HALO Guideline

Harmonised coordination
of the Atmosphere, Land and Ocean integrated
projects of the GMES backbone
WP 3210 and 3310

Written by	Responsibility-Office-Company	Date	Signature
Jean-Marc PECHINOT Fabrice LEVY Johannes FLEMMING Johannes Kaiser	ASP Project Manager ASTRIUM Project Manager ECMWF Project Manager ECMWF Project Manager	21/04/05	
Verified by			
Approved			

DOCUMENT CHANGE RECORD

Issue	Date	Author	Pages	Description of Changes
2.09	14/11/2005	Kaiser	2	Document Change Record included
2.10	22/11/2005	Kaiser	1-57	smooth text, explain RMDCN CIR

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 Introduction

This document aims at providing the necessary inputs for the Work Packages 3X10 (Candidate solutions, Atmosphere-Ocean and Land-Atmosphere). The objective is to analyze interfaces concerning the Ocean/Atmosphere (resp. Land/Atmosphere) IPs (data acquisition, sharing, dissemination) and, by successive iterations, to propose candidate solutions for common interfaces with scientists and co-ordinated by ECMWF.

This Work Package is fourfold:

- Characterization of data and data users
- Analysis of data and users requirements
- Collection of Data services
- Analysis of Candidate solutions

The first step will consist in gathering data and users from MERSEA, GEMS, GEOLAND and related projects. It is described in this document.

The document is based on the previous version of this guideline provided by ASP and ASTRIUM.

1.1 Structure of the Document

The Document is divided in five chapters including this one, and an appendix:

1. Introduction
2. Scope of the document
3. General HALO scope
4. Architecture overview
5. Requirement Analysis
   - Data flows analysis
   - Data and product analysis
   1. Function Analysis
6. Appendix: Interacting parts of the IPs

Applicable Documents

Ref	Title	Author	Date
AD1	HALO Part B Forms	HALO team	
AD2	MERSEA Information Management (MIM) High level requirements	IFREMER	11 Oct 2004
AD3	MERSEA data and products for GMES	MERSEA Team	15 Nov 2004
AD4	HALO draft report on interacting parts of GEMS, MERSEA and geoland	ECMWF	7 jan 2005
AD5	HALO Guideline WP3210	ASTRIUM	3 March 2005
AD6	HALO Infrastructure Candidate Solution Overview	ASTRIUM	14 Nov 2005

1.2 Reference Documents

Ref	Title	Author	Date
RD1	EUMAREX Centres of Expertise Report - 2004	ROYAL BELGIAN INSTITUTE OF NATURAL SCIENCES	
RD2	BICEPS, Building an European information capacity for environment and security	GMES Action Plan (2002-2003)	

2 Scope of the document

Chapters 1to 4 of this document are a guideline for filling in the following three templates contained in chapter 5:

- Data Flow Analysis (Table 5-1, Table 5-2, Table 5-3)
- Data and Product Analysis(Table 5-4, Table 5-5, Table 5-6)
- Function Analysis (Table 5-7, Table 5-8,Table 5-9)

The tables aim at proposing the necessary inputs for the analysis of candidate solution for infrastructures on the themes Land, Atmosphere, and Ocean and considering the architecture functions related to Data acquisition, sharing and Dissemination.

In a general way, for each of these items, the existing situation (data and products providers, related services) shall be considered as a baseline for the candidate solutions. This document is based on MERSEA, GEOLAND and GEMS project with the target to settle a common, globally shared approach. The classes have been populated by the representatives of each project after circulation of chapters 1to 4. The resulting document will be used for the definition of the architecture. At least three candidate solutions will also be proposed at the end of this Work Package. Keeping in mind that purpose, the document starts with a system overview of MERSEA, GEOLAND and GEMS.

The logic of work shall also follow the figure below :

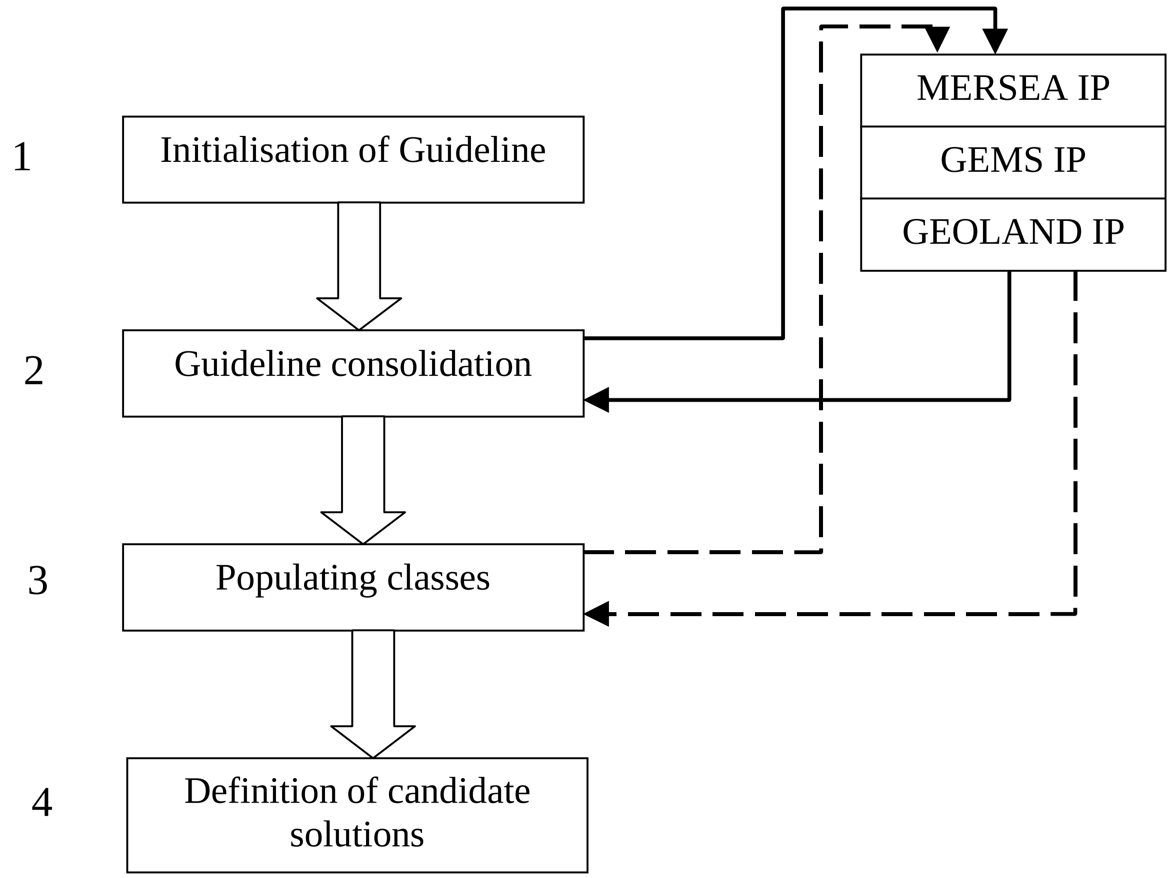


Figure 2-1 : Logic of work for HALO WP 3310

1) Initialization of Guideline

Three documents presenting the basic components for each Integrated Projects have been issued.

2) Guideline consolidation

The three previous documents have been used to issue the present coherent document, that synthesize as templates or tables, data flows, data and product, functions to be populated.

3) Populating classes

During the "Guidelines consolidation" representatives of each IP have fulfilled the templates. The objective is to take into account as exhaustively as possible, the existing components, the non-existing and required components.

4) Definition of candidate solution

The last stage will consist in infering some appropriate candidate solutions at functional and system level from the previously populated tables in this document.

The following guidelines aim at providing the necessary input for defining the candidate solutions. The necessary input is provided in form of the tables in chapter 5:

- Data Flow Analysis (Table 5-1, Table 5-2, Table 5-3)
- Data and Product Analysis(Table 5-4, Table 5-5, Table 5-6)
- Function Analysis (Table 5-7, Table 5-8,Table 5-9)

These tables and the way they have been fulfilled are described in the rest of the preceding chapters. Their fulfillment has taken into consideration the existing background as well as the needs of the users.

3 General HALO scope

First of all, to be sure of a common understanding this chapter synthesize our comprehension of the HALO WP 3210/ WP3310 objectives.
HALO wants to support the transition of the IPs to operational status. Indeed, operational commitment imposes time constrains of the data production, transfer and storage. Common data needs or product exchanges are a strong link between the IP because a **shared solution** could be beneficial for all IP.
Therefore, WP 3210/3310 aims at defining candidate solutions for both common data needs provision and products exchanges interface between the IPs. Once populated, the four inventory tables defined by WP 3310 should give a clear overview of all potential common data needs and inter-IP product exchanges.
As defined in the “HALO draft report on interacting parts of GEMS, MERSEA and GEOLAND” we suggest using the following criteria and categories for data and products characterisation:
Interaction and communality categories:

- ✓ direct product exchange,
- ✓ common data,
- ✓ un-accomplished data

The data groups will then be classified in terms of their **origin:**

- ✓ Observation (In-situ, Satellite),
- ✓ Model and assimilation products (global, regional)

And their associated **operational constraint:**

- ✓ Operational mode - real time (RT) and near real time (NRT)
- ✓ Off-line or Re-analysis mode
- ✓ Research mode

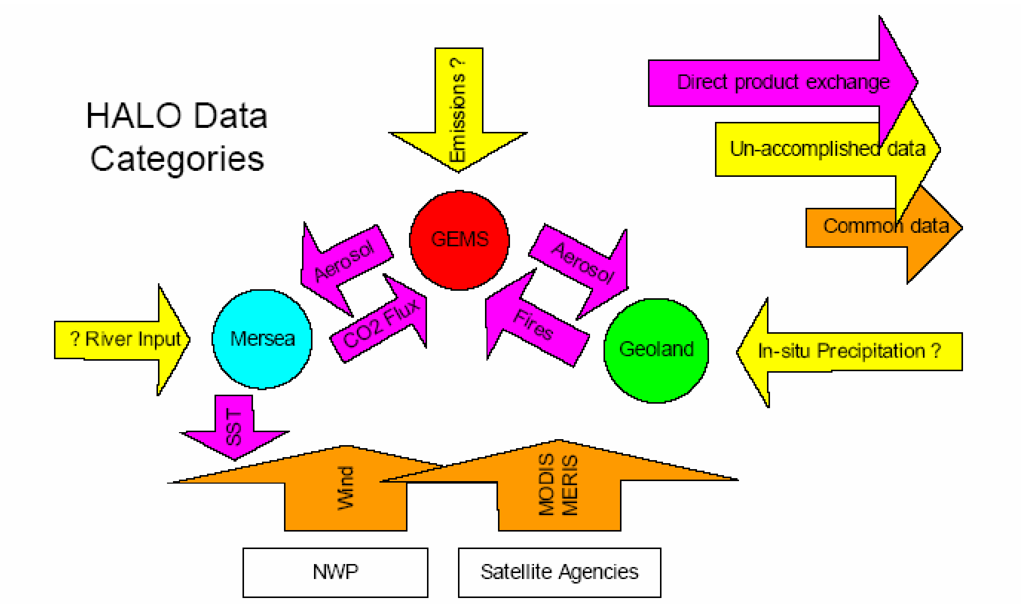


Figure 3-1: Flow structure of different data categories covered by HALO

According to the definition of the HALO work scope, only the global and continental or basin scale model and data activities with operational commitment will be considered (e.g: Geolands - ONC and CSP; MERSEA global and ocean scale model; GEMS global production system), **regional activities are not covered so far**.

Moreover, the exchange between the IPs will be mainly between GEMS and GEOLAND (ONC and CSP). **Since both ONC and GEMS models will be hosted at ECMWF, no external link will have to be established for their interaction**. Nevertheless, direct exchange between CSP and GEMS will have to be established.

4 Architecture overview

This section is an overview of the various architectures of the systems under development in the integrated projects MERSEA, GEOLAND, and GEMS in order to define some generic concepts which are key drivers for the candidate solutions to be issued from this work package. This architectural approach shall give a point of view of the system overall architecture definition.
In the scope of HALO, the candidate solutions to be issued will also be based on generic concepts, taken from infrastructure IPs such as WIN, and use as far as possible the existing operational infrastructure in accordance with their application fields.

