System Specification Document Phase 2 Year 3

Prepared by: Glaciers_cci consortium
Contract: 4000109873/14/I-NB
Name: Glaciers_cci-D3.1-SSD-Ph2Yr3
Version: 0.6
Date: 21. 02. 2017

Contact:
Frank Paul
Department of Geography
University of Zurich
Winterthurerstr. 190
CH-8057 Zurich
frank.paul@geo.uzh.ch

Technical Officer:
Stephen Plummer
ESA ESRIN

Document status sheet

Version	Date	Changes	Approval
0.1	07. 11. 2014	First draft GAMMA	
0.2	24. 11. 2014	Contributions from UZH, GUIO, SEEL	
0.2	04. 12. 2014	Final Review SL	
0.3	10. 03. 2015	Revised after TO feedback by SL and SE	
	24. 03. 2015	Revised by SL	
	16. 06. 2015	Revised after TO feedback by SL and SE	
0.4	16. 12. 2015	Revised by SL and SE	
0.5	21. 02. 2017	Revised by SL and SE	
0.6	19. 04. 2017	Revised after TO feedback by SL and SE	

The work described in this report was done under ESA contract 4000109873/14/I-NB. Responsibility for the contents resides with the authors that prepared it.

Author team:
Andreas Wiesmann, Tazio Strozzi (Gamma), Kate Briggs, Andrew Shepherd (SEEL), Frank Paul, Philipp Rastner (GIUZ), RobertMcNabb, Andreas Kääb (GUIO), Thomas Nagler (ENVEO)

Glaciers_cci Technical Officer at ESA:
Stephen Plummer

Table of Contents

1. Introduction

1.1 Purpose and Scope

This document is deliverable D3.1 Phase 2 Year 3 of the System Specification Document (SSD) of the Glaciers_cci project requested in the Statement of Work (SoW) [AD 1]. The SSD incorporates the requirements described in the System Requirements Document (SRD) [RD 9] 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 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 of Phase 1 are applicable to Phase 2 of CCI with their main content, the processing system (PS) is further developed based on the Phase 1 SPD to meet the specifications provided in this updated SSD.

1.2 Applicable and Reference Documents

List of Applicable Documents

ID	Title	Issue	Date
[AD 1]	ESA Climate Change Initiative Statement of Work RFQ/3-13904/13/I-NB	1.0	17.05.2013
[AD 2]	Glaciers_cci proposal	revised	2.09. 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-D1.4_DARD-Ph2Yr1	0.5	18.10. 2014
[RD 3]	Glaciers_cci_ph2-D2.1_ATBD-Ph2Yr1	0.2	17.08. 2014
[RD 4]	Glaciers_cci IODDv1	0.5	21.11. 2012
[RD 5]	Glaciers_cci DPMv1	0.5	8.11. 2012
[RD 6]	Glaciers_cci-D1.2_PSD-Ph2Yr1	0.5	18.10. 2014
[RD 7]	Glaciers_cci SVR	0.4	18.03. 2013
[RD 8]	Glaciers_cci-D1.1-URD-Ph2Yr1	0.4	17.10. 2014
[RD 9]	Glaciers_cci SRD	1.0	12.07. 2012
[RD 10]	Glaciers_cci SPD	0.5	13.02. 2013

1.3 Document Structure

This document is organised 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.
- 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.
- Section 7 gives the system requirements of [RD-9] traceability and evolution.

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

As depicted in the general overview of Fig. 2.1, 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 scales. The PS will be used by the Glaciers_cci consortium but can also be applied by others, i.e. those processing the EO data and contributing them to GLIMS and/or WGMS (the science community in Fig. 2.1) 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 glacier-related products such as glacier outlines (GO), elevation changes, (SEC) and velocity (IV) 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).

