7th Framework Programme
ENV.2010.4.1.2-2
Integrating new data visualisation approaches of earth Systems into GEOSS development

Project Nr: 265178

QUAlity aware VIsualisation for the Global Earth Observation system of systems

Deliverable D7.1
Document pilot case studies in a standard format

Version 1.0

Due date of deliverable: 01/07/2011
Actual submission date: 31/07/2011

Document control page				
Title	D7.1 Document pilot case studies in a standard format			
Creator	ES_UAB			
Editor	ES_UAB and MN_UAB			
Description	Document pilot case studies in a standard format			
Publisher	GeoViQua Consortium			
Contributors	GeoViQua Partners			
Type	Text			
Format	MS-Word			
Language	EN-GB			
Creation date	01/06/2011			
Version number	1.0			
Version date	31/07/2011			
Last modified by	MN_UAB			
Rights	Copyright © 2011, GeoViQua Consortium			
Dissemination level		CO (confidential, only for members of the consortium)		
	X	PU (public)		
		PP (restricted to other programme participants)		
		RE (restricted to a group specified by the consortium)		
	When restricted, access granted to:			
Nature	X	R (report)		
		P (prototype)		
		D (demonstrator)		
		O (other)		
Review status		Draft	*Where applicable:*	
	X	WP leader accepted		Accepted by the PTB
		PMB quality controlled		Accepted by the PTB as public document
	X	Coordinator accepted		
Action requested	X	to be revised by all GeoViQua partners		
	X	for approval of the WP leader		
		for approval of the PMB		
	X	for approval of the Project Coordinator		
		for approval of the PTB		
Requested deadline				

Revision history			
Version	Date	Modified by	Comments
0.0.1	22-03-2011	MN_UAB	Initial datasets descriptions (PD_CREAF, PS_UAB, MN_UAB, GM_UAB, JB_UREAD, DN_52N, JS_S&T) and thesaurus (AZ_UAB, MN_UAB, XP_UAB)
0.1	01-06-2011	ES_UAB	Created the basic content of the deliverable
0.2.0	10-06-2011	MN_UAB	Structuring the content
0.2.1	15-06-2011	ES_UAB	Adding and deepening contents (AZ_UAB, MN_UAB, PS_UAB, IS_CREAF, PD_CREAF)
0.2.2	20-06-2011	MN_UAB	Minor edits
0.3.0	21-06-2011	MN_UAB	Review of datasets descriptions (JB_UREAD, DN_52N, PP_CEA, JS_S&T)
0.3.1	06/07/2011	ES_UAB	Second PTB meeting. Workshop comments.
0.4.0	15/07/2011	ES_UAB	Revision by GeoViQua partners (VG_ESA, DC_AST, JB_UREAD)
0.5.0	20/07/2011	ES_UAB	Revision (XP_UAB, JM_CREAF) and minor edits (AZ_UAB, GM_UAB, PS_UAB, AR_UAB)
0.6.0	25/07/2011	ES_UAB	Revision by GeoViQua partners (KY_UREAD, VG_ESA, DC_AST, DN_52N)
0.7.0	28/07/2011	ES_UAB	Introduction of UML Models (JM_CREAF)
0.7.1	29/07/2011	MN_UAB	Carbon cycle dataset description (PP_CEA)
0.7.2	30/07/2011	ES_UAB	Minor edits
1.0	31/07/2011	JM_CREAF	Coordinator acceptance

Institution	Contributors
52N	DN “Daniel Nüst”
AST	DC “Dan Cornford”,
CEA	PP “Philippe Peylin”
CREAF	IS “Ivette Serral, JM “Joan Masó”, PD “Paula Díaz”
ESA	VG “Veronica Guidetti”
S&T	JS “Joost Smeets
UAB	AR “Anna Riverola”, AZ “Alaitz Zabala”, ES “Eva Sevillano”, GM “Gerard Moré” MN “Miquel Ninyerola”, PS “Pere Serra”, XP “Xavier Pons”
UREAD	JB “Jon Blower”, KY “Kevin Yang”

Table of Contents

Abstract

This deliverable is the first focused on Work Package 7 (WP7) "Pilot case studies" in the GeoViQua 7th Framework Programme ENV.2010.4.1.2-2 project framework. WP7 activities will be based on the collection of a varied sample of datasets that will undergo scientific testing of concrete aspects and demands of the project, and from which a limited set of illustrative scenarios will be generated, thus validating the complete integrated results of the project. These targets require a deep knowledge of the diversity of alternatives implied in the choice of datasets. Only a thoughtfully selected sample will allow for testing and validating of results in accordance with the specifications of GeoViQua.

The rationale of pilot cases selection is depicted, leading to the gathering of suitable and controlled datasets (*e.g.,* thorough comprehension of provenance, associated errors, peculiarities, potential use, highlights). To achieve these aims, a specific datasets description template has been purposely designed embedding relevant aspects of geospatial databases, including data sources, processing methods and quality elements. These main areas are defined by means of the corresponding dictionary tables, carefully regarding international standards (suchlike ISO standards) with particular contributions when applicable. In addition, relations to concepts currently under development within the Global Earth Observation System of Systems (GEOSS) and other work packages in GeoViQua are considered (*e.g.,* user requirements, use cases, quality elicitation and encoding). Scenarios address integration of the instruments and outcomes in GeoViQua, considering components within a background beyond the features of the datasets in themselves, embracing a dynamic flow of processes, the interaction with the communities of users, providers and industry, interfaces, and definition of system boundaries. Although some gaps have been identified, the analysis of the datasets gathered at date reveals that the sample is representative of the current state-of-the art in Earth Observation, covering a wide range in sources, methods, quality indicators and scales. This results in the documentation of pilot cases testing datasets in a standard format and sketching of scenarios design.

The reader has to realise this deliverable is envisaged for the early stages of the project. Many relevant components of WP7 are still under development. Hence, the workflow and ensuing stages are barely outlined: suitable datasets from the selected sample are to be subjected to experimental testing of the innovative technologies and tools to be developed in GeoViQua, and derived scenarios will provide an overview of the services that will be demonstrated.

1. INTRODUCTION

This deliverable is focused on selecting pilot cases and characterizing them in a standardized form. Using suitable datasets, work package 7 (WP7) in GeoViQua aims at testing several aspects of the project, such as:

Quality parameterization (subsequently to WP2, WP3 and WP6 results)
Quality search components (subsequently to WP4 results)
Quality aware visualisation components (subsequently to WP5 results)
GEO Label (subsequently to WP6 results)

These experiences will provide practical data on testing real datasets.

The pilot cases will be selected to cover the Global Earth Observation System of Systems (GEOSS) community topics spectrum and especially the main testable aspects of geospatial data.

Testing datasets have been selected among those in the GEOSS Common Infrastructure (GCI) and some other sources available from the partners. GeoViQua is a European project framed within GEOSS acting as a testing platform, evaluating GEOSS progress and providing reports on usability, visualisation, parameterisation and search with a special remark on quality issues. Thus, continuous inputs and feedbacks are implied between GEO tasks and GeoViQua tasks and even inside GeoViQua's tasks and work packages, assuring coherence between projects, targets and tasks.

The GEOPortal (http://www.geoportal.org/web/guest/geo_home) provides convenient access to the full range of GEOSS data and information, linking to GCI where components and services are being deployed and catalogued. The current GEOSS Web portal is provided on the basis of an ESA's activity. GeoViQua solutions are expected to be integrated into current GEOSS Web portal logic.

Briefly, WP7 targets are to be fulfilled carrying out the following tasks:

Task 7.1
Explore GCI services and components and GEOPortal search and visualisation tools. Evaluation of usability and detection of gaps. Selection of pilot cases.
This task therefore revises GEOSS datasets (available through the GEOPortal) and, when available, associated quality data. Pilot cases will come directly from GEOSS and/or from sources provided by GeoViQua partners, accomplishing selection criteria.

Task 7.2
Quality product parameterisation. Classification of main factors of geospatial datasets, with special emphasis in quality components, and subsequent analysis of selected datasets.
Previous results from WP2 (user requirements detection), WP 3 (quality elicitation; in particular, tasks 3.1, 3.2, 3.4 and 3.5) and WP6 (encoding) are inputs for this task. Outputs will provide a valuable feedback for WP2 and WP3 and, in turn, an input to task 7.4.

Task 7.3 Validation of search services and visualisation (datasets and quality data) employing a heterogeneous sample of pilot cases. Inputs and feedback from and to respectively WP4 (geosearch) and WP5 (visualisation).

Task 7.4 Validation of GEO Label. This task provides an input to Geo Task ST-09-02 (which aims at engaging the research communities in GEOSS, promoting awareness and benefits of GEO and multidisciplinary interactions.

Partners involved in WP7 are CREAF, UAB, 52N, AST, UREAD, CEA, ESA and S&T. This work package has been assigned 15.5% of the total workload in GeoViQua, as defined in the Description of Work. It is regarded as a crucial step, implying scientific testing of real datasets. As stated in the tasks description above, other work packages will receive feedbacks from preliminary results of WP7. Final results are expected to improve quality awareness issues and be of practical appliance in GEOSS.

The present deliverable (**D7.1 Document pilot case studies in a standard format**) reports the progress on WP7 especially focusing on Tasks 7.1 and 7.2.

At present, pilot cases have been either provided or facilitated by partners, some of them already catalogued and integrated in the GEOSS GCI, linked to Societal Benefit Area's (SBA) and Communities of Practices (CoP) as defined in GEOSS and selected with the aim of coping with a varied sample of parameters (*e.g.,* temporal and spatial resolution, data sources, preprocessing and processing methods, quality parameters) in order to enable testing of a wide range of situations. This favours the design of illustrative scenarios that will integrate the surrounding environment of the selected corresponding datasets, suchlike practical appliance of quality elicitation, search and visualisation tools to be developed in GeoViQua, linking components and services, at the same time embedding other issues (*e.g.,* GEO Label, interaction with communities or system boundaries). By means of collecting a varied sample, a correct matching between particular experiments and suitable datasets is guaranteed. It is clear that the diversity of features of the datasets imply different approaches. This is to say that not all datasets will be valid for all experimental tests in GeoViQua, as certain features will make of some of them better candidates for a specific test.

As for WP7, a thorough control of the testing dataset is essential for the success of the tasks to be undertaken. This is to say, a thorough understanding on the data peculiarities, including provenance, guaranteeing traceability and, if available, a certain control of associated quality measures must be carefully watched by each dataset provider in this work package. The contributions of partners to WP7 enrich the testing dataset, at the same time generating positive synergies in workflow and interactions with other work packages in GeoViQua.

In the second Project Technical Board meeting held at ESA-ESRIN facilities in Frascati (Italy) it was agreed that these criteria in the selection of pilot cases testing datasets were determinant for the scientific success of GeoViQua, constituting a convenient way to provide solid results coming from a comprehensive validation and experimental tests of datasets and associated quality data.

This document describes:

1. Framework of WP7 in GeoViQua (section 1)
2. Classification of the pilot case activities (section 2).
3. Selection criteria of the pilot case testing datasets (section 2).
4. The design of a pilot cases description sheet of relevant parameters of datasets and documentation of the relevant metadata (section 3).
5. The analysis of selected datasets and associated quality data and notes on methodological design of pilot case scenarios derived from them (section 4).
6. Workflow on WP7: tasks, roles and contributions (section 5).
7. Up-to-date conclusions of WP7 (section 6).

A common glossary has been annexed in order to ease understanding of frequently employed terms as defined in the context of GeoViQua as a whole, and thereby appearing in this document (also acting as acronyms list). In this sense, a detailed GeoViQua dictionary (http://twiki.geoviqua.org/twiki/bin/view/GeoViQua/GeoviQuaTerminology) is being implemented in GeoViQua’s site, which can also be consulted.

2. SELECTION OF PILOT CASES

From the beginning of the project, the partners realise the significance of the decision making process in the selection of pilot cases for the success of GeoViQua. WP7 is a work package in which the experimental component is crucial. Therefore, as scientists well acknowledge, the quality of the datasets and their suitability for the targets of the project determines the scope and reliance on the results. In addition, in the frame of GEOSS potential data providers, clients and, in the end, users of geospatial data must also be taken into account.

Three main aspects will be considered:

-the dataset in itself (sources, methods, parameters, topics covered, social impact)

-associated quality data of the dataset (quality measures of raw data and products)

-data chain (*e.g.,* discovery time, propagation, services, delivery, stakeholders)

A first classification approach of the pilot cases in WP7 generated the following table, derived from the Description of Work, although extended in further GeoViQua discussions:

Pilot Cases	Main dataset providers
Remote sensing	S&T and UAB
Carbon cycle	CEA-LSCE
Climate	UREAD and UAB
Water cycle	UAB
Disasters	*To be developed*
Agriculture	UAB
Air quality	52N
Land use	CREAF
Marine	ESA and UREAD

This table includes a first draft of partners responsible for providing datasets. The next step was to explore whether some GEOSS resources could be used for this task in GeoViQua and, in the end, decide on the contributions on real datasets coming from partners.

Considering the wide range of factors implied and the variety of available samples, this initial classification evolved into the definition of pilot case testing datasets and pilot case scenarios. Not to loose sight of GeoViQua scope and means, it seemed reasonable to:

1. Define criteria to adequately select the dataset sample to be subjected to experimental testing in GeoViQua
2. Attach when possible to terms and references used in GEOSS (*e.g.,* SBAs)
3. Introduce the idea of selecting scenarios from several testing datasets, with dissemination purposes

Moreover, each partner, having a comprehensible idea of the project requirements, commented on the initial classification and accordingly checked availability of datasets.

The revision of the compilation of datasets confirmed the pertinence of defining a standardized classification form.

The definition of selection criteria was in fact reflected on a pilot case dataset description template, designed on purpose to ease decision-making on dataset selection. The template, which will be explained in section 3, contains description fields for geospatial datasets, integrating the corresponding dictionary tables (assimilated to UML code lists). Once having revised the template contents of some datasets, the appropriateness of the outlined selection criteria was more tangible.

2.1 Pilot cases rationale

As has been aforementioned, the pilot cases description template and several discussions within GeoViQua generated the definitive selection criteria for the pilot case testing datasets:

-provenance and traceability: prioritizing contributions from partners and/or datasets registered in GEOSS Global Common Infrastructure (GCI) Component and services Registry (CSR).

In this context, it is to be noted that not all datasets are already registered in GEOSS. Nonetheless, the purpose is to achieve this goal during the span of GeoViQua. Due to administrative procedures, depending on the entity status as regards representing data ownership, sharing, restrictions and licenses, it takes some time to be registered in the GEOSS Components and Services Registry (CSR). This circumstance affects some partners in GeoViQua. At the time being, GeoViQua had evaluated a possible contribution from WP7 to AIP4 during 2011, as the objective this year was the registry of datasets in GEOSS. However, it was not feasible to match the timing alongside AIP4 procedures.

-contributors expertise: guarantee comprehension of the dataset (control on sources, processes, data peculiarities, data description, potential use)

-availability (or else retrieval system) and reliability of associated quality parameters

-suitability for experimental testing

-variety: cover the wider range of fundamental aspects (*e.g.,* data sources, methods, quality measures, scales, parameters) in order to assure an optimized match between the dataset features and the corresponding experimental tests

-scenarios building capacity

-suitability for dissemination purposes

-topic and social impact: cover a wide range of thematic areas as derived from the Description of Work including parameters highlighted in the critical Earth Observation priorities list, sticking to GEOSS SBAs and CoPs

A related issue is that the selected datasets were to be in accordance with the Critical Earth Observation priorities list, derived from the GEO Task US-09-01a. This task seeks to identify Earth Observation needs across a full spectrum of user types and communities both within each SBA and across all SBAs, including observation needs from a variety of geographic regions and significant representation from developing countries. This list includes ground, in situ, airborne, and space-based observations. The focus is on both observed and derived observation parameters as well as model products. A full description of the most relevant parameters in this list is shown in the glossary.

As a result, a complex sample will be subjected to several experimentation tests related to WP7 objectives (quality parameterization, validation of search components, quality aware visualisation, and validation of the GEO Label). This variety pretends being representative of situations commonly faced by all potential users (scientists, technicians, decision makers, politicians, disseminators, teachers, students or general public) in their daily activities with geodatasets, setting an example of real cases.

This sample prioritizes heterogeneity of pilot cases, aiming at covering when feasible a wider range in methods, data sources and quality measures. Scientifically, this is more appropriate to GeoViQua's target. Moreover, it is specifically contemplated in GeoViQua's Description of Work. As was aforementioned, some datasets will be more valid for a specific test. For instance, different statistic approaches will be needed to adequately assess quality measures when dealing with quantitative parameters in comparison to discrete, categorical or ordinal parameters. Likewise, inputs from work packages will probably need to be addressed separately in WP7: the testing of search tools developed in GeoViQua is expected to require a different dataset sample than the one demanded by the testing of visualisation tools. Thus, the selected sample will desirably permit extracting the most appropriate datasets for each methodological design in GeoViQua. Datasets in WP7 must be able to flexibly cope with this open spectrum of circumstances. In addition, the scenario's design will set demonstrative examples of integration of testing results and tools developed in GeoViQua, at the same time introducing the background of the processes, actors, and system features involved, in general and within the frame of GEOSS components and services.

2.2 Testing datasets sample and scenarios

Keeping in mind GeoViQua serves GEOSS purposes, the topics specified in the GEOPortal -related to the SBAs and CoPs- were taken into account in the Description of Work, from which the first classification of pilot cases derived. However, the operational capacity of WP7 led to the adopted classification.

Thus, current classification of pilot cases is still somehow related to SBAs and CoPs although, the WP7 approach used in GeoViQua will focus on:

-Selected testing dataset

-Illustrative scenarios, following a scenario-design method based on several testing dataset extracted from the sample, inserted in the environment of components and services within the system

The latter will be selected according to factors suchlike social impact, potential users, topic, global or local coverage or dissemination potential, appropriateness for implementation in GEOSS (*e.g.,* fulfilment of data sharing policies).

In agreement with the partners involved in WP7, Carbon cycle, Disasters, Agriculture, Air quality, and Marine are candidates that have been drafted as illustrative scenarios topics. The experimental testing and coordinated tasks with other work packages will naturally drive the final scenarios selection and methodology design at later stages in the project.

The following section depicts the dictionary tables of the main fields of the pilot cases testing datasets description table in detail.

3. FORMALISING PILOT CASES DESCRIPTION

3.1 Design of pilot cases description template

Once having revised GEOSS structure (see section 5 below), from a practical perspective related to scientific objectives of GeoViQua, a wide variety of datasets was analysed.

From this point of view, the necessity of standardization arises. Thus, the first step was to design a way of classification of the main aspects of geospatial data. The components and service tables of GEOSS CSR inspired this template, although the design was adapted to WP7 targets.

With this purpose in mind, a template containing the main parameters characterizing geospatial datasets was created. Fulfilling the template for each dataset enables a first analysis on metadata contents. Later on, this table proved to be remarkably useful. In this sense, it is to be noted that at present, the template has been designed for the analysis of GeoViQua pilot cases. Very likely, a standardized version of general appliance would content inputs of the results of WP7 and other GeoViQua work packages (*e.g.,* user requirements, search components, visualisation). Depending on obtained results, a polished version of standardized forms, adapted to GEOSS community needs, might be derived in order to complement the current CSR tables. If appropriate, a modified template will be presented to the Architecture and Data Committee in GEO.

