PURPOSE AND SCOPE	5
1.2 APPLICABLE AND REFERENCE DOCUMENTS	5
1.3 DOCUMENT OVERVIEW	6
2. OVERVIEW OF THE GLACIERS_CCI PROCESSING SYSTEM	7
2.1 PURPOSE	7
2.2 CONTEXT	7
2.3 REQUIREMENTS	8
2.4 FUNDAMENTAL OPERATIONS	8
2.5 THE MODULES OF THE PS	9
2.6 HIGH-LEVEL ARCHITECTURE OF THE PS	10
2.7 MAJOR CONSTRAINTS	11
3. OPERATIONAL SCENARIOS.....	12
3.1 ROLES.....	12
3.2 USER INFORMATION AND DATA ACCESS	12
3.3 PROCESSING.....	13
3.3 VALIDATION	14
3.5 ALGORITHM IMPROVEMENT	14
4. INFRASTRUCTURE	16
4.1 DISTRIBUTED OR CENTRAL	16
4.2 CLUSTER OR CLOUD	16
4.3 SHARING WITH OTHER ECVs	17
4.4 RECOMMENDATIONS	17
5. FUNCTIONAL DESIGN	18
5.1 USER INTERFACE / SERVICES FOR THE USER	18
5.2 DATA PROCESSORS	20
5.3 DATA MANAGEMENT	20
5.4 PROCESSING MANAGEMENT	21
5.5 UPDATE MANAGEMENT	21
5.6 DOCUMENTATION MANAGEMENT	22
6. DEVELOPMENT, LIFE CYCLE, PERFORMANCE, COST	23
6.1 DEVELOPMENT	23
6.2 LIFE CYCLE	24
6.3 PERFORMANCE	25
6.4 COST	28
7. REQUIREMENTS TRACEABILITY	30
ACRONYMS	33

Executive Summary

The Glaciers_cci System Specification Document (SSD) specifies the characteristics of an operational ECV production system from a developers point of view. It reflects on the requirements of the Glaciers_cci System Requirements Document (SRD) and thus takes a look on the Glaciers_cci processing system in a potential second phase with potential for the time beyond.

The Glaciers_cci processing system covers different processing modules that are technically self-standing. The modules rely on different input data and processing steps and the available prototype implementations are on different implementation levels. Furthermore the products have different production intervals and varying needs on the infrastructure.

To best address the modular structure of the processing system, a distributed processing system is proposed to produce the glacier area, elevation change and velocity products in the next development phase. The methodology is based on the existing prototypes.

Different implementation approaches are foreseen for the products. Glacier outlines are created by a community effort with the analysts experience being key to a high-quality product. While the Glaciers_cci EO-team will follow the workflow described in the SPD and IODD / DPM, other approaches exist and are applied by the community (see PVASR). In a next step, the processor should be developed towards a higher degree of automation (also considering Landsat 8 and Sentinel-2 data), and illustrated guidelines should be produced for and by the community to increase product consistency.

For the glacier elevation change (DEM differencing), a toolbox performing the co-registration of two DEMs is foreseen to aid in the consistent compilation of such data from the various satellites and techniques used. For the equivalent product from altimetry sensors, the prototype will be used to process the full archive of RA and ICESat altimeter data. In addition, the system can be used to process future Sentinel-3 and ICESat-2 data and will potentially be extended to process Cryosat 2 data (which has rather different requirements).

For glacier velocity, an on-demand service is foreseen. The user shall be able to decide himself from which image pairs (and consequently where and for which period) velocity information shall be produced and made available. The processing algorithms will be made available to the community in a special package, increasing the consistency in product generation. The data distribution is foreseen through the web portal of Glaciers_cci (as an intermediate solution), as well as via the GLIMS (glacier area and velocity) and WGMS databases (elevation change products).

1. Introduction

1.1 Purpose and Scope

This document is deliverable 5.3, the version 1 of the System Specification Document (SSDv1) of the Glaciers_cci project requested in the Statement of Work (SoW) [AD 1]. The SSDv1 incorporates the requirements described in the System Requirements Document (SRD) [AD 4] and specifies the characteristics of an operational ECV production system from a developer's point of view.

The system design is based on experience with the prototype developed and applied in phase 1 of Glaciers_cci. The prototype is documented in the System Prototype Description (SPD) [RD 6], the Input and Output Data Definition (IODD) [RD 4], and the Detailed Processing Model (DPM) [RD 5]. Important information about the system can also be found in the System Verification Report [RD 7]. While the IODD and the DPM will also be applicable to Phase 2 of CCI with their main content, the processing system (PS) will be further developed based on the SPD to meet the specifications provided in this SSD. The use of the prototype within the Glaciers_cci PS is an additional topic of this SSD.

According to the SoW [AD 1], the SSD shall include:

- a specification of the purpose of an operational ECV production system and its intended use
- an overview of the context of the system, defining all significant interfaces among system components and crossing the system's boundaries
- a definition of the fundamental operations to be performed within the system to accept and process the inputs and to process and generate the outputs
- a description of major constraints of the system
- a description of operational scenarios for the system including data sources, valid ranges of values, timing considerations, operator requirements, and special interfaces
- specification of the environmental characteristics of where the system will be installed
- specification of the growth, expansion, and capability characteristics of the system.
- description of the life cycle sustainment activities to be executed during the life cycle of the system
- a trade-off analysis of different technical solutions for a system concept taking into the account the requirements in the SRD, the prototype development described in the SPD and cost, performance, and operational constraints

1.2 Applicable and Reference Documents

List of Applicable Documents

ID	Title	Issue	Date
[AD 1]	ESA Climate Change Initiative Statement of Work	1.4	09.11.2009
[AD 2]	Glaciers_cci proposal	revised	26.07. 2010
[AD 3]	Glaciers_cci URD	1.1	11.10. 2011
[AD 4]	Glaciers_cci SRD	1.0	12.07. 2012
[AD 5]	Glaciers_cci SPD	0.5	13.02. 2013

List of Referenced Documents

ID	Title	Issue	Date
[RD 1]	ESA CCI Project Guidelines, EOP-DTEX-EOPS-TN-10-0002	1.0	5.11. 2010
[RD 2]	Glaciers_cci DARD	1.0	20.11. 2011
[RD 3]	Glaciers_cci ATBDv2	1.1	13.02. 2013
[RD 4]	Glaciers_cci IODDv1	0.5	21.11. 2012
[RD 5]	Glaciers_cci DPMv1	0.5	8.11. 2012
[RD 6]	Glaciers_cci PSD	1.0	20.11. 2011
[RD 7]	Glaciers_cci SVR	0.4	18.03. 2013

1.3 Document Overview

This document is organised along the requirements given in the SoW as follows: Section 2 gives an overview of the Glaciers_cci processing system. It describes its purpose and intended use as well as the main requirements, functions and components. Section 3 shows the main operational scenarios. Section 4 discusses the necessary infrastructure and Section 5 highlights the functional design from different perspectives, the users, system operators and developers view. Section 6 summarises information about the system life cycle design, implementation and maintenance costs and performance. Finally Section 7 connects this document to the SRD by tracing the system requirements of the SRD to sections of this document.