4.1 MERSEA Architecture concepts

The MERSEA system is the set of equipment, process, organization, operators ... that contributes to the fulfilment of the needs of system users. The needs shall be expressed by the users towards the service or data providers (delay of acquisition, type and quantity of data, various information for programming and so on…). They will result in infrastructure requirements, such as type and number of sub-network, as well as their capacity, development of portals in order to share the necessary input between users and providers. Such portals may be thematic oriented, with themes as general as Ocean, Land, Atmosphere, but also with more refined themes such as climate change, seasonal forecasting, oil spill, marine safety, pollution,…
Thematic Portals or TEP, shall be considered as tools for users to access services. Beyond their function as portals, providing data management, the TEPs are also viewed as Thematic Service Providers for data (remotely sensed, in situ, and forcing fields) and monitoring and forecasting centres.
The sub-network is the way to physically connect users and providers to share data, services and information.
The concepts of TEP and sub-network shall be considered independently as various sub-systems. Anyway, they will both play a role in the defining the components of the overall architecture, by considering the various technical constraints, as well as users requirements at local, regional, national or European level. An overview of the Requirements and related Services for a European Capacity is also provided in BICEPS project

Another important aspect is the consideration of the existing infrastructure components, including local, regional or global sub-network, Thematic portal on various applications.

MERSEA aims to define the system as a two-fold approach:

- First, take into consideration the existing components, available data or services, providers, users, their needs as well as unaccomplished data needs, services.

- Derive from the previous analysis the technical requirements, the functions of the system, and then the list of components or sub-system in answer to the requirements.

As a preliminary analysis, the following figure gives an overview of the system involving the various actors. The TEP and sub-networks will be introduced afterwards in the architecture description. The so-described architecture components and the MERSEA system will be considered as stand-alone system. The systems GEOLAND and GEMS will be considered as external systems also described in a similar way.

The architecture of a common global infrastructure will also be the result of the cross-analysis of each various system, consisting in defining the common or independent sub-networks or portal, their interfaces, the data flows.

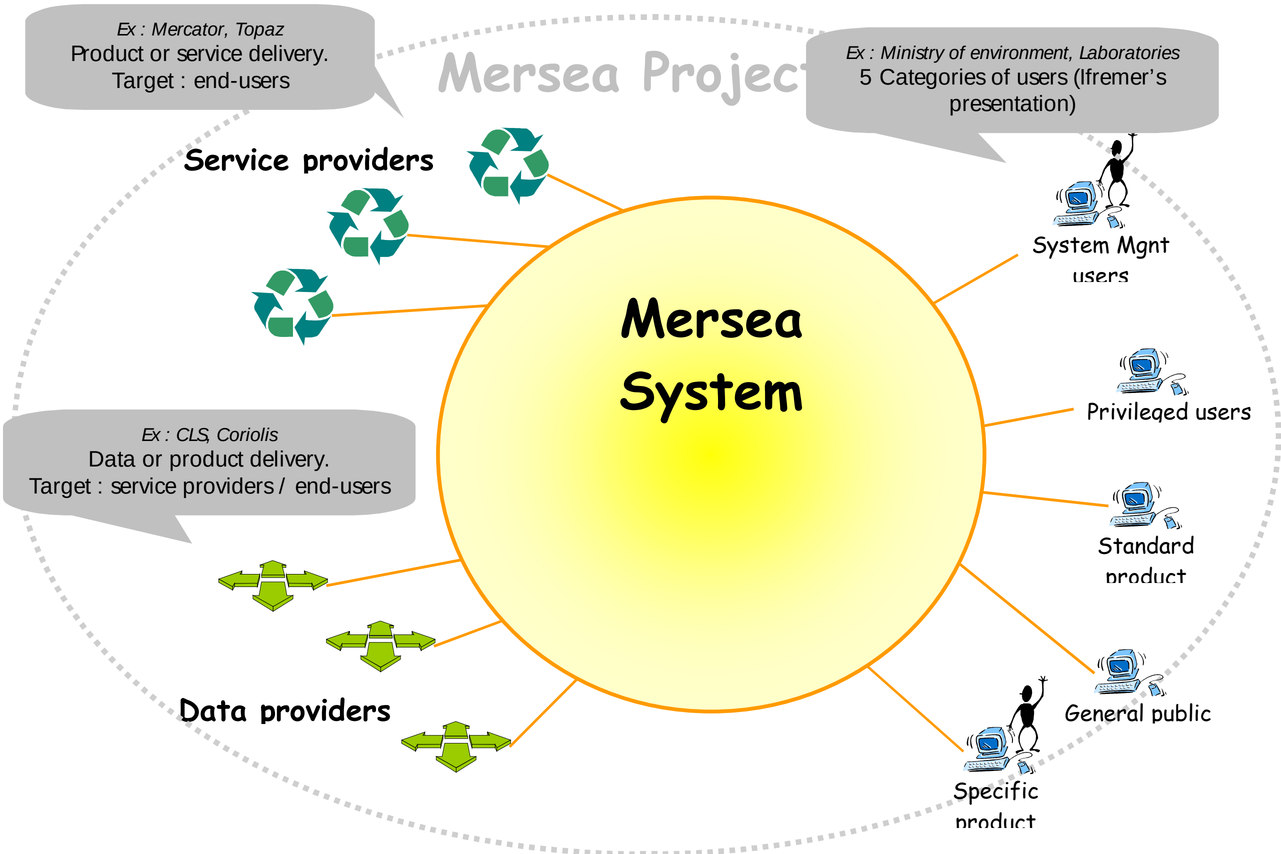


Figure 4-1 : Users classification

Considering the TEP (ThEmatic Portal), the MERSEA system can be decomposed in three sub-systems :

- Assimilation TEP

This TEP is used by the service providers for producing MERSEA assimilation models. They involve service providers for value-added products. Request may be formulated by users through Mersea Portal.

- Data TEP

This TEP is responsible for archiving, distributing data and products from data providers. Request may be formulated directly by users through Mersea Portal or by service Providers through Assimilation TEP.

- Mersea Portal

All the categories of users can formulate their data request on any Portal. The requests can also derive towards Data or Assimilation TEPs.

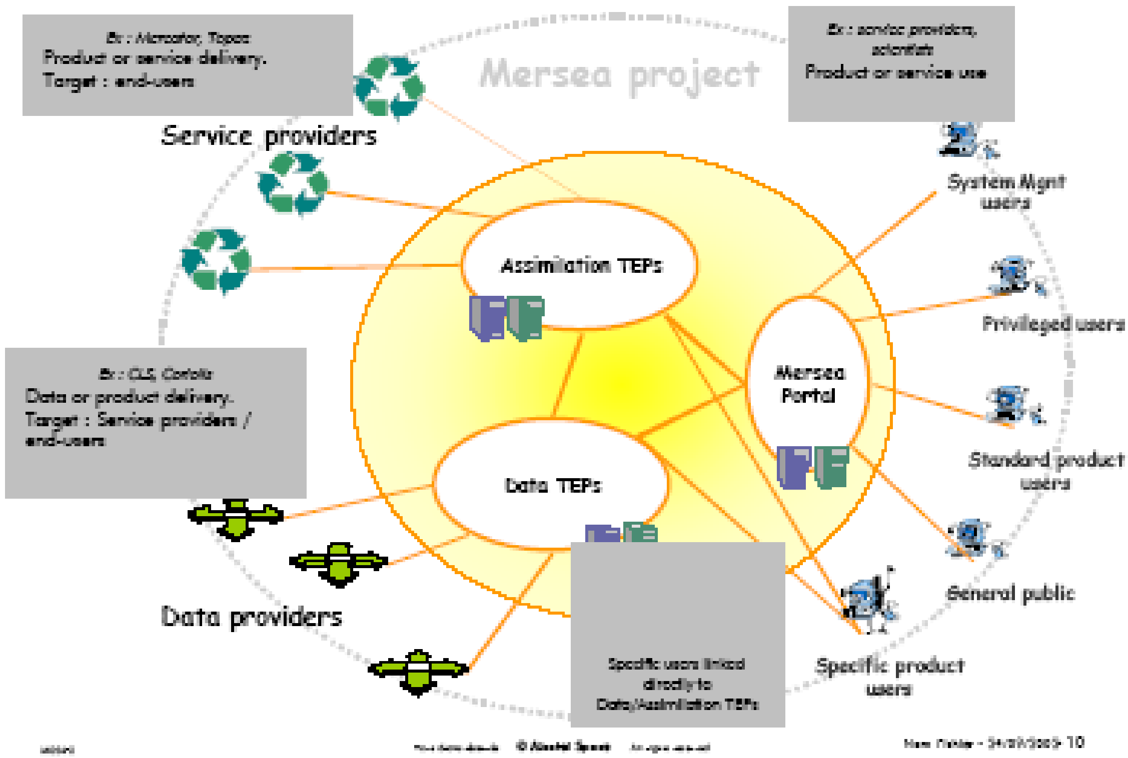


Figure 4-2 : Data Distribution (from MERSEA Project)

The above figure refers to tree sub-systems, Data ThEmatic Portal (TEP), Assimilation TEP, General Portal.

The next figure shows the corresponding architecture in the MERSEA system, by instantiating the various detailed sub-systems, including TEP and Data sub-networks, the data centres, and the Data flows.

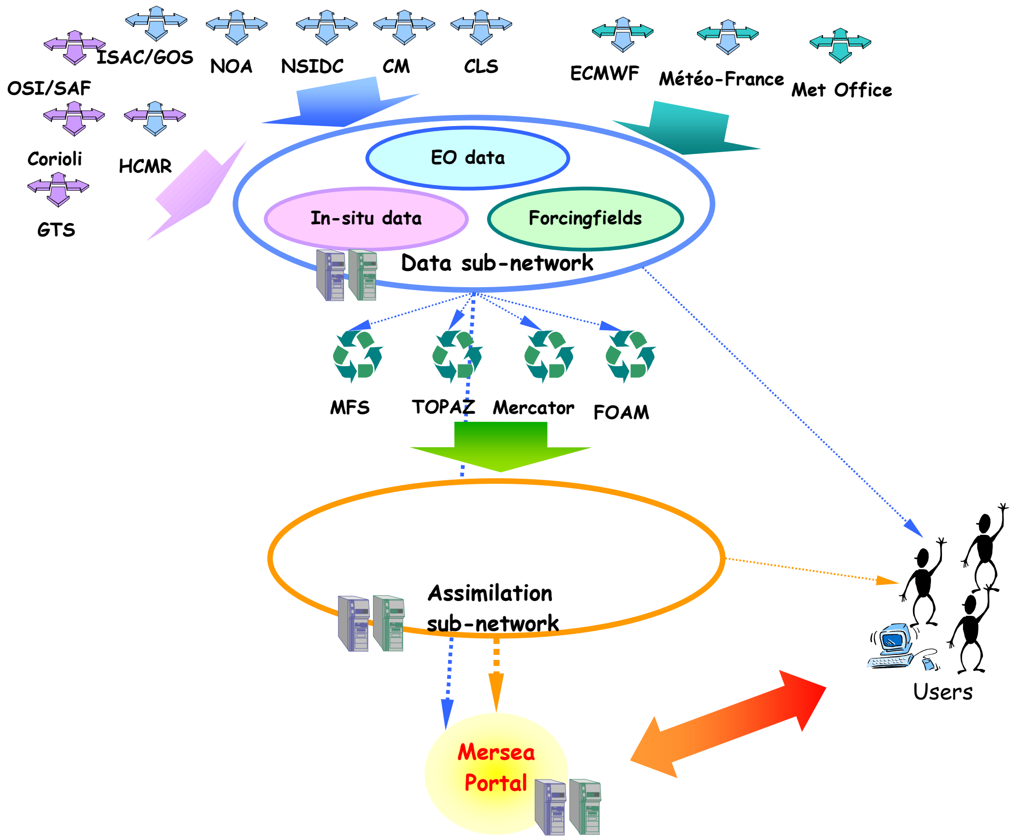


Figure 4-3 : Sub-networks Decomposition

The sub-networks have to be defined explicitly. They may be chosen in accordance with various criteria, such as the type of data (for a first level of decomposition of data sub-networks), but also the capacity of the network, its maximum bit rate, its capacity (number of users, number and size of simultaneous data), its policy (local, regional, global, free or paying access, for urgency or not), and so-on. The sub-networks can also be application oriented.
A sub-network sample is presented below :

TEP	Sub-Networks	Providers	Users
Data TEP	EO SN	DP	SP
	Forcing Field SN	DP	SP
	In situ SN	DP	SP
Assimilation TEP	Global SN	SP	MIM
	Arctic SN	SP	MIM
	Baltic SN	SP	MIM
	Atlantic SN	SP	MIM
	Mediterranean SN	SP	MIM
MERSEA Portal (MIM)			All Users

Table 4-1 : Sub-Networks decomposition

The following nomenclature is also presented in the next section

Data providers (DP)
Service Providers (SP)
System management users (SM)
Privileged users (PU)
Standard product users (STP)
Specific product users (SPP)
General public (GP)

4.2 GEOLAND Architecture Overview

4.2.1 GEOLAND AREA OF INTEREST with respect to HALO objectives

GEOLAND aims to provide and establish geo-information products and services to support GMES, using available Earth Observation resources, and integrating them with existing models into pre-operational end-user applications. The GEOLAND products and services aim at monitoring land cover and vegetation addressing the GMES priorities "Global Vegetation Monitoring" and "Land Cover Change in Europe", "Environmental Stress in Europe".