In short, the experiences on WP7 will provide practical data on testing real datasets covering the main aspects of geospatial data that are contained in the template:

-Data sources
-Methods
-Quality measures
-Temporal and spatial resolution/extent
-Product models

The first three fields, involving proper dictionary tables, due to their relevance as regards precise vocabulary in Task 7.2 will be described in detail in specific sub-sections (respectively sections 3.2, 3.3, 3.4). First, the template and other fields therein contained are briefly explained.

Table 1. shows an empty form to be used as template, containing information on how to be filled in. The fields in blue are linked to a drop-down menu, specifying the possible options to be chosen for these fields.

Table 1. Pilot cases testing datasets description table template.

Pilot case*	*Options: Remote sensing, Carbon cycle, Climate, Water cycle, Disasters, Agriculture, Air quality, Land use, Marine.*
Societal Benefit Areas (SBA)	*Options: Health, Energy, Climate, Water, Weather, Ecosystems, Agriculture, Biodiversity, Disasters.*
Dataset title*	*Entitle the dataset.*
Dataset description (and a feasible user story)	*Provide additional information (not covered by any other field) useful to describe the dataset as well as its purpose.*
Methods* (options)	*In order to conceptualize the methods we provide a classification. Please select among the options.*

Methods details			*If found useful, provide a deeper explanation about the methodology and processes applied to the original data. Provide references if needed.*
Main data sources* (options)			*Select among the options.*
Data sources details*			
Product data model*			*On the next rows, select the appropriate data models ("X" if used, "-" if not on the left column). If necessary, provide details on the right column (e.g., number of layers, number of vector entities, etc.).*
	X or -	Rasters	*Explain details if needed.*
	X or -	Points	*Explain details if needed.*
	X or -	Lines or Arcs/Nodes	*Explain details if needed.*
	X or -	Polygons	*Explain details if needed.*
Thematic attributes* (enter "(units)" or type "(categorical variable)")			
Spatial resolution / Equivalent scale			
Spatial extent description (including total covered area)			
Temporal resolution			*Daily, monthly, annual...*
Temporal extent			
Available quality measures* (options)			*Does this dataset include some quality indicators? If yes, please select among the options.*
Spatialized quality (Y/N)*			*Y: there exists an additional layer with spatial uncertainty information. N: there only exists global uncertainty information or quality measures are non-existent.*
Quality available as			*Text files, xml files, raster files,...*
Desired quality measures			*Although it is possible that the dataset has some quality indicators, it could be interesting to compute new parameters or improve the existing ones. Describe which ones if appropriate.*
Desired spatialized quality (Y/N)			
Desired quality available as			*Text files, xml files, raster files,...*
Partner responsible*			
Organizations (if different)			*Fill it when dataset comes from a non-partner organization.*
Partner contact e-mail*			*Contact to facilitate the data access to other* GeoViQua *partners that wish to use this dataset as pilot case.*
Restrictions / license			*None / free registration (all public) / free registration (depending on user profiles) / registry fees.*
Data access			*Provide the link of the map server.*
Data services			*Is this dataset related/embedded with/in some Internet map service? If yes, define the service and provide its link.*
CSR (GEOSS) Record			*It would be desirable that all the datasets used as pilot cases would be registered in the Component and Services Registry of GEOSS. Please provide the link.*
Other CSR (GEOSS) or FP7			*Provide related datasets that currently exist in the Component and*

related information		*Services Registry of GEOSS or in other FP7 projects.*
Role in GeoViQua		
	Requirements: testing/gathering (Y/N)	*Different datasets could be useful for different objectives: Will this dataset be useful for requirements testing? Has this dataset generated requirements? Liaison with other WPs in GeoViQua.*
	Project results testing (Y/N)	*Different datasets could be useful for different objectives: Will this dataset be used to test results? Liaison with other WPs in GeoViQua.*

**indicates mandatory information. Blue text depicts information that has to be selected from a predefined list (see the "options" link).*

General comments on the template's fields are:

-Pilot case was meant to ease classification of pilot cases in big areas as the ones suggested, although it was finally considered that probably some datasets would apply to several of these and more specific keywords or several areas are accepted. A proper specific name field complementing this is "dataset title".

-Societal Benefit Areas directly apply to GEOSS nomenclature and it's therefore useful for implementation in GCI.

-User stories: In WP7, analogously to WP2 definition although adapted to the description of the pilot cases testing datasets, it applies to a written description of the story used for planning and details highlight. In addition, the "Role in GeoViQua" field in the template pretends sketching the links between testable requirements, user stories, scientific objectives or user groups identified in WP2, and the datasets that will be used in the corresponding tests.

-GeoViQua aims at standardizing quality measures. A frequent situation when revising geospatial datasets is that the metadata are absent or incomplete. In particular, quality parameters lack in specification. In this context, "desired quality parameters" in a dataset and its standardization, reflects a main target of the scientific approach in WP7.

-Data sharing, and related issues (*e.g.,* restrictions, licenses, access, services) were discussed over the second Project Technical Board meeting. Partners must specify if the datasets provided are only to be used with experimentation purposes in WP7, available for sharing among GeoViQua partners or freely available for potential users (*e.g.,* to be registered in the CSR, should it not be already catalogued or some samples used with dissemination purposes in GeoViQua).

The GEOSS 10-Year Implementation Plan explicitly acknowledges the importance of data sharing in achieving the GEOSS vision and anticipated societal benefits. The Plan, endorsed by nearly 60 governments and the European Commission at the 2005 Third Earth Observation Summit in Brussels, highlights the following GEOSS Data Sharing Principles:

• There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation
• All shared data, metadata and products will be made available with minimum time delay and at minimum cost
• All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

At the same time, it is important to underscore the fact that GEOSS is composed of voluntarily contributed systems and data, which are governed by pre-existing laws, policies and practices that may not be fully compatible with the principles. The adherence to the Data Sharing Principles is not legally binding. The

Principles will gain acceptance and importance through good-faith voluntary adherence, which may also be accompanied by legal and policy changes at the national or international levels. When registering data in GEOSS, the contributor should present any restrictions arising from relevant international instruments and national policies and legislation, and the duration of each restriction that is applicable to the exchange of the data, metadata and products submitted.

The Implementation Guidelines states that data, metadata and products made available through the GEOSS are made accessible with minimal time delay and with as few restrictions as possible, on a non-discriminatory basis, at minimum cost for no more than the cost of reproduction and distribution.

The successful implementation of the Data Sharing Principles will depend upon successfully promoting the benefits of full and open access to GEOSS data through a process that engages directly both data providers and data users, including their respective contributions and perspectives. Moreover, it should be recognized that GEOSS is a dynamic system in a continual state of development and evolution.

Because the value of data lies in their use, the users of GEOSS data would therefore need the flexibility to reuse and re-disseminate the resulting shared information in order to maximize their own uses, as well as the relevant secondary applications of such data and information for the broad societal benefits.

-Role in GeoViQua pretends a first sketch of the potential use of the dataset in GeoViQua. It is intended to collect a brief outline of the relationship of the dataset with specific aspects of other work packages, suchlike linking the dataset to requirements that are being defined in WP2, and other work packages (namely WP3, WP4, WP5 and WP6), with the aim of identifying the most appropriate dataset for testing concrete tools or methods developed in GeoViQua. This is a dynamic process: a priori some links can be defined, whether others will only be tangible from the experimental testing and/or scenarios development (*e.g.,* introduction of user requirements derived from working with the pilot cases, defined dissemination and communication issues will likely arise from work in WP7).

The description table and linked dictionary tables were uploaded on GeoViQua's site (together with an example of a filled in template corresponding to a real testing dataset) for internal discussion. An open period for commenting on the design followed. This issue was also discussed during the second Project Technical Board meeting. The fields of the template were revised and the first sample of draft templates was checked. After the meeting, the selection of pilot cases datasets to be subjected to experimental testing was consolidated. Thus, in the end, the following partners provided datasets to be used in WP7 and consequently filled in definite versions of the corresponding templates: 52N, CEA, CREAF, ESA, S&T, UAB, UREAD. The content of the templates of the datasets finally selected are analysed in section 4.

The testing datasets description table enables analysing the three determinant aspects of datasets as regards quality testing: data sources, methods and quality measures.

3.2 Data sources

There is a remarkable range in the types of data used for GIS analysis, reflecting the varied goals of the systems themselves. Even within a single GIS project, the range of source materials employed can be overwhelming. An attempt of inventory and comprehensive classification of data sources is embedded on the template.

Not so long ago, most GIS projects had to rely almost exclusively upon data available only in printed or paper form. Data already produced in digital format certainly ease the work and speed the process of developing GIS, but only if users learn how to employ these new sources effectively. In addition, users must often search for materials in different places. The internet is used to distribute data to a greater extent, and this means knowing where to look and how to search the networks. This is in fact one of the main concerns in GEOSS, and in WP4 in GeoViQua.

The data sources list from which to scroll-down in the template contains the following options:

- Paper maps vs. digital maps (including digital elevation models (DEM) and/or derived products as slope, etc.)
- In-field collected data from humans (*e.g.,* forest inventories, surveys, volunteer geographic information gathering). If relevant for spatial, thematic or temporal data quality, please indicate tools (GNSS, total stations, etc.)
- Pre-existing data (*e.g.,* census data on databases, historical literature)
- Active sensors (*e.g.,* radar, lidar) vs. passive sensors (*e.g.,* conventional aerial photographs)
- Analogical sensors (*e.g.,* ancient aerial photographs) vs. digital sensors (*e.g.,* present-day sensors)
- In the case of sensors, what is recorded: electromagnetic waves / sound waves / gravity fields / chemical compounds / temperature / rainfall / snowfall / evapotranspiration / humidity / wind speed / barometric pressure
- In the case of sensors recording electromagnetic waves which spectral regions are recorded (select all that apply): ultraviolet / visible / near infrared / middle infrared / thermal infrared / microwave sensors
- In the case of sensors recording electromagnetic wave: singleband sensors / multispectral sensors / hyperspectral sensors
- In the case of sensors, position: handheld sensors / fixed on ground sensors (*e.g.,* weather radar, meteorological stations, flux towers) / airborne sensors / spaceborne sensors / sensors free on the medium (*e.g.,* weather balloons)

Some of these terms are explained in the glossary. Finally, it is to be noted that all data sources have strengths and limitations. It is important to understand their characteristics, costs, and benefits before using them. Although data source is a field commonly known to users, there could be some improvement in the delivering of information on intrinsic characteristics related to the sources of products delivered by data providers (*e.g.,* raw data and sensor biases at discovery time).

3.3 Methods

Likewise, the table proposes a classification of methods frequently employed in the process of transforming raw data recorded by the sensor into a geospatial product that, in turn, some final users manage as input data in diverse applications.

The methods contained in the menu are mainly photointerpretation, classification, spatial interpolation, generalization (including scaling), modelling, time series analysis, land use and land cover dynamics, landscape metrics, re-projections and changes in data model.

- Photointerpretation: consists in examining photographic images for the purpose of identifying objects and judging their significance. The main elements of image interpretation are: location, size, shape, shadow, tone/colour, texture, pattern, height/depth and site/situation/association.

- Digitizing:

Digitizing is the representation of an object, image or a signal (usually an analogical signal) by a discrete set of its points or samples. The result is called digital representation or, more specifically, a digital image, for the object, and digital form, for the signal. Strictly speaking, digitizing means simply capturing an analogical signal in digital form. For a document the term means to trace the document image or capture the "corners" where the lines end or change direction.

McQuail (2000) identifies the process of digitization having immense significance to the computing ideals as it allows information of all kinds in all formats to be carried with the same efficiency and also intermingled.

- Digital image classification

The intent of the classification process is to categorize all pixels in a digital image into one of several land cover/use classes, or "themes". This categorized data may then be used to produce thematic maps of the land cover present in an image (mono or multitemporal). Normally, multispectral data are used to perform the classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization (Lilesand et al. 2003).

Considering statistical procedures, the most commonly employed classification strategies include supervised and unsupervised classification. In supervised classification, spectral signatures are developed from specified locations in the image. These specified locations are given the generic name of “training sites” and are defined by the user. This approach involves the selection of areas on the image which statistically characterize the informational categories of interest, while the unsupervised approach attempts to identify spectrally homogeneous groups within the image that are later assigned by the user to informational categories (Chuvieco 2002). Something in between, a hybrid classification is an automatic and objective process that involves both unsupervised classification and training areas (collected as in the first stage of a conventional supervised classification).

As for results of the classification, the so-called hard classifiers assign each pixel in the scene a discrete value or category based on the training sites in the vector layer. For example, if four different land use types were identified with the training sites, each pixel in the scene will take on the value of one of those four land uses. Instead, if soft classifiers are used, probabilities may be listed as to which category each pixel may or may not belong to. In other words, definitive decisions about land categories are not made. More specifically speaking about classifiers, there are many types that could be categorized as non parametric, parametric, probabilistic, or distant dependent classifiers, to name a few.

- Spatial interpolation from categorical data

- Spatial interpolation from quantitative data

Spatial interpolation is the procedure by which a surface is created estimating the value of properties at unsampled sites within the area covered by existing observations. In almost all cases the property must be interval or ratio scaled. There are several forms of interpolation, each of which treats the data differently, depending on the properties of the data set.

- Upscaling and map generalisation (includes coarsing the spatial resolution in raster data)

- Downscaling (includes densifying in raster data)

Generalization has a long history in cartography as an art of creating maps for different scale and purpose. Cartographic generalization is the process of selecting and representing information of a map in a way that adapts to the scale of the display medium of the map. In this way, every map has, to some extent, been generalized to match the criteria of display. This includes small-scale maps, which cannot convey every detail of the real world. Cartographers must decide and then adjust the content within their maps to create a suitable and useful map that conveys geospatial information within their representation of the world.

Generalization is meant to be context-specific. That is to say, correctly generalized maps are those that emphasize the most important map elements while still representing the world in the most faithful and recognizable way. The level of detail and importance in what is remaining on the map must outweigh the insignificance of items that were generalized, as to preserve the distinguishing characteristics of what makes the map useful and important.

Scaling can be considered nowadays an entity in its own right in map generalization. In fact, scaling procedures are commonly applied.

Scaling procedures enable converting a raster dataset at one resolution to a raster dataset at another resolution. If cell resolutions are divisible then conversion is straightforward. When cell resolutions are not divisible, or when converting to a raster dataset in another geometric orientation a resampling method must be used (*e.g.,* nearest neighbour, bilinear interpolation, cubic convolution).

Thus, respectively upscaling refers to producing a new derived dataset with a larger cell size/cell resolution whereas downscaling refers to producing a derived dataset with a smaller cell size/cell resolution. Upscaling aggregates data based on different cell size (most frequent value –nominal data-, or interpolation of neighbourhood of values –radio data-). It must be noted it is not possible to create information by down-scaling, as simple cell assignment is based on source cell (*i.e.,* coarser resolution). This adds epistemic (lack of knowledge) uncertainty to the resulting layers. However, data compression techniques reduce the data storage requirements for raster datasets in some situations.

- Physical modelling (*e.g.,* DEM-based solar radiation, collinearity equations for geometric corrections of aerial images)

- Process-based modelling (*e.g.,* climatic scenarios derived from general circulation models, water stress indices)

- Statistical and empirical-based modelling (*e.g.,* species spatial distribution models, multicriteria analysis)

A model is a simplified representation of a phenomenon or a system. The types considered in GeoViQua attempt to reflect the gradient between theoretical models and empirical models, the latter built on direct observations.

In this sense, physical (theoretical or analytical) models are the most accurate and sensitive, but demand a complete knowledge of the physical properties of the system. They represent the synthesis of the individual components that affect the modelled variable, including the complex interactions between various factors and their spatial and temporal variability. Physical models theoretically speaking are consistent, purely deterministic and require very few observation data inputs.

Physically-based descriptions of processes may be often justified where there is sufficient information, and there are examples of success in the application of these models. However, physically-based descriptions do not necessarily lead to realistic simulations because of difficulties in fulfilling the information requirements. In sparsely gauged regions, there are additional troubles in fulfilling information requirements for key aspects of physically based models. To resolve this dilemma, more conceptual models have been developed for conditions of poor data availability. Often such models are not able to correctly incorporate new process understanding or measurement techniques and rely on heavy calibration of parameters because they have poor process conceptualizations. However, when models properly conceptualize the processes, not only are parameters relatively stable over time and space, but they can incorporate new process information and measurements to improve predictions. These process-based models have great potential for ungauged prediction. Process modelling is based on the mechanisms inherent to the studied variable, which incorporate a large amount of information in the response functions. Hence, process models assume at least part of the mechanisms inherent to the studied parameter is known, so the model is expected not to contradict predetermined facts although some kind of calibration is required.

This is to say process models seek primarily to describe data using key mechanisms or processes that determine an object's internal structure, rules, and behaviour. In contrast, empirical models seek principally to describe the statistical relationships among data with limited regard to an object's internal structure, rules, or behaviour (Korzukhin et al. 1996). Thus, statistical and empirical models predictions are based on probabilities derived from observations. Empirical models have constraints of applicability limited to conditions similar to those from which data were used in their development.

In addition, mechanisms included in process models are general enough that they can maintain some degree of relevance for new objects or conditions (mechanism constancy), while empirical models tend not to be tied to any specific mechanism, so that derived model parameters must remain constant (parameter constancy) for new objects or conditions.

In the end, the use of physical models is rare, as completely relying only on theoretical models is not a common practice in most of the tasks implied in Earth Observations data processing. Even when using them, models are often validated by using empirical data in model fitting. On the other hand, the use of too many empirical data is expected to affect quality somehow (*e.g.,* error propagation is hardly traceable in most cases when several data sources are combined in modelling). This fact is of relevance in the framework of GeoViQua.

- Time series analysis

A time series is a sequence of data points, measured typically at successive times spaced at uniform time intervals. Time series analysis comprises methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data. A typical application would be time series forecasting, which uses a model to forecast future events based on known past events to predict data points before they are measured.

Time series data have a natural temporal ordering. This makes time series analysis distinct from other common data analysis problems, in which there is no natural ordering of the observations. Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations. A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values.

Methods for time series analyses may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and recently wavelet analysis. The latter include auto-correlation and cross-correlation analysis.

- Analysis of land use & land cover dynamics

Land-cover refers to the physical characteristics of earth's surface, captured in the distribution of vegetation, water, soil and other physical features of the land, including those created solely by human activities (*e.g.,* settlements). Land-use refers to the way in which land has been used by humans and their habitat, usually with accent on the functional role of land for economic activities. It is the intended employment of management strategy placed on the land-cover type by human agents, and/or managers (Baulies and Szejwach 1997).

Examples of techniques in analysis of land use and land cover dynamics would be detection techniques such as image differencing, image ratioing, vegetation index differencing and image regression

- Landscape metrics

The measurement, analysis and interpretation of spatial patterns receive much attention in landscape ecology (Haines-Young and Chopping, 1996). Landscape metrics and landscape indices are employed for the characterization of landscape structure and various processes at both landscape and ecosystem level.

The common usage of the term “landscape metrics” mostly refers to indices developed for categorical map patterns (McGarigal et al. 2002), but it is sometimes also used for topographic measures (Iampietro et al. 2005, Vivoni et al. 2005) that characterize landscape, or it may also just refer to some combination of several characteristics that are important to a particular species (Schils 2006, Fernández et al. 2007).

- Changing map projections

Geospatial data are encoded with certain units, the spatial components of each feature (lines, points) are projected using a mathematically defined coordinate system transformation. Map projection involves taking spatial data defined on the three-dimensional curved surface of the earth and transforming it to the flat two-dimensional surface of a map. A map projection is the mathematical algorithm used for this transformation. It is to be noted that changing the coordinate system of a data frame does not alter the coordinate system of the source data contained in it. Different projections are used for different types of maps because each projection particularly suits specific uses. For example, a projection that accurately represents the shapes of the continents will distort their relative sizes.