2. Overview of the Glaciers_cci processing system

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 Purpose

The Glaciers_cci PS generates products and supports the process of algorithm improvement, reprocessing and validation. It provides products and services to the glacier and hydrology community supporting their climate change impact assessment over a wide range of spatial scales. The PS is used by the Glaciers_cci consortium but can also be applied by others, e.g. those processing the EO data and contributing them to GLIMS. The PS is specified to provide glacier related products such as glacier outlines, elevation changes, and velocity 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 and validation reports.

The products are hosted by the two key science bodies WGMS and GLIMS within the Global Terrestrial Network for Glaciers (GTN-G) and will mainly be used by the glaciological and hydrological modelling communities in their applications. The creation of a globally complete and detailed glacier inventory was a major action item (T.2.1) in the GCOS implementation plan that has largely been achieved in Phase 1 with the compilation of the Randolph Glacier Inventory (RGI). This product has already served the UNFCCC by providing baseline data for various modelling approaches (e.g. their contribution to sea-level) that have been included in the fifth assessment report of the IPCC (AR5). The PS is designed for further improving and extending the existing datasets and establishing a monitoring service (for all glaciers and the three products) based on the distributed processing approach.

2.2 Context

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 (Fig. 1). There are interfaces to the different user communities, which receive products and can provide feedback. The user communities comprise the WGMS and GLIMS communities, hydrological modelling communities and the other ECVs. Another interface is with the EO data providers. Depending on the module, EO data are obtained from the providers at Level 1 or 2 and are ingested into the PS [RD 4]. Another interface is towards third-party sources to receive ancillary and validation datasets.

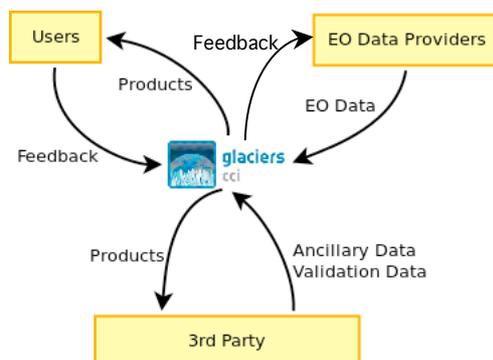


Fig. 1: Principle context of Glaciers_cci processing system.

2.3 Requirements

System requirements are compiled in the Glaciers_cci SRD [AD 4]. 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.

High level requirements are to generate the Glaciers_cci products (GL-FUN-0010 , GL-FUN-0060), in collaboration with the GLIMS community (GL-INT-0030). The processing line shall be well defined and flexible for future updates and adaptations (better algorithms, new input data) (GL-FUN-0040). The available data shall be frequently reported and properly disseminated to the interested user communities (GL-INT-0050).

The main scenario is different for the different products and linked modules. While for glacier area the focus has been on the production of a comprehensive global glacier inventory, the other products are addressing more regional needs. All products will benefit from future available sensors such as the Sentinels and therefore will need to incorporate this new data in the development. Already existing historical data can be processed with the currently available processors. The functional scope of the system is not restricted to the processing, reprocessing, validation and improvement cycle, though this is its main purpose. Also functions to make output products and documentation available to users are in the scope.

2.4 Fundamental Operations

Requirements in this section are:

- GL-INT-1010 Long-term storage
- GL-INT-1140 Self-standing documentation
- GL-FUN-2210 Reprocess products
- GL-FUN-2211 Reprocess with new or so far unprocessed data
- GL-OPE-6410 Glacier Inventory
- GL-OPE-6411 Complete set of parameters

The PS system hosts input data, performs pre-processing, classification processing, supports validation, and serves the output to users. 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 support test and development of new algorithms and continuous improvement.

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

- Production
- Dissemination
- Life Cycle Management

In the following we will discuss the fundamental operations of the PS with regard to these 3 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;
5. Comparison with reference data;
6. Documentation of the processing using meta-data;
7. Ingestion of new input data and auxiliary data.

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 glaciological and hydrological but also on the climate modelling community. The products are distributed through existing and in development platforms of WGMS (<http://www.wgms.ch>) and GLIMS (<http://www.glims.org/>). Functions are:

- Online Data Access
- Data and Processing Information
- Project Information
- Long-term Preservation
- Feedback Handling
- Validation Support
- For velocity a processing initiation function must exist

Life Cycle Management is crucial 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 (e.g. atmospheric windows for optical sensors, wavelength of LIDAR sensors). 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

Unlike most of the other CCI projects, Glaciers_cci produces different products that rely on completely different processing chains in terms of input and ancillary data and the processor. The processing chains for each product are organized in modules that are part of the PS. In [AD 5] the following modules are part of the PS:

1. Glacier area
2. Elevation change: DEM Differencing
3. Elevation change: Altimetry
4. Velocity from microwave sensors
5. Velocity from optical sensors

The ESA strategy is to develop the most efficient, consistent and sustainable system addressing the needs of the corresponding community. We 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 have to address both levels in the SSD, the module level and the system level.

2.6 High-level architecture of the PS

The functions and modules listed in the previous section are implemented as functional components. In this section we show the high-level architecture of Glaciers_cci on this subsystem level. It follows the diagram shown in Fig. 1 of [AD4] and repeated below as Fig. 2.1 Uptake is either through the CCI Common Data Access portal or the web-based services of satellite data providers (e.g. EOLI-SA, GLOVIS) and those listed in [RD-2]. Data are then processed in the product generation modules and distributed through the CCI Common Basic Services that are in the case of Glaciers_cci pointing to the GLIMS/WGMS databases. Within this document we focus on the product generation modules (see Section 3.2). Here 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.

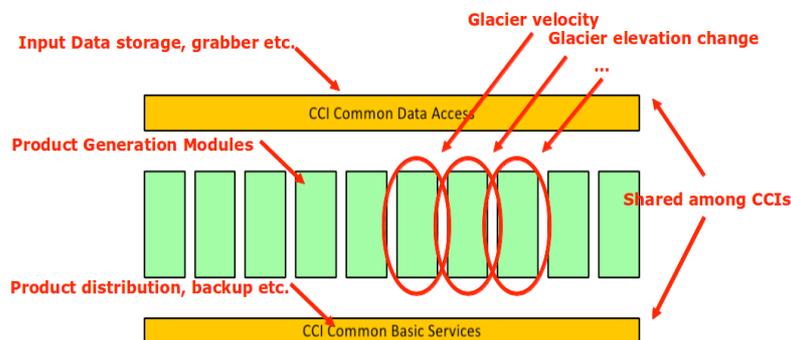


Fig. 2.1: Common data access and de-centralized production environments (from [AD-4]).

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 need another archive accessible online for data download. Finally the long-term archive is optimised for reliable long-term storage of the data and all representation information required to use it and is part of the data stewardship.

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. A complete replacement is not foreseen for some of the products (e.g. area), as the quality of the derived products is largely independent of the way of processing the data (apart from the post-processing).

User services subsystem

This subsystem consists of at least three components: web presentation, data access by users, and a catalogue. The web presentation includes a user forum and issue tracking. The catalogue is used for product search and metadata access. In addition, users can get read access on the processor repository, documentation and validation.