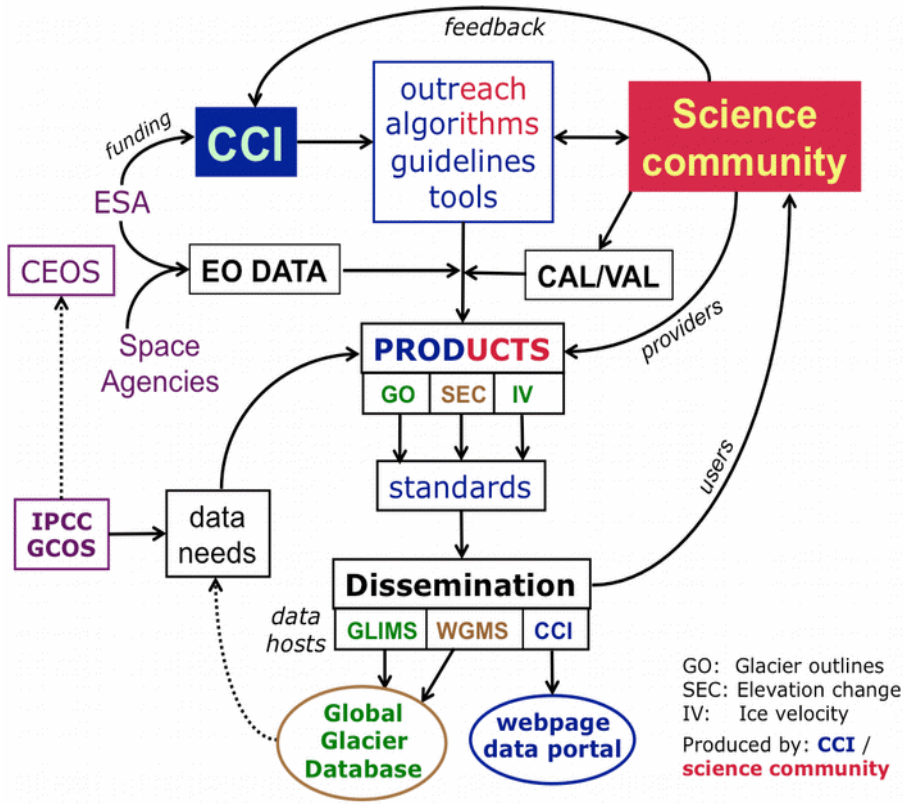

Fig. 2.1: Schematic overview of the Glaciers_cci processing system and its relation to developments by the science community for the individual products and their dissemination possibilities (colour-coded for each product).

The products products generated by Glaciers_cci 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. They will also be made available via the Glaciers_cci website and the CCI data portal). 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 through contributions to 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. the sea-level contribution of glaciers and ice caps) that have been included in the fifth assessment report of the IPCC (AR5). The new purpose of the PS in Phase 2 is its application for further improving the existing datasets, covering multiple epochs and establishing a monitoring service (for all 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. 2.2). 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 (see also RD-8). Another interface is with the EO data providers. Depending on the module, EO data are obtained from the providers at CEOS level 1b or 2 and are ingested into the PS [RD 4]. 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.

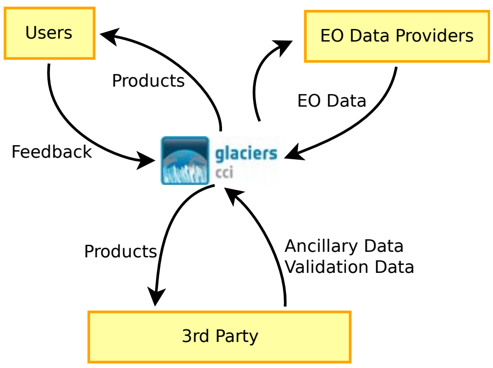

Fig. 2.2: Principle context of Glaciers_cci processing system.