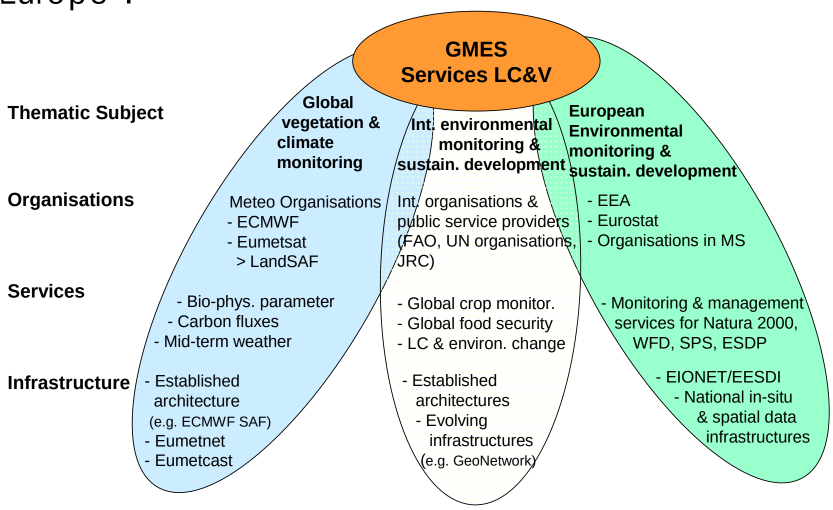
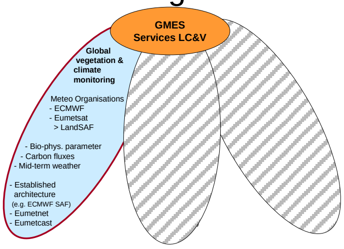


Figure 4-4: GEOLAND main families of service & user segments

The project is organized into three regional and three global observatories, plus two core services which support the observatories with cross cutting products.

From the point of view of HALO, the global Observatory Natural Carbon Fluxes (ONC) and global Core Service Biophysical Parameters (CSP) have to be predominantly covered due to their strong interaction with atmosphere and Ocean IPs and their more developed operational maturity. Indeed, the diversity of issues addressed and the different status of operational maturity make it difficult to find a common summary of the used and exchanged data and products on local observatories.

In this context GEOLAND analysis should be restricted to the left hand side= “Global Vegetation Monitoring”:



As an evidence of the possible synergy with GEMS, we can notice that, in practice, this service area of GEOLAND is essentially managed by meteorological organisations (ECMWF, Eumetsat). The main characteristics of this “Meteo-world” are**:**

- ✓ Large European institutions
- ✓ Operating centralised facilities to provide basic parameters (ECMWF model)
- ✓ Or service networks for downstream value adding addressing dedicated applications/user groups (EUMETSAT SAF model).

4.2.2 GEOLAND preliminary functional architecture

The figure below gives a preliminary overview of GEOLAND functional architecture. The general idea of the data flow structure in GEOLAND is that the **two core services** provide the main input data sets of the observatories. The input data of the global observatories are mainly based on satellite observations whereas in-situ observations are mainly used for validation purposes.

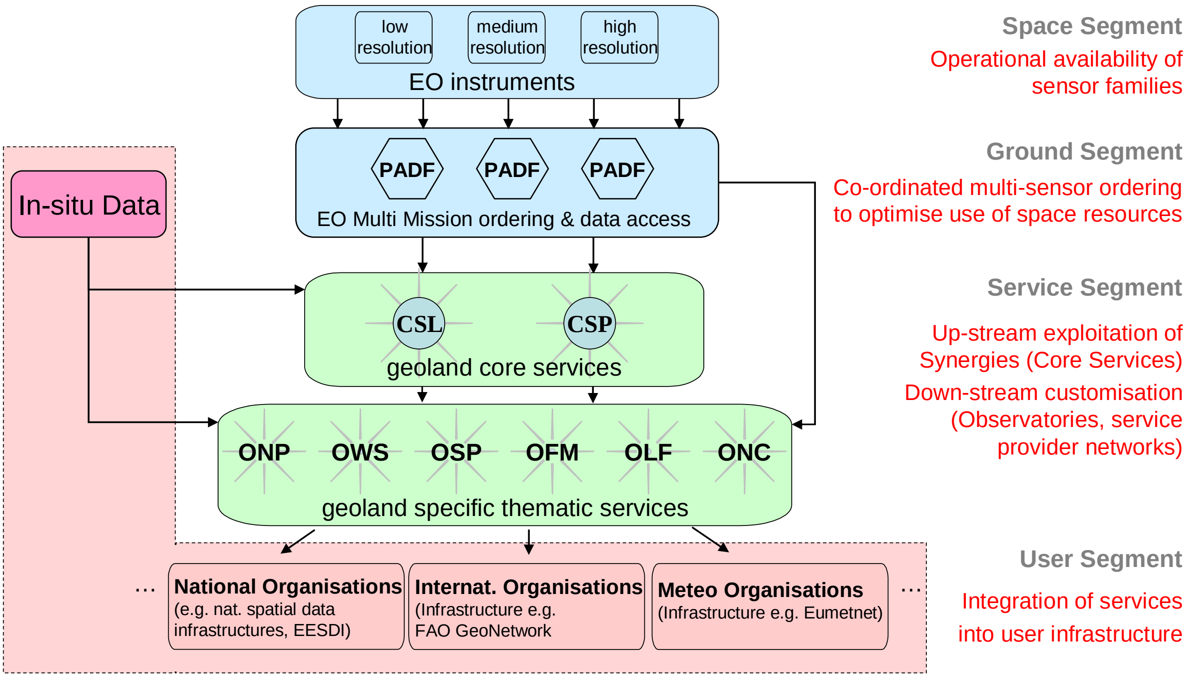


Figure 4-5: GEOLAND preliminary functional architecture

One key element of this architecture in terms of global efficiency to order and to get access to Earth Observation data products is the PADF.
The "PADFs" (**P**rocessing **A**rchiving & **D**issemination **F**acility) would take place between the **observation data collection** systems (space or aerial) and the Service providers (core services and observatories). They would offer to any of these service providers a simplified and standardized access to advanced Earth Observation data products. Several PADFs would be deployed in the global system, dividing up tasks and workloads, ensuring robustness and availability of product provision service, allowing proximity with GMES Service providers and **adaptation to their needs**. The PADFs would propose the minimum of basic functions that are required at this level: **provision of standardized data products** with adapted interfaces for Service providers.

4.3 GEMS Architecture Overview

4.3.1 Overview of GEMS system architecture

The GEMS project (Global and regional Earth-system Monitoring using Satellite and in-situ data) consists of sub-projects covering the following aspects of atmospheric composition monitoring (see Figure 4-6).

- Global greenhouse gases (GHG)

- Global reactive gases (GRG)
- Global aerosols (AER)
- Regional air pollution (RAQ)

The thematic sub-projects are complemented by a subproject dealing with the implementation of an operational production system and a sub-project on validation. Figure 4-6 shows the building block of the GEMS system and their interaction.

According to the definition of the HALO work scope, predominantly the global operational activities, which will be integrated in ECMWF operational system, have to be considered in HALO. GEMS integration at ECMWF ensures that the existing ECMWF infrastructure including telecommunications, high performance computing and archive facilities can serve GEMS. Nine institutions in Europe, applying 11 models, conduct GEMS regional air quality modelling activities.

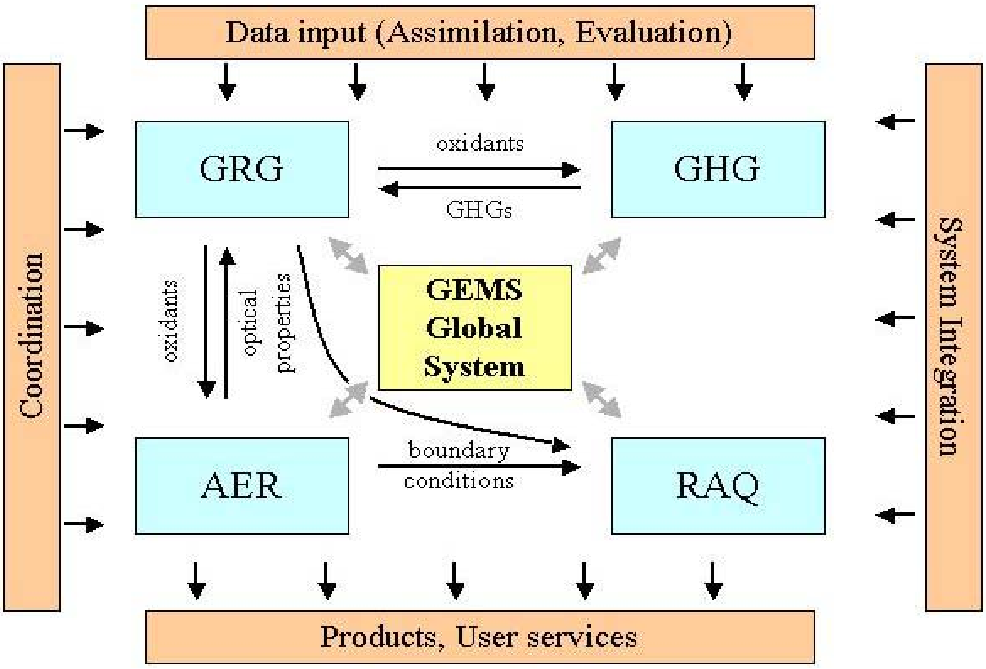


Figure 4-6 Schematic of GEMS building blocks, GHG - Greenhouse gases, GRG - Global reactive gases, AER - Aerosol, RAQ - Regional air quality.

GEMS is going to provide fields of the atmospheric composition and will improve knowledge about sources and sinks. GEMS uses satellite data and applies atmospheric models to obtain information about the atmospheric composition in global and regional scales. Satellite data are exploited either by assimilation in atmospheric models or by independent retrieval algorithms. Ground based and airborne in-situ measurements will be used mainly for validation purposes but for assimilation and inverse modelling as well.

As part of the ECMWF forecast system, GEMS will develop and implement a comprehensive, validated, and novel operational global data assimilation / forecast system for atmospheric dynamics and composition, which combines all available remotely sensed and in-situ data to achieve global monitoring of the dynamics and composition of the atmosphere from global to regional scales (50 km) and covering the troposphere and stratosphere.

The deliverables will include operational monitoring and forecast of three-dimensional global distributions (four times daily with a horizontal resolution of 50 km) of greenhouse gases, global reactive gases, and global aerosols. Further, GEMS will provide retrospective analysis of the global atmospheric chemical composition for the period 2000-2005.

The global assimilation / forecast system will provide initial and boundary conditions for nine operational regional air-quality and 'chemical weather' forecast systems. This will provide improved operational real-time air-quality forecasts. It will also provide a methodology for assessing the impact of global climate changes on regional air quality.

The GEMS system will use as far as possible the existing infrastructure provided by (i) WMO's World Weather Watch and (ii) European resources in information technology.
By far the largest international flows of Earth Observation data (both in-situ data and space-based data on meteorology, oceanography, atmospheric composition and hydrology) are the daily operational global observation and forecast exchanges (and the corresponding monthly global exchanges of climate summaries) by means of the WMO operational World Weather Watch System (see Figure 4-7).

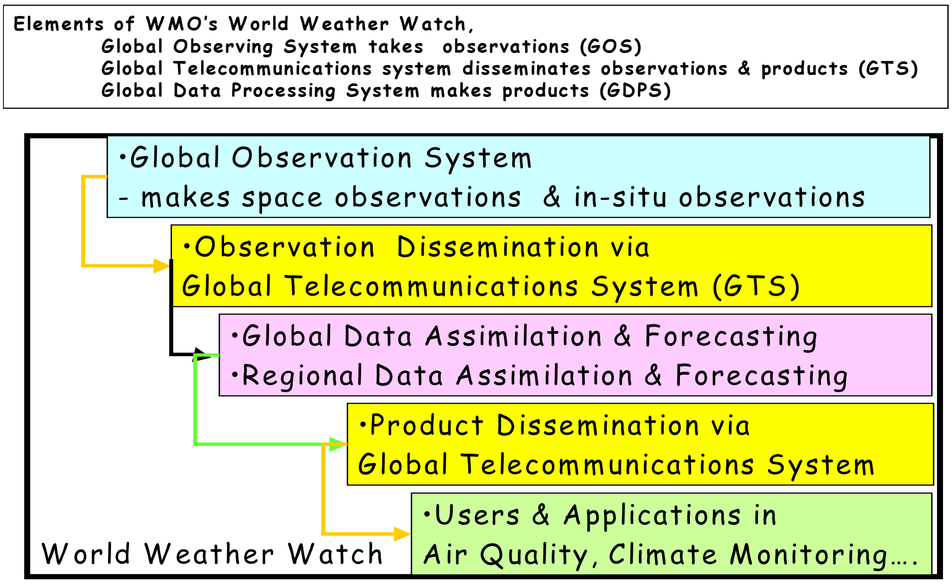


Figure 4-7 The main elements of the WMO operational World Weather Watch System

According to present planning, data transfer in GEMS will be based on the following networks:

- GTS/RMDCN (WMO)
- Standard internet/ FTP
- Dedicated Networks such as EUMETCast or GEANT

GTS (Global telecommunication system) is the network of the meteorological community. The access to GTS is restricted and governed by WMO rules. RMDCN, which is part of GTS, is the European Regional Meteorological Data Communications Network between European meteorological centres including ECMWF and EUMETSAT. The GTS consists of an integrated network of point-to-point circuits, and multi-point circuits which interconnect meteorological telecommunication centres. Data on GTS are encoded in GRIB (meteorological fields) or BUFR (observations) format.