Data Stewardship

Its main function is the long-term data preservation and bulk data provision on request. Depending on whether this function is provided externally or not, a component for the preservation of inputs and outputs may have to be foreseen within the system.

2.7 Major constraints

Here we have to distinguish between constraints related to the technical feasibility and the status of the implementation. We will indicate areas where we see potential to change the constraint level through development. Potential major constraints are:

- Performance of the system in terms of processing time and/or data needs
- System Portability
- Input Data availability
- System development status
- Funding

2.7.1 Performance of the system in terms of processing time and/or data needs

The processing time is mainly critical for the glacier area product because there are constraints on the reprocessing time of the glacier area product. By design, the PS module is distributable so that the processing can be parallelized. Development of the processor towards an automated machine-based processing module will help reducing the production time and allow large scale reprocessing to test improved algorithms or to repeat the inventory more frequently. The data volume needs are small at this stage and are not considered as a constraint.

2.7.2 System portability

The glacier area module needs to be portable between different operators. Consequently constraints to be considered are the operator skills and reliability, operator infrastructure (hardware and software). The toolboxes must be designed portable to work on common hard and software platforms. The on demand systems should also be designed portable so that they can be ported easily on different servers (e.g. as a virtual machine).

2.7.3 Input data availability

Input data availability has two aspects, access and access time. For glacier area the data to be processed (Landsat, Sentinel) are freely available and an up-to-date internet connection for the operator is sufficient. For the velocity product the access (and acquisition) of very high resolution SAR data (e.g. Terra_SAR-X) is an issue as these sensors are commercial and scientific data access is restricted/limited.

2.7.4 System development status

The current PS modules are tested. However, they need more or less operator intervention for data injection, export and decision making. Nominal effort is necessary to raise the implementation level to a reasonable degree of autonomy for large scale processing.

2.7.5 Funding

Glacier_cci Phase 2 will provide funding to develop the system towards a sustainable one. this includes the further development of the PS modules and product generation. For the operation of the user services and on-demand processing, a collaboration with GLIMS and WGMS is planned. At the moment the funding situation of GLIMS to run these services is open.

3. Operational scenarios

This section covers the operational scenarios, namely the user data and documentation access, processing and validation of the CDR, and algorithm improvement. The roles are:

- User (GL-INT-0030)
- Developer (Development Team) (GL-OPE-6430)
- Operator
- Validator (Validation Team)

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:

- GLIMS/WGMS users
 - interested in best existing glacier products
 - skilled in glaciology, provide feedback and proposals
 - data format compatible with their communities
- Development team
 - mandated to run Glaciers_cci
 - in dialogue with users
 - further development of the PS
 - issues product versions
- External validation experts
 - international community
 - support development team
 - provide local expertise
 - feedback on the products
- ESA
 - project supervision

3.2 User information and data access

Users access the Glaciers_cci data products using the GLIMS/WGMS web sites as well as the Glaciers_cci homepage (GL-INT-0050, GL-INT-5020, GL-INT-5010). These are also the entry points for meta-data, documentation (GL-INT-1140), a catalogue, data services, and the DEM-differencing toolbox (GL-FUN-0010). The website provides also a forum (GL-FUN-0160) and an issue tracking system. This approach is common for all Glaciers_cci products and needs to be designed and implemented in the future. For the velocity modules, the website also provides the possibility to select a satellite scene pair that will be used to produce the velocity product based on a process initiation form. Important to mention is the use of the GLIMS and WGMS mailing lists for dissemination/announcements.

- The development team announces a new version of the Glaciers_cci product on the Glaciers_cci website and WGMS and GLIMS mailing lists.
- A user accesses the Glaciers_cci web site as an entry point and gets general information on Glaciers_cci. It also provides links to software, services and resources.

- The user follows links to the Glaciers_cci products (potentially hosted on GLIMS and WGMS servers) and views them online.
- The user optionally registers himself to mailing lists.
- The user uses the catalogue to find Glaciers_cci products. The catalogue also provides metadata and quicklooks.
- The user downloads product documentation (e.g. the Product User Guide).
- The user enters the Glaciers_cci forum and reads contributions by other users and answers from the development team. The user adds a question to the forum that is answered by the development team as well as other users.

3.3 Processing

Here processing covers all steps from data retrieval, pre-processing, classification, and product generation. Data retrieval covers extraction of the input (and auxiliary) data from the data provider, data format conversion for transfer to the PS, and geoid corrections where applicable (e.g. for ICESat data). Pre-processing covers all steps necessary for the later processing such as re-projection and coordinate transformations, mosaicking, etc. Finally, product generation includes the calculation of values in the selected region, filtering, output format generation, meta-data production, etc.

As mentioned above, the processing for the different Glaciers_cci products is done in specific modules. Common for all modules is that the design is based on the prototypes described in the SPD [AD 5] and therefore not repeated here. However, it has to be noted that at this stage the prototypes cover mainly the core of the processing, the classification part, while the data retrieval and the standardised product generation is done manually or is not implemented yet. The related development effort and strategy is discussed in Section 6.1.

For glacier area and elevation change (altimetry) the processor modules process a set of input data selected by the analyst to produce the corresponding products that are made available to the community through the product archive. The processing is initiated and controlled by an operator.

The elevation change (DEM differencing) module is a toolbox that can be downloaded from the website and run on the clients computer using data the user already owns (GL-FUN-0010). This allows the user to produce products also from licensed elevation data that cannot be shared or used outside his facility. This approach also takes into account the large variability in characteristics of available elevation data that requires user / expert interaction at various processing stages. Provision of results will differ on a case-by-case basis.

The velocity modules follow a different strategy. To allow interested clients to choose the time span of interest for a glacier/region the production is initiated at the client site by selecting an image pair from a candidate list. After the processor is applied, the resulting products are stored in the respective product archives (GLIMS), to be also accessible to other clients.

Generalized requirements to run the processing workflow as an operator:

- Receive temporal coverage, processor versions and processor parameters from the development team
- Provide the results to the development team for quality assessment
- Report on the production

- Re-runs certain parts of the production if necessary to fix data issues
- Manage data storage
- Ensures the processor software is frozen

During CCI, the EO team is in charge of the processor development and local implementation while the system engineers are taking care for a sustainable solution (e.g. taking care that algorithms work across platforms). It also implements improvements such as new versions of processors and if necessary a modified workflow. The team tests and validates new algorithms and decides about upgrades to be implemented in the PS. Activities contributing to a processor upgrade are:

- Identification of new requirements
- Resulting in a new or modified processor
- Reprocessing
- Validation and comparison with the former product versions

3.4 Validation

The validation team is in charge of the Glaciers_cci system and product validation (GL-RAM-3110). The validation is done separately for the different PS modules. It covers:

- System (module) validation (after an upgrade or new installation)
- Product validation and quality control
- In case of multi-sensor usage cross-comparison

The PS provides tools that facilitate these tasks such as

- Benchmark test data (GL-OPE-6611, GL-FUN-6710)
- Test tools (GL-OPE-6610)
- Verification Report (GL-RAM-6612)
- Feedback functionality

External validation experts use the test tools and their knowledge of the region and help build up a benchmark test dataset. This has yet to be implemented in the already existing website. The validation team also compares the Glaciers_cci products with validation data.