2.3 Requirements

System requirements are compiled in the Glaciers_cci Phase 1 SRD [RD-9]. 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 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 GO the focus has been on the production of a comprehensive global glacier inventory, the other products are currently less well spatially distributed, are hence focused on more regional needs and thus generated for key regions only. However, global-scale data production with related assessments of trends is also evolving in the science community for the elevation change and velocity products. These activities are mostly based on research projects that would benefit from some coordination. All products benefit from recently available sensors such as the Sentinels and have incorporated 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, although this is its main purpose. Also functions to make output products and documentation available to users included are in the scope of the system.

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 hosts input data, performs pre-processing, classification processing, supports validation, and serves the output to users ('Dissemination' in Fig. 2.1), the part that is often missing in other scientific studies (see 2.1). 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;

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 service for the glaciological and hydrological community. The products are consequently distributed through existing and in development platforms of WGMS (http://www.wgms.ch) and GLIMS (http://www.glims.org/). Further product platforms are the dedicated project website for the CRDP https://glaciers-cci.enveo.at/crdp2/index.html and the CCI data Portal. Functions are:

1. Project Information/Introduction
2. Processing Information (including velocity)
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 (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 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 [RD-10] the following modules are part of the PS. Here with updated product names as defined in [RD-6]:

1. Glacier Outlines (GO), Glacier Inventory (GI)
2. Elevation Change DEM Differencing (ECD or SEC-DD)
3. Elevation change Altimetry (ECA or SEC-ALT)
4. Velocity from microwave sensors (IV)
5. Velocity from optical sensors (IV)

Within the framework of the Glaciers_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 architecture of the PS

The functions and modules listed in the previous section are implemented as functional components. In this section we outline the high-level architecture of Glaciers_cci on this subsystem level. Data-uptake is either through the mostly web-based services of satellite data providers (e.g. EOLI-SA, GLOVIS/Earthexplorer, S1/2 Science Hub) and those listed in [RD-2]. 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 that in the case of Glaciers_cci point to the databases of GLIMS (for outlines/inventories and possibly also for velocity) and WGMS (for elevation change products) as indicated by the colour-coding in Fig. 2.1. 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.

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.

2.6.1 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).

2.6.2 User services subsystem

This subsystem consists of at least three components: web presentation, data access by users, and a catalogue, all accessible via the GTN-G homepage (gtn-g.org). 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.

2.6.3 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

One of the long term objectives of the project was to help establish a structure in GLIMS (executive board, core group) that would facilitate long-term steward ship (e.g. regarding data production, distribution and documentation by the PS). This has already been established and will be used in the future to coordinate the work and provide an improved service for the community at various levels. The legacy of Glaciers_cci is also already visible due to its contribution to data production for RGI / IPCC purposes and several publications (algorithms, accuracy assessment) that will be important milestones for the community in the years to come.

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

The processing time is mainly critical for the GO product because there are constraints on the reprocessing time of the GO inventory. By design the PS module is distributable so that the processing can be parallelised. Development of the processor towards a more automated machine-based processing module helps to reduce the production time and allow large scale reprocessing to test improved algorithms or to repeat the inventory more frequently. Although the related development is enforced by the science community, basic issues such as product consistency have still to be addressed (e.g. debris). For the other products data processing is already largely automated, but some manual steps in the pre- and post processing stages remain (e.g. selection of filters or matching windows) to obtain high-quality products. The data volume needs are small at this stage and are not considered as a constraint.

2.7.2 System portability

The GO module needs to be portable between different operators. Consequently constraints to be considered are the operator skills and reliability, and operator infrastructure (hardware and software). The toolboxes must be portable in design so as to work on common hard- and software platforms. The on demand systems should also be designed to be 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 GO and optical IV the data to be processed (Landsat, Sentinel) are freely available and an up-to-date internet connection for the operator is sufficient. The higher data volumes to be expected from the Sentinels might not automatically increase the processing volume as external (clouds, seasonal snow) and internal (e.g. has a region to be processed again?) constraints reduce the amount of useful data for both products. For the microwave 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. For altimetry, the input data is freely available from the agencies and the computing infrastructure of the relevant project partners is sufficient to meet the download needs.