4.3.2 Data volumes and reception times of satellite data

The main challenge for operational production might become the acquisition of satellite data in a timely manner. This paragraph communicates anticipated and current satellite data volume numbers for the GEMS project as well as typical reception times of selected satellite data.

Figure 4-8 shows a summary of the estimated daily transfer rates of satellite radiance data needed for GEMS. Figure 4-9 shows the transfers summed up for retrieved products used in the GEMS subprojects.

Much of the high volume sounding data required for GEMS (IASI, 1500 Mb/day) is already being received in real-time, or will be received in real-time at ECMWF. The main challenge will be the transfer of MODIS (2 x 600 MB/day) and MERIS radiance data. Most of the satellite data to be used in GEMS will come from satellite agencies (ESA, EUMETSAT, NASA) and in the case of some products from research centres (KNMI) and universities (Universities of Bremen, Heidelberg).

The operational production in real or near real time mode requires data reception within a given time window. The end of this time windows is called cut off time since data arriving later can not be used anymore in the model and assimilation run. The reception of the satellite data is controlled by the dissemination performance of the provider and the network capabilities. Satellites products tend to be more slowly distributed since time for processing is needed. Figure 4-10 shows average arrival times at ECMWF of AIRS, MSG, ASAR and Meteosat data for the 0 UTC model run with a time window from 21 to 3 UTC (cut off). In the given example about 90 % of the MSG (Meteosat second generation) and Meteosat data, 80% of the ASAR (Envisat) and 50% of the AIRS (Aqua) data reach the centre in time. The cut off time of 3 hrs is typical for numerical weather prediction but the GEMS production might have longer cut off times.

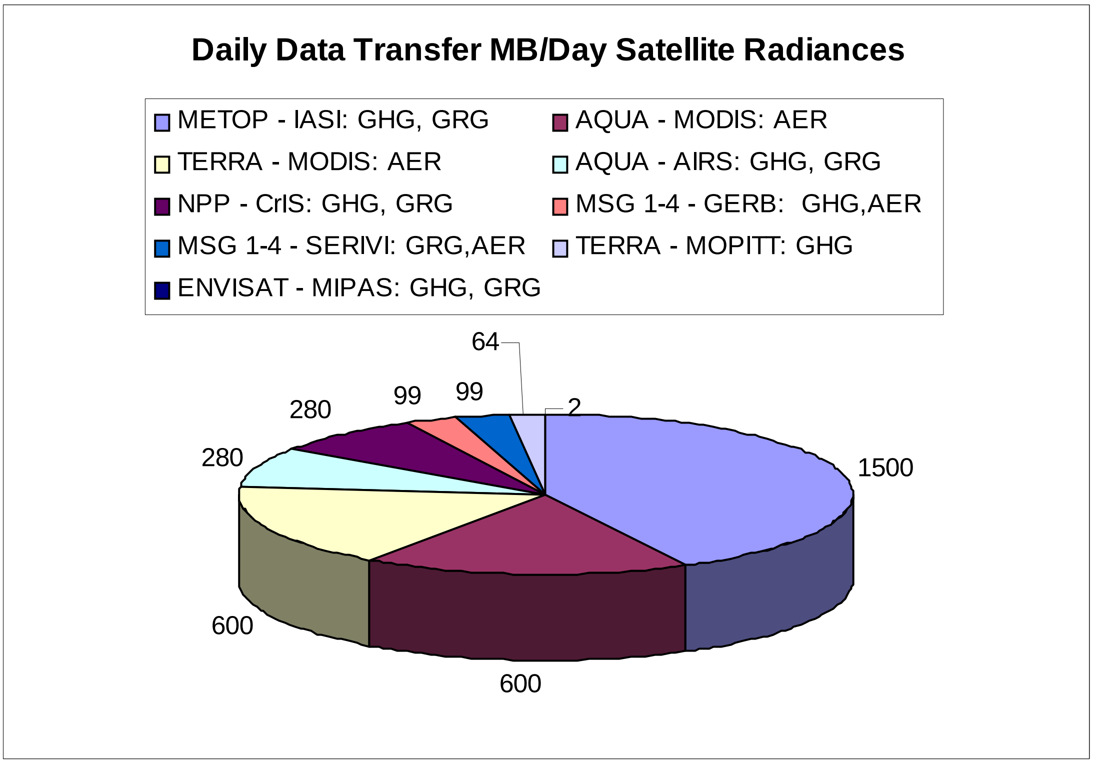


Figure 4-8 Estimated daily transfer in MB/day of satellite radiance data for GEMS.

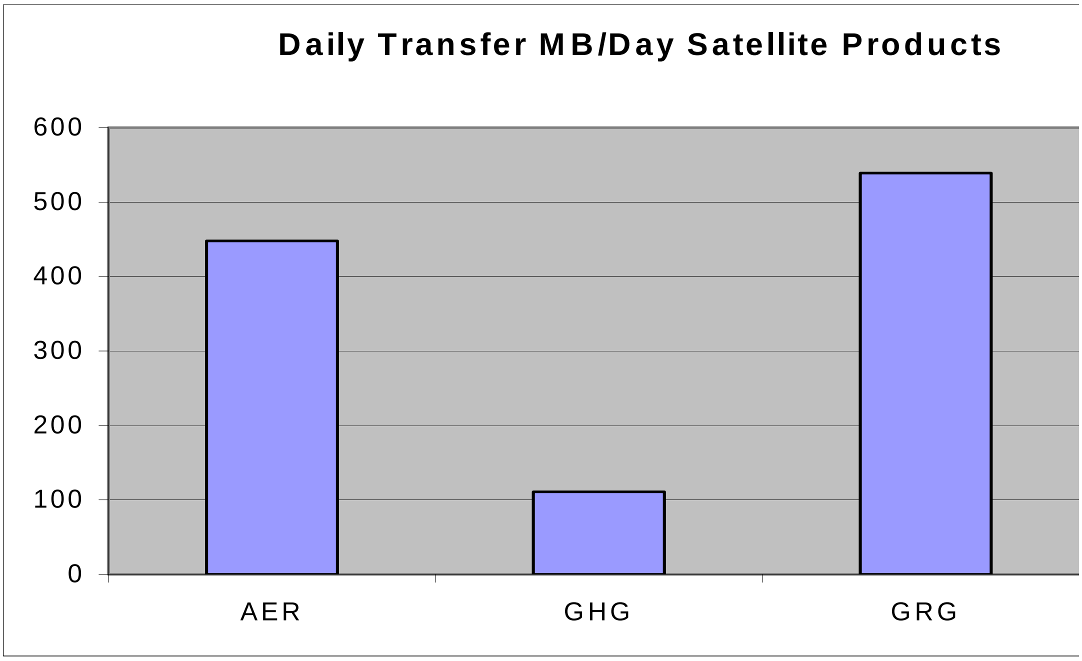


Figure 4-9 Estimated daily transfer in MB/day of satellite products need in GEMS subproject Aerosol (AER), Greenhouse gases (GHG) and Global Reactive Gases (GRG).

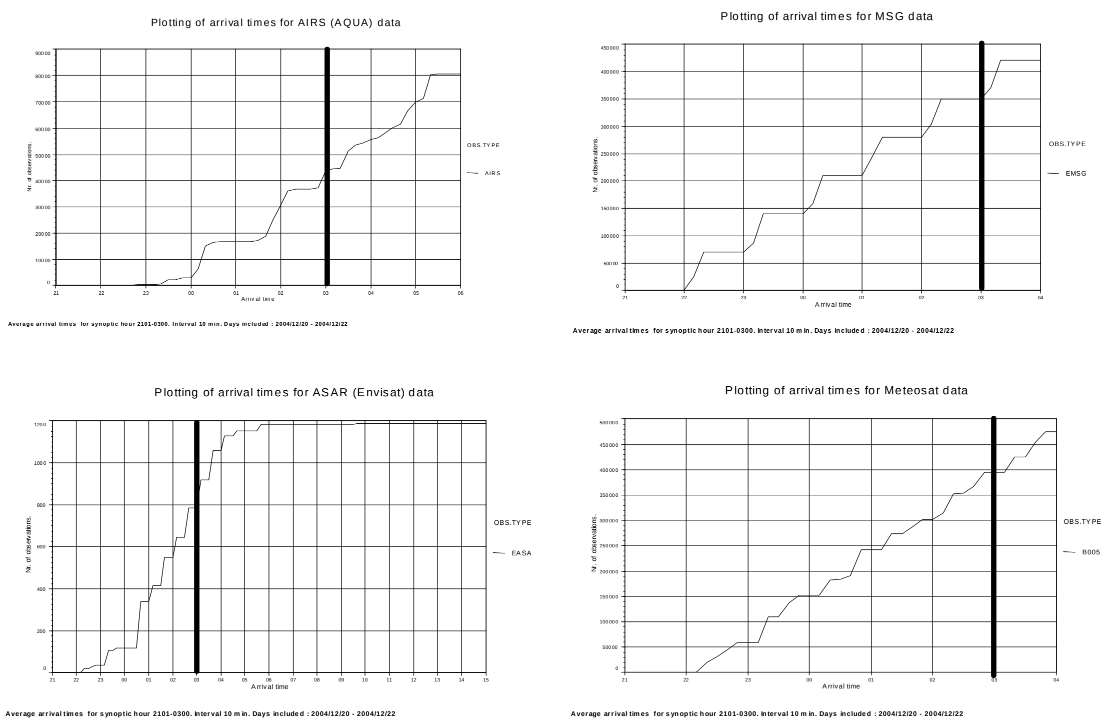


Figure 4-10 Arrival times for satellite data at ECMWF for the 0 UTC cycle from AIRS, MSG, ASAR and Meteosat. The time window lasts from 21 to 3 UTC (cut off). Data arriving after cut off could not be used in this cycle.

4.3.3 Access to GEMS products

Global Analyses and Forecasts

GEMS will provide a number of different access mechanisms for GEMS users, depending on user capabilities, user requirements, and the type of the products needed. It is expected that most users will require access via the internet to the validated analysis and forecast datasets provided on the GEMS. Other users will require dissemination for pre-operational testing, in which case the distribution mechanism may be either the Internet or the dedicated meteorological circuits. Finally, researchers will need access for experimentation on the GEMS data assimilation and forecast systems, through user interfaces accessed over the Internet. The access and distribution mechanisms represent are based on considerable software investments made in earlier projects funded by the EU, and others funded by the meteorological community.

The ERA-40 Reanalysis project (http://www.ecmwf.int/research/era/) funded by the 5th Framework programme has built substantial software systems to undertake extended reanalyses of historical data and to provide access services to the validated results. These capabilities will be used in GEMS to produce the GEMS reanalyses and to provide access to the outputs and to visualise the outputs.

GEMS Regional Products

Validated GEMS regional products will be archived and displayed on the GEMS web-site with a common graphics package and in a common format, for ease of comparison and to facilitate common verification protocols.

GEMS Diagnostic products

Validated GEMS diagnostic will be archived and displayed on the GEMS web-site with a common graphics package and in a common format, for ease of comparison and to facilitate common verification protocols.

Management & Dissemination of Experimental NRT Global Runs

In the last year of the GEMS project it is planned to undertake experimental Near-Real-Time (NRT) and pre-operational running of the full GEMS system including global analyses and forecasts and regional analyses and forecasts. National Institutes and Research Institutes of all EU Member States will be welcome to participate in this experimental work. GEMS will appoint a GEMS dissemination manager at the beginning of year 3 of the programme to prepare the real-time aspects of the experiment. The telecommunications means most appropriate for each participant will be used, either RMDCN or the Internet.