3.5 Algorithm improvement

The development team decides about features or processes to be improved in order to meet user requirements in agreement with ESA. 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 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.

4. Infrastructure

There are some fundamental decisions that determine how the PS and the corresponding infrastructure will develop and look like:

- To what extent shall the PS use and build on the module prototypes?
- To what extent shall the PS be unified or distributed?
- Which PS and module functions shall be shared among the modules?
- Shall the system be implemented in an existing infrastructure or an existing data centre?
- Shall the system run completely virtualised in a cloud?
- Shall the middleware be used, and which one?
- Which functions and subsystems are candidates for sharing with other ECVs?

This section describes the trade-off for these questions. The discussion may not be complete in its alternatives but tries to discuss solutions that are viable. Some of the questions above depend on each other such as an existing infrastructure of a dedicated data centre will exclude the deployment in a cloud. The following subsections describe for each alternative how it will look like, what has to be done in order to realise it, and the pros and cons.

4.1 Distributed or central

The Glaciers_cci PS is based on selfstanding modules addressing to some degree different communities. The prototype PS is based on distributed processing. Consequently a distributed processing system can be directly derived from the prototype. Advantage of the distributed system is also that the corresponding experts have local access to the processing and outputs. This is beneficial for those modules that require expert decisions in the production or are based on a community effort in the production such as the current version of the glacier area module.

The advantage of a central infrastructure is that it minimises data transfer and allows to optimise computer resources. However, the development effort to convert the distributed nature of the Glaciers_cci PS to a centralised system is nominal (though currently not foreseen for the distributed processing of the glacier area product). Currently the PS is developed as a distributed system (cluster solution) but the development aims at more operator independence, in particular for velocity.

4.2 Cluster or cloud

Advantage of the cluster is that no overhead is introduced and the infrastructure can be optimized for the processing tasks. Advantage of a cloud solution is that flexible capacity can be rented.

The cluster requires that the hardware to process the data is physically available and the cluster computers are connected to the data sources through a reasonable data link. The scalability is given by procuring more hardware if further resources are needed. That makes investments necessary in hardware and support labour necessary. An advantage of the cluster solution is that it better fits the operational level of the modules.

The cloud requires that the data is in the cloud, too, which is extremely expensive for large amounts of EO data. A cloud infrastructure abstracts from the physical hardware and provides

services like storage and virtual machines. Cloud infrastructures are available commercially from different providers (Amazon as an example). They provide computing power in form of virtual machines, selectable classes regarding CPU power and RAM (e.g. Amazon EC2), storage as virtual disks, mountable on virtual machines (e.g. Amazon EBS), network storage accessible via the Web (e.g. Amazon S3). Virtual disks are more expensive (magnitude of 100 Euros per TB and month) than network storage (magnitude of 50 Euros per TB and month) such that network storage will be used for the processing archive. Computing power is either paid on an hourly basis or with an advanced payment and a smaller hourly rate (magnitude of 2500 Euros per year for a 16GB/8 core machine). (All numbers are rough estimations to understand the relations). In order to use the flexibility a virtual machine template for the PS must be stored in network storage and can be launched on demand.

Practical experience with cluster solutions and the operational level of the current module implementations speaks for a cluster solution, i.e. decentralized or distributed computing using a standard set of algorithms and quality control mechanisms for the operator-based stages.

4.3 Sharing with other ECVs

Advantage of sharing is the re-use of input products and the option for the respective use of outputs as auxiliary data or for inter-comparison. Disadvantage of sharing may be that the computing platform may not be available at any time, and that a data policy other than free and open access is more difficult to implement, e.g. to protect third party input data.

Arguments for sharing are

- Saves space for inputs used by several ECVs
- Operations and maintenance required only once for the shared functions
- Synergy by immediate availability of outputs and data exchange among ECVs,
- Sharing provides functions to ECVs with less infrastructure

But

- Harmonisation is work, some prototypes are already derivatives of existing elaborated systems
- Dependencies increase the risk of failure
- User communities and the interaction with users are different for different ECVs. A common user service may be not flexible enough
- Data with licence restrictions can not be shared

The discussion above speaks for an independent implementation approach. There is not much synergy in joint implementations and the focus at the current stage should be on improving the operational aspects and merging of the Glaciers_cci modules.

4.4 Recommendations

The following recommendations are given:

- aim for a distributed solution
- aim for a cluster solution
- focus on the Glaciers_cci community needs and merging and developing of the Glaciers_cci modules

5. Functional design

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

5.1 User interface / Services for the user

The Glaciers_cci website is the entry point for users and external evaluators to assess information, meta-data and data services. The provided user services are the functions and interfaces that a user needs, to be able to interact with the PS. User services provide information on different levels. Besides the data services the web site shall act as a central starting point to explore achievements, data and other resources available to the users. It shall provide basic information for users not familiar with the project, in depth access to resources for the users of the data with static and dynamic content, and finally access to administrative pages and tools for the system operators.

Requirements addressed by this section are:

- GL-FUN-0050 Report available data
- GL-FUN-0160 User feedback functionality
- GL-FUN-5010 Website
- GL-FUN-5020 Data access through internet

The Glaciers_cci user services includes the functional aspects:

- Access to the data generated by the PS
- Access to tools provided by the PS
- Access to the document archive
- Access to the velocity processing initiation

While the Glaciers_cci website is an entry point for all data while the project is running (and at least three years beyond), the data and services shall make use of existing services as much as possible. At this stage it is foreseen to host and distribute the Glaciers_cci products through the GLIMS (Fig. 5.1) and WGMS data archives and websites (Fig. 5.2). It has to be defined to what degree also the other services for the users can be hosted at these places. Also feedback functionality and the documentation archive shall make use of the community resources as much as possible. The portal is preferably implemented using a Content Management System (CMS). Some of the desired functionality is directly available in the CMS portal software (forum, documentation management, etc.) while other components can be stand-alone services or even remote web-resources. Modern FOSS CMS provide typically:

- Separation of content and layout, corporate web site layout using CSS
- Support for authoring by separated creation and publishing, dedicated approval step
- Management of links independent from web pages, links to services and data access
- User rights management, LDAP interface (see user management below)
- Document database
- News feeds
- Forum

All of the distributed functional components are connected using links from the central portal. The other user-accessible services are catalogues, data access via FTP and/or other protocols

and the version control system (software repository). The LDAP-based user management is necessary for access authentication allowing the same credentials for all provided user services. Both communities GLIMS and WGMS take use of mailing lists for communication that might also be beneficial for the Glaciers_cci PS.