2.7.4 System development status

The current PS modules are tested. However, they need more or less operator intervention for data ingestion, 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 has provided 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, collaboration with GLIMS and WGMS has been established. GLIMS has recently integrated the RGI into its database and a executive board for improved coordination of all activities has been set-up. Support of the latter as well as continued data production could be established in the framework of the Copernicus Climate Change initiative.

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 (client)
    - are interested in best existing glacier products
    - are skilled in glaciology
    - provide feedback and proposals
    - request data format compatible with their communities
- Development team
    - mandated to run Glaciers_cci
    - in dialogue with users
    - develops the PS further
    - issues product versions
- External validation experts
    - are part of the international community
    - support development team
    - provide local expertise
    - feedback on the products
- Auditor
    - project supervision

3.2 User information and data access

Users access the Glaciers_cci data products using the GLIMS/WGMS web sites (see colour-coded overview in Fig. 2.1) as well as the Glaciers_cci homepage http://www.esa-glaciers-cci.org/ (GL-INT-0050, GL-INT-5020, GL-INT-5010). New data access points are currently implemented at WGMS for glacier fluctuation datasets (WGMS Fluctuations of Glaciers Browser at http://www.wgms.ch/fogbrowser/) and data documentation (glacier app at http://wgms.ch/glacierapp). Access to the GLIMS database is also under revision (http://www.glims.org/maps/gtng). These are also the entry points for meta-data, documentation (GL-INT-1140), the catalogue, data services, and tools such as the determination of glacier flow velocities from image cross-correlation http://www.mn.uio.no/geo/english/research/projects/icemass/cias/ or the co-registration of two DEMs before calculating elevation changes in a freely available publication (Nuth and Kaab, 2011) (GL-FUN-0010). The websites provide also a forum (GL-FUN-0160) and an issue tracking system. This approach is common for all Glaciers_cci products. Updates to the system are disseminated via the email lists for GLIMS and through Cryolist.

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

- The development team submits new products to the respective databases (GLIMS/RGI or WGMS) where data are reformatted and ingested. New versions of the RGI or major updates of the WGMS database are announced via mailing lists (GLIMS list and Cryolist).
- 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 registers themselves 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 (if applicable), 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, mosaicing, 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 [RD-10] 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 (if applicable), 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 GO and SEC-ALT the processor modules process a set of input data selected by the user 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 SEC-DD module is a toolbox that can be downloaded from the web http://www.mn.uio.no/geo/english/research/projects/icemass/cias/ and run on the users 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, depending on the possibilities to calculate mean specific values per glacier (which is scientifically challenging by the variable presence of data voids and artefacts).

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

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 the local implementation of the code. 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
- Development of a new or modified processor
- Reprocessing
- Validation and comparison with the former product versions

3.4 System Verification

The system operators are in charge of system (module) verification. Findings are reported in the System Verification Report (SVR). 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 (GL-OPE-6611, GL-FUN-6710)
- Test tools (GL-OPE-6610)
- Verification Report (GL-RAM-6612)

3.5 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:

- Product validation and quality control
- In case of multi-sensor use, cross-comparison

The PS provides tools that facilitate these tasks such as

- Feedback functionality

External validation experts use the test tools (e.g. the description on how to determine product accuracy) and their knowledge of the region and build up their own benchmark test dataset. This has yet to be standardized by the community. It is not yet done in a standardized way but scientists recognize the importance in general and use a range of validation datasets they have at hand or access to. A standard validation (overlay in Google Earth) is performed for the glacier outlines during database ingest. As glaciers are constantly changing, validation datasets will be created by the community on a case-by-case basis, e.g. in the framework of a publication. Usually the latest available high-quality datasets are used for comparison with their own results. A collated validation database should be considered for the future.