- Changes in grid origin (raster data)

Both methods (changing map projections and changes in grid origin) could be somehow considered to be reprojections. In practice, interpolation is involved.

- Changes in data models (includes rasterization and vectorization)

Data restructuring can be performed to convert data into different formats. For example, a GIS may be used to convert a satellite image map to a vector structure by generating lines around all cells with the same classification, while determining the cell spatial relationships, such as adjacency or inclusion.

More advanced data processing can occur with image processing, a technique developed to provide contrast enhancement, false colour rendering and a variety of other techniques including use of two dimensional Fourier transforms. Since digital data is collected and stored in various ways, the two data sources may not be entirely compatible. A GIS must be able to convert geographic data from one structure to another.

3.4 Quality measures

Quality measures in the template are supported by ISO standards complemented with strict statistically improving remarks introduced in the frame of GeoViQua. Actually, the content of this section is a preamble of related deliverables on the matter, connected to the Quality elicitation mechanisms (*i.e.,* WP3, Deliverable 3.1 Metadata extraction quality component) and Delivery of solutions to end users (*i.e.,* WP6, Deliverable 6.1. Data quality encoding as a best practice paper) work packages.
Special attention has been paid on this issue, as it constitutes the main pillar of GeoViQua.

First, lists of typified common quality parameters (defined as "quality elements" in ISO standards –ISO 19113-) for geographical datasets were revised. Quality parameters are related to aspects of datasets quality: positional accuracy, logical consistency, temporal accuracy and thematic accuracy.

In turn, quality parameters can be quantified using several measures (*e.g.,* root mean square error, error rate, etc.). Following ISO standards, quality parameters and quality indicators correspond to elements and sub-elements, extracted from ISO 19113.

Quality measures in this context apply to the actual quantification (*e.g.,* root mean square error, value at 95% confidence level) of quality parameters or indicators, obtained from ISO19114 and ISO19138 (ISO-TC211 2006). Thus, these three concepts related to quality elements comprise a hierarchy, from quality parameters (aspects), to quality indicators (sub-elements) and finally quality measures (test applied to evaluate data quality elements).

Table 2 summarizes the results of the revision of the abovementioned ISO standards, including some complementary concepts introduced in GeoViQua.

Table 2. Classification and definition of quality parameters and indicators

Quality parameter	Quality indicator	Definition
Completeness	Commission	Excess data present in a dataset
	Omission	Data absent from a dataset
Logical consistency	Conceptual consistency	Adherence to rules of the conceptual schema
	Domain consistency	Adherence of values to the value domains
	Format consistency	Degree to which data is stored in accordance with the physical structure of the dataset
	Topological consistency	Correctness of the explicitly encoded topological characteristics of a dataset
Positional accuracy	Absolute or external accuracy	Closeness of reported coordinate values to values accepted as or being true
	Relative or internal accuracy	Closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true
	Gridded data position accuracy	Closeness of gridded data position values to values accepted as or being true
Temporal accuracy	Accuracy of a time measurement	Correctness of the temporal references of an item (reporting of error in time measurement)
	Temporal consistency	Correctness of ordered events or sequences
	Temporal validity	Validity of data with respect to time
Thematic accuracy	Classification correctness	Comparison of the classes assigned to features or their attributes to a universe of discourse (*e.g.,* ground truth or reference dataset)
	Non-quantitative attribute correctness	Correctness of non-quantitative attributes
	Quantitative attribute accuracy	Accuracy of quantitative attributes

These quality parameters are very typical of geospatial data sets, usually conforming important fragments of the metadata, and allowing users to judge whether the data set is fit for their purpose. However, there is a bias towards vector data within the list.

In general it is expected that all data sets should be logically consistent - if not the case, it would not make sense to use them. In this sense, there is a difference between discovery time and propagation of the information in data processing. Thus, the difference between quality information required for dataset discovery and the information required during the use of the dataset should be stated clearly.

With regards the associated quality indicators, it must be remarked that these indicators have a singular utility in particular scenarios. For example, if two datasets are to be used in a single project, being aware of only their respective internal accuracy is not sufficient, as the fact of how they are related remains unknown. In that case, it is not possible to combine them in a principled framework. Therefore, it is worth noting that accuracy measures reporting "the 'closeness' of a value to that which is accepted as being true (reality)" are the most useful, since these allow true information interoperability in the sense that the data can be combined with other data, processed and integrated into a coherent decision making framework.

The list of quality measures (see Table 3) is obtained from ISO 19138 (ISO-TC2112006) although introduces slight modifications linked to the UncertML dictionary (https://wiki.aston.ac.uk/foswiki/bin/view/UncertWeb/UncertMLDictionary) and other contributions from GeoViQua, guaranteeing statistical consistency. For instance, it should rather be considered the value at a given confidence level, instead of uncertainty at a defined significance level. This list is a preliminary stage of in progress activities from WP3 and WP6, and will be further developed in a proper deliverable, related to task 6.1 (Encodings for data quality) in GeoViQua. Moreover, it must be remarked ISO 19138 (ISO-TC211 2006) is being revised. The ISO 19157 (ISO-TC211 2011), currently under development, is prone to provide updated insights in this context. Nevertheless, it seemed appropriate to give an advance on what WP7 datasets provide related to this issue, reporting back to WP3 and WP6 the items to be solved/upgraded in ensuing updates of this initial list.

Table 3. Quality parameters, quality indicators and quality measures.

Origin	Quality parameter	Quality indicator	Quality measure
ISO19138_1	Completeness	Commission	Excess item
ISO19138_2	Completeness	Commission	Number of excess items
ISO19138_3	Completeness	Commission	Rate of excess items
ISO19138_4	Completeness	Commission	Number of duplicate feature instances
ISO19138_5	Completeness	Omission	Missing item
ISO19138_6	Completeness	Omission	**Number of missing items**
ISO19138_7	Completeness	Omission	Rate of missing items
ISO19138_8	Logical consistency	Conceptual consistency	Conceptual schema noncompliance
ISO19138_9	Logical consistency	Conceptual consistency	Conceptual schema compliance
ISO19138_10	Logical consistency	Conceptual consistency	**Number of items noncompliant to the rules of the conceptual schema**
ISO19138_11	Logical consistency	Conceptual consistency	Number of invalid overlaps of surfaces
ISO19138_12	Logical consistency	Conceptual consistency	Non-compliance rate with respect to the rules of the conceptual schema

ISO19138_13	Logical consistency	Conceptual consistency	Compliance rate with the rules of the conceptual schema
ISO19138_14	Logical consistency	Domain consistency	Value domain non-conformance
ISO19138_15	Logical consistency	Domain consistency	Value domain conformance
ISO19138_16	Logical consistency	Domain consistency	Number of items not in conformance with their value domain
ISO19138_17	Logical consistency	Domain consistency	**Value domain conformance rate**
ISO19138_18	Logical consistency	Domain consistency	Value domain non-conformance rate
ISO19138_19	Logical consistency	Format consistency	**Physical structure conflicts**
ISO19138_20	Logical consistency	Format consistency	Physical structure conflict rate
ISO19138_21	Logical consistency	Topological consistency	Number of faulty point-curve connections
ISO19138_22	Logical consistency	Topological consistency	Rate of faulty point-curve connections
ISO19138_23	Logical consistency	Topological consistency	Number of missing connection due to undershoots
ISO19138_24	Logical consistency	Topological consistency	Number of missing connections due to overshoots
ISO19138_25	Logical consistency	Topological consistency	Number of invalid slivers
ISO19138_26	Logical consistency	Topological consistency	Number of invalid self intersect errors
ISO19138_27	Logical consistency	Topological consistency	Number of invalid self overlap errors
ISO19138_28	Positional accuracy	Absolute or external accuracy	Mean value of positional uncertainties
ISO19138_29	Positional accuracy	Absolute or	Mean value of positional uncertainties

		external accuracy	excluding outliers
ISO19138_30	Positional accuracy	Absolute or external accuracy	Number of positional uncertainties above a given threshold
ISO19138_31	Positional accuracy	Absolute or external accuracy	Rate of positional errors above a given threshold
ISO19138_32	Positional accuracy	Absolute or external accuracy	Covariance matrix
ISO19138_33	Positional accuracy	Absolute or external accuracy	Linear error probable
ISO19138_34	Positional accuracy	Absolute or external accuracy	Standard linear error
ISO19138_35	Positional accuracy	Absolute or external accuracy	Uncertainty at 90% significance level
ISO19138_36	Positional accuracy	Absolute or external accuracy	Uncertainty at 95% significance level
ISO19138_37	Positional accuracy	Absolute or external accuracy	Uncertainty at 99% significance level
ISO19138_38	Positional accuracy	Absolute or external accuracy	Near certainty linear error
ISO19138_39	Positional accuracy	Absolute or external accuracy	**Root mean square error**
ISO19138_41	Positional accuracy	Absolute or external accuracy	Absolute linear error at 90% significance level of biased vertical data
ISO19138_42	Positional accuracy	Absolute or external accuracy	Circular standard deviation (probability of 39.4%)
ISO19138_43	Positional accuracy	Absolute or external accuracy	Circular error probable (probability of 50%)
ISO19138_44	Positional accuracy	Absolute or external accuracy	Circular map accuracy standard (probability of 90%)
ISO19138_45	Positional accuracy	Absolute or external accuracy	Circular error at 95% significance level
ISO19138_46	Positional accuracy	Absolute or external accuracy	Circular near certainty error

ISO19138_47	Positional accuracy	Absolute or external accuracy	Root mean square error of planimetry
ISO19138_49	Positional accuracy	Absolute or external accuracy	Absolute circular error at 90% significance level of biased data
ISO19138_50	Positional accuracy	Absolute or external accuracy	Uncertainty ellipse
ISO19138_51	Positional accuracy	Absolute or external accuracy	Confidence ellipse
ISO19138_52	Positional accuracy	Relative or internal error	Relative vertical error
ISO19138_53	Positional accuracy	Relative or internal error	Relative horizontal error
ISO19138_54	Temporal accuracy	Accuracy of a time measurement	**Uncertainty at 68.3% significance level**
ISO19138_55	Temporal accuracy	Accuracy of a time measurement	Uncertainty at 50% significance level
ISO19138_56	Temporal accuracy	Accuracy of a time measurement	Uncertainty at 90% significance level
ISO19138_57	Temporal accuracy	Accuracy of a time measurement	Uncertainty at 95% significance level
ISO19138_58	Temporal accuracy	Accuracy of a time measurement	Uncertainty at 99% significance level
ISO19138_59	Temporal accuracy	Accuracy of a time measurement	Uncertainty at 99.8% significance level
ISO19138_60	Thematic accuracy	Classification correctness	Number of incorrectly classified features
ISO19138_61	Thematic accuracy	Classification	**Misclassification rate**

		correctness	
ISO19138_62	Thematic accuracy	Classification correctness	**Misclassification matrix**
ISO19138_63	Thematic accuracy	Classification correctness	Relative misclassification matrix
ISO19138_64	Thematic accuracy	Classification correctness	**Kappa coefficient**
ISO19138_65	Thematic accuracy	Non-quantitative attribute correctness	Number of incorrect attribute values
ISO19138_66	Thematic accuracy	Non-quantitative attribute correctness	Rate of correct attribute values
ISO19138_67	Thematic accuracy	Non-quantitative attribute correctness	Rate of incorrect attribute values
ISO19138_68	Thematic accuracy	Quantitative attribute correctness	**Uncertainty at 68.3% significance level**
ISO19138_69	Thematic accuracy	Quantitative attribute correctness	Uncertainty at 50% significance level
ISO19138_70	Thematic accuracy	Quantitative attribute correctness	Uncertainty at 90% significance level
ISO19138_71	Thematic accuracy	Quantitative attribute correctness	Uncertainty at 95% significance level
ISO19138_72	Thematic accuracy	Quantitative attribute correctness	Uncertainty at 99% significance level
ISO19138_73	Thematic accuracy	Quantitative attribute	Uncertainty at 99.8% significance level

		correctness	
GeoViQua_1	Thematic accuracy	Quantitative attribute correctness	**Coefficient of determination (R^2)**
GeoViQua_2	Thematic accuracy	Quantitative attribute correctness	List of case residuals

In the column of "Quality measure" some usual measures have been highlighted. A more detailed explanation of quality parameters, indicators, and measures can be found in the glossary.

The quality measures presented above are a small subset of the possible approaches to representing quality of data. Although two of them are shown in Table 3 (GeoViQua_1 and GeoViQua_2), additional approaches that had been added to this table as GeoViQua contributions are currently under revision, and will be further defined during the project span. Some concepts that are being analysed are, to name a few, contingency tables, which rely on a whole family of indicators (similar to kappa statistic), or the trust skill statistic. The target is that all, quantitative, categorical and binary data, are covered by appropriate quality measures.

This topic generated complex discussions in GeoViQua, from which the fusion of ISO standards complemented with a hint to other statistical concerns was agreed in the second Project Technical Board meeting. A comprehensive list of quality measures definitions and vocabularies are currently underway in the project web and tasks are being coordinated with WP2, WP3 and WP6, directly linked to the subjects contained in this section. The corresponding deliverables will further extend this section. For example, deliverable “D6.1 Data Encoding as a best practice paper” represents an updated version of these efforts.

Remarkably, albeit the relevance of associated quality when related to the use of datasets, clearly determining achievable accuracy of results, it is to be noted that frequently quality parameters are unknown and not explicit in the metadata. This fact is shown in the template designed for WP7 datasets selection (*e.g.,* desired quality measures). Data providers are able to specify in these fields the quality parameter that fits best according to their knowledge of the properties of the dataset.

3.5 Other relevant fields

Temporal and spatial scales: resolution and extent

Based on two commonly accepted components of scale (grain and extent), scale space is shown to be at least two-dimensional, and spatiotemporal scale space is four-dimensional.

The terms “grain” and “extent” have been used quite often in descriptions of scale (Weins 1989, Turner et al. 1989, Allen and Hoekstra 1991, King 1998). Grain is the fundamental unit by which a phenomenon is measured or described, and extent denotes the spatial area or temporal duration of phenomena. Weins (1989) has likened grain and extent to the mesh size and the over-all size of a sieve. At times, grain and extent are artefacts of observation, and at other times they are attributes of phenomena. Often, the determination of their values is both important and difficult. In land-use/land-cover classification, for instance, different grain sizes yield different research results. One square meter is too fine an area to contain the complex criteria for a particular land use, and ten square kilometres may be too broad. In any case, the concept of scale functions as a sort of container in space or time for heterogeneous phenomena and processes, which have form and dynamics (Pereira 2002).

Granularity is a common method of offering expression to this heterogeneity. Continuous functional methods of representing heterogeneity are also valid. There are at least two forms of grain. One might be defined as the spacing between adjacent measurements (the resolution), and the other is the integration volume of each measurement.

In the context of remote sensing image analysis, Cao and Lam (1997) have classified scales into four categories: cartographic, observational, operational, and resolution. Cartographic scale characterizes the spatial relation between a landscape and its map or physical model. Resolution scale is equivalent to grain. Observational scale corresponds to extrinsic extent. Operational scales correspond to the characteristic scales of processes, which are functions of their intrinsic extents. As Cao and Lam point out, operational scales are not necessarily equivalent to observational scales. That is, the scales over which processes operate do not necessarily correspond to the spatial extent of the observation of such processes.
If we understand scale to consist of both grain and extent, these two definitions become complementary rather than contradictory, with Cao and Lam (1997) defining the grain and Clark (Clark 1987) defining the extent of a characteristic scale.

In GeoViQua, the spatial scales as for the pilot cases testing datasets, have been classified as global, continental, regional and local. This is of major importance as the approaches in methodologies and quality assessment are drastically different from the global to the local scale.

Temporal scales enable dynamic analysis, both for categorical (*e.g.,* land use and land cover dynamics) and quantitative variables (*e.g.,* time series analysis). From the scientific point of view, natural variability, processes, perturbances, disasters and simply evolution have a determinant temporal component. In order to reach realistic models and interpretation, updating of datasets, serious comparison from past to present and future predictions, and specially for rapidly changing environments, multitemporal datasets are required.

Product data model

This epigraph refers to GIS features: vector attributes (points, lines or arcs/nodes, polygons) versus grids.

The selected testing datasets comprise all these types.

Raster and vector are the two basic data structures for storing and manipulating images and graphics data on a computer. All of the major GIS (Geographic Information Systems) and CAD (Computer Aided Design) software packages available today are primarily based on one of the two structures, although some extended functions might be incorporated to support other data structures.

The template intends recollecting both type of product data model and quantification (*e.g.,* whether a point layer contains 20 points or 20,000 points is useful information for GeoViQua).

Thematic attributes

Integrated thematic attributes in combination with geometrical and topological information make GIS a strong tool in the spatial analysis. Thematic attribute may be anything that is alphanumerical ascertainable, representing elements that are not geometrical, such as names, measurements, properties, etc. Thematic attributes enable queries, analyses and visualisation of spatial data.

In WP7, pilot cases testing datasets include both numerical (continuous and discontinuous) and categorical attributes.

Data availability: restrictions/licenses

Finally, although technically not relevant, a primary issue nowadays is data availability, restrictions and licenses. Actually, GEOSS has defined documents on data sharing policies and an implementation guide. GeoViQua has integrated these recommendations in the project.

The partners in GeoViQua have access to and in fact some of them directly generate datasets from field measurements (understood in a broad sense), which is essential for the purposes of the project: control of quality data associated to each dataset all along the chain of recording and proceeding transformations between measurements and final users, therefore implying traceability. Thus, pilot cases include specific datasets concerning this issue. In this sense, the partners provide paramount datasets both from the datasets intrinsic characteristics (*e.g.,* as regards diversity of topics, GEOSS scope and up to date resources) and the particular quality issues involved in GeoViQua.

Nonetheless, as explained in the introduction, some datasets in the selected sample are already registered in GEOSS CSR and are freely available. Moreover, under the illustrative scenarios scope, some of the resources and results of WP7 will be openly displayed. The members of the consortium in the framework of GeoViQua provide datasets in accordance with the GEOSS data sharing principles. In fact, ARGO float is shown in the GEOSS Data Sharing Action Plan.
More information on data sharing in the frame of GEOSS can be found in *http://twiki.geoviqua.org/twiki/bin/view/GeoViQuaIntranet/TechnicalDocumentshttp://www.earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Rev2.pdf*

The following diagram depicts the interrelationship between the data model and the quality model, summarising data sources, methods and quality elements in WP7.