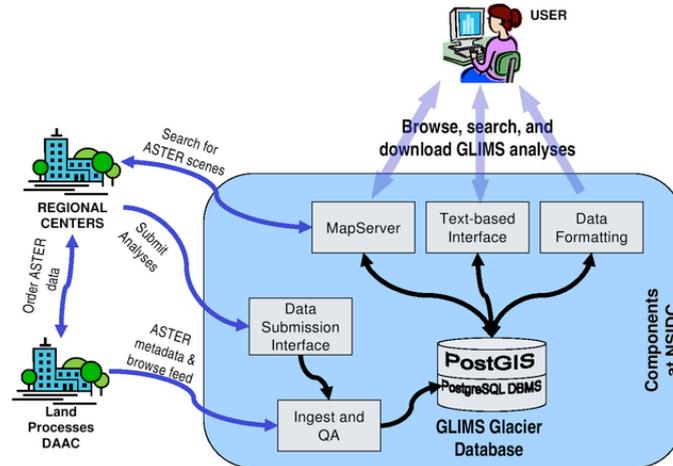


Figure 5.1: The GLIMS Glacier Database Architecture and Environment for glacier outlines and velocity data with ASTER data as an example for EO input data.

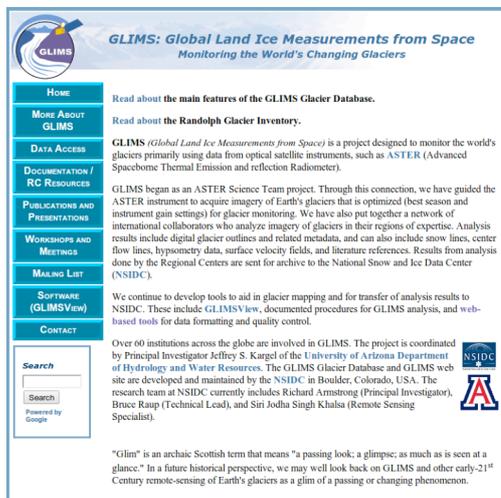


Figure 5.2: The entry pages of the GLIMS website www.glims.org (left) and the WGMS website www.wgms.ch (right).

User management is an important aspect and is based on roles and necessary access rights:

- Restrict administrative access
- Access Bandwidth/ Traffic shaping
- Identify authors e.g. in forums, mailing lists etc.
- Upload access to a provider of data

The role-based access control is based on user roles with different responsibilities and use cases as discussed before:

- GLIMS and WGMS community access the complete product archive as data contributors as well as a user group. They will also provide functional feedback.
- Other scientific users interested in parts of or the complete dataset. They are also interested in the algorithms, inter-comparison with other sensors, tools. Feedback is expected from them, too.
- Interested public seeking product examples, information about the project.
- The development team comprises algorithm and system developers and validators. This role has the same interests as the science community and has furthermore the mandate to develop the PS.
- The system operator is responsible to maintain the servers, manage the forum/ mailing lists, do user and web content management, maintain the software, tools and data services provided by the user services, delegation of questions to the appropriate experts. They will receive support from additional specialists when necessary.

5.2 Data processors

The data processors cover the necessary tools to produce the different Glaciers_cci products. The Glaciers_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:

- GL-FUN-0040 Flexible production
- GL-FUN-6020 Data overwrite

Requirements:

Component	Purpose	Content	Implementation
Data processor Glacier Area	Generate L2/L3 Glaciers_cci Area Product	Input Data	Existing processor, derived automated processor
Data Processor Elevation Change Altimetry	Generate L2/L3 Glaciers_cci Elevation Change Product	Input Data	Extension of existing processor
Data Processor Elevation Change DEM differencing	Generate L2/L3 Glaciers_cci Elevation Change Product	Input Data	Toolbox
Velocity microwave	Generate L2/L3 Glaciers_cci Velocity Product	Input Data	Extension of existing processor
Velocity optical	Generate L2/L3 Glaciers_cci Velocity Product	Input Data	Extension of existing processor

5.3 Data management

The data management includes the components, archive and an inventory. Furthermore a production control component is needed.

Requirements:

- GL-FUN-0040 Flexible production
- GL-FUN-1011 External data connection check
- GL-FUN-1020 Unique identifier
- GL-FUN-1030 Store data in structured way
- GL-FUN-1050 Data loss
- GL-FUN-2120 Data format
- GL-RAM-2230 Traceability

Component	Purpose	Content	Implementation
Archive	Stores products, meta-data, processing logs, auxiliary data, validation, processor software bundles, in a structured directory tree, makes them available	Products Meta-data, Logs Validation Data Software	WGMS/GLIMS database
Inventory	Handles product entries and collections, attributes of products like QA information, extensional collections (lists) of product entries and intentional collections (logical selection criteria like type and time)	Product entries	WGMS/GLIMS database
Access	Provides means to select and access products	Products	WGMS/GLIMS website

5.4 Processing management

Production requires automated workflows and requests with status and reporting to the operators. It includes production and quality control steps. The overall workflow is described in more detail in Section 3. The PS is being developed from the prototype processors which are described in Section 5.2. The production control is a function that initiates and controls data processing activities of the system. Data product quality checks are necessary. The quality check function supports automated and operator-performed quality checks. Input products are screened automatically for product consistency (format, file size, content). Corrupted products are marked in the inventory and are removed from the processing storage. Output products are quality checked and validated before they are stored in the product archive.

Requirements addressed by this section are:

- GL-RAM-2160 Configurable
- GL-RAM-3110 Quality Control during processing
- GL-FUN-4010 PS has logging mechanism
- GL-RAM-4040 Event reporting
- GL-INT-4110 Processing Status
- GL-RAM-5520 Data import logging

Component	Purpose	Implementation
Production control	Handles production requests, manages workflows, manages resources, processing capacity and storage space	Extension of processor modules
Quality check	Checks product integrity and content, adds quality attributes to inventory entries, generates quicklooks, provides tools for systematic visual screening and for product inspection	Product Validation Tool, WGMS/GLIMS website

5.5 Update management

Continuous improvement is an important aspect in the CCI projects. To ensure a transparent process Software Modularity, Software Version Concept, Version Control, 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:

- GL-OPE-6330 Development under version control
- GL-OPE-6340 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 parameterised and that generates a (higher level) output product of a certain type from one or several input products of certain types. A PS module consist of the sequential call of processors. Each processor has its own version information. Parameters are usually provided as command line arguments, environment variables or as information in a parameter file. Feedback is received by a return code, messages on stdout/stderr and in log files.

Component	Purpose	Content	Implementation
Software Repository	Stores all versions of the processor code in a transparent way, with branches, authorship	Code	Subversion Redmine Tools

5.6 Documentation management

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.

Requirements addressed by this section are:

- GL-FUN-1140 Product Description
- GL-RAM-6420 Self-standing documentation

Component	Purpose	Content	Implementation
Documentation Management	Stores documentation in a structured and transparent way	Documentation	WGMS/GLIMS platform

The PS documentation includes requirement documents, design and interface control documents, test documents, manuals, and maintenance information. Glaciers_cci deliverables to name here are ATBD and the PUG. The SRD and SSD define requirements and design of the system. They most likely will be updated during CCI phase 2.

6. Development, life cycle, performance, cost

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. Historic data from other sensors (e.g. Korona/Hexagon, MSS, JERS) will be processed on a regional basis as they do not provide global coverage.