3.6 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. For example, Landsat 8 OLI and Sentinel 2 MSI can still be processed with the same PS (methods and workflow). A key innovation that is currently in development is related to the processing of raw data in Google Earth Engine, allowing stack processing of entire time series without downloading the individual scenes. Another one is the processing of 2 m DEMs from WorldView satellite data. When this is released in a few years there will be a further jump in innovation after SRTM (2000) or the GDEM (2008).

4. Infrastructure

There are some fundamental decisions that determine how the PS and the corresponding infrastructure will develop and what it will 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 answers to these questions for the Glaciers_cci modules

4.1 Distributed or central

The Glaciers_cci PS is based on self standing 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 products. 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 GO module. 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 in particular for a modular system. Glaciers_cci follows a cluster solution, i.e. decentralized or distributed computing using a standard set of algorithms and quality control mechanisms for the operator-based stages. A cloud solution would introduce nominal overhead and increased development costs. However, it is currently tested for the glacier outline product for other purposes, e.g. rapid change assessment that only considers clean ice mapping.

4.3 Sharing with other ECVs

Glaciers_cci follows an independent implementation approach. Focus is on improving the operational aspects and merging of the Glaciers_cci modules. Synergies with other CCI projects such as Ice_sheets_cci, Land_cover_cci, Sea_level_cci are followed as possible, for example in regard to algorithms (velocity), input data (e.g. for specific regions), and joint science (e.g. closing the sea level budget).

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 product 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 product services the web site shall act as a central starting point to explore achievements, products 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 products 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 include the functional aspects:

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

Whereas the Glaciers_cci website is an entry point for all products while the project is running (and likely a few years beyond), the products 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. Links should be established to connect the different sites to ease user access. 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 product 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 make use of mailing lists for communication that might also be beneficial for the Glaciers_cci PS.

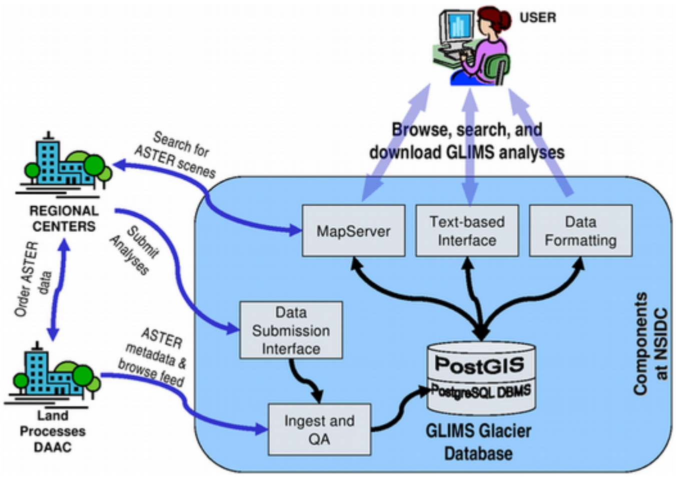

Figure 5.1: The GLIMS Glacier Database Architecture and Environment for glacier outlines and velocity data.