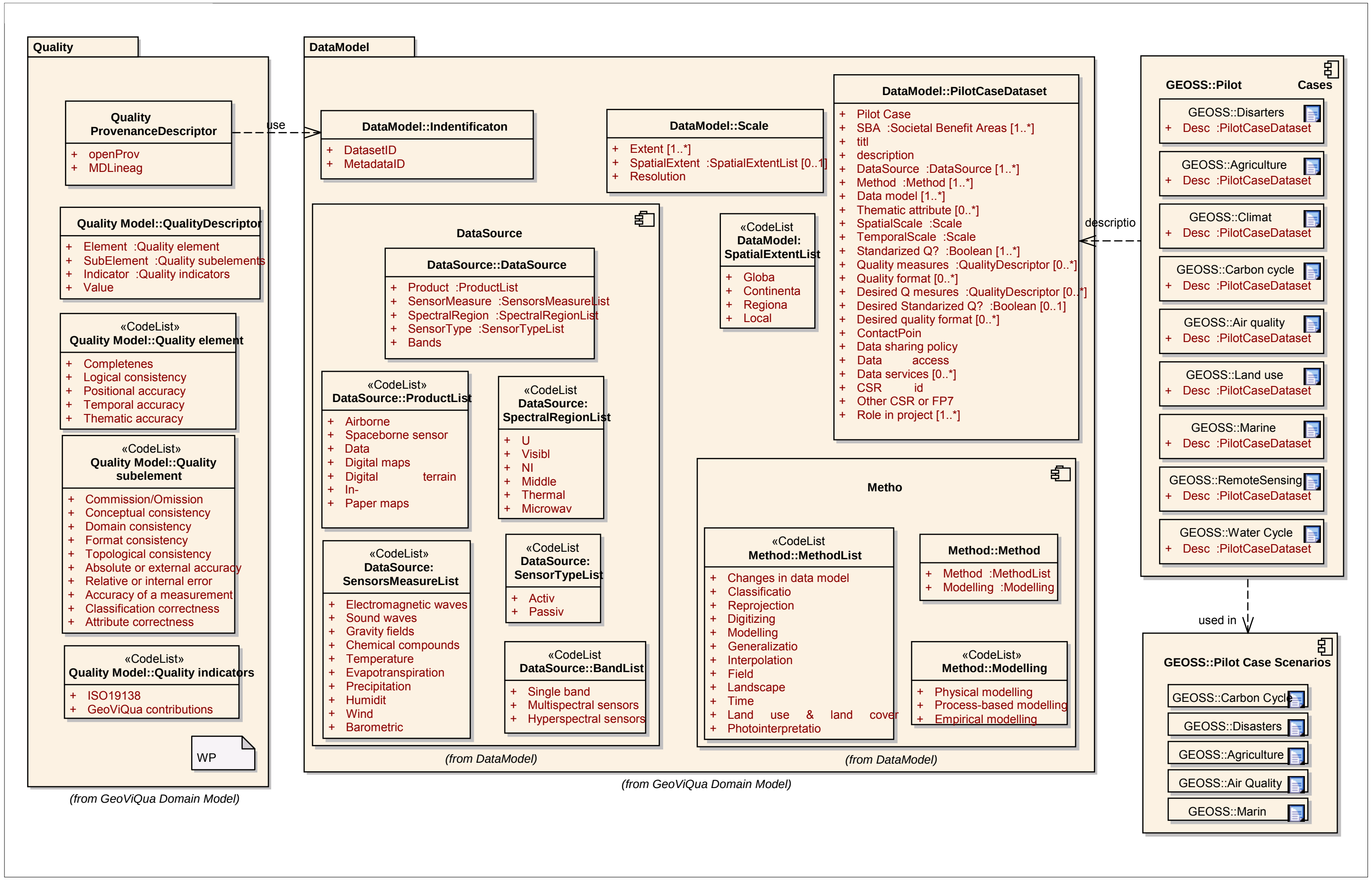

Figure 1. Data sources, methods and quality elements in the pilot cases datasets description table.

4. ANALYSIS OF SELECTED TESTING DATASETS

The following table shows a summary of the 16 selected testing datasets that will be analysed in this section.

Table 4. Pilot cases testing datasets.

Pilot case	Provider	Dataset title
Remote sensing	UAB	Geometric correction of Landsat series
Remote sensing	UAB	Radiometric correction of Landsat series
Remote sensing	S&T	Intercomparison of atmospheric data measurements from remote sensing instruments
Carbon cycle	CEA	Estimates of global carbon fluxes from model data fusion
Climate	UAB	Digital Climatic Atlas of the Iberian Peninsula
Water cycle	UAB	Snow surface map of Catalonia
Water cycle	UAB	Water vapour image of Catalonia
Water cycle	UAB	Water index image of Catalonia
Disaster		*Input not yet received from the appropriate provider*
Agriculture	UAB	Crop maps of northern Catalonia
Agriculture	UAB	Vegetation indices from northern Catalonia
Agriculture	UAB	Monitoring environmental measures using remote sensing
Agriculture	UAB	Crop water monitoring using remote sensing
Air quality	52N	Air quality in-situ
Land use	CREAF	Land cover map of Catalonia
Land use	CREAF	SIOSE
Marine	UREAD	Argo ocean floats

The selected testing datasets will be revised in section 4.1.

Nonetheless, a compilation of the data sources in this sample includes: digital maps, passive sensors, analogic and digital sensors. Concerning sensors positioning, the sample comprises fixed on ground sensors, airborne sensors, spaceborne sensors and sensors free on the medium. In the case of sensors, the recorded phenomena include electromagnetic waves, temperature, rainfall, snowfall, evapotranspiration, humidity, pressure and chemical compounds. The electromagnetic waves have been recorded by multispectral sensors and include wavelengths of visible, near infrared, middle infrared and thermal infrared radiation. Revising the options described in the template regarding data sources, this means that five out of the six products list (*i.e.,* 83%), seven out of ten (*i.e.,* 70%) in the sensors measures, and four out of six (*i.e.,* 66.6%) of the spectral region list are represented in the sample of selected datasets in GeoViQua.

The compilation of methods covered by the pilot cases testing datasets includes photointerpretation, digitizing, classification using hard and hybrid classifiers (supervised and unsupervised), spatial interpolation from quantitative data, physical modelling (collinearity equations for geometric corrections of aerial images, radiometric corrections), process-based modelling (water stress indices, vegetation indices, snow indices), statistical and empirical-based modelling, generalization (upscaling), field survey, time series analysis, land use and land cover dynamics, map projections and changes in data models (rasterization and vectorization). The coverage of the sample selected for testing in WP7 reaches 91% of the methods in the template. This is to say eleven out of twelve methods, implying all the range considered except landscape metrics. Nonetheless, the datasets in the sample of the land use pilot case are successfully used in local landscape metrics analysis.

The sample of pilot cases testing datasets covers the following combination of quality parameters, indicators and measures:

- Positional accuracy: absolute or external accuracy (root mean square error)
- Thematic accuracy: quantitative attribute correctness (uncertainty at 68.3% significance level; uncertainty at 95% significance level; coefficient of determination; covariance) and non-quantitative attribute correctness (rate of correct and incorrect attribute values)
- Completeness: omission and commission errors and number of missing items
- Logical consistency: domain consistency (value domain non-conformance), conceptual consistency (conceptual schema compliance), format consistency (physical structure conflicts) and topological consistency (number of invalid self overlap errors)

The covariance as a thematic accuracy quality measure refers to a measure of the correlation in function of a variable within the profile (*e.g.,* from lower atmosphere to upper atmosphere a pattern of the correlation of error with altitude could be expected).

Although some of the quality measures of the sample are far more frequent, the selected datasets are considered to adequately cover the main quality measurements (for diversity of quality parameters and indicators). In addition, the datasets cover several spatial and temporal resolutions.

Hence, the sample adequately covers a representative range of methods, sources and quality elements. The more relevant gaps refer to data sources, in which it would be desirable to include datasets from an active sensor (*e.g.,* radar, lidar) and a hyperspectral sensor. As for spectral regions uncovered, the ultraviolet and microwave regions are currently missing, and as for sensor measures, records of sound waves, gravity fields or wind speed remain untracked. Some of these are seldom employed in earth observation applications. Finally, although precipitation is already contemplated in some datasets, a weather radar remote sensing record of precipitation would be an asset.

4.1 Testing datasets

This section will study each of the selected pilot cases testing datasets, in order to point out their peculiarities and emphasizes the contribution of the case in GeoViQua, especially as regards quality parameterisation.

Remote sensing

This GeoViQua pilot case study provides processed images that will serve to validate the proposals of the project and will aid to go deep inside the knowledge about how to organize and use metadata and data in order to improve the tools that deal with provenance.
This pilot cases category does not apply to a particular thematic area. Instead, one of its main characteristics is a cross-SBA feature, as remote sensing products are used in all of them and most CoPs are implied. It is a mature technology, continuously improved, and widely spread in recent years. Its weight in an Earth Observation project like GeoViQua is beyond doubt. In this context, remote sensing refers to sensors, raw products and pre-processing, from time of discovery to delivery of products. This is essential in GeoViQua as it guarantees control of data and associated quality elements traceability. There are many users employing

products derived from remote sensing scenes preprocessing (*e.g.,* geometric, radiometric corrections) as primary inputs in applied research. The radiance measurements (encoded as digital numbers) are the basis of several of the cross-SBA critical Earth Observation priorities list (*e.g.,* NDVI, land cover, surface humidity, vegetation cover, sea surface temperature). Delimiting quality issues at this stage is therefore crucial.

Dataset title
Geometric correction of Landsat series
User story: Started in 1972, the LANDSAT program has been providing a continuous record of Earth Observation. Spatial, spectral, and temporal resolutions of LANDSAT, acquired over a reasonably sized image extent, result in imagery that can be processed to represent land cover over large areas with an amount of spatial detail that is very attractive for monitoring, management, and scientific activities. However, it seems that the use of these images has not totally reached its potential due, in part, to difficulties in data access by non specialist users. Geometric correction is a basic procedure needed to overlay images from different dates, sensors or maps.
Contribution in GeoViQua
Primary processing from raw data (geometric corrections). LANDSAT images are widely used. Wide temporal extent, over 35 years, from early remote sensing imagery (1972-2011) Statistical-empirical modelling. High number of images corrected.
Sources: LANDSAT scenes. Ground control points (1:5000 ortophotomaps). Digital maps (15 m resolution DEM).
Methods: Statistical-empirical models: geometric correction with polynomials taking relief into account. Hundreds of ground control points (for fitting and for test) are obtained through an automatic matching procedure between LANDSAT and ortophotomaps (correlation analysis).
Product data model: Raster (more than 700 LANDSAT scenes), vector, points (ground control points).
Quality: Positional accuracy, absolute or external accuracy, RMSE. Available as metadata, report, points (X error, Y error, XY error). Quality testing potential due to the amount of data (LANDSAT scenes) and thorough knowledge (traceability control from raw data).
Scales: Regional (40,000 km^2). Spatial resolution: 15 m, 30 m, 60 m, 120 m depending on the sensor and band. Temporal resolution: 16 days (LANDSAT 4-7)/18 days (LANDSAT 1-3). Temporal extent: 1972-2011.
Parameter measured: Radiance, emissivity in the thermal band (digital number).
GEOSS: Cross SBA. Registered in CSR. Open source.
Missing gaps: Spatialized quality available as raster (X error, Y error, XY error).
Relation with other WP in GeoViQua: Functional requirement GeoViQua-137 “Framework for quantitative data quality estimation”

Dataset title
Radiometric correction of Landsat series
User story: A previous and essential step before working with remote sensing imagery is to process these images applying geometric and radiometric corrections. The radiometric correction process is especially important in multitemporal studies where large amount of imagery is used. Considering facts like the opening of the USGS LANDSAT archive world widely, thus offering to the scientific community the opportunity to use a huge number of images from the historical LANDSAT series (MSS, TM and ETM+), it is still more necessary to find a way for processing all these data, through a method providing realistic and consistent results between different images and the different sensors of the series, therefore avoiding too simplistic methods. Moreover, there is also the need for a standardized protocol to generate the LANDSAT products.
Contribution in GeoViQua
Primary processing from raw data (atmospheric and topographic corrections). MODIS and LANDSAT images are widely used. Wide temporal extent, over 35 years, from early remote sensing imagery (1972-2011). Physical modelling. High number of images corrected.
Sources: MODIS scenes, georreferenced LANDSAT scenes. 15 m resolution DEM.
Methods: Automatic protocol of atmospheric and topographic effects corrections (Pons and Solé 1994): the method estimates the radiance received by the sensor (from an area where only atmospheric contribution exists), and the atmospheric optimal depth, using pseudoinvariant areas with known reflectances values. Pseudoinvariant areas were obtained using MODIS scenes and verified with LANDSAT test imagery. Differences between estimated reflectance and the reference value for 3000 test pseudoinvariant areas are low and consistent (varying from -2% to + 2% in reflectance).
Product data model: Raster (more than 700 LANDSAT scenes), vector, polygons (pseudoinvariant regions).
Quality: Positional accuracy, absolute or external accuracy, RMSE. Thematic accuracy, quantitative attribute accuracy, accuracy of quantitative attributes, uncertainty at 68.3 % significance level. Available as technical report. Quality testing potential due to the amount of data (LANDSAT scenes) and thorough knowledge (traceability control from raw data).
Scales: Regional (40,000 km^2). Spatial resolution: 15 m, 30 m, 60 m, 120 m depending on the sensor and band. Temporal resolution: 16 days (LANDSAT 4-7)/18 days (LANDSAT 1-3). Temporal extent: 1972-2011.
Parameter measured: Reflectance (non-thermal bands) (%).
GEOSS: Cross SBA. Registered in CSR. Open source.
Missing gaps: Spatialized quality
Relation with other WP in GeoViQua: Functional requirement GeoViQua-137 “Framework for quantitative data quality estimation”

Dataset title
Intercomparison of atmospheric data measurements from remote sensing instruments
User story: Monitoring of the earth atmosphere using space-borne remote sensors has been performed for several decades through different missions (ERS, ENVISAT, EOS-AURA, METOP). These sensors typically deliver global coverage of atmosphere properties such as total column and/or altitude-dependent trace gas concentrations (water vapour, ozone, carbon monoxide, and nitrogen dioxide), pressure and temperature. Important aspects affecting data quality are uncertainties in calibration parameters (also known as Key Data Parameters), some of which have been measured on-ground only, and instrument degradation. The magnitude of some of these uncertainties can be estimated using the retrieval models that also derive the atmosphere properties from the raw instrument data. However such models may not be sufficient for detecting systematic errors in instrument performance. Hence, numerous initiatives for intercomparing measurement results from different space-borne instruments are employed, as well as comparing measurement results to ground/balloon/aircraft-based measurement results.
Contribution in GeoViQua
Variety of data sources (several spaceborne sensors) methods (interpolation, modelling, time series analysis). Spatialized quality, inclusion of covariance as quality measure for some measurements. Global scale. Atmospheric measurements.
Sources: Spaceborne sensors (multispectral, hyperspectral). Proposed: ENVISAT/GOMOS, MIPAS, SCIAMACHY Level 2 products.
Methods: Spatial interpolation from quantitative data, process-based modelling, time series analysis, physical modelling.
Product data model: Vectorial, points (measurement results are often expressed at a specific time, longitude, latitude, altitude), polygons (*e.g.,* SCIAMACHY total column density is expressed versus time on a 4 point-polygon defining longitude, latitude.
Quality: Positional accuracy, absolute or external accuracy, RMSE (*e.g.,* GOMOS). Completeness, omission, number of missing items, nearly always present (explicit or implicit). Thematic accuracy, quantitative attribute correctness, uncertainty at 95% significance level (could be 68% for some quantities), covariance (*e.g.,* also available for MIPAS). Spatialized quality: available as separate dataset in data product, uncertain values accompanying the data, and in quality disclaim manifests that are published on the web.
Scales: Global (sun-synchronous orbit yielding partial global coverage each day; full-global coverage through approx. 30 day repeat cycle). Single measurements typically correlate to several (3-30) km. Temporal resolution: single measurement approx. 1 to 10 s. Temporal extent: several years (*e.g.,* ENVISAT is operational since 2002).
Parameter measured: Atmospheric trace gas concentrations (*e.g.,* ppm NO_2, ppm O_3, ppm CO), pressure (Pa), temperature (K)
GEOSS. Related to CoPs rather than to SBA (although it could be considered cross-SBA). Top-ranked parameters in the critical Earth Observation priorities list.
Missing gaps: Overall consolidated a-posteriori accuracy for each measurement after intercomparison/assimilation (desirable as absolute value). The following data quality aspects can be studied: - Methods for propagation of uncertainty from raw data measurement to (Level 1/2) measurement result. - Inclusion of data quality information in the measurement data product

- Current methods for Informing the user on data quality for remote sensing datasets (quality disclaimers, Quality Information and Action Protocol) - Visualizing data quality of data products - Intercomparing/assimilating data from various remote sensors
Relation with other WP in GeoViQua: User story GeoViQua-72 "Users want to visualize data uncertainties both at the map level (overall accuracy) and at the pixel level (spatially distributed accuracy)".

Carbon cycle

Integrated Global Carbon Cycle Observations (IGCO) can provide the capacity to support the monitoring, reporting, and verification of the information required by future regulatory frameworks for the inclusion of natural CO_2 fluxes in post-Kyoto climate agreements.
This builds upon the efforts of the efforts of the IGCO CoP in GEO for timely provision of integrated carbon cycle observations world-wide required for their routine use in close collaboration with national governments, space agencies, and relevant technical experts. This pilot study will demonstrate this capability via establishment of robust methodologies, and a series of regional pilot studies, which will provide a template for consistent and reliable global carbon flux visualisation system. This will ensure permanent support for development of carbon Earth Observation visualisation interfaces and carbon tracking simulation models.

Dataset title
Estimates of global carbon fluxes from model data fusion
User story: Several estimates of the natural carbon fluxes exchanged between the land/ocean and the atmosphere, estimated either from "atmospheric inversion" (inverse procedure using atmospheric CO_2 measurements to estimate surface fluxes) or from a land/ocean ecosystem models are used to study the fraction of CO_2 initially released by anthropogenic emissions that is taken up by land and ocean ecosystems, its spatial and temporal variability. The different estimates can be used as an ensemble to assess the errors associated to the different approaches.
Contribution in GeoViQua
Sources: In-situ measurements have been developed to address the issue of time discontinuity: depending on the technique used, instruments provide only measurements for one gas (*e.g.,* infra-red non dispersive spectrometry) or for several green house gases (*e.g.,* gas chromatography). Such instruments can be placed at the surface or attached to a tall tower with several height of sampling in order to increase the fetch of the observatory. Since the late 1990s, aircraft sensors sample the planetary boundary layer (first kilometres of the atmosphere). Punctual campaigns occurred but also regular flights of small aircraft over precise locations, and commercial flights for a few routes around the world. More recently, Fourier Transform Infrared spectrometry (FTIR) has provided first retrievals of greenhouse gases columns in the atmosphere. Finally, satellite retrievals of atmospheric columns are already available for some gases (carbon monoxide, methane, ozone, nitrous oxide, formaldehyde) and will become available for CO_2 in the next years. These data, although not as precise and accurate as surface measurements, will fill the spatial gaps of the surface networks.
Methods: -For the inversions: assimilation of atmospheric CO_2 data within an inverse procedure (inversion of an atmospheric transport model) in order to optimize an *a priori* estimate of the CO_2 fluxes. -For the ecosystem model simulations: Process-based ecosystem models that described the absorption and release of carbon by the ecosystems (photosynthesis and respiration) for the land and the absorption

of CO_2 and the chemistry for the ocean.
Product data model: Rasters. 7-10 different flux estimates (1ºx1º)
Quality: Uncertainty estimates for few large regions (corresponding to the fluxes aggregated over these regions). Spatialized quality only for uncertainty estimates associated to two flux inversions.
Scales: Global. All fluxes are re-mapped onto a 1ºx1º degree global grid. Monthly fluxes (higher resolution for few estimates). Temporal extent variable depending on the estimates (from the last 7 years to the last 20 years): common period 1995-2005.
Parameters measured: Fluxes of CO_2 (KgC/m^2/hour)
GEOSS: Related to the Climate and Ecosystems SBAs.
Missing gaps: Uncertainties based on estimated variance-covariance matrix and ensemble of estimates. Spatialized quality available as uncertainty measure.
Relation with other WP in GeoViQua: User group GeoViQua-7 “Carbon CoP (former IGOS Carbon theme-IGCO)”, Functional requirement GeoViQua-105 “The system should support recording, editing, reviewing and subsequent querying of dataset provenance.”, User story GeoViQua-77 “Users want to visualise or access the data provenance information from each data source using workflow”

Climate

This pilot case will enhance the usefulness of the corresponding datasets by devising a means to provide information about associated data quality in a consistent fashion. The quality information already supplied with the data will be augmented with further information about context (*e.g.,* satellite images or aerial photographs of the location of the met station), and past experiences of expert users. The datasets involved enable practical applications of visualisation and quality elicitation components in GeoViQua. Since the combination of models and observations can reveal problems with data sources that were previously concealed, provenance is a key issue for models.
The climate pilot case constitutes a GEOSS SBA. Nowadays, the relevance of climatic issues is beyond doubt, and the influence of climatic parameters over ecosystems, health, agriculture or biodiversity shows the link to the rest of the SBAs and CoPs. The climatic change is indeed a hot topic. Thus, the inclusion of a climate pilot case in the GeoViQua pilot cases is mandatory. It is expected that both a local-regional and a continental-global scale approaches will be subjected to experimental testing in GeoViQua.