Requirements addressed by this section are:

- GL-FUN-0110 Sentinel-1/2/3, LDCM
- GL-FUN-0120 TanDEM-X, new global DEM
- GL-OPE-6620 Freeze prototype
- GL-OPE-6320 Minimize maintenance and cost
- GL-OPE-6330 Development under version control
- GL-OPE-6340 Development decoupled from research
- GL-OPE-6410 Development plan
- GL-OPE-6411 Development
- GL-OPE-6430 Science team

6.1 Development

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

Requirements addressed by this section are:

- GL-RAM-2160 Variables
- GL-OPE-6340 Development decoupled from research
- GL-OPE-6620 Freeze prototype
- GL-FUN-6710 Verification of implementation

In the table below we summarise the tools that can be adapted, configured and integrated as well as need to be developed for the PS in the next phase (to be complemented in SSDv1).

Name	Usage	Remarks
Subversion	Version control	FOSS
Redmine	Issue tracking	FOSS
Glacier area processor prototype	Glacier outline production	Based on closed source COTS but implementable for any COTS
	Base for the standardization of data processing	Implementation from scratch needed
DEM differencing processor prototype	Base for DEM differencing toolbox	DEM co-registration tool free/public from GUIO; rest to be developed.

Altimetry processor prototype	Base for altimetry processor	Needs to be further developed to process CS2, S3 and ICESat2 data
Velocity (optical) processor prototype	Base for velocity (optical) processor	Algorithm prototype free from GUIO; refinement and implementation needed.
Velocity (microwave) processor prototype	Base for velocity (microwave) processor	Currently COTS with source code, online version envisaged

6.1.1 Glacier area

The aim of the glacier area product is a complete worldwide inventory of glacier outlines. The prototype covers all necessary processing steps. It is operator and COTS software based. The necessary processing resources are gained exploring community resources by following a distributed processing approach based on a people network. Their work is currently project based and requires better coordination, i.e. data are provided as projects permit. One of the main objectives in Phase 2 will be to make this process community coordinated for regular updating. To address the new upcoming data of the Sentinels, the module prototype is used to implement a processing line that is able to produce glacier outlines on a more regular basis.

6.1.2 Elevation change: DEM Differencing

The aim of the DEM differencing module is to provide a toolbox for the community that allows to produce elevation change products by DEM differencing. The toolbox is based on the module prototype and distributed through the GLIMS/WGMS website.

6.1.3 Elevation change: Altimetry

The aim of the Altimetry module is to process the altimetry data of the full archive of RA and ICESat altimeter data for which the prototype module is built and reprocess the RA dataset using data from the ESA REAPER project once it is available.

In a second step that processor is enhanced to utilise CS2 data and later also S3 and ICESat2 data and to allow cross-calibration of the various sensors.

6.1.4 Velocity from optical and microwave sensors

The aim of the velocity modules is to produce the glacier velocity map from a given radar or optical image pair. The production of velocity products can be started on the Glaciers_cci website by selecting an image pair from the respective archives (integrated by hyperlinks).

6.2 Life cycle

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 (e.g. Landsat 8, Sentinel-1,-2,-3)
- 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 phase 2 of Glaciers_cci it is reasonable to decide for a given hardware/software environment and to keep this constant. But for the longer perspective also renewal of hardware and optional change of software layers must be expected. The PS design is prepared for this by the modularity of its functional components. The last item is not so relevant for Glaciers_cci at the moment as the dependence on other CCI projects is minor.

Requirements:

- GL-FUN-1010 Long-term storage
- GL-FUN-1020 Unique identifier
- GL-FUN-1030 Structured storage
- GL-FUN-2210 Reprocess Products
- GL-FUN-2211 Reprocess Products
- GL-OPE-6330 PS shall be under version control
- GL-RAM-6610 Test tools
- GL-RAM-6611 Verification

6.3 Performance

In the SRD, 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, but might grow to more substantial TB values with new sensors becoming available (e.g. the Sentinels).

There exist requirements on disk space that are modest:

- GL-SIZ-1110 Space for input data
- GL-SIZ-1120 Space for auxiliary data
- GL-SIZ-1130 Space for output data
- GL-FUN-2220 Storage not bottleneck
- GL-FUN-2310 Reprocess within 10 years

A dramatic increase has to be taken into account when moving to a more frequent observation schedule and with the availability of large Sentinel archives for the glacier velocity product.

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

Glacier area

We estimate the number of satellite scenes (Landsat coverage) required to map all glaciers globally at a minimum of 1200 (resulting from a digital intersection of the Landsat footprints with the global glacier coverage). The processing time per scene is about 30 min for the pre-processing (image selection, format conversion, etc.) and 30 min for the main processing (threshold selection, raster-vector conversion, etc.) stage. Post-processing demands per scene are highly variable and depend on the required correction for debris cover, seasonal snow and shadow regions, among others. In a best case only one hour is required, in a worst case about a week (50 hours). A suitable mean value on a global scale is maybe 3 days (25 hours) per scene for a complete analysis (this is just a mean value assuming a normal distribution of the given range). This gives a global reprocessing workload of 30,000 hours or 10.3 years for 365 days with 8 hours or 16.3 effective years (230 working days). This is in good agreement with an earlier calculation (Paul and Kääb, 2005) that has assumed a 10 minute workload per glacier or 30,000 hours for 180,000 glaciers.

Theoretically, 30 persons can produce a global update of the entire inventory in 1/2 year. Hence, in terms of costs for personnel, 15 man years are required. Using a 100 k Euro PDRA salary per year, global reprocessing requires 1.5 million Euros (or 150 k per year if an update is made every 10 years). Additional costs might be required for computer infrastructure (hardware and software licences) which might not exceed 5 k Euro per year (or about 3% of the total).

Storage of the data during processing is part of the hardware and storage requirements for the final dataset (currently about 1 GB for the entire RGI, uncompressed) are negligible compared to the raw data (about 1 GB per scene for Landsat 5/7 TM/ETM+ and 2 for Landsat 8 OLI). So the 1200 scenes mentioned above might need about 1-2 TB of storage space. This capacity is currently held by the respective data repositories (e.g. ESA EOLI, USGS LPDAAC) and must not be provided by the users. Typically not more than 100 scenes are worked with at one place, requiring about 100-200 GB of disk space. The final product (in vector format) is much smaller, about 300 MB for the entire uncompressed inventory.

The required auxiliary datasets such as DEMs will require additional storage space. Depending on their spatial resolution (25 m to 100 m) this might vary by an order of magnitude (factor 10) for the same region. Roughly, the required storage space is in the same order of magnitude as for the satellite data (including processed products such as slope, aspect, sine and cosine grids in real format). Costs for raw data (Landsat, Sentinel sensors) and DEMs (SRTM, GDEM) are zero under current data access policies.

Other products:

The Glaciers_cci products beside glacier area are processed on demand and customer driven. No definition of a globally complete dataset exists for these products as so far only regional scale assessments have been made and data gaps exist for several regions. The focus here is on a solution that can be distributed as a toolbox so that the user can run it on their own computers, or as a service that the user can select (or provide) the input data and receive the corresponding output data. In both cases it is foreseen to store the products that are derived in an archive for open access. Results will not be forwarded automatically, as an analyst-based pre- and post-processing step is required (e.g. matching window size, filter thresholds) to obtain the best product quality. Consequently, the driving factor is less the need for a full dataset for these products but an upscaling and individual handling of single product requirements.