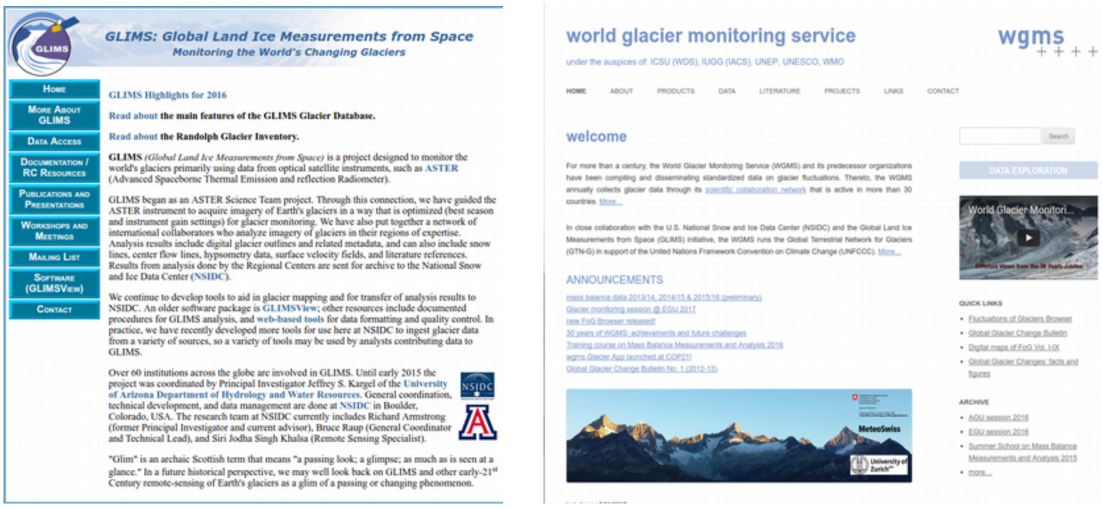

Figure 5.2: The entry pages of the GLIMS website www.glims.org (left) and the WGMS website www.wgms.ch (right) as of 21. February 2017.

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, or 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. He might get support from additional technical specialists.

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 GO	Generate L2/L3 Glaciers_cci Area Product	Input Data	Existing processor, derived automated processor
Data Processor SEC-ALT	Generate L2/L3 Glaciers_cci Elevation Change Product	Input Data	Extension of existing processor
Data Processor SEC-DD	Generate L2/L3 Glaciers_cci Elevation Change Product	Input Data	Toolbox
Data Processor IV-MW	Generate L2/L3 Glaciers_cci Velocity Product	Input Data	Extension of existing processor
Data Processor IV-OPT	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, 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:
→ 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 consists 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.

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, Landsat 8, Cryosat 2[1]
→ 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 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:
→ 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 were used, adapted, configured and integrated during development of the PS within the CCI and beyond.

Name	Usage	Remarks
Subversion	Version control	FOSS
Redmine	Issue tracking	FOSS
GO processor prototype	Glacier outline production	Based on closed source COTS but implementable for any COTS
	Base for the development of an automated processor	Implementation from scratch needed
SEC-DD processor prototype	Base for DEM differencing toolbox	DEM co-registration tool free/public from GUIO; rest to be

1Sensors changed

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

6.1.1 Glacier Outlines (GO)

The aim of the GO module is to produce glacier outlines and inventory data - feeding into the global inventory of glacier outlines (RGI/GLIMS) with a sufficient quality for regional-scale applications. The prototype covers all necessary processing steps. It is currently 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 not coordinated, i.e. data are provided as projects permit. During Phase 2 of Glaciers_cci an Executive Board has been set-up for GLIMS that will in the future aim at getting this contribution more actively requested and then also provided by the community for the regions of interest. 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 from DEM differencing (SEC-DD)

The aim of the DEM differencing module is to provide a tool for the community that allows the production of elevation change products from DEM differencing. The tool is concerned with the accurate co-registration of two DEMs in a relative sense and - when using ICESat data - also in an absolute sense. The tool is available in the form of an Excel spread sheet and other formats.

6.1.3 Elevation change from altimetry (SEC-ALT)

The aim of the Altimetry module is to process the altimetry data of the full archive of EnviSat, CryoSat-2 and ICESat altimeter data for which the prototype module is built.

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

6.1.4 Ice Velocity (IV)

The aim of the velocity modules is to produce glacier velocity information from a given microwave or optical image pair. The production of velocity products can be started on the CIAS website (http://www.mn.uio.no/geo/english/research/projects/icemass/cias) by selecting an image pair from the respective archives. To be able to compare optical and microwave products, results are also provided in slope parallel flow geometry. Merging of the products from both sensors is not conducted as they generally refer to different time periods.