Dataset title
Digital Climatic Atlas of the Iberian Peninsula
User story: The Climatic Atlas contains 65 surface map models generating a high quality cartographic product with a rather complete time resolution. Climatic variables influence geochemical processes as well as species physiological traits and ecosystems dynamics. For instance, vegetation distribution is constrained by climatic factors. Precipitation surface models are adequate for the comprehensive prediction of extreme rainstorm events and floods. On the other hand, temperature and solar radiation surface models are relevant inputs for fire risk monitoring. Climatic surface maps are an essential source of information to include in GIS systems. The Climatic Digital Atlas results from the implementation of the respective surface maps in a WMS, which is an accessible platform that enables easy downloading and queries. The WMS is a standard map server that constitutes a helpful tool for the dissemination of contents to be used as inputs in diverse disciplines

research. In general, climatic maps are needed by researchers working in disciplines related with Earth sciences (climatology, hydrography, forestry sciences, agronomic sciences, vegetation sciences, etc.), as well as by land planning, and decision and policy makers, applying both to experts and general users (*e.g.,* journalists, citizens).
Contribution in GeoViQua
Precipitation and temperature are respectively first and third top-ranked in the cross-SBAs Critical Earth Observation priorities list. 50 years temporal extent. Open source dataset widely used in applied research.
Sources: Ground meteorological stations.
Methods: Spatial interpolation is widely used. Sufficient measurements for interpolation testing (*e.g.,* comparison between associated errors of calculations based on daily measures/average values).
Product data model: 65 raster layers (multiband raster).
Quality: Thematic accuracy quantitative correctness (uncertainty at 68.3% significance level, coefficient of determination). Available as XML metadata.
Scales: Regional extent (583,564 km^2). Spatial resolution 200 m. Temporal resolution: monthly and annual. Temporal extent 1950-2000.
Parameters measured: temperature, precipitation, solar radiation.
GEOSS: Parameters ranked in Earth Observation priorities list. Related to Weather SBA and Water cycle CoP. Registered in CSR (open source).
Missing gaps: Spatialized quality (likely, interpolation methods imply spatial and/or temporal error patterns that could be calculated and spatially visualised. Desirable format: multiband raster). Updating versions standards including associated quality.
Relation with other WP in GeoViQua: Functional requirement GeoViQua-155 “Spatially display the distribution of the uncertainty of the data values”, GeoViQua-47 “To establish a web-based Quality Control (QC) System”, Functional requirement GeoViQua-106 “The system should support the comparison of different model runs for a dataset to allow a user to establish a measure of uncertainty for the dataset.”

Water cycle

Water cycle as pilot case in WP7 is, as it happens with agriculture, a GEOSS SBA on its own right. The specific datasets in GeoViQua are also linked to weather somehow.

Some of the parameters measured in the water cycle datasets are employed in generation of other products, as inputs combined with other sources. For instance, the daily stress resulting from a combination of LANDSAT and water vapour image (within the water cycle testing datasets sample) is used in the Crop water monitoring using remote sensing dataset (see agriculture pilot case below). This fact could allow for tracing error propagation or related quality issues in the experimental testing to be developed in WP7, if feasible.

Dataset title
Snow surface map of Catalonia
User story: Water resources and its management are essential in many alpine mountainous areas. Snow cover monitoring in Catalonia requires obtaining accurate snow cartography. Snow data can be obtained by field campaigns, although this way usually does not provide a spatial and temporal cover of enough detail and quality. Alternatively, Remote Sensing could provide better snow cover estimation due to its spatial and temporal resolution. Accurate snow surface determination is essential to analyse its spatial and temporal dynamics and to estimate the volume of water derived from snow melting. Moreover, it is also important to determine the emissivity of this cover, an extremely useful data because it is used as an input parameter in thermal band atmospheric correction, and in species distribution modeling.
Contribution in GeoViQua
Combined use of MODIS and LANDSAT scenes, both widely used. Hard classifiers, ISODATA algorithm. Use of NDSI (snow index). Upscaling. Local scale. Recent data (2007-2011).
Sources: MODIS and LANDSAT-5. Digital maps.
Methods: Geometric and radiometric corrections (the latter implying physical modelling), NDSI (normalized difference snow index) for LANDSAT images. Unsupervised classification using hard classifiers (ISODATA algorithm) for MODIS. Changes in map projections. Process-based modelling: NDSI (snow index). Upscaling. Time series analysis/Land use & land cover dynamics.
Product data model: 300 raster layers.
Quality: Positional accuracy, absolute or external accuracy, RMSE. Available as txt files in metadata.
Scale: Regional (32,000 km^2). Resolution (500 m/pixel). Temporal resolution and extent: twice a week from 2007-2011.
Parameters measured: snow surface (km^2).
GEOSS: Related to water and weather/climate SBA. Registry in CSR (in process). Original scenes available in WMS.
Missing gaps: Compare with snow cover maps considering different territorial scales and methodologies (GIS based models versus remote sensing classification).
Relation with other WP in GeoViQua: User group GeoViQua-122 “Remote Sensing from satellite, the complete data chain Pilot Case”, User story GeoViQua-25 “An earth observation data user wants to know thematic accuracy of categorical maps”.

Dataset title
Water vapour image of Catalonia
User story: The MODIS Precipitable Water product consists of column water-vapour amounts. During the daytime, a near-infrared algorithm is applied over clear land areas of the globe and above clouds over both land and ocean. Over clear ocean areas, water-vapour estimates are provided over the extended glint area. An infrared algorithm for deriving atmospheric profiles is also applied both day and night for Level 2.

MODIS is the first space instrument to use near-IR bands together with the traditional IR bands to retrieve total precipitable water. Experience in this retrieval is based on an AVIRIS instrument onboard an ER-2 aircraft. Atmospheric water vapour should be determined with an accuracy of 5-10%.
Contribution in GeoViQua
Variety in methods (map projection, interpolation, data aggregation). Particular type of interpolation (reclassifications of flag values and application of selective filter to fill in all the NODATA values within the image). Recent data (2000-2011). MODIS scenes are widely used. Process-based modelling. High number of product data models (3,400 rasters).
Sources: MODIS (level 2 product).
Methods. Changing map projection. Temporal data aggregation (merging water vapour images from the same day into a single one). Particular type of interpolation (reclassifications of flag values and application of selective filter to fill in all the NODATA values within the image). Process-based modelling.
Product data model: Approximately 3,400 layers, composed by raster images.
Quality: Positional accuracy, absolute or external accuracy, RMSE. Metadata file.
Scales: Regional (32,000 km^2). Spatial resolution 1000 m. Temporal resolution: daily (2000-2011).
Parameter measured: total column of precipitable water vapour, NIR retrieval (km^2).
GEOSS: Related to water and weather/climate SBA. Registry in CSR (in process). Original LANDSAT scene available in WMS.
Missing gaps: Spatialized quality.
Relation with other WP in GeoViQua: User group GeoViQua-8 “Air Quality CoP”.

Dataset title
Water index image of Catalonia
User story: The water images are the result of the application of the Gao index developed in 1996; using the Landsat bands, NIR (Near-infrared, 0.760 – 0.900µm) and SWIR (Short-wavelength infrared, 1.550 – 1.750 µm). This is called Normalized Difference Water Index (NDWI). The SWIR reflectance reflects changes in both the vegetation water content and the spongy mesophyll structure in vegetation canopies, while the NIR reflectance is affected by leaf internal structure and leaf dry matter content but not by water content. The combination of the NIR with the SWIR removes variations induced by leaf internal structure and leaf dry matter content, improving the accuracy in retrieving the vegetation water content (Ceccato et al. 2001).
Contribution in GeoViQua
LANDSAT is widely used. Primary processing from raw data. Fine resolution. Wide temporal extent (20 years). Use of NDWI. Parameters related to soil (top-ranked in the cross-SBA critical Earth Observation priorities list).
Sources: LANDSAT-5.

Methods: Geometric and radiometric correction (the latter implying physical modelling). Changes in map projections. Process-based modelling: NDWI (normalized difference water index).
Product data model: 228 rasters.
Quality: Positional accuracy, absolute or external accuracy, RMSE. Metadata file.
Scales: Regional (32,000 km^2). Spatial resolution 20 m. Temporal resolution and extent: fortnightly (1989-2009).
Parameter measured: soil moisture (5 cm of soil surface)
GEOSS: Related to water, ecosystems, biodiversity and weather/climate SBA. Parameter ranked in Earth Observation priorities list. Registry in CSR (in process). Original LANDSAT scenes available in WMS.
Missing gaps: Spatialized quality.
Relation with other WP in GeoViQua: User group GeoViQua-122 “Remote Sensing from satellite, the complete data chain Pilot Case”.

Disasters

GeoViQua realise that disasters constitute a unique type of pilot cases, due to the peculiarities involved. Quality issues in this kind of scenario are subjected to other relevant factors: real time data collection and rapid response. It is widely recognised that early access to data is essential to provide scientific understanding of hazards necessary for improved risk assessment, forecast, and development of appropriate mitigation and adaptation strategies.

In this sense, GeoViQua evaluated considering a thorough post-response quality evaluation to a past event, or alternatively, a pilot case adapted to circumstances in real time scenarios, in the line of real time stations. If suitable, a pilot case dataset of this type will be added to the sample.

Agriculture

In its own right a GEOSS SBA, from the analysis of agriculture pilot case it is clear that working at a local scale allows for enhanced accuracy of results. GeoViQua is evaluating the possibility of include a global irrigation map (*e.g.,* GIAM) in the dataset sample, although the suitability of global agriculture maps for testing in WP7 needs further assessment. A particular contribution of these datasets in GeoViQua is providing categorical and ordinal variables to the sample. Both temporal dynamics (*i.e.,* time series analysis of quantitative variables and analysis of land use & land cover dynamics associated to categorical parameters) are covered. As for quality, a combination of remote sensing imagery, training areas and adequate field survey, together with statistical approaches (*e.g.,* error matrix) would be highly desirable.

Dataset title
Crop maps of northern Catalonia
User story: In land management or in agriculture, mapping or monitoring dynamics of crops may be chiefly helpful. In this sense remote sensing data may be a very useful tool in crop detection. For example, a very common task is the discrimination and monitoring of crop phenology (irrigated or dry crops) using satellite images. Nevertheless, this objective may be rather challenging especially in case of fragmented areas as typical Mediterranean landscapes. In time series analysis, the presence of clouds is

a recurrent problem, making of crop phenology continuous monitoring a hard task. Another issue would be the pixel size of satellite images: in general, a higher spatial resolution allows for a higher accuracy in crops classification.
Contribution in GeoViQua
Primary processing from raw data (physical modelling). Hybrid classifiers. Categorical attributes. In agriculture, working at a local scale favours accuracy of output results. Multitemporal recent data.
Sources: 9 LANDSAT images (multitemporal). Ground-control points.
Methods: Physical modelling (not strictly). Classification: hybrid classifiers (starting with the unsupervised process to obtain many classes and using the supervised process to objectively assign these classes to the final categories). Land use & land cover dynamics. Ground-control points (training areas).
Product data model: Raster, vector (polygons).
Quality: Positional accuracy, absolute or external accuracy, RMSE. Thematic accuracy, classification correctness, error matrix (rate of correct attribute values). XLS files.
Scales: Local. Pixel size 20 m. Annual temporal resolution. Temporal extent 2002-2010.
Parameter measured: Crop type (categorical variable).
GEOSS: Agriculture SBA. Land cover and vegetation cover (top-ranked in the critical Earth Observation Priorities list).
Missing gaps: Intensifying of field surveys (training areas). Spatialized quality.
Relation with other WP in GeoViQua: User story GeoViQua-25 “An earth observation data user wants to know thematic accuracy of categorical maps”

Dataset title
Vegetation indices from northern Catalonia
User story: Vegetation indices are the most widely used technique for monitoring vegetation density of plant, phenology and vigour. Among all, the Normalized Difference Vegetation Index (NDVI) is the most common measurement. NDVI is also used as an indicator of drought. The estimation of this index over a local area of northern Catalonia has been essential in order to monitor agriculture crops, determine phenological state and water stress level. Moreover, NDVI has been a reliable indicator of crop type automatic classification in time series for categorical map production but overall NDVI has been valuable information for water management decision makers given its drought estimation property.
Contribution in GeoViQua
LANDSAT scenes and NDVI vegetation index are widely used in applied research. High number of raster products (346). The use of NDVI provides an example of reliability. NDVI top-ranked in the critical Earth Observation priorities list. Recent data. In agriculture, working at a local scale favours accuracy of output results. Multitemporal recent data.
Sources: 9 LANDSAT images (multitemporal).

Methods: Primary processing from raw data (radiometric corrections: physical modelling). Process-based modelling. Indicators of crop stage and vigour included in the automatic classification. Time series analysis.
Product data model: 346 rasters.
Quality: Positional accuracy, absolute or external accuracy, RMSE. The use of NDVI provides an example of reliability. Available as metadata.
Scales: Local. Pixel size 20 m. Temporal resolution: annual. Temporal extent 2002-2010.
Parameter measured: Crop stage and wetness (based on NDVI calculations).
GEOSS: Agriculture SBA. NDVI (top-ranked in the critical Earth Observation Priorities list).
Missing gaps: Accuracy of results limited to NDVI saturation susceptibility. Spatialized quality calculation and visualisation. Field surveys could improve associated quality. The use of NDVI on its own commonly implies the impossibility of delimiting product accuracy.
Relation with other WP in GeoViQua: Functional requirement GeoViqua-135 “Quantitative estimates of data quality”.

Dataset title
Monitoring environmental measures using remote sensing
User story: South-west of Catalonia is environmentally singular given the wetland formed due to delta of Ebre's river. This delta is part of the Ramsar convention on wetlands. With the purpose of verifying the flooded rice crop stage required in order to receive the agroenvironmental financial assistance, a multitemporal flooding dynamics monitoring of the fields was carried out. This monitoring allows the classification of crops into different flood types and flood stages according to several thresholds. The categorical maps produced during the time series analysis are essential for the local Ramsar agents to determine the whole time of flooding. Apart from that this measurements are interesting for studies on habitat distribution according to flooding type and its time evolution.
Contribution in GeoViQua
LANDSAT scenes are widely used in applied research. Primary processing of raw data (radiometric corrections: physical modelling). Particular type of categorical variable (ordinal). Multitemporal dynamics of rice fields based on recent data. In agriculture, working at a local scale favours accuracy of output results.
Sources: 6 LANDSAT images. Ground-control points.
Methods: Radiometric corrections (physical modelling). Discriminant analysis of corrected images (statistical and empirical modelling). Land use & land cover dynamics.
Product data model: 10 raster layers.
Quality: Positional accuracy, absolute or external accuracy, RMSE. Thematic accuracy, classification correctness, error matrix (rate of correct attribute values).
Scales: Local. Pixel size 20 m. Temporal scale and extent: annual images from 2005-2010.

Parameter measured: Flood stage. Ordinal categorical variable: qualitative quantification range (*e.g.*, very dry, moderately dry, flooded, moderately flooded, highly flooded).
GEOSS: Agriculture SBA, but also linked to Ecosystems. Top-ranked in the critical Earth Observation Priorities list).
Missing gaps: Intensifying of field surveys (training areas). Spatialized quality.
Relation with other WP in GeoViQua: Scientific objective GeoViQua-12 "Statistical framework to quantify and manage uncertainty".

Dataset title
Crop water monitoring using remote sensing
User story: Since 2002 to present, the Department of Geography through an agreement with the Catalan Water Agency is conducting follow-up irrigation campaigns for irrigation communities in the Lower Ter and Muga basins through Remote sensing imagery from satellites Landsat-5 TM and Landsat-7 ETM +. This information is used to to know which crops are active and, therefore, watered, and its consumption. Thus, these data allow us to perform a calculation of consumption of water for the communities with the aim to control spending on water and irrigation to make recommendations to improve management. For these purposes the Metric model to retrieve actual evapotranspiration and the crop water stress have been implemented.
Contribution in GeoViQua
Primary processing from raw data. Comparison between LANDSAT 5 and LANDSAT 7 images (different spatial resolutions). Thermal bands. Variety in methods. Physiological parameter measure elaborated with physical and statistical models, in contrast with employment of NDVI. Variability in temporal resolution. Recent multitemporal data. Lack of quality is common to many datasets: alternative approach of quality assessment.
Sources: LANDSAT 5, LANDSAT 7. Thermal bands.
Methods: Physical modelling in images corrections. Spatial interpolation from quantitative data. Process models. Physiological parameter measure elaborated with a physical model (in turn implying sub-models and a high number of independent variables) combined with a statistical model.
Product data model: 41 evapotranspiration and 232 crop water stress raster layers.
Quality: Not available for evapotranspiration. Positional accuracy, absolute or external accuracy, RMSE for intermediate products (net solar radiation, global solar radiation, air temperature). Available as metadata.
Scales: Local (1,986 km^2). Pixel size 120 m (LANDSAT 5) and 60 m (LANDSAT 7). Temporal resolution: 16 days during growing season (evapotranspiration), 16 days all year round (crop water stress). Temporal extent: 2002-2010 (*i.e.*, average but detailed temporal resolution).
Parameter measured: Crop evapotranspiration (mm day^{-1}), crop water stress (° C) (*i.e.*, physiological parameter)
GEOSS: Agriculture SBA, but also related to climate/weather. Linked to several top-ranked cross SBA critical Earth Observations priorities list. Restricted to internal use in GeoViQua, only with testing

purposes.
Missing gaps: Evapotranspiration is lacking quality. Desirable quality measures: thematic accuracy quantitative correctness (uncertainty at 68.3% significance level, coefficient of determination, RMSE), spatialized quality.
Relation with other WP in GeoViQua: Functional requirement GeoViqua-135 “Quantitative estimates of data quality”.

Air quality

This pilot case is directly related to the Air Quality CoP and the Health SBA. Although a single dataset, it is a complex one. It includes diverse parameters, real time data collection and almost real time distribution using interoperable interfaces. Users of earth observations related to the health impacts of air quality comprise the general public, air quality managers, and a variety of scientific disciplines. The characteristics of this dataset make it a suitable candidate for developing a scenario in GeoViQua.
The GEO CoP is a self-organized voluntary group that fosters the application of earth observations to air quality management and science by leveraging synergies of sharing and collaboration. Community activities include collaborative website, workshops, gathering user requirements, sharing tools and best practices as well as aiding other communities (data portals, science teams and decision support activities). A persistent output facilitated by the Air Quality CoP is a network of air quality/atmospheric composition data servers that use OGC-standard protocols and the GEOSS interoperability process.