In the following table the data budget for historical EO input data and for new data of future missions per scene and in total for full dataset production is provided. For elevation change related sensors (altimetry) and velocity (SAR) the data budget for the full archive is given.

Data	Product	Time Span	Historical Data	New Data
raw Landsat/Sentinel 2 scene	Area	1984-2012	1 GB (TM/ETM+)	2 GB (OLI/MSI)
Landsat pan / Sentinel 2 scene	Velocity (opt.)	1999-	0.5 GB	1 GB (OLI/MSI)
DEMs, per scene	Elev. change	var.	0.5 GB (60 m)	1 GB (30 m)
SRTM	Elev. change	2000	70 MB/tile (90 m)	-
ASTER GDEM	Elev. change	2000-present	200 MB/tile (30 m)	-
ERS-1, ERS-2	Velocity (SAR)	1991-2003	173 GB	N/A
ENVISAT	Velocity (SAR)	2002-2012	80 GB	N/A
ICESat-GLAS	Elev. change (ALT)	2003-2009	23 GB	N/A
CryoSat-SIN	Elev. change (ALT)	2010 –	17 GB	4.5 Gb/yr
SARAL Altika (K-Band)	Elev. change (ALT)	2013–2017 (est)	11 GB	5.5 Gb/yr
Total			306 GB	10 Gb/yr

The volume of Glaciers_cci output products is small (vector data). The following table is an estimation of the data budget for all Glaciers_cci products (one full dataset). It does not consider products that will only be derived regionally (e.g. historic Landsat MSS scenes for change assessment) and/or not available for free on a global scale (e.g. SPOT DEMs).

Product	Time Span	Historical Data	New Data
Glacier area			
outlines (one snapshot)	var.	300 GB (globally)	300 GB (globally)
Elevation Change (dDEM)			
national DEMs	1950s-present	<GB range per DEM	>GB range per DEM
SRTM (90m)	Feb. 2000	30-60 GB (global)	
ASTER (30m)	2000-present	150 GB (global)	
TanDEM (30/90 m)			150-300 GB (global)
Elevation Change (ALT)			
ERS-1, ERS-2	1991-2003	66 MB	N/A
ENVISAT	2002-2012	60 MB	N/A
ICESat-GLAS	2003-2009	18 MB	N/A
CryoSat-SIN	2010 –	12 MB	3 MB per year
SARAL Altika	2013 –	6 MB	3 MB per year
Velocity (optical)			
Landsat/ASTER (15m)	1984 - present	600 GB (global)	2 TB
Sentinel-2 (10 m)	from 2015 on	-	3 TB (global)

Velocity (SAR)			
ERS1/2/Envisat	1991-2010	200 GB	-
ALOS	2006-2012	200 GB	-
TerraSAR-X	2008-present	60 GB	100 GB
Sentinel-1	from 2014 on	-	3 TB

The hardware storage budget must foresee some spare for redundancy, concurrently kept product versions, intermediates and test results. We assume here a figure of a factor of 4-5 in the storage budget.

The following table summarizes the processing load budget for product reprocessing (per product) DEM Diff and optical/SAR Velocity (for 1 scene pair), and elevation change (per year of data).

Product	Pre-processing	Main-Processing	Post-Processing	Total (mean)
Glacier area	0.5 hr	0.5 hr	1-50 hr	2 -50 hr
DEM Differencing	0.5 hr	0.5 h	1 hr	2 hr
ERS-1 / 2/ Envisat	2 hr	4 hr	20 hr	26 hr
IceSat / CS2 / Altika	4 hr	4 hr	20 hr	28 hr
Optical Velocity	0.5 hr	2.5 hr	1 hr	4 hr
SAR Velocity	0.5 hr	2.5 hr	1 hr	4 hr

6.4 Cost

According to the following overview table, costs for storage, processing hardware, network, development and integration, operations, dissemination result in a theoretical amount of about 500,000 € per year for the activities planned for the three years of the second phase.

Requirements addressed by this section are:

→ GL-OPE-6320 Min 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 GLIMS/WGMS infrastructure.

The cost estimates per year in kilo Euros are summarized in the following table. Costs from operational production of on-demand services move from the products to the user services after initial setup.



Contract: 4000101778/10/I-AM
System Specification
Document Version 1 (SSDv1)

Name: Glaciers_cci-D5.3_SSDv1
 Version: 0.8
 Date: 17.04. 2014
 Page: 29

Product	Component	per Year	Year 2	Year 3
User Services	Hardware	10	8	8
	Development	150	75	75
	Operations	150	150	150
	Infrastructure	40	40	40
	Total	350	273	273
Glacier Area	Hardware	1	n/a	n/a
	Software	2	n/a	n/a
	Operations	150	n/a	n/a
	Infrastructure	from University	n/a	n/a
	Total	150	n/a	n/a
DEM Diff	Development	75	0	0
Altimetry	Hardware	5	0	0
	Development	75	0	0
	Operations	75	75	75
	Infrastructure	from University	n/a	n/a
	Total	155	75	75
Velocity Optical	Hardware	5	0	0
	Development	75	50	50
	Operations	75	0	0
	Infrastructure	from University	n/a	n/a
	Total	155	50	50
Velocity SAR	Hardware	5	0	0
	Development	75	50	50
	Operations	75	0	0
	Infrastructure	at company	at company	at company
	Total	155	50	50

7. Requirements traceability

Id	Title	Reference
GL-FUN-0010	The Glaciers_cci system provides tools to generate a global glacier inventory. Tools should allow to complement and improve the existing inventory but also to revisit the inventory every 10 years.	§3
GL-INT-0020	Its creation shall be coordinated with advice from a strategic operations team. This team need to be in close contact with all relevant high-level organizations (e.g. GTN-G (WGMS, GLIMS, NSIDC), GTOS/GCOS, CEOS, GCW).	§3
GL-INT-0030	The global GLIMS community shall play an active role in its creation according to a given set of guidelines and advice from the strategic operations team. They shall also give feedback from implementation to the strategic team.	§3
GL-FUN-0040	The system shall also 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).	§5.2, §5.3
GL-INT-0050	The available data shall be frequently reported and properly disseminated to the interested user communities.	§5.1, §3.1
GL-FUN-0060	The Glaciers_cci system shall generate 3 types of products: 1. Glacier Area (AREA) 2. Glacier Elevation Change (ELC) 3. Glacier Velocity (VEL)	§2
GL-FUN-0110	The PS shall include Sentinel 2, LDCM and Sentinel-1 products.	§6
GL-FUN-0120	The PS shall include a global elevation data set and TanDEM-X products when available.	§6
GL-FUN-0160	The PS shall provide a user feedback functionality.	§3.1, §5.1
GL-FUN-1010	All data stored in the system shall be available for the long-term (at least 15 years).	§2.4, §6.2
GL-FUN-1011	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.	§2.4, §5.3
GL-FUN-1020	Product shall be uniquely identified.	§5.3, §6.2
GL-FUN-1030	PS shall store data in a structured way using type, revision, date.	§5.3, §6.2
GL-RAM-1050	PS shall provide means against data loss of its input and output products.	§5.3
GL-SIZ-1110	The PS shall provide storage space for its input products of about 5 TB.	§6.3
GL-SIZ-1120	The PS shall provide storage space for its auxiliary data of about 2 TB.	§6.3
GL-SIZ-1130	The PS shall provide storage space for its output products of about 5 TB.	§6.3
GL-FUN-1140	To facilitate the use of these data by the climate research community a self-standing 6-8 pages explanation of the products shall be generated. This shall detail the algorithm, input data, description of the processing steps, geophysical data product content, flags, meta-data, data format, grid, software tools for decoding and exploiting the data.	§2.4, §3.1, §5.6
GL-FUN-2110	The PS shall produce glacier outlines (inventory data) compliant with the GLIMS database (GDB) format specifications.	§2.2
GL-FUN-2111	The PS shall produce elevation changes in agreement with WGMS (sheets D and EEE) requirements.	§2.2