4.4 Overview of Relevant Meteorological Telecommunication Networks

The meteorological community uses dedicated networks for the operational transfer of both observations and model output. The main advantages are bandwidth and availability guaranties. The so-called Regional Meteorological Data Communication Network (RMDCN) is the most relevant for HALO as it connects the European National Meteorological Services and ECMWF. The dissemination of operational products by ECMWF to its member states is one of the major tasks of the RMDCN. An overview of the connections is given in Figure 4-11. The connection speeds, Committed Information Rate (CIR) in kilobit per second (kbps) of France, the UK, and ECMWF are detailed in Figure 4-12 to Figure 4-14. The CIR may be asymmetrical, e.g. the CIR from France to the UK is 128 kbps, the one from the UK to France is 384 kbps. Daily usage statistics can be found at
<http://www.ecmwf.int/services/computing/cos/d/charts/rmdcn/daily/>.

The RMDCN may be a means of communication between the three IPs involved in HALO. In particular, the dissemination of meteorological fields and other GEMS products belongs to the tasks of RMDCN. However, modifications, e.g. in bandwidths may become necessary.

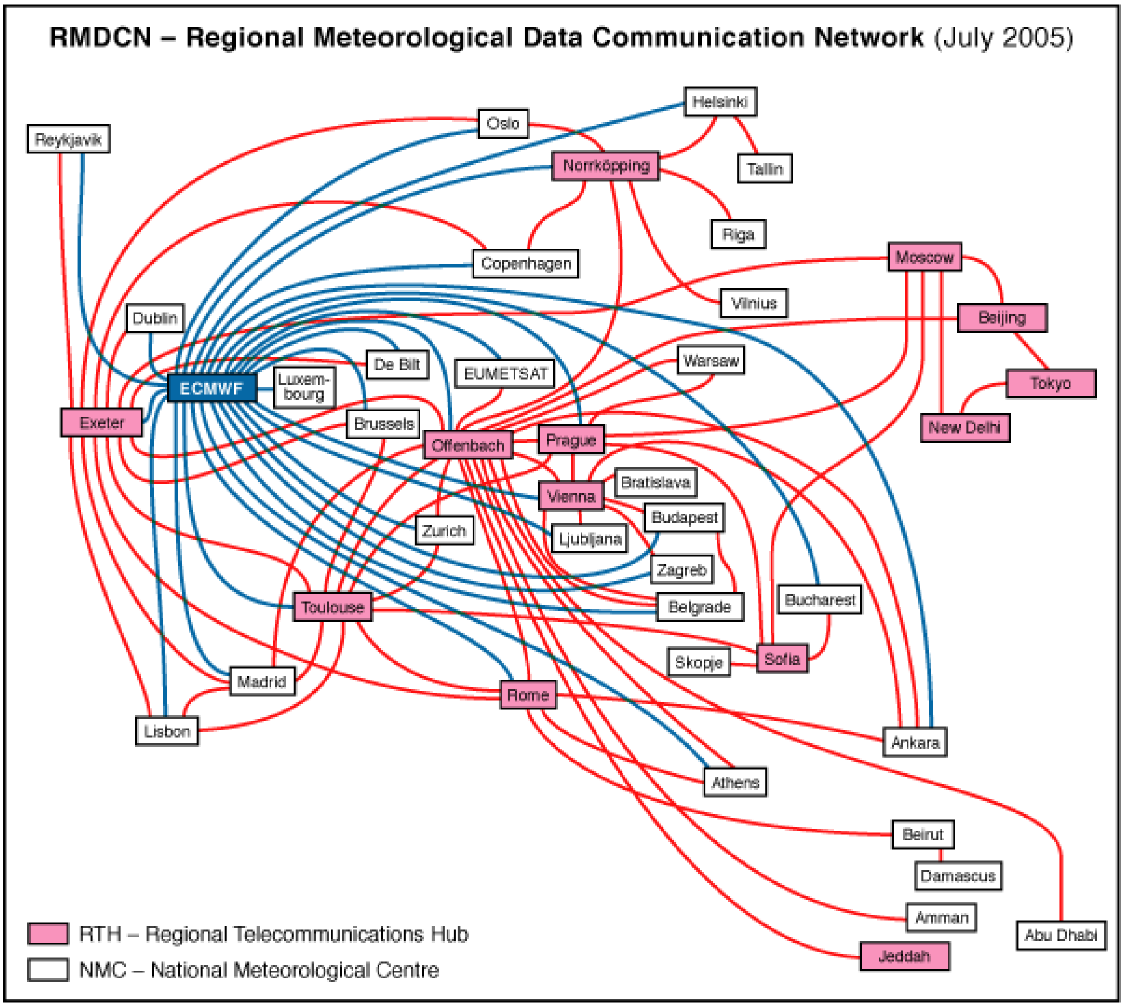


Figure 4-11: Overview of RMDCN connections (source: www.ecmwf.int)

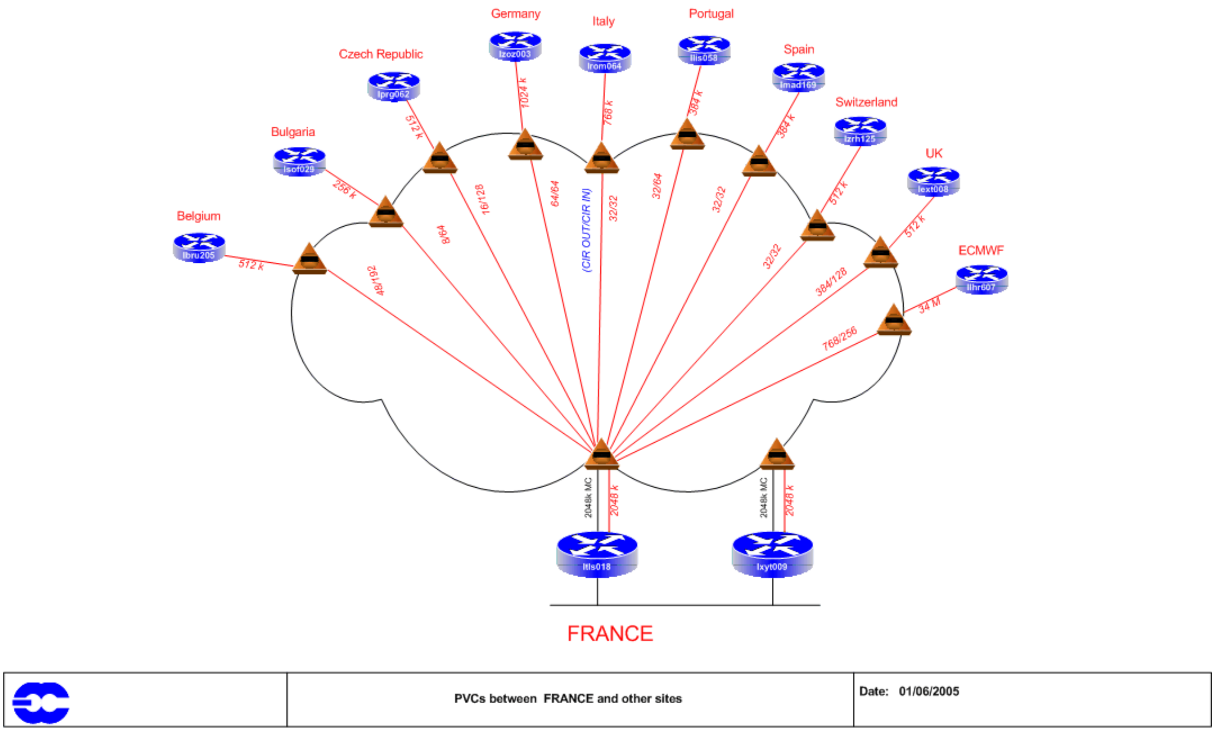


Figure 4-12: RMDCN Connections of France (source: www.ecmwf.int)

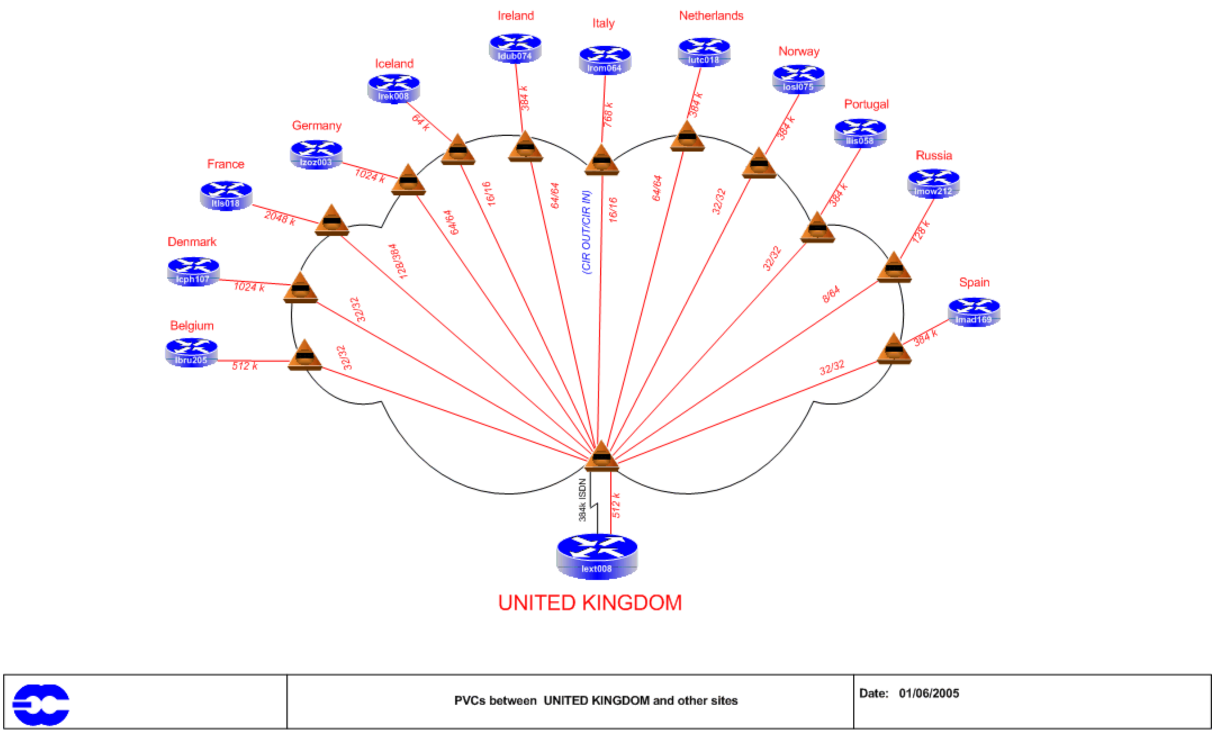


Figure 4-13: RMDCN connections of the UK (source: www.ecmwf.int)

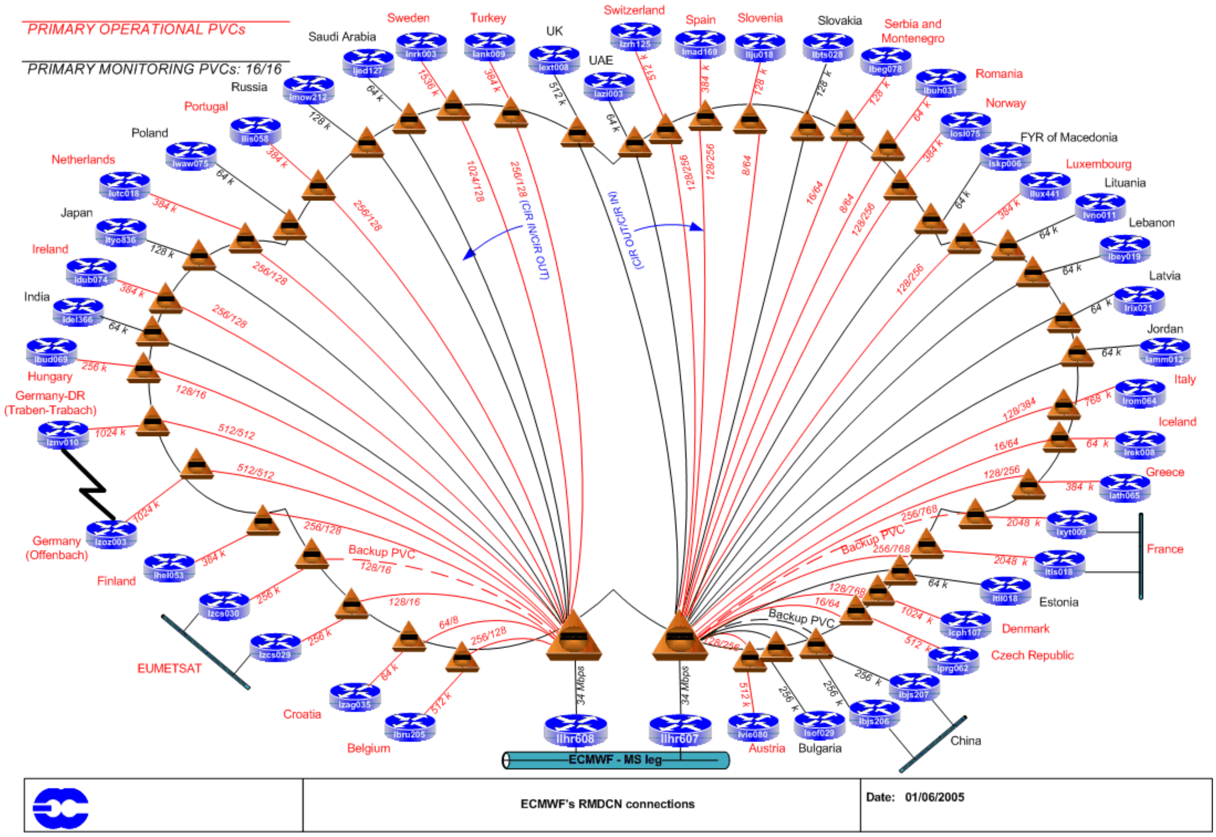


Figure 4-14: RMDCN connections of the ECMWF (source: www.ecmwf.int)

The RMDCN is part of WMO’s Global Telecommunication System (GTS). The GTS is organised in three hierarchy levels: Main Telecommunication Network (MTN), Regional Networks, and National Networks, see www.wmo.intweb/www/TEM/gts.html. RMDCN is the regional network for region VI, Europe, in the GTS framework. Figure 4-15 shows the commited data rates of the RMDCN in the context of the GTS. Connections are dedicated to both the Regional Network for Europe and the MTN.