Additionally, the air quality pilot case study will involve remote sensing earth observation products that are to be determined during the project. They comprise direct remote data, like meteorological data such as wind, and indirect, secondary data such as land-use classifications. These additional data sources are used as input for a forecasting and calibration system, but also as an augmentation of the primary dataset presented in the table below. However, that system is yet to be developed.

Dataset title
Air quality (in-situ)
User story: AirBase is a public air quality database containing air quality monitoring information for more than 30 participating countries throughout Europe. Every year countries report air quality measurement data for a set of pollutants at a representative selection of stations. Statistics calculated on the reported data are also stored in AirBase. AirBase is managed by the European Topic Centre on Air and Climate Change (ETC/ACC) on behalf of the EEA. AirBase data comes in two flavours: As a validated dataset provided with a 1.5 to 2 year delay, and as unvalidated real-time data.
Contribution in GeoViQua
In-situ data provided in near real-time. Complex product involving variety of sources at different levels (in-situ, remote data, ancillary data). Continental scale. User interface, low band-width mobile devices. Temporal extent (around 40 years). Top-ranked in the critical Earth Observation priorities list.
Sources: Stationary digital sensors recording concentrations of pollutants. -Direct in-situ data (used for calibration): raw data, eventually modelled (*e.g.,* transport model) as interpolated data provided via a WMS interface in near real-time. -Direct remote data (to be calibrated): satellite instruments aggregate measurements for the whole atmosphere. Sources: channels in the GEONATCast and ESA-DDS data streams. -Complementary earth observation data: forecasting system input, (*e.g.,* land use layer, weather

conditions).
Methods: Spatial interpolation. Assimilation module, putting together the measured values and the model values, doing the dynamic calibration. The model will be a forward model, like AUSTAL 2000.
Product data model: Vector point layer (in-situ measurements by more than 6000 stations).
Quality: Not available for real time data. Validated data provided with a quality check report.
Scales: Continental. Temporal resolution and extent: annual, 1969-2008.
Parameter measured: SO_2, NO_2, NO, NO_x, O_3, CO, particulate matter (hourly values, daily values, daily 8-hours maxima, annual mean, percentiles, maximum). Wind speed.
GEOSS: Related to Health SBA, to a lower extent to weather. Parameters top-ranked in the critical Earth Observation priorities list. Registred in CSR. European Environment Agency standard re-use policy: permitted provided that the source is acknowledged.
Missing gaps: Discovery problem exists, even more so on mobile devices. In-situ data could be based on mobile sensors, or on stationary sensors extended by mobile sensors where necessary. Visualization use case could be an expert flying around in a helicopter looking at the augmented reality map with the air quality in the area under him. Spatialized quality available as metadata.
Relation with other WP in GeoViQua: User stories GeoViQua-79 “Users want the system to provide a quality-aware augmented reality browser including mobile devices”, GeoViQua-67 “Users want to find certain error distribution pattern via visualisation of error estimates”.

Land use

The land use datasets in GeoViQua are related to ecosystems, agriculture and biodiversity SBAs. This pilot case is directly connected to landscape dynamics.
An interesting remark is that the method of photointerpretation and digitizing are unique to this pilot case. The GeoViQua sample includes photointerpretation of ortophotomaps and satellite imagery. The latter is used in widely employed datasets (*e.g.,* CORINE land use). Other relevant contribution is a detailed and hierarchic categorical variables datasets the different scales of the datasets (local and regional). In particular, as it was the case with some datasets in the water cycle pilot case, the land cover map is used in applied research related to landscape metrics. Thus, some traceability issue could be derived. In addition, for the SIOSE dataset, quality is available for a local area within the dataset region whereas quality measures are lacking in the regional dataset as a whole. This fact is interesting towards quality approaches. It is not common to find quality reports in photointerpretation datasets.
Errors and uncertainties are propagated when combining diverse data models (*e.g.,* raster vs. vector), combining data with different resolution, combining data from different dates and with different transect models or combining categorical data where categories do not nest or correspond. Analysis that requires data fusion (*e.g.,* a land cover change map which results from combining two maps generated through different methodologies) has to cope with the incomplete metadata and unknown data quality that will be propagated through the process. This pilot case will provide data that can be used to verify quality protocols and visualisation.

Dataset title
Land cover map of Catalonia
User story: The Land cover map of Catalonia is a high-resolution thematic map about the main land covers of Catalonia (forests, crops, urban areas, etc.). Because of the high spatial and thematic detail, this product is used in order to improve the landscape knowledge and also to value the land use from an ecologic and economic point of view.
Contribution in GeoViQua
Photointerpretation. Digitizing. Field survey, ancillary use of SPOT images. Categorical variables. Highly detailed hierarchic legend (241 categories). Four editions enable comparison between ancient/recent versions and technical improvements (four levels in legend detail, starting from 4 classes). Land use & land cover dynamics. This map is used in landscape metrics. Local scale. Fine spatial resolution. High amount of multitemporal layers.
Sources: Colour ortophotomaps (including NIR ortophotomaps when possible). Ancillary data: habitat maps of Catalonia (1:5000), forest inventories, SPOT images in the latest version.
Methods: Photointerpretation of ortophotomaps (1:25000, 1:5000, 1:2500) and ancillary data. Digitizing. Land use & land cover dynamics.
Product data model: 2 m pixel size raster (detailed legend level). Polygons (with topological structure).
Quality: Positional accuracy, absolute or external accuracy, RMSE. Thematic accuracy, non-quantitative attribute correctness, rate of incorrect attribute values. Metadata (text describing thematic errors are about 3%).
Scales: Regional (32,000 km^2). First edition (1:25000), 2nd and 3rd editions (1:5000), 4th edition (1:5000 but using ortophotomaps at 1:2500). Temporal resolution: 1993, 2000-2002, 2005, 2009. Fine spatial resolution. High amount of multitemporal layers.
Parameter measured: Land cover classes structured in a hierarchic legend. Also containing statistics related to classes, landscape indices at sheet and cover level (number of polygons, minimum area, maximum area, mean area, Shanon index, equity index).
GEOSS: Land and vegetation cover are top-ranked in the cross SBA critical Earth Observation priorities list. Open source. CSR Registered.
Missing gaps: Misclassification matrix. Desired quality available as accuracy for cover and for sheet.
Relation with other WP in GeoViQua: User story GeoViQua-76 “Users want to visualize the quality information through desktop and web-based GIS tools”.

Dataset title
SIOSE (Information system on land cover of Spain)
User story: The objective of SIOSE is to integrate the information of the land cover and land use of Spain. SIOSE project is used as a tool for environmental planning and management, with potential use in: -Land use dynamics studies -Causes and consequences of natural or artificial processes, such as desertification -Environmental impact assessment

-Ecological variables monitoring -Land use planning -Environmental modelling -Coastal zones monitoring
Contribution in GeoViQua
Photointerpretation of SPOT scenes (fine resolution). Digitizing. Categorical variables. Complex hierarchic legend with four different levels and 45 composite categories and 7 simple categories. Quality reports available for a control local area within the study area (local scale) providing a wide array of quality elements (parameters, indicators, measures).
Sources: False colour composite of SPOT bands.
Methods: Photointerpretation of spaceborne sensor scenes. Digitizing
Product data model: Vector layer (polygons) linked to a geodatabase.
Quality: Quality reports only available for Catalonia (checking control items). Positional accuracy, absolute or external accuracy, RMSE. Logical consistency, domain consistency, value domain non-conformance, naming convention, coherence between adjacent sheets. Logical consistency, conceptual consistency, conceptual schema compliance, reference system, minimum size of the polygons, polygon attributes. Logical consistency, format consistency, physical structure conflicts. Logical consistency, topological consistency, number of invalid self overlap errors. Completeness, omission, missing items, metadata completeness coherence. Thematic accuracy, non-quantitative correctness, rate of incorrect attribute values. Available as technical report (not distributed with the data). It is not frequent to find quality reports in photointerpretation datasets.
Scales: Regional (504,000 km^2). Spatial resolution 1:25000. Temporal resolution: 2005, 2009, 2011. Spatial resolution of SPOT scenes is higher than those of LANDSAT/MODIS.
Parameter measured: Land cover classes. Complex hierarchic legend with four different levels and 45 composite categories and 7 simple categories.
GEOSS: Satellite imagery photointerpretation is the base of widely spread datasets (*e.g.,* CORINE land use cover). Land and vegetation cover are top-ranked in the cross SBA critical Earth Observation priorities list (land cover, vegetation cover).
Missing gaps: Quality assessment for the whole extent. Implementation of associated quality in appropriate format to the dataset characteristics.
Relation with other WP in GeoViQua: Functional requirement GeoViQua-155 “Spatially display the distribution of the uncertainty of the data values”.

Marine

A marine pilot case would cover the particularities involved when dealing with water, as a complement to terrain and air, already mentioned. Indeed, a great extent of the Earth surface is covered by sea water.
This pilot case has potentially a high impact to the Coastal Zone CoP since the data evaluated for the FAO's study area is relevant to many of the CoP's objectives and activities for managing coastal zones, fisheries and aquaculture. This pilot case will provide a validation of the project's technical developments and

achievements. The context lies under the umbrella of activity of the International Ocean-Colour Coordinating Group (IOCCG).

Dataset title
Argo Ocean floats
User story: Argo is a global array of 3,000 free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection.
Contribution in GeoViQua
Vertical profile of measurements. Global. 10,000 new profiles per month. Sea surface temperature is top-ranked in the cross-SBA critical Earth Observation priorities list. Related to two SBA: Climate, Water, and CoPs: Coastal zone, Water cycle, Biodiversity.
Sources: Instruments attached to floats sampling ocean water (3,000 floats).
Methods: Spatial interpolation.
Product data model: Vector layers (point), though actually vertical profiles.
Quality: Flags (qualitative).
Scale: Global.
Parameters measured: ocean temperature, salinity, other chemical components, up to 1500 m depth.
GEOSS: Example in the Data Sharing Action Plan. Parameter ranked in Earth Observation priorities list. Open source.
Missing gaps: Not started yet.
Relation with other WP in GeoViQua: Functional requirement GeoViQua-155 “Spatially display the distribution of uncertainty of the data values”

Finally, the following table summarizes the main characteristics covered by the pilot cases testing dataset sample.

Table 5. Summary of aspects covered by the pilot cases testing datasets collection

	Techniques												Source						SBA									Time				Scale[1]			
	Reprojection	Changes in data models	Digitizing	Classification	Modelling	Generalization	Interpolations	Field survey	Landscape metrics	Time series analysis	Land use and land cover	Photointerpretation	Satellite imagery	In-field	Airborne sensors	Digital maps	Paper map	Digital models	Disasters	Health	Energy	Climate	Water	Weather	Ecosystems	Agriculture	Biodiversity	Near real time	Daily-weekly	Annual	Multiyear	Global	Continental	Regional	Local
Remote sensing		X			X		X			X			X			X		X	Cross SBA										X	X	X	X			X
Carbon cycle	X	X			X			X					X	X	X							X			X					X	X	X			
Climate		X					X							X		X		X				X	X	X	X	X	X		X	X	X	X		X	
Water cycle	X	X		X	X	X	X						X			X						X	X	X					X		X				X
Disasters																																			
Agriculture		X		X	X		X	X		X	X		X	X									X	X		X			X	X	X				X
Air quality		X			X		X						X	X		X				X		X		X				X	X	X	X			X	X
Land use		X	X					X			X	X	X	X	X	X									X	X	X			X	X			X	X
Marine								X						X								X	X				X					X			

[1] Spatial scales: Global/Continental/Regional/Local. See glossary

It is to be noted that remote sensing images are SBA neutral and can be used everywhere but do not exclusively affect any of them.

4.2 Demonstration scenarios

The user-oriented approach related to demonstrative scenarios in WP7 uses this term understood as platforms to learn from, as tools for use in tasks development and as media for interacting with other people, as stated by Carroll (1999).

According to this author, scenario-based design addresses five technical challenges:

-Scenarios evoke reflection in the content of design work, helping developers coordinate design action and reflection.

-Scenarios are at once concrete and flexible, helping developers manage the fluid of design situations.

-Scenarios afford multiple views of interaction, diverse kind and amounts of detailing, helping developers manage the many consequences entailed by any given design move.

-Scenarios can also be abstracted and categorized, helping designers to recognize, capture, and reuse generalizations, and to address the challenge that technical knowledge often lags the needs of technical design.

-Finally, scenarios promote work-oriented communication among stakeholders, helping to make design activities more accessible to the great variety of expertise that can contribute to design, and addressing that designers and clients often distract attention from the needs and concerns of the people who will use the technology.

This approach seeks to exploit the complexity and fluidity of design by trying to learn more about the structure and dynamics of the problem domain, by trying to see the situation in many different ways, and by interacting intimately with the concrete elements of the situation.

It is a coherent and concrete vision, not an abstract goal or a list of requirements. Elements of the envisioned system appear in the scenario embedded in the user interactions that will make them meaningful to the user. The scenario exposes not only the functionality of the system, but specific claims about how the user will access that functionality and what the user will experience in doing so.

At this stage some examples have been reviewed on possible methods for developing the scenarios design in GeoViQua (Koy et al. 2011). The idea is to include attractive and dynamic visualisation tools, downloadable material, comprehensive contents and foster community interaction.

Upon the thematic areas of scenarios in GeoViQua, a first draft was discussed over in the second Project Technical Board meeting. Although some specific topics and relationships with adequate user stories of available datasets were pointed out, the final selection of datasets and, specially, development of scenarios is to be carried out in later stages of WP7. The development and concretion of scenarios is, at the present stage, a step behind the compilation and analysis of available datasets presented in section 4.1. Nevertheless, the scenarios approach requires a maturer development of other activities in GeoViQua and will be built integrating services, tools and technologies under development in other WP, as well as after a deeper analysis of the involved interactions with providers, users, system boundaries, etc. Indeed, and as in other work packages, WP7 expects to reach the user of the outcomes generated in GeoViQua. This methodology of design of scenarios is intended to follow this line. Feedbacks from users and other work packages would be highly desirable. The design of scenarios would ease this interaction and exchange of experiences.

In agreement with Carroll (1999), scenarios are work-oriented, design objects. A design process in which scenarios are employed as a central representation will *ipso facto* remain focused on the needs and concerns of users (Carroll and Rosson 1990).

According to the advances in the very recent second Project Technical Board meeting, a first proposal of especially interesting scenarios to be developed is:

- Carbon cycle
- Disasters
- Agriculture
- Air quality
- Marine

Our approach tries to facilitate design actions informed by reflection on multiple levels and from multiple perspectives, including direct collaboration among team members. We argue that making scenarios of use a focal design object serves this variety of purposes. Summarising, as has been aforementioned in previous sections, in GeoViQua, scenarios go beyond the datasets features (*e.g.,* experimental testing limited to tackle specific quality measures in a sample of datasets), providing practical examples of the relationships of the datasets involved with the surrounding environment. In this sense, GeoViQua scenarios are aimed at providing a unique background where outcomes, results and tools developed in GeoViQua are integrated, considering interaction with communities, system boundaries and the flow of processes. In short, scenarios will provide an overview of the services that will be demonstrated.

5. WORKFLOW AND CONTRIBUTIONS

Integration of geographical information science fields, geoservices, standardization, metadata, applied work and software development are also part of the daily tasks of partners involved in WP7.

UAB acknowledges a wide practical experience on geospatial datasets analyses, and consequently plays a coordinator role in WP7. Other partners involved significantly contribute or will contribute to the general design of the WP7, for instance, providing experience in services and interaction with wider communities, thus polishing the experimental testing, the relationships with the surrounding environment, and assisting in the building of demonstrative scenarios.

It is to be noted that the involvement of partners is essential in this WP. As it is, GeoViQua requires updated data, varied sources and testing of methods, covering diverse topics at different scales. Moreover, some partners produce data themselves, which is crucial in control from time discovery of provenance, traceability and over the propagation (modelling, processing, and delivery of products), and the adequate definition of scenarios.

Apart from their respective contributions to the pilot cases datasets, subgroups were defined for the development of specific issues and scenarios. For instance, sharing experience in delivering global coverage products, CEA and ESA would elaborate an example on carbon cycle, checking possible alignments on the climate case with UREAD.

This report particularly applies to a compliance of tasks 7.1 and expressly 7.2. Nevertheless, it is also linked with several tasks in WP7, constituting an input for tasks 7.3 and 7.4 and other tasks in GeoViQua.

5.1 GCI services/integration (task 7.1)

There is a deliverable specifically related to task 7.1. This task has already been started, and embodies the analysis of GCI components and services, as well as the GEOPortal search and visualisation tools. In fact, this deliverable and the selection criteria of the pilot cases testing dataset, the main body of WP7, results from the exploring of GEOSS resources. The exploration of GEOSS components and services from a GeoViQua perspective will be thoroughly explained in the corresponding deliverable: D7.2 Report on current GCI and identification of issues in integration with GEOSS.

As was stated, among the selection criteria, there was the prioritization of selecting datasets already integrated in GEOSS. In the introduction, it was explained that the timing of this process does not necessarily match GeoViQua's flow.

At this point, the template designed records the link to the registry of the proposed dataset in the components and services of GCI as well as links to other related registered datasets that could benefit from the inclusion of the dataset in the selection of pilot cases of GeoViQua and support building of related scenarios.

5.2 Quality parameterisation (task 7.2)

This report is partially focused on task 7.2. Thus, the weight of quality parameters, indicators and measures in GeoViQua, a particular issue of WP7, is shown in the tables and has been explicitly described in this deliverable.

The fulfilment of the template for each pilot case dataset accomplishes this task, at the same time embedding results from WP2 (user requirements definition), WP3 (quality elicitation), and applying the specifications of WP6 to all pilot case studies.

Indeed, most partners are involved in this task: 52N, CEA, CREAF, ESA UAB, UREAD, and S&T. In fact, the latter was not initially compromised. However, the potential of datasets they could provide for WP7 and their expertise in quality related issues in geospatial data resulted in volunteering contributions related to this task.

In addition, ASTON plays a coordinator role within GeoViQua, introducing in WP7 applied statistical concepts, terms and modelling, thus securing accuracy of results and the proper use of quality measurements adequate to the targets of GeoViQua. As was stated in section 3.4, the corresponding deliverables linked to WP2, WP3, and WP6 will further develop quality parameterisation.

In short, task 7.2 enables gathering the correct description of the associated quality for each pilot case considering spatial, thematic and temporary parameters.

Illustrative scenarios together with the implementation of the selected datasets in GEOSS encourages participation on user feedback, promoting an environment to collect expert/volunteer valuable information for the pilot case studies.

5.3 Validation of visualisation, search services and GEO Label (tasks 7.3 and 7.4)

Subtler, at least in the sense of overlapping and current development, results the interaction of the present deliverable with tasks 7.3 (validation of quality aware visualisation and search services) and 7.4 (validation of the GEO Label). However, the selection of datasets, resulting from this deliverable is the key input to be used in these tasks, related to WP5 (quality aware visualisation) and WP4 (search services). In addition, the development of scenarios is the appropriate frame in which these tasks will be implemented.

The GEO Label is related to a particular activity that the Scientific and Technical Committee in GEOSS is planning to implement through the ST-09-02 task. The work content is described in the Task Sheet of ST-09-02 as: “Develop a concept for a GEO Label related to the scientific relevance, quality, acceptance and societal needs for activities in support of GEOSS as an attractive incentive for involvement of the Scientific and Technological communities”. In GeoViQua, The GEO Label is linked to WP6 and could be part of WP7 activities. GeoViQua is expecting the formalisation of a draft concept, which is to be proposed liaising with existing major Earth Observation data providers.