Id	Title	Reference
GL-FUN-2112	The PS shall produce velocity information over glaciers in agreement with GDB or the WGMS requirements. (TBD)	§2.2
GL-FUN-2120	The PS shall produce the area product (global map of glacier-covered area) in netCDF format and compliant with the CF meta-data standards.	§5.3
GL-FUN-2130	The PS shall use the input data as outlined in the DARD.	§6
GL-FUN-2140	The glacier area product shall be produced according to the guidelines given in Paul et al. (2009).	§2.2
GL-OPE-2150	A hierarchical approach shall be taken to the production of the glacier products based on their complexity, data availability and contribution to the worldwide glacier inventory. Priority is on data availability, contribution (community request), complexity.	§5.4
GL-RAM-2160	Where possible, calibration and other values should be configurable to facilitate easier processing updates	§5.4, §6.1
GL-FUN-2170	The PS shall include the generation of consistent quantified errors and biases per pixel for the subsequent use of the glacier products in climate impact studies and water resource management models.	§5.2
GL-FUN-2210	The PS shall have the capability to reprocess already successfully processed products as well as products generated with errors in a transparent and comparable way.	§2.4, §6.2
GL-FUN-2211	The PS shall have the capability to reprocess already successfully processed products with new data for change assessment.	§2.4, §6.2
GL-FUN-2220	The PS shall store its input data optimised for reprocessing, i.e. in such a way that storage is not a bottleneck for reprocessing.	§6.3
GL-RAM-2230	All output products will contain sufficient information to ensure full traceability of any product to all inputs involved in its generation.	§5.3
GL-FUN-2310	The PS shall allow processing of the full AREA product within 10 years.	§6.3
GL-RAM-3110	Strict quality control procedures shall be followed during processing: the production shall be interrupted and the implementation checked and corrected if the resulting products do not meet previously agreed (scientific) quality standards. This shall include internal quantitative validation tests for each processing step.	§5.4
GL-FUN-4010	The PS has a logging mechanism	§5.4
GL-RAM-4020	The following events and parameters must be reported per task: a) start of processing b) end of processing c) significant processing events d) termination status (terminated safely, aborted etc)	§5.4
GL-RAM-4040	The following significant processing events shall be reported: a) input data missing, corrupt or invalid b) product cannot be fully produced c) product generation failed	§5.4
GL-INT-4110	The PS provides information of the processing status: a) Status (in progress, finished, stopped) b) Progress c) Errors	§5.4
GL-INT-5010	The PS provides a Web site presenting the objectives of the project and describing data access.	§3.1, §5.1
GL-INT-5020	Data access shall be through the Internet.	§3.1, §5.1

Id	Title	Reference
GL-INT-5510	There is a possibility to inject new input data into the system	§6.1
GL-RAM-5520	Data import is logged.	§5.4
GL-FUN-6010	The PS has an interface for commanding all the subsystems, including archive and data services. This commanding interface can be a command line interface (CLI) or graphical user interface (GUI).	
GL-FUN-6020	The operational processor shall not overwrite existing data. Versioning shall be used instead.	§3.2, §5.2
GL-OPE-6310	Software re-use shall be limited as much as possible to Public Domain software	§6.1
GL-OPE-6320	The PS design shall ensure minimal maintenance and operational costs.	§6.4
GL-OPE-6330	Development of the PS shall be under version control.	§5.5, §6.2
GL-OPE-6340	The system should be decoupled from the own research.	§5.5
GL-INT-6340	The PS shall have the capability and interfaces to extend for future adaptations.	§6.1
GL-OPE-6410	Development of the system shall be based on the outcomes of Task 2 and the requirements specified in Task 1 and used to generate the baseline products for the worldwide glacier inventory.	§2.4
GL-OPE-6411	The system shall be developed including all the necessary steps for the production of each product with the potential to produce the complete set of parameters for each glacier.	§2.4
GL-RAM-6420	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.	§5.6
GL-OPE-6430	The PS development shall be overseen by a science team that drives the development process interacts with the GLIMS community and is using the system to improve and evaluate methods and algorithms.	§3, §6.1
GL-OPE-6610	Each PS installation includes a set of test tools, data and benchmark data to test PS integrity (end-to-end, interfaces)	§3.3, §6.2
GL-OPE-6611	The verification is regarded as successful, when all tests agree within TBD limits.	§3.3, §6.2
GL-RAM-6611	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.	§3.3, §6.2
GL-OPE-6620	If a module is based on a prototype, the prototype state has to be frozen until it is implemented.	§6.1
GL-FUN-6710	Verification of the correct implementation of the prototype system against the algorithms developed in Task 2 is a fundamental part of the process.	§3.3, §6.1

Acronyms

AD	Applicable Document
ALT	Altimetry
ATBD	Algorithm Theoretical Basis Document
CCI	Climate Change Initiative
CDR	Climate Data Record
CMS	Content Management System
DARD	Data Access Requirements Document
dDEM	DEM differencing
DEM	Digital Elevation Model
ECV	Essential Climate Variable
ELC	Elevation Change
ELCSS	European Committee for Space Standardisation
EO	Earth Observation
ESA	European Space Agency
ESRIN	European Space Research Institute
FOSS	Free and Open Source Software
GCOS	Global Climate Observing System
GDB	GLIMS Database
GLIMS	Global Land Ice Measurements from Space
IPR	Intellectual Property Rights
LDAP	Lightweight Directory Access Protocol
LPDAAC	Land Processes Distributed Active Archive Center
MW	Microwave
NSIDC	National Snow and Ice Data Center
OC	Ocean Colour
OPT	optical
PS	Processing System
PSD	Product Specification Document
RA	Radar Altimeter (ERS-1 and ERS-2)
RD	Reference Document
SAR	Synthetic Aperture Radar
SLC	Single Look Complex Radar Image
SoW	Statement of Work
SRD	System Requirements Document
URD	User Requirements Document
USGS	United States Geological Survey
UUID	Universally Unique Identifier
VEL	Velocity
WGMS	World Glacier Monitoring Service