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 was reasonable to decide for a given hardware/software environment and to keep this constant. 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 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 Outlines
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 RGI. Various places are foreseen to store the resulting products: The glacier outline and velocity products will be stored at GIUZ (internally, long term), at ENVEO (CRDP web page,mid-term), at ESA (data portal, hopefully long term) and at GLIMS (USGS LPDAAC, long term). The elevation change products will also be stored at WGMS (long term).

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 GO are processed on demand and customer driven. No definition of a globally complete dataset exist for these products as so far only regional scale assessments have been made and data gaps exists 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 products. 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 a user-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 up-scaling 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, per scene	GO	1984-2012	1 GB (TM/ ETM+)	2 GB (OLI /MSI)
Landsat pan, 2 scenes	IV-OPT	1999-	0.5 GB	1 GB (OLI/MSI)
DEMs, per scene	SEC	var.	0.5 GB (60 m)	1 GB (30 m)
SRTM	SEC	2000	70 MB / tile (90 m)	-
ASTER GDEM	SEC	2000-present	200 MB / tile (30 m)	-
ERS-1, ERS-2	IV-MW	1991-2003	173 GB	N/A
ENVISAT	IV-MW	2002-2012	80 GB	N/A
ICESat-GLAS	SEC-ALT	2003-2008	23 GB	N/A
EnviSat	SEC-ALT	2002-2010	50 GB	N/A
CryoSat-SIN	SEC-ALT	2010 –	17 GB	4.5 Gb/yr
Total			345 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, KH-9 Hexagon).

Product	Time Span	Historical Data	New Data
GO			
outlines (one snapshot)	variable	300 GB (globally)	300 GB (globally)
SEC-DD			

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)
SEC-ALT			
ENVISAT	2002-2010	60 MB	N/A
ICESat-GLAS	2003-2008	18 MB	N/A
CryoSat-SIN	2010 –	12 MB	3 MB per year
IV-OPT			
Landsat/ASTER (15m)	1984 - present	600 GB (global)	2 TB
Sentinel-2 (10 m)	from 2015 on	-	3 TB (global)
IV-MW			
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. A good ballpark figure seems to be a factor of 4-5 in the storage budget.

The following table summarizes the processing load budget for product reprocessing (per product) SEC-DD and IV (for 1 scene pair), and elevation change (per year of data).

Product	Pre-processing	Main-Processing	Post-Processing	Total (mean)
GO	0.5 hr	0.5 hr	1-50 hr	2 -50 hr
SEC-DD	0.5 hr	0.5 h	1 hr	2 hr
Envisat	2 hr	4 hr	20 hr	26 hr
IceSat / CS2	4 hr	4 hr	20 hr	28 hr
IV-OPT	0.5 hr	2.5 hr	1 hr	4 hr
IV-MW	0.5 hr	2.5 hr	1 hr	4 hr

6.4 Cost

Costs for storage, processing hardware, network, development and integration, operations, dissemination comprise about 500,000 € per year of the three years 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.

Product	Component	Year 1	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
GO	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
SEC-DD	Development	75	0	0
SEC-ALT	Hardware	5	0	0
	Development	75	0	0
	Operations	75	75	75
	Infrastructure	from University	n/a	n/a
	Total	155	75	75
IV-OPT	Hardware	5	0	0
	Development	75	50	50
	Operations	75	0	0
	Infrastructure	from University	n/a	n/a
	Total	155	50	50
IV-MW	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 and history

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

Id	Title	Reference
GL-FUN-2111	The PS shall produce elevation changes in agreement with WGMS (sheets D and EEE) requirements.	§2.2
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

Id	Title	Reference
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
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
DD	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
GO	Glacier Outline
IPR	Intellectual Property Rights
IV	Ice velocity
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
SEC	Surface Elevation Change
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