In these tasks, the experimental tests will be defined and carried out. Results from these validation tests will probably shed some light on the concepts and parameters proposed in the GeoViQua context, confirming their usefulness and applicability. Therefore, the interaction (two-sided inputs and feedbacks) of WP7 with WP4, WP5 and WP6 is patent. It must be kept in mind that specific datasets will be extracted from the selected sample depending on the experimental testing requirements. The success of GeoViQua can be achieved only by collecting adequate datasets for the precise features to be tested. Likewise, scenarios complement this approach, supporting the implementation of results and tools to be developed in a proper frame in which the surrounding background is taken into account (*e.g.,* interactions with communities, system boundaries, interfaces, services and components).

5.4 Relationships with other work packages

As was aforementioned in the introduction, WP7 is a complex work package implying inputs and feedbacks from and to the rest of work packages in GeoViQua.

From the template, the relationship with WP2 becomes evident in the epigraph stating whether the dataset assists in testing/gathering user requirements as defined in GeoViQua. This field reflects the usefulness of the dataset for serving different objectives. User requirements and uses cases, which are being defined in WP2, might be testable in WP7, previous to the implementation. Furthermore, user requirements and use cases might be identified and described by the pilot cases datasets and/or the scenarios, providing a specific input from WP7 to WP2. In particular, the user groups' requirements are straightforwardly connected to the pilot cases.

Thus, in WP7, a prior step is the on-going selection of user requirements susceptible of experimental testing. In the case of use cases, which are in first instance abstract concepts not directly linked to real datasets, the approach in WP7 is linked to the demonstration scenarios: use cases are assessed in WP7 by means of scenarios where one or several use cases at a time can take a more defined shape when combined with real geospatial data.

Some notes from WP3 on quality measures state that to achieve the optimal information interoperability we should focus on accuracy measures. In general it is not necessary to separate temporal, spatial and thematic accuracy. The accuracy of the value can be defined as the relation of the value given to the value in reality (where reality is well defined - the target and the spatial and temporal support of the observation / variable). It is very unlikely that we can provide the accuracy uniquely. On the contrary, it will be achievable to define its statistical properties. Else, if more could be defined, then it would be possible to correct for it, and remove the bias / error.

As a consequence, the measures we should use to characterise accuracy are actually measures of uncertainty, indeed most often epistemic uncertainty (uncertainty due to our lack of knowledge, not some intrinsic randomness). Thus, the natural set of measures to use to describe accuracy would not be strictly the ones depicted in the ISO19138 (ISO-TC2112006) (in addition currently being revised by the ISO19157, under development, from which WP7 is expecting new insights on this issue). A polished version of the template would be complemented by definitions contained in the UncertML dictionary, provided by ASTON. The UncertML describes in a very clear manner how to encode uncertain values using Statistics, Distributions and Realisations. All of the types within UncertML have rigorous statistical definitions, which means they can be used interoperably in further processing of the data.

As for the time being, WP7 directly deals with tasks 3.1 (quality elicitation in WP3) and 6.1 (encoding in WP6):

-Task 3.1. Metadata extraction quality component.
This task is carried out by CREAF, UAB and S&T. CNR is also involved since this partner uses a system implementing this task, which is being integrated into GEOSS.

In turn, this directly links to Task 6.1.

-Task 6.1. Encodings for data quality.
AST is the leader partner in this task, necessarily coordinated with WP7 and in particular with the work explained in this document.
S&T and 52N are nominally involved too. Ideally the encodings should respect the requirements established in WP2.

-Tasks 3.2, 3.3 and 3.4. Spatial error indices, cal/val, thematic indices,
AST will lead these tasks, collecting ideas and comments on the proposals transferred to GeoViQua.

-Task 3.5 User feedback and quality assessment of data sets in GEOSS.
Task 3.5 is led by FRAUN.

As explained above, other tasks in WP7 run closer to other work packages, in a more practical way: tasks 7.3 and 7.4 actually validate and carry out the testing procedures for WP4 and WP5 (this is to say, search and visualisation).

Further details on the interaction of tasks and work packages are found in the Description of Work and other deliverables. GeoViQua's site keeps updated and dynamic records of each work package involved. Teleconferences and meetings are part of the workflow in GeoViQua. The two Project Technical Board meetings held have been productive and clarifying as regards synergies between work packages and tasks in GeoViQua. Progress in WP7 has been reported in plenary sessions during the second Project Technical Board meeting. Moreover, a proper split session was dedicated to WP7, which resulted in valuable conclusions, to be directly applied in WP7 planning and development. At the time of closing this deliverable, telecons on the ensuing steps and experimental testing examples, and defining scenarios related to WP7 are scheduled as preliminary discussion points that together with the standardization of pilot cases testing datasets constitute refined material for the coordination of tasks in GeoViQua within work packages in the next face to face meeting, to be held in November.

Finally, an overview of WP7 in GeoViQua is shown in Figure 2.

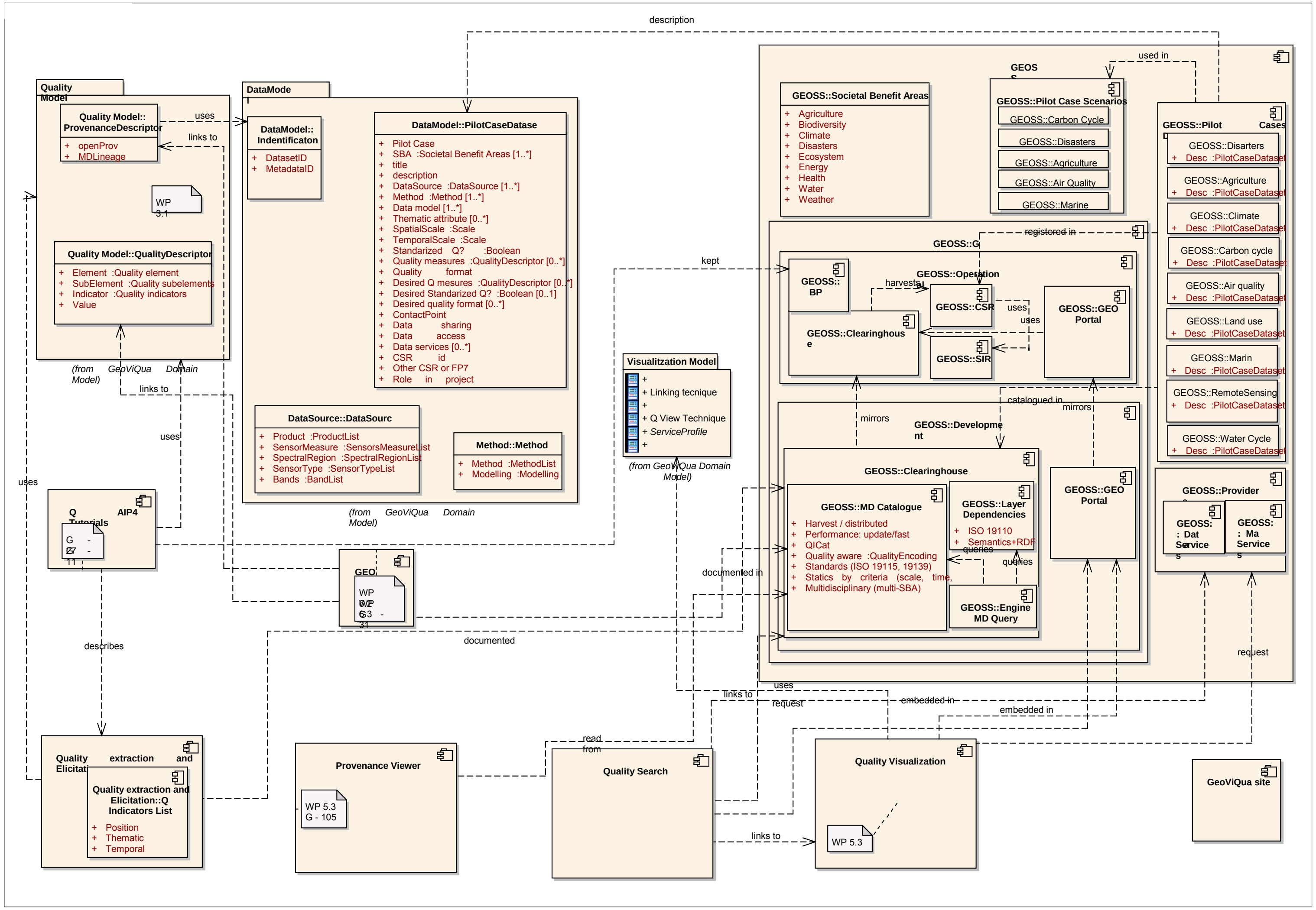

Figure 2. Work package 7 in GeoViQua.

6. CONCLUSIONS

The process of selecting pilot cases datasets for experimental testing in WP7 is dynamic. The initial pilot cases classification outlined in the Description of Work evolved into a comprehensive pilot cases selection adapted to detected needs in GeoViQua (*i.e.,* pilot cases testing datasets, pilot cases scenarios).

The pilot cases testing datasets sample has been carefully thought over to cover a wide spectrum of geospatial datasets features, aiming at constituting a representative sample of real states. Part of the results might provide appropriate material for a scenario-based design with dissemination purposes, adequately selected considering factors suchlike social impact or illustrative features. Thus, experimental testing will scientifically approach specific issues related to the datasets (*e.g.,* quality measures statistic refinement) whether scenarios will focus on provide illustrative examples of integration of results and tools developed in GeoViQua within the surrounding background beyond datasets features (*e.g.,* design of interfaces, implementation of search or visualisation tools, interaction with communities, system boundaries).

The template design, adapted to the framework of GeoViQua, has proved useful in dataset characterization. The main aspects of geospatial datasets are included in the formal dataset description table: sources, methods, quality elements and scales. Other elements related to GeoViQua context are also included (*e.g.*, GEOSS SBA, CSR, data sharing conditions, role in GeoViQua). A polished version of the template could be used in the design of a standardized description form for geospatial datasets (*e.g.,* adapted to GEOSS community needs). In this sense, the description table will be uploaded in the website. Open accessibility and feedback mechanisms will be provided.

Fulfilling the description table and the definition of selection criteria ease decision-making on pilot cases testing datasets selection for GeoViQua purposes. Thus, the documentation of pilot cases datasets in a standard format as stated in the Description of Work is achieved. At the same time, the description table provides a homogeneous frame of formalisation of geospatial datasets features. The systematization makes evaluation of each dataset straightforward, enabling testing design and supporting building of scenarios in subsequent WP7 tasks.

Whenever possible, international standards must be used (*e.g.,* ISO standardized quality measures, OGC, GEOSS SBAs, critical Earth Observation priorities list). However, in the specialized frame of GeoViQua, not strictly sticking to standards might be preferred. Particularly, when seeking improvement in certain approaches, agreed specific clarifications or additional measures are allowed (*e.g.,* quality measures suitable for specific cases not currently contemplated in the quality ISO standards, statistically refined UncertML conceptual contributions related to quality indicators and quality assessment). It must be remarked that the main target of GeoViQua is the enhancement of quality issues (*e.g.,* elicitation of quality parameters availability, quality aware visualisation and search tools, establishment of quality certification procedures).

The analysis of the pilot cases testing datasets description tables confirms the sample covers particular aspects relevant for experimental testing in WP7. Although some gaps have been detected, the state-of-the-art is represented. If possible, WP7 will try to enlarge the sample introducing relevant sources (*e.g.,* active sensors, hyperspectral sensors, microwave or ultraviolet spectral regions records). Some of the missing items are indeed rarely found when dealing with geospatial datasets and geoprocessing (*e.g.,* gravity fields, sound waves); in fact, certain challenges in GeoViQua imply fostering the use of some items (*e.g.,* quality measures availability, spatialized error patterns). The variety of datasets will favour extracting the most appropriate sample for each experimental design requirements (*e.g.,* GEO Label testing is likely to demand both categorical and quantitative parameters datasets; a global scale approach significantly differs from a local scale approach; validity of quality indicators testing necessarily implies using a wide sample).

Workflow is dynamic with key partners playing an active role, essential features in WP7 evolvement. Next steps will imply actual experimental testing, as required from WP4, WP5 and WP6, using the most suitable datasets for each circumstance. WP2 and WP3 provide input features essential to the realisation of tasks in WP7. The exhaustive exploration of the GCI is an on-going process in WP7 that might add some contributions to the pilot cases testing datasets sample, at the same time guaranteeing coherence with GEOSS development. Inputs and feedbacks between work packages are described in the GeoViQua components model, a conceptual scheme of the whole project at the same time providing a dynamic tool (which is shared by all work packages to favour coordination of tasks and interactions).

In summary, the present deliverable reports progress in WP7 and liaisons with other work packages in GeoViQua. The consortium agrees that the steps undertaken are adequate to the targets defined in the Description of Work. WP7 is accomplishing the role initially intended, constituting a sound base for further project development in total harmony with GEOSS perspective. The pilot cases testing datasets have been documented in a standard description table, formalising geospatial datasets main features. The template designed enables analysis of real datasets that will be subjected to subsequent testing and will lead to the composition of demonstration scenarios.

7. GLOSSARY

-AIP4 stands for the **4th phase of the GEOSS Architecture Implementation Pilot** (AIP-4). The leadership corresponds to the Open Geospatial Consortium, Inc. (OGC®).

AIP-4 will improve access to GEOSS datasets that support the "Critical Earth Observation Priorities" that have been identified by the GEO User Interface Committee. It will increase the use of these data by building on the accomplishments of prior AIP phases. AIP-4 aims to:

- Increase on-line access to "Critical Earth Observation Priorities Data Sources"
- Ensure datasets are discoverable through the GEOSS Common Infrastructure
- Demonstrate effectiveness of general and specialized software tools for using data.

GeoViQua is linked to AIP4: GEO announced a Call for Participation (CFP) in AIP-4 and GeoViQua is responding to the call with a coordinated effort to produce a tutorial on data quality. The team is formed by GeoViQua, QA4EO, and ESIP IQ cluster. GeoViQua is waiting for a new AIP5, where our project could maybe contribute in a clearer way.

-Architecture and Data Committee (ADC). The Architecture and Data Committee supports GEO in all architecture and data management aspects of the design, coordination, and implementation of the Global Earth Observation System of Systems (GEOSS) for comprehensive, coordinated, and sustained Earth observations.

Objectives:

- Enable GEO, based upon user requirements and building on existing systems and initiatives, to define the components of GEOSS, and to converge or harmonize observation methods, and to promote the use of standards and references, intercalibration, and data assimilation.
- Enable GEO to define and update interoperability arrangements to which GEO Members and Participating Organizations agree to adhere, including technical specifications for collecting, processing, storing, and disseminating shared data, metadata and products.
- Enable GEO to facilitate data management, information management, and common services, and will help to promote data sharing principles in support of the GEO Plenary for the full and open sharing and exchange of data and information, recognizing relevant international instruments and national policies and legislation.

-Community of Practice (CoP). A Community of Practice (CoP) is a user-led community of stakeholders, from providers to the final beneficiaries of Earth observation data and information, with a common interest in specific aspects of societal benefits to be realized by GEOSS implementation.

-**Critical Earth Observation priorities list**

Some Earth Observations may be critical to a particular field, such as disasters or agriculture. On the other hand, some observations may be commonly needed by many fields. The activity, known as Task US-09-01a, examined needs across a range of users to identify critical and common observations. Precipitation, Soil Moisture, and Surface Air Temperature were the highest-ranked priority observations. This analysis resulted in a list defined for each of the SBAs, and a summarising list of overall, cross-SBAs priorities. The latter contains the following parameters:

- **Precipitation.** This parameter includes observations of the phase, amount, frequency, and duration of precipitation; precipitation in thunderstorms; and extreme precipitation events.
- **Soil Moisture.** Soil Moisture is on the list of priorities for the 8 SBAs in the Cross-SBA analysis. Observation needs include surface and sub-surface soil moisture.
- **Surface Air Temperature.** Surface Air Temperature is on the list of priorities for the 8 SBAs in the Cross-SBA analysis. The related Land Surface Temperature is on the list of priorities for 5 SBAs and is within the 30 highest-ranked parameters.
- **Surface Wind Speed.** Surface Wind Speed is on the list of priorities for 7 SBAs. Users typically distinguish between a need for surface wind direction vs. speed. The parameter Upper Level Winds is related and is a priority for 4 SBAs.
- **Land Cover.** Land Cover is on the list of priorities for 6 SBAs. As a derived parameter, Land Cover draws on multiple data sources. The parameter Urbanization is related and is within the 30 highest-ranked parameters.
- **Surface Humidity.** Surface Humidity is on the list of priorities for 7 SBAs. The related parameter of Upper Level Humidity is within the 30 highest-ranked parameters. Together, these two parameters are critical for global numerical weather prediction models.
- **Vegetation Cover.** Vegetation Cover is on the list of priorities for 5 SBAs, including Agriculture, Ecosystems, Health, Water, and Weather. Some users require additional information such as Vegetation Type and Vegetation Indices.
- **Surface Wind Direction.** Surface Wind Direction is on the list of priorities for 6 SBAs. In the overall rankings shown here, this parameter is slightly less critical than Surface Wind Speed.
- **Normalized Difference Vegetation Index (NDVI).** Normalized Difference Vegetation Index is on the list of priorities for 5 SBAs. NDVI is often used as a surrogate for vegetation parameters such as biomass, status, greenness, and moisture, which may be the underlying user data needs.
- **Sea Surface Temperature (SST).** Sea Surface Temperature is on the list of priorities for 5 SBAs. SST is critical for seasonal to long-range forecasts. Other ocean parameters such as Ocean Topography and Sea Level also are within the 30 highest-ranked parameters.

-**CSR. Components and Services Registry**.

The GEOSS Components and Services Registry provides a formal listing and description of all the Earth observation systems, data sets, models and other services and tools that together constitute the Global Earth Observation System of Systems. These various components are being interlinked using standards and protocols that allow data and information from different sources to be integrated. The components and services listed on the Registry can be searched and explored by decision-makers, managers and other users of Earth observations via the GEO Portal.

The success of the GEOSS Components and Services Registry will depend on the commitment of GEO Members and Participating Organizations to input and update their registration details on a regular basis. Registration involves completing a standardized form on-line.

-**GCI. Global Common Infrastructure** in GEOSS (Global Earth Observation System of Systems). The GCI consists of web-based portals, a clearinghouse for searching data, information and services, and registries containing information about GEOSS components, standards, and best practices.
The goals of GCI in 2011 seek a facilitated access to datasets and other resources supporting the Earth Observation Priorities identified in UIC report, enable discovery or and facilitate access to data-core contributions.

To achieve convincing and demonstrable benefits at the GEO-VIII Plenary, a short term action is underway (ADC, GCI providers and Others) to collaborate bi-laterally with a few data providers to identify datasets supporting the Critical Earth Observation Priorities.

To mutually adapt interfaces to enable effective data discovery and access ('fewer clicks to the data'), better usability/user experience.

-GEO. Group on Earth Observations.

GEO is an intergovernmental organization working to improve the availability, access, and use of Earth observations to benefit society. It is composed of a voluntary partnership of 148 governments and international organizations, launched in response to calls for action by the 2002 World Summit on Sustainable Development and by the G8 (Group of Eight) leading industrialized countries. GEO is coordinating efforts to build a Global Earth Observation System of Systems (GEOSS).
http://earthobservations.org/about_geo.shtml.