The MTN connections in the regions VI (“Network II”) and IV (“Network I”) are also shown in Figure 4-16. It becomes obvious that data exchanged between these regions is routed through the UK, Russia, or Japan. The suitability of the MTN connection form Washington to Exeter for the transfer of commonly required American satellite observations has to be assessed in HALO. The guaranteed data rate of the connection given in Figure 4-15 is 64 kbps.

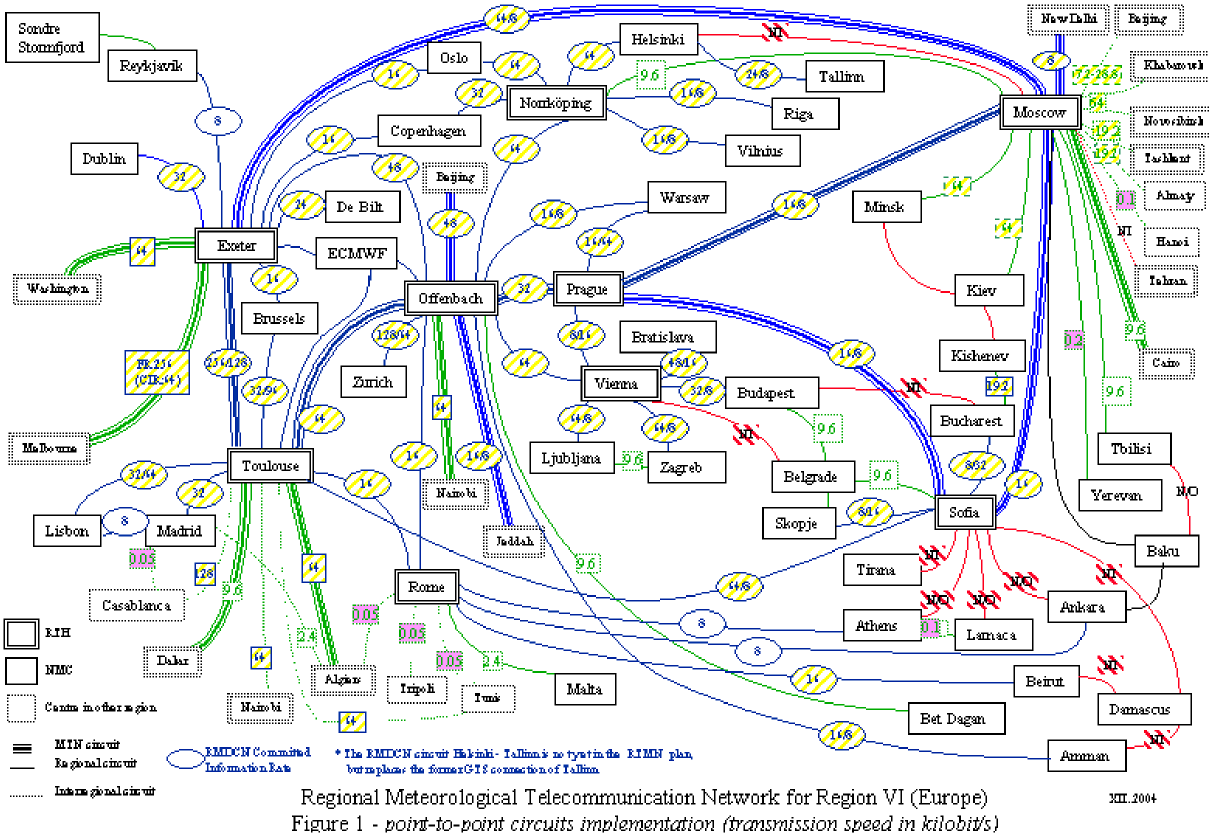


Figure 4-15: Regional Meteorological Network for Europe (source: www.wmo.int)

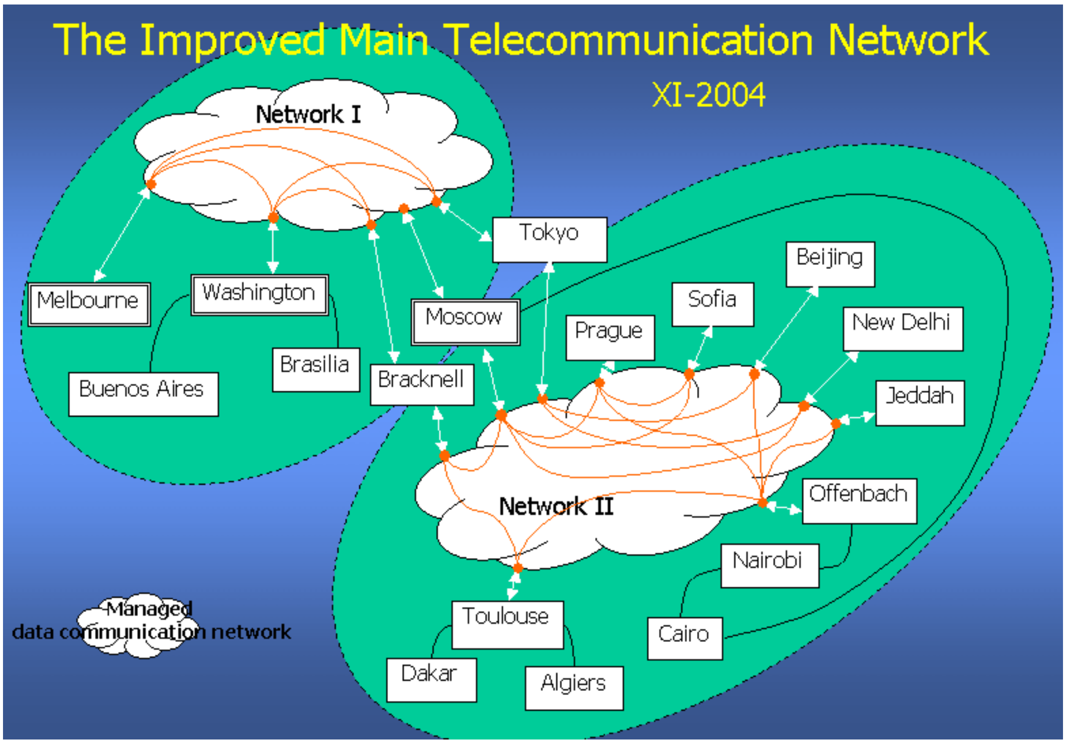


Figure 4-16: MTN in Europe and North America (source: www.wmo.int)

5 Requirements Analysis

The architecture must be designed to fulfil the requirements of the interacting parts of the system components.

They can be considered as high level requirements or low level requirement depending on some criteria that are of major importance, such as the author, the application targeted, the scale (National, Local, Regional…) of an application.

The requirements will be provided by the various actors of the processing chain (value-adders, experts …). Right now, actors are using operational services, current data centres for data delivery and ordering for scientific or operational purpose. The requirements shall also refer, on the one hand, to the unfulfilled needs, on the other hand, to some general top level requirements also covered partly, fully or not at all by all existing systems. The matrix shall also refer to that information by providing the data used, the technology already used (server, portal …), the percentage of fulfilling of the requirement. This last information shouldn't be neglected since it is of major importance for the system architecture definition.

Data Flows Analysis, Data and Product Analysis, Functions Analysis templates, as referred to in the following sections, are also required in order to size the system, to get homogeneous information from various themes (ocean, land atmosphere). Cross-correlation shall be maintained between these three templates, and exhaustivity of the information can be held, for instance, by maintaining independently a statistics Matrix, referencing the users and the services (wrt themes), with an inventory for each node of the list of requirements from the Functions Analysis template.

The industrial partners are aware of the fact that perhaps not all entries can be complete, or that other categories need to be introduced. Therefore, the templates have been considered as reminders of the type of information the industrial partners need for their work.

5.1 Data Flows Analysis

The objective of this section is to list the Data Flows in and between the various Integrated Projects, illustrating the exchanges of Data between Land and Atmosphere and between Ocean and Atmosphere, respectively. The Data flows may refer to a specific Theme or may be used for one or more end products or value-added product. This specified in the columns 6 and 7. The individual Product IDs are defined section 5.2.
The columns are populated as explicitly as possible with existing flows in accordance with the activity of each interviewed IP.

Columns 1 to 3 *are characterizing the Data Flow with a title and source and Destination*

4-Author
This columnl refers to the author of the Data flow, for the traceability.

5-Criteria
This column is of major importance for the dimensioning of the architecture since it provides all the constraints or requirements related to this Data flows.

6-Theme/Product
This column refers to the theme or service associated to the requirement and can give some example of output product also linked with the related service.

Each table is subdivided into three parts:

- **Interacting** describes data flows between the IPs.
- **External** *describes data flows from a third party to one or several of the IPs.*
- **Internal** *describes data flows between different parts of the same IP.*

1	2	3	4	5	6	7
Data flow	**Source**	**Destination**	**Author**	**Criteria**	**Theme/Product**	**Affected Product IDs**
interacting						
Meteorological forcing fields	ECMWF	Ocean Model Centre	Desaubies	Regular distribution, real-time analysis and forecasts, Regional High resolution models	Meteorological forecast/NWP Bulletin	O6-8, 12-20
GEMS global aerosol products	ECMWF	Mersea retrieval centres	Flemming	to be checked, initially research mode only	Atmospheric Aerosol data for atmospheric corrections in retrieval	O16-17 A6-7
external						
Satellite data	ESA, EUMETSAT, NASA, NOAA	MERSEA Satellite TEP	Desaubies	Regular	Along track, validated	O2-6
Satellite products	SAT -TEP, GHRSST, SSALTO, OSI/SAF	Ocean Model Centre	Desaubies	Regular	Merged, gridded, validated products	O7
In-situ observations	GDAC, RDAC, ARGO, GTSPP, DBCP,	In-situ Data Centre	Desaubies	Regular + On-demand	High quality controlled, merged gridded products, climatology	
In-situ observations in real time	ARGO	In Situ - TEP	Kaiser	Real Time flow	ARGO data in real -or near real -time, with QC flags	

internal						
In-situ observations in real time	In Situ - TEP (from ARGO)	Ocean Model Centre	Desaubies	Real Time flow	ARGO data in real -or near real -time, with QC flags	O9
In-situ observations in real time	In Situ - TEP	?	Kaiser	Real Time flow	GSUD / VOS, Ocen time series / BBCP	O10-11

Table 5-1: Mersea System (TEPs) Data Flow Analysis

1 **Data flow**	2 **Source**	3 **Destination**	4 **Author**	5 **Criteria**	6 **Theme/Product**	7 **Affected Product ID**
			interacting			
Meteorological forcing fields for land surface models	ECMWF	Geoland-ONC	Calvet	Regular	Air temperature/humidity, wind speed, precipitation, incoming radiation (short and longwave)	
Geoland Global products	Geoland-CSP	GEMS	Leroy	Regular + On-demand	Generic Land Cover (300 m – 1 km resolution	L5
Geoland CSP-OFM vegetation CO2	GEOLAND-OFM	GEMS @ ECMWF	Calvet	to be checked, initially research mode only	Land use change and forest fires	L2-3, 9
geoland ONC vegetation CO2	GEOLAND-ONC @ ECMWF	GEMS @ ECMWF	Flemming	to be checked, initially research mode only	Vegetation data as input for emission models (biogenic and fires): CO2 fluxes, above-ground biomass, stomatal conductance	L12-13
GEMS global aerosol products	ECMWF	geoland retrieval centres	Flemming	to be checked, initially research mode only	Atmospheric Aerosol data for atmospheric corrections in retrieval	A6-7
			external			
Satellite data	ESA EUMETSAT NOAA / NASA	Geoland-CSP	Leroy	Regular + On-demand	Satellite observation to infer information about the land surface, in three areas : vegetation, radiation, water	L4, 6, 11

in-situ data	Meteo	Geoland-CSP	Leroy	Regular + On demand	Rainfall	L10
In-situ data	Research labs	Geoland-CSP	Leroy	On demand	Validation data for Vegetation, radiation, soil moisture products	
Satellite data	SPOT Image, NASA	Geoland-CSP	Leroy	On demand	Validation data for Vegetation & Land cover products	L5
Satellite data to be assimilated	ESA EUMETSAT NOAA/NASA CNES	Geoland-ONC	Calvet	Regular + On-demand	Satellite observation to infer information about the land surface and the vegetation status.	L5, 6, 11
In-situ data for validation	Fluxnet	Geoland-ONC	Calvet	On-demand	CO2 and water fluxes	
In-situ data for validation	GAW	Geoland	Kaiser	On-demand	radiative surface fluxes	
internal						

Geoland GlobalGobal products	Geoland-ONC @ECMWF	GEMS @ ECMWF	Leroy	Regular + On-demandTo be checked, initially research mode only (TBC)	Biogeophysical Parameters (Rainfall for water cycle, burned area, active fire and LAI for trace gas emission)Vegetation data as input for emission models (biogenic and fires) (TBC)	
Satellite forcing fields for land surface models	Geoland-CSP	Geoland-ONC	Calvet	Regular	Improved precipitation fields and incoming radiation (short and longwave)	L4

Table 5-2: geoland Data Flow Analysis

1	2	3	4	5	6	7
Data flow	**Source**	**Destination**	**Author**	**Criteria**	**Theme/Product**	**Affected Product ID**
interacting						
Geoland CSP-OFM vegetation CO2	GEOLAND-OFM	GEMS @ ECMWF	Calvet	to be checked, initially research mode only	Land use change and forest fires	L2-3, 9
geoland ONC vegetation CO2	GEOLAND-ONC @ ECMWF	GEMS @ ECMWF	Flemming	to be checked, initially research mode only	Vegetation data as input for emission models (biogenic and fires): CO2 fluxes, above-ground biomass, stomatal conductance	L12-13
GEMS global aerosol products	ECMWF	Mersea retrieval centres	Flemming	to be checked, initially research mode only	Atmospheric Aerosol data for atmospheric corrections in retrieval	A6-7
GEMS global aerosol products	ECMWF	geoland retrieval centres	Flemming	to be checked, initially research mode only	Atmospheric Aerosol data for atmospheric corrections in retrieval	A6-7 O16-17
Meteorological forcing fields	ECMWF	Ocean Model Centre	Desaubies	Regular distribution, real-time analysis and forecasts, Regional High resolution models	Meteorological forecast/NWP Bulletin	O6-8, 12-20

Meteorological forcing fields for land surface models	ECMWF	Geoland/ONC	Calvet	Regular	Air temperature/humidity, wind speed, precipitation, incoming radiation (short and longwave)	
Geoland Global products	Geoland-CSP	GEMS	Leroy	Regular + On-demand	Generic Land Cover (300 m – 1 km resolution	L5
external						
Satellite data	ESA, EUMETSAT, NOAA / NASA (UNI-BREMEN, UMW)	ECMWF	Flemming	operational	Raw radiances and satellites products on atmospheric species concentration and fire count/ burnt area	A9
in-situ data	Scattered provider (NILU, EEA, national and regional authorities)	ECMWF MPI KNMI RAQ Centres	Flemming	regular	In situ observation for validation	
CO2 concentration	www.cmdl.noaa.gov, gaw.kishou.go.jp	GEMS @ ECMWF	Kaiser	on demand	validation data for CO2 assimilation. open access on the internet.	
internal						
GEMS global products	ECMWF	GEMS RAQ Centres (6)	Flemming	operational	Boundary conditions for reginal air pollution models	A1-7

Table 5-3 : GEMS Data Flow Analysis

5.2 Data and Products Analysis

The objective of this section is to list all the Data required for the exchange between the interacting parts of Land and Atmosphere, and between Ocean and Atmosphere, respectively . We also indicate the data and Product that are dimensioning on an architectural point of view by considering for example some huge quantity of data, very high rate required for data transfer or low delay of acquisition. In addition to the data and product templates remote sensing products shall be characterized following this template:

The mandatory criteria and categories for data and products templates are the columns 1 to 4:

2-Origin

- *Observation (In-situ, Satellite)*
- *Model and assimilation products (global, regional)*
- *Service information products for users from model products (aggregated information or for special purposes such as oil spill forecast)*

3-Operational constrains of provision

- *Operational mode - real time (RT) and near real time (NRT)*
- *Off-line or Re-analysis mode*

- *Research mode*

4-Interaction and communality

- *Direct product exchange between IPs*
- *Unaccomplished data needs from individual Ips*
- *Common data needs*

5-Security (Optional)

- *Restricted access or not (to which actors' category, DP, SP, ...).*

The other criteria are optional and significantly explicit.

6-Local, Regional or Global (Optional)
Gives the scale of the data, to be used at local, regional or global level.

7-Service
Gives the services associated to the data, if explicit.

8-Provider
Give the identity of the data provider of the data.

***Columns 9 to 14** are related to the characterization of the data itself.*

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Data / Product ID	**Origin**	**Operational Constraints**	**Interaction and communality**	**Security[1]**	**Local, Regional or Global**	**Service[2]**	**Provider**	**Format[3]**	**Temporal Coverage**	**Spatial Coverage**	**Access Delay**	**Revisit Delay**	**Access Time**
O2: SST	NOAA, ESA, EUMETSAT, NASA	NRT, re-analysis, research	High, direct exchange	NR	Global, with regional products		SAT -TEP		2000 to present	Global to regional	daily	daily	
O3: Altimetry	NOAA, ESA, EUMETSAT, NASA	NRT, re-analysis, research	Low	NR	Global, with regional products		SAT -TEP		1992 to present	Global to regional	daily	Daily to 10 days (for merged products	
O4: Ocean Color	NOAA, ESA, EUMETSAT, NASA	NRT, re-analysis, research	Medium	NR	Global, with regional products		SAT -TEP		2004 to present	Global to regional	daily	Daily to 10 days, but cloud dependent	

[1] Security levels (i.e. data policy) are to be defined for the system to be delivered. It is expected to make all data available freely for **research purposes**, and in compliance with Resolution 40 of WMO, notwithstanding specific restrictions from some providers.
[2] The MERSEA system plans on *classes of users* for whom different *services and products* will be available.
[3] Preferred data format is netcdf, because of its ability to carry meta-data; this issue is part of the ongoing design work plan (MERSEA Information Management).

O5: Sea Ice	NOAA, ESA, EUMETSAT, NASA	NRT, re-analysis, research	High	NR	Polar regions		SAT -TEP		?	Polar regions	daily	3 to 10 days	
O6: Wave / winds	ESA, EUMETSAT + NWP	RT, re-analysis	High	NR	Global, with regional products		NWP service		Real time	Global to coastal	6 hrs	6 hrs	
O7: NWP	Assimilation Models	RT, re-analysis, forecasts	NA	?	Global, with regional high resolution		NWP service		ERA- 40 (40 years)	Global, with regional high resolution	6 hrs	6 hrs	
O8: Meteorological field Data	ECMWF, NWP	RT, re-analysis, forecasts	Ocean corrected fluxes	?	Global, and regional		NWP service, Forcing TEP		ERA-40 (40 years)	Global, and regional	6 hrs	6 hrs	
O9: Argo / GTSPP	In-situ TEP	RT, and delayed mode	Low	NR	Global		In situ TEP, ARGO GDAC		Since 1992	Global	24 hrs	10 days	
O10: GOSUD / VOS	In-situ TEP	RT, and delayed mode	Medium	NR	Global + regional		In situ TEP, GOSUD, GDAC		Since 1985	Global + regional	24 hrs to weekly	monthly	
O11: Ocean time series/BBCP	In-situ TEP	RT, and delayed mode	high	NR	Global, regional		In situ-TEP; OTS , GDAC		Since 2001	Point observatories	24 hrs to weekly	daily	
O12: GHG forecast	ECMWF	operational											

O13: GHG analysis	ECMWF	off line											
O14: GRG forecast	ECMWF	operatio nal											
O15: GRG analysis	ECMWF	off line											
O16: AER forecast	ECMWF	operatio nal											
O17: AER analysis	ECMWF	off line											
O18: GRG forecast	ECMWF	operatio nal											
O19: GRG analysis	ECMWF	off line											
020: RAQ forecast	9 RAQ Centres	operatio nal											

Table 5-4: Data and Product Analysis Mersea

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Data / Product ID	**Origin**	**Operational Constraints**	**Interaction and communality**	**Security**	**Local, Regional or Global**	**Service**	**Provider**	**Format**	**Temporal Coverage**	**Spatial Resolution**	**Access Delay**	**Revisit Delay**	**Access Time**
L1: Leaf Area Index (LAI)	Satellite observations Model & assimilation	NRT Re-analysis Research	Direct exchange	Non restricted (NR)	Global	Geoland	CSP		–1980 up to present	8km/ 1 Km	1 day	10 days	1 hour
L2: Burned areas	Satellite observation	NRT Re-analysis Research	Direct exchange	NR	Global	Geoland	CSP		1998 up to present	1 Km	1 day	10 days	1 hour
L3: Active fires	Satellite observation	NRT Re-analysis Research	Direct exchange	NR	Global	Geoland	CSP		1998 up to present	1 km	1 day	10 days	1 hour
L4: Rainfall	Satellite observation	NRT Re-analysis Research	Direct exchange	NR	Global	Geoland	CSP		1998 up to present	50 Km	1 hour	1 day	1 hour
L5: Land Cover	Satellite observation	Differed time delivery	Direct exchange	NR	Global	Geoland	CSP		2000 up to present	300 m – 1 km		5 years	
L6: Land & Vgt Monitoring	Satellite images high resolution	Research	data need partly common	NR	Local	Geoland	, Spot Image, NASA			10 – 30 m			

L7: TBD	In-situ Products	Research	TBD									
L9: Burned areas	observation		TBD		Regional	Geoland	OFM		2000-2003	1 Km		
L10: Rainfall	observation		TBD		Global/ Regional	Geoland	CSP			5 Km		
L11: Land & Vgt Monitoring	observation		data need partly common			Geoland	ESA, Eumetsat, Spot Image, NASA			60km-earth disk		
L12: CO_2 and Water fluxes	model & assimilation	off line	perhaps exchange	no	global	geoland	ONC ECMWF	GRIB	2000-05			
L13 : Carbon storage	model & assimilation	reanalysis	perhaps exchange	no	global	geoland	ONC ECMWF	GRIB	2000-05			

Table 5-5: Data and Product Analysis geoland

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Data / Product ID	**Origin**	**Operational Constraints**	**Interaction and communality**	**Security**	**Local, Regional or Global**	**Service**	**Provider**	**Format**	**Temporal Coverage**	**Spatial Coverage**	**Access Delay**	**Revisit Delay**	**Access Time**
A1: GHG analysis + forecast	model & assimilation	operational	perhaps exchange	access if needed (a)	global	GEMS	GEMS ECMWF	GRIB	starting 2008	global	RT		?
A2: GHG re-analysis	model & assimilation	off line	perhaps exchange	open access for scientific community (oa)	global	GEMS	GEMS ECMWF	GRIB	2000-05	global	avail. 2007		?
A3: GHG surface fluxes	variational inversion	off line	perhaps exchange	a	global	GEMS	GEMS ECMWF	GRIB	2000-05	global	avail. 2007		?
A4: GRG analysis + forecast	model & assimilation	operational	perhaps exchange	a	global	GEMS	GEMS ECMWF	GRIB	starting 2008	global	RT		?
A5: GRG re-analysis	model & assimilation	off line	perhaps exchange	oa	global	GEMS	GEMS ECMWF	GRIB	2000-05	global	avail. 2007		?
A6: AER analysis + forecast	model & assimilation	operational	perhaps exchange	a	global	GEMS	GEMS ECMWF	GRIB	starting 2008	global	RT		?

A7: AER re-analysis	model & assimilation	off line	perhaps exchange	oa	global	GEMS	GEMS ECMWF	GRIB	2000-05	global	avail. 2007		?
A8: RAQ analysis + forecast	model & assimilation	operational		?	regional	GEMS	GEMS regional partners)	TBD		regional			?
A9: Satellite data and products	Satellite	off line and operational	data need partly common	?	global		ESA, Eumetsat	GRIB, NetCDF tc	from project start	global	operational and off line		

Table 5-6: Data and Product Analysis GEMS

5.3 Functions Analysis

The objective of this section is to help prepare a good overview of the requirements of the systems concerning the HALO candidate solutions. Templates for communicating the requirements are defined. These are filled in in a separate document (AD6).

The categories provided should be considered as a guide. They do not have to be used if they do not seem appropriate. Alternatively the IPs can specify their own categories if they whish to do so. The requirements are to be fulfilled by the interviewed IP in its related thematic. A requirement can be specific, related to one specific need in one specific theme or generic related to the overall infrastructure. In that way, the actor shall list the requirement that he consider to be key driving for the architecture or its activity.

The requirements shall refer to the three main functional groups as identified in HALO and other general groups

- Data acquisition
- Data sharing
- Data dissemination

Table 5-7 to Table 5-9 are the templates for the three IPs. The meanings of the columns is explained below. lists the possible function classes to used in column 1.

1	2	3	4	5	6	7	8
Function	**Requirement Description (Criticity 1..5)**	**Theme**	**Author**	**From Project**	**Data and Product ID**	**User(s)**	**Operational/User constraints**
Data acquisition	Increase frequency for data collect (5)						
	Visualize media on land (3)						
	Visualize intervention area (4)						
Data sharing							
Data dissemination							

Table 5-7 : Function Analysis Mersea

1	2	3	4	5	6	7	8
Function	**Requirement Description (Criticity 1..5)**	**Theme**	**Author**	**From Project**	**Data and Product ID**	**User(s)**	**Operational /User constraints**
			CSP				
Data acquisition	Archiving Access Log Book Data mining Data transformation Catalogues autonomy Storage media Data updating management Indexation Data processing traceability	Architecture	CSP	GEOLAND	Level 1 satellite data, all satellites observing land	CSP	
Data sharing	None						
Data dissemination	Data updating management Data validation and referencing Indexation Scalability Data processing traceability	Architecture	CSP	GEOLAND	All CSP products	GEMS ONC, OFM, OLF Science community	
			ONC				

Data acquisition	Archiving Access Log Book Data mining Data transformation Catalogues autonomy Storage media Data updating management Indexation Data processing traceability	Architectur e	ONC	GEOLAND	Level 3 satellite products (vegetation radiation, water quantity, land cover change)	ONC	
Data sharing	None						
Data disseminatio n	Data updating management Data validation and referencing Indexation Scalability Data processing traceability	Architectur e	ONC	GEOLAND	Level 4 products (derived from process models)	National / European Environmen t Agencies Water, civil proction, agriculture / forest agencies Science community	

Table 5-8 : Function Analysis geoland

1	2	3	4	5	6	7	8
Function	**Requirement Description (Criticity 1..5)**	**Theme**	**Author**	**From Project**	**Data and Product ID**	**User(s)**	**Operational /User constraints**
Data acquisition	Timely acquisition of high volume satellite data (high volumes)						
	Timely acquisition of in situ data from many different providers						
Data sharing	Archive facilities (storage and quick access) for the large output of GEMS re-analysis and operational activities						
Data disseminatio n	Timely distribution of global GEMS global data to the regional RAQ centres						

Table 5-9 : Function Analysis GEMS

2 - Requirement Description

This column can be populated using functions as listed below but other requirements might be more appropriate

Categories	Function
DATA DISSEMINATION	Data Updating Management
	Data Validation and Referencing
	Indexation
	Scalability
	Data processing Traceability
DATA SHARING	Interoperability with Catalogues
	Voice Communication
	Video Communication
	Communication with Mobile
	Web Interface
	Telecom Interface
	Collaborative Work
DATA ACQUISITION	Archiving
	Access Log book
	Data Mining
	Data Transformation
	Billing
	Alert Services (on event)
	Multi-lingual communication
	Catalogues Autonomy
	Archiving
	Storage media
	Normalization
	Data Updating Management
	Data Validation and Referencing
	Indexation
	Scalability
	Data processing Traceability
	Collaborative User directory

Table 5-10 : Example of Functions Classes

The criticity is

1 minor but not really useful right now, maybe later
2 minor
3 Medium
4 Major but not mandatory (Work around exists
5 Mandatory

3-Theme

Give the theme or service for which is applicable the requirement. If the requirement is general, for architecture design, it will be set as Architecture.

The columns 4 and 5 are related to the traceability of the requirement

4-Author

Identify the author of the requirement

5-From Project

From which project is the requirement originated if issued from one EC project or else.

6-Data and Product ID

what are the data and product considered if Applicable (see templates) :
which user(s) is (are) involved.
Operational/User constraints :

7-Users

Shall refer to which kind of user is concerned for the requirement

8-Operational/User constraints

Shall give an inventory of any constraints for fulfilling the requirement, for example refer to any missing component of the infrastructure such as fast network, dedicated thematic portal...
This column can be populated using functions as listed below

6 Appendix: Interacting parts of the IPs

This appendix lists preliminary elements of an assessment of the interacting parts of the land, ocean and atmosphere segments in terms of data (considering data provider - data user) and infrastructure (technical interfaces between IPs, telecommunications ...). It is build on the "interacting part of GEMS, MERSEA and Geoland" report (AD4) but formulated from an industrial point of view in complement to the IPs perception presented in this report.

First of all, let's summarise the foreseen limited HALO work scope. According to the HALO proposal, only the global and continental or basin scale model and data activities with operational commitment will be considered. In particular these activities are:

- MERSEA's core models (Mercator, FOAM, MFS and TOPAZ)
- Geoland's global observatories, mainly Observatory Natural Carbon (ONC) and Core Sevice Geophysical parameters (CSP)
- GEMS operational and data assimilation system at ECMWF from the GHG[5], GRG[6] and AER[7]-subprojects.

Therefore, the challenge for HALO will be to propose a sufficient link between GEMS and MERSEA's modelling activities as well as between GEMS and an operational follow up on geolands CSP. A smooth interaction of ONC and GEMS activities in the framework of ECMWFs model has to be ensured.

The figure below present the structure and transfer lines of the interacting parts of the three IP extracted from the "interacting part of GEMS, MERSEA and Geoland" report. Input data is provided by satellite agencies, weather services and in-situ observation systems. The raw in-put data is process and retrieved by data centres within the IP, symbolized by a star-like shape, and passed over to the model centres, symbolized by circles.

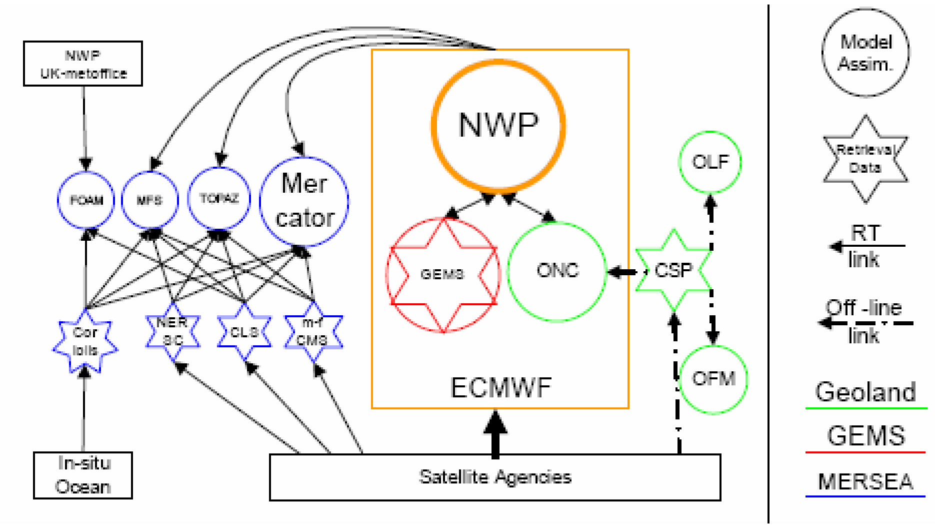


Figure 6-1: System structure and data flow of the operational model activities in MERSEA, Geoland and GEMS. (ECMWF requires in-situ observations, too.)

In term of infrastructure, IPs Interaction could be separate into two main categories:

1. Common data needs: satellite, in-situ

   Although IP use often the same instruments, the approach of the IPs towards satellite data processing differs: MERSEA has parameter-specific retrieval centres which provide satellite products for assimilation to the modelling centres. GEMS fosters the assimilation of direct (raw) radiance data in its operational model at ECMWF without the intermediate retrieval. The Core Service Geophysical Parameters in geoland acts a single data and retrieval centre for the demands of the three global observatories. For instance, a common approach for the use of operational meteorological satellites of the MSG, MetOp, MTG, NPP and NPOESS series might be beneficial (Eumetsat EO data are provided via **RMDCN/GTS** and/or the **EUMETCast facilities**).

   All IPs use in situ data to validate the modelling and retrieval activities. The challenge of the in-situ data is their collection from a large variety of data providers with different operational commitment. The HALO report on GEMS lists about 40 different sources for in-situ data used in GEMS.

2. Direct product exchange between IPs

   Coordinating envisaged direct product exchange between IPs has the highest priority within the HALO activities. A good example of the inter-dependencies of the IPs is their contribution to the carbon cycle as presented in the "interacting part of GEMS, MERSEA and Geoland" report.

Direct product exchange solutions should be scalable to enable creation of new multi-IP products/services in the future.

Since different IPs have to be satisfied under the constraint of the existing infrastructure we can not assume to build a single overall infrastructure replacing the existing ones. The idea is rather to define tools that could be usable by any service providers to manage all external data flows.

Is this context, another representation of the data flow between IP could be:

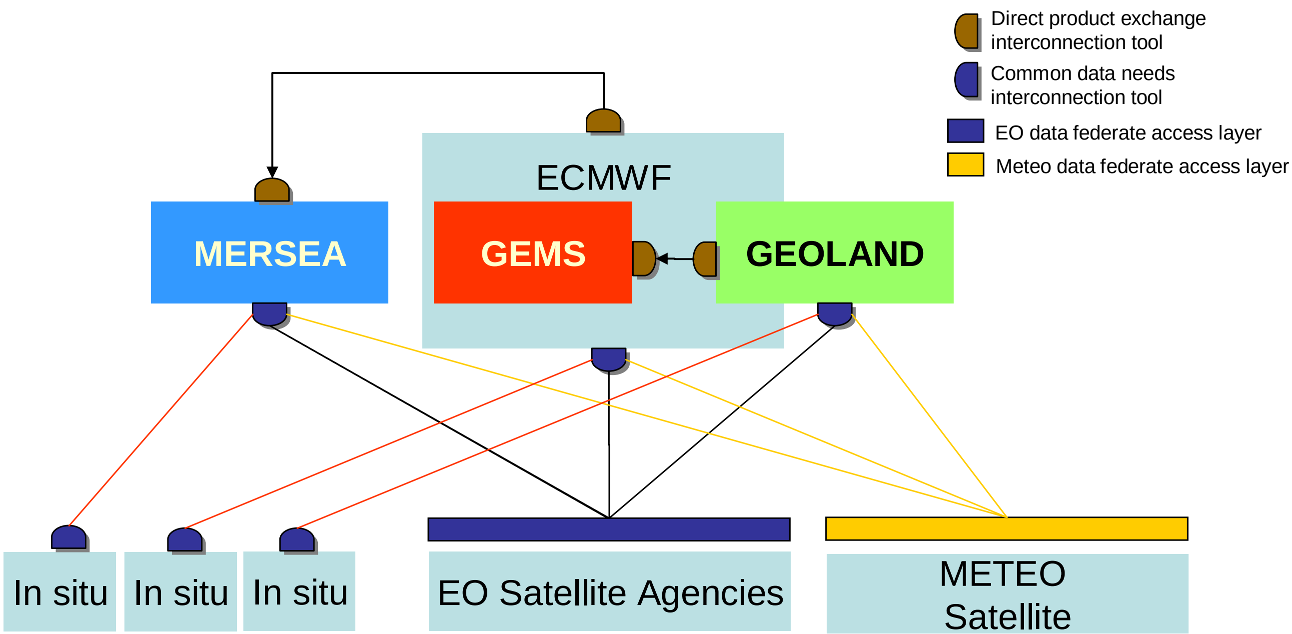


Figure 6-2: interconnection tools & layers to manage data flow between IPs

These tools and layers will have to be designed (through candidate solution) taking into account every specific requirements expressed by IP representatives with the template tables defined in this document.