-GEO Label

The Science and Technology Committee committed itself to developing the concept of a voluntary GEO Label. This was intended to:
encourage scientists, researchers, and others to contribute their data and systems to GEOSS by offering an accepted voluntary label that provides recognition that their contribution is valued by the GEO community.
differentiate components, data and products delivered through GEOSS and provide a "trusted brand" to GEOSS users; member governments may base their decisions on data/products of such contributions.
highlight the importance of GEOSS to those previously unaware they were reliant on this initiative for their data or product.
The active task states that such a label should assist the user to assess the scientific relevance, quality, acceptance and societal needs of the components. These parameters clearly contain a mix of objective and subjective assessments.
GeoViQua follows the evolution of this task and actually participates with some people involved in the GEO Label group, linked to other institutions and projects connected to this activity (*e.g.,* EGIDA project). Specifically, members from ASTON, CNR, CREAF and UAB, led by ESA, constitute the GEO Label teamwork in GeoViQua. As regards this deliverable, some contributions from GeoViQua will come from WP7 (*e.g.,* task 7.4 might include testing of GEO Label features, gathering user requirements, design of surveys, etc.).

-GEOPortal. The GEO Portal is a website that provides convenient access to the full range of GEOSS data and information.

Operated by the European Space Agency (ESA) and the Food and Agriculture Organization (FAO) of the United Nations (UN), it provides a web-based interface for searching and accessing data, information, imagery, services and applications. It connects users to a variety of databases, services and portals that provide reliable, up-to-date, integrated and user-friendly information – vital for the work of decision-makers, managers and other users of Earth observations.

The content available via the Portal continues to expand at a rapid rate and promises to reach a critical mass in the near future.

-GEOSS. Global Earth Observation System of Systems.
GEOSS builds on national, regional, and international observation systems to provide coordinated Earth observations from thousands of ground, in situ, airborne, and space-based instruments. GEO is focused on enhancing the development and use of Earth observations in nine SBAs: Agriculture, Biodiversity, Climate, Disasters, Ecosystems, Energy, Health, Water, and Weather.

-GeoViQua. GeoViQua stands for a FP7 project related to Quality aware Visualisation for the Global Earth Observation System of Systems (GEOSS). The focus is on adding quality specifications to spatial data in order to improve reliability in scientific studies and policy decision-making. To achieve our targets, our team is collecting pilot cases spread over the whole Earth Observation chain.

Work packages in GeoViQua:

-WP1: Project management
-WP2. Requirements for data quality visualisation
-WP3: Data quality elicitation mechanisms
-WP4: Enhanced geo-search components
-WP5: Quality aware visualisation components
-WP6: Delivery of solutions to end users
-WP7: Pilot case studies
-WP8: Dissemination and Capacity building

The Description of Work is the reference document which contains a complete account of GeoViQua project and related work packages.

-**Lidar. Light Detection And Ranging**, also LADAR, is an optical remote sensing technology that can measure the distance to, or other properties of a target by illuminating the target with light, often using pulses from a laser. LIDAR technology has application in geomatics, archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing and atmospheric physics, as well as in airborne laser swath mapping (ALSM), laser altimetry and LIDAR Contour Mapping.
The acronym LADAR (Laser Detection and Ranging) is often used in military contexts. The term "laser radar" is sometimes used even though LIDAR does not employ microwaves or radio waves and is not therefore radar even though both systems employ electromagnetic radiation.

-**Metadata.** This term refers to data complementary information, providing details about data features. In the case of geographic information data (*e.g.,* remote sensing scenes, categorical maps resulting from the application of digital classification techniques to remote sensing scenes, etc.), metadata would cover for instance identification aspects, discovery date, accuracy, distribution conditions, horizontal reference system, etc. There are several metadata standards defining metadata contents (*e.g.,* ISO19115).

-OGC. Open Geospatial Consortium.
The Open Geospatial Consortium, Inc. (OGC®) funded in 1994 is an international consortium of more than 420 companies, government agencies, research organizations, and universities participating in a consensus process to develop publicly available geospatial standards, GIS data processing, or data sharing. OGC Standards empower technology developers to make geospatial information and services accessible and useful with any application that needs to be geospatially enabled. Visit the OGC website at http://www.opengeospatial.org. Among the most well-known standards, the Web Map Server (WMS) that enable remote sensing images and maps visualisation, the Web Coverage Service (WCS) that is providing downloadable images servers, or the Web Feature Service Interface Standard (WFS) that provides and interface allowing requests for geographical features across the web using platform-independent calls, can be cited. Some of the OGC standards (*e.g.,* WMS) have been accepted as ISO standards.

-**Quality.** The definition contained in the ISO multilingual glossary states quality is the degree to which a set of inherent characteristics fulfils requirements (ISO 9000:2005).
It is recommended that quality conformance levels based on user requirements refer to the data quality elements defined in this standard.

Quality parameters (in bold) and associated quality indicators (dashes)

Completeness
Completeness is defined as the presence and absence of features, their attributes and relationships. It shall be described by applicable data quality elements from the following list:

- commission: excess data present in a dataset;
- omission: data absent from a dataset.

Logical consistency
Logical consistency is defined as the degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical). If these logical rules are documented elsewhere (for example in a product specification) then the source should be referenced (for example in the data quality evaluation). It shall be described by applicable data quality elements from the following list:

- conceptual consistency: adherence to rules of the conceptual schema.
- domain consistency: adherence of values to the value domains.
- format consistency: degree to which data is stored in accordance with the physical structure of the dataset.
- topological consistency: correctness of the explicitly encoded topological characteristics of a dataset.

Spatial accuracy
Spatial accuracy is defined as the accuracy of the position of features in relation to Earth. It shall be described by applicable data quality elements from the following list:

- absolute accuracy: closeness of reported coordinate values to values accepted as or being true.
- relative accuracy: closeness of the relative, spatial positions of features in a dataset to their respective relative, spatial positions accepted as or being true.
- gridded data position accuracy: closeness of gridded data spatial position values to values accepted as or being true.

Thematic accuracy
Thematic accuracy is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. It shall be described by applicable data quality elements from the following list:

- classification correctness: comparison of the classes assigned to features or their attributes to a universe of discourse (*e.g.,* ground truth or reference dataset).
- non-quantitative attribute correctness: measure of if a non-quantitative attribute is correct or wrong.
- quantitative attribute accuracy: closeness of the value of a quantitative attribute to a value accepted as or known to be true.

Temporal quality
Temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. It shall be described by applicable data quality elements from the following list:

- accuracy of a time measurement: closeness of reported time measurements to values accepted as or known to be true.
  NOTE: Time measurement may be either a defined point in time or a period.
- temporal consistency: correctness of the order of events.
- temporal validity: validity of data with respect to the format and calendar specified for the dataset.
  NOTE March 33 is an example of invalid data where the specified format is ISO 8601 compliant.

Usability

Usability is the degree of adherence to a specific set of data quality requirements.
Usability shall be used to describe specific quality information about a dataset's adherence to a particular application or requirements.
If the other data quality elements listed in this International Standard do not sufficiently address a component of quality, Usability shall be used.
NOTE For example, with this element, a data producer can show for one dataset, with quantitative elements, how it fits different identified usages. This element may be used to declare the conformance of the dataset at a particular specification.

Quality measures: A data quality measure is a test applied to evaluate data quality elements. Quality measures in GeoViQua are based on ISO19138 (ISO-TC211 2006) standards.

-Partners

GeoviQua is supported by the participation of 10 partners, including several universities, research centres, companies and the European Space Agency, from different European countries (France, Germany, Italy, Netherlands, Spain, and United Kingdom).

- 52° North GmbH (Germany): GeoViQua acronym: 52N

52° North GmbH (52N) is an SME that forms the back office and service centre of an international research and development network aiming at fostering the process of innovation in the field of geoinformatics. This network is driven by leading research organizations and individuals of the international GIS community. Partners participate in the development of innovative technologies and their transformation into practical solutions. Principal participants and shareholders of the 52° North GmbH are: con terra - Gesellschaft für Angewandte Informationstechnologie mbH, Münster (Germany), International Institute for Geo-Information Science and Earth Observation (ITC), Enschede (The Netherlands), Institute for Geoinformatics in University of Muenster (Germany), and Environmental Systems Research Institute, ESRI Inc., Redlands (California, USA).
The participants have a long and outstanding record in the domain of geoinformatics, spatio-temporal modelling, spatial data infrastructures, software architectures, and standardisation processes. 52° North and its participants are actively involved and leading the OGC standardisation process for web based geoprocessing (Chair of the OGC WPS 2.0 Standards Working Group), sensor web enablement (Chair of the SOS 2.0 OGC Standards Working Group) and digital rights management (Chair of the OGC GeoRM 1.0 Standards Working Group) as well as in INSPIRE drafting teams.

- Aston University (UK): GeoViQua acronym: AST

Aston University is a technology and application focussed University with areas of international excellence. Within Aston University, the School of Engineering and Applied Sciences (SEAS) has roughly 80 members of academic staff working in areas including Computer Science, Sustainability and Mathematics. Within SEAS, the world-class Neural Computing Research Group (NCRG) has for many years been at the forefront of developments in machine learning. The developing Knowledge Engineering Group (KEG) works within the areas of knowledge representation and management, database integration and interoperability, and software engineering. The Sustainable Environments Research Group (SERG) focuses on the application of GIS and modelling methods to address broad sustainability issues including conservation planning, renewable energy and disease monitoring.

- Commissariat a l'Energie Atomique (CEA). LSCE -Laboratoire des Sciences du Climat et de l'Environnement (France). GeoViQua acronym : CEA

The Institute LSCE (Laboratoire des Sciences du Climat et de l'Environnement) is a joint research unit of the Centre National de la Recherche Scientifique (CNRS-INSU), the Commissariat à l'Energie Atomique (CEA), (two major funding agencies in France) and the University of Versailles-Saint Quentin. For this project, CNRS will not participate and CEA alone will carry out and be responsible for the work on this project through LSCE personal. The LSCE is part of the Institut Pierre Simon Laplace (IPSL) in Paris and covers numerous fields of research related to climate and environment. The experience and qualification of LSCE researchers for the Global Carbon Project comprise a unique expertise in atmospheric composition monitoring, development of process-based models over land and ocean, and inversion methods to quantify sources and sinks of greenhouse gases using atmospheric, space-borne, and in-situ observations. The LSCE researchers also played a leading role in the development of Earth System modelling in France. The organizational experience of the LSCE in the European research area is reflected through its active participation in several EU funded projects, GEMS-IP, CARBOEUROPE-IP, QUANTIFY, CARBOOCEAN, CARBOAFRICA and NITROEUROPE-IP. The LSCE coordinates GEOMON-IP dealing with groundbased atmospheric observation complementary to satellite, and IMECC and ICOS dealing with infrastructure carbon cycle measurements. The LSCE is strongly involved in the GMES core services for land, atmosphere and ocean.

- Consiglio Nazionale delle Ricerche (CNR). IMAA - Istituto di Metodologie per l'Analisi Ambientale (Italy). GeoViQua acronym : CNR.

The CNR-IMAA (Institute of Methodology for Environmental Analysis) of the National Research Council of Italy (CNR) is a research Institute, whose research concerns the study of the atmosphere and of the Earth's surface, and of human impacts on these, using remote sensing and environmental and geophysical monitoring. It is a centre of competence of the Italian Civil Protection Department of the Prime Minister's cabinet. The Earth and Space Science Informatics Laboratory (ESSI-Lab) of CNR-IMAA was established to facilitate the effective and seamless provision of Earth and Space resources to Information Society applications. ESSI-Lab research focuses on the application of information and communications technologies to manage, share and harmonize Earth and Space Science data, information, services and knowledge in the framework of Geospatial Information technologies and infrastructures. A major research activity concerns crosswalks between the Earth and Space Science and Geospatial Information communities, investigating interoperability solutions. CNR-IMAA is a member of the OGC (Open Geospatial Consortium).
CNR-IMAA has specific knowledge and experience on international framework and standard related to Earth Science Systems resources access and interoperability.

- CREAF: Centre for Ecological Research and Forestry Applications (Spain). GeoViQua acronym: CREAF

The Centre for Ecological Research and Forestry Applications (CREAF) is a public research institution that was created in 1987 by the Generalitat (Autonomous Government) of Catalonia, the Autonomous University of Barcelona (UAB) the Institute of Catalan Studies (IEC) and the University of Barcelona (UB), to promote basic and applied research in terrestrial ecology. Some of these contributions are the publication of numerous scientific papers in international academic journals and the development of numerous scientific methodologies and technological tools such as the GIS&RS software MiraMon.

- ESA: European Space Agency (France): GeoViQua acronym: ESA

The European Space Agency (ESA) is Europe's gateway to space. Its mission is to shape the development of Europe's space capability and ensure that investment in space continues to deliver benefits to the citizens of Europe and the world. ESA's purpose shall be to provide for, and to promote, for exclusively peaceful

purposes, cooperation among European States in space research and technology and their space applications, with a view to their being used for scientific purposes and for operational space applications systems.
ESRIN, known as the ESA Centre for Earth Observation, is one of the five ESA specialised centres situated in Europe. Earth Observation data has grown in importance as more and more international and national agencies recognise the many uses to which it can be put. Satellites for Earth observation keep a constant watch over the Earth and the data they provide help to safeguard the planet in which we live. Since 2004, ESRIN has been the headquarters for ESA's Earth Observation activities.

- Fraunhofer Institut Graphische Datenverarbeitung -IGD- (Germany): GeoViQua acronym: FRAUN

The Fraunhofer Gesellschaft (FhG) is the leading organization of applied research in Germany, undertaking contract research on behalf of industry, the service sector and the government. At present, the organization maintains 80 research establishments at 40 locations throughout Germany with around 13,000 employees. The Fraunhofer Gesellschaft will contribute to the GIGAS Project through Fraunhofer Institute for Computer Graphics Research (IGD). FhG focuses on the development of product prototypes (hard- and software) and the realisation of concepts, models, and solutions for computer graphics and its adoption to specific application requirements. The work is rounded off by object-oriented basic research projects and the realization of single devices and computer graphics systems with pilot character (*e.g.,* virtual and augmented reality, mobile computing, and security technologies). The tasks of the Graphic Information Systems (GIS) department at FhG are research, development and services in the field of Graphic Information Systems, especially on 3D GIS, location based services and spatial data infrastructure technologies. The activities cover research as well as end user and producer-relevant services.

- S[&]T Corporation (The Netherlands): GeoViQua acronym: S&T

S&T Corporation (S&T stands for Science & Technology) is an SME conducting high-tech projects and consultancy related to technical software engineering and analysis. S&T started in 2000 and has grown to a diverse group of 30 scientists and engineers. Typical customers and participants are the European Space Agency (ESA), TNO, EADS, ASTRIUM, VEGA, Siemens, SNECMA, Logica and others. S&T has a broad experience in building data quality monitoring, control and visualisation systems for various EO-missions, such as ENVISAT, GOCE, and Radar Altimetry. Typically these systems involve extraction of data from both instrument raw data and processed product flows, processing the data into quality data, conditional handling of EO data based on data quality analysis results, and visualisation and reporting data quality to users. The systems also allow quality engineers to analyse these data interactively in order to perform in-depth investigation.
In addition to its quality monitoring software development activities, S&T is currently involved in various standardisation projects for data quality, including investigating application of QA4EO to its software. S&T aims will contribute the lessons learnt from the implementation of industrial data quality monitoring systems and from its current quality information standardisation activities to the GeoViQua project.

- UAB: Universitat Autònoma de Barcelona (Spain): GeoViQua acronym: UAB

The Universitat Autònoma of Barcelona (UAB) was founded in 1968 and is 2nd in the national ranking of universities (2004-2005) and recently has been recognized as a campus of international excellence by the Ministries for Education (MEC) and for Science and Innovation (MICINN) of the Spanish Government. The research group in the Department of Geography aims to propose new algorithms, methodologies and tools for GIS, Remote Sensing, Cartography and land dynamics. UAB has broad experience in image processing of remote sensors, having worked with images of low, medium and high spatial resolution from both satellite and airborne sensors. The members of the research group have published about 150 scientific papers and obtained a "Research, innovation and development" award from the Catalan Government.

- University of Reading (UK): GeoViQua acronym: UREAD

The University of Reading is ranked as one of the UK's 10 most research intensive universities and one of the top 200 universities in the world, with a world-class reputation for teaching, research and enterprise. The University's research in environmental science is rated among the very highest in the UK, and in 2006 the University received the Queen's Anniversary Prize for Higher Education for exceptional contributions to Meteorology. The University enjoys very strong scientific and technical links with the UK Met Office and the European Centre for Medium-Range Weather Forecasting. The UK National Centre for Earth Observation (NCEO) is coordinated from the University of Reading.
The Reading e-Science Centre, formed in 2003, is an interdisciplinary group of scientists and technologists at the University that aims to apply modern computing techniques to environmental science. The ReSC is known internationally for its work in applying OGC-based techniques to problems of environmental data sharing and visualisation. Its open-source ncWMS software (an implementation of the Web Map Service protocol for multidimensional gridded data) is used by research groups, government agencies and private industry around the world.

-Pilot case. Understood in the context of this deliverable, pilot cases conform thematic areas in which datasets are classified in WP7 in GeoViQua. The pilot cases are: Remote sensing, Agriculture, Climate, Water cycle, Land use, Carbon cycle, Air quality, Marine, and Disasters. These pilot cases result initially from the DoW, and the needs detected in WP7 since the beginning of the project, intended to adapt GEOSS SBAs and CoPs, to the particular targets of GeoViQua.

Due to the development of WP7, Pilot cases have evolved into the following:

-Pilot cases testing datasets. Specific datasets that have been collected with the aim of covering a representative range of methods, sources, quality indicators, scales and main features of geospatial databases. Validity of the scientific experiments to be carried out in WP7 in GeoViQua requests a varied sample, in order to be able to optimize the match between datasets and test requirements (*i.e.,* extraction of the most appropriate materials for each of the methodology design involved in specific tests). For instance, categorical parameters imply different statistical approaches than quantitative parameters. GeoViQua targets can only be addressed by coping with this diversity of geospatial aspects.

-Pilot cases scenarios. It is likely that experimental testing and the integration of results from WP7 and other work packages in GeoViQua lead to a proper development of scenarios, that will contribute to the demonstration of GeoViQua results. Scenarios will naturally derive from practical work with real datasets, and will be accordingly selected, considering aspects suchlike social impact, potential use, display of representative results in GeoViQua, etc. The most representative characters of the design of scenarios in GeoViQua will be addressing the implementation of tools and outcomes developed in GeoViQua within the surrounding environment. This is to say, beyond the experimental testing of specific features of the dataset, scenarios comprise other equally relevant aspects related to geospatial datasets: dynamic flow of processes, interfaces, system boundaries, interaction with the community, technical assistance and support, involving a wide range of users and providers.

-PRODUCT MODEL
Raster.

Data model used for representing and organizing spatial information, consisting in the space division following a square regular grid (only rarely rectangular grid) resulting in individual units of information (*i.e.,* tiles or pixels).
Raster image comes in the form of individual pixels, and each spatial location or resolution element has a pixel associated where the pixel value indicates the attribute, (*e.g.,* colour, elevation, ID number). A raster

image is normally acquired by optical scanner, digital CCD camera and other raster imaging devices. Its spatial resolution is determined by the resolution of the acquisition device and the quality of the original data source.

Vector.
Data model used for representing and organizing spatial information consisting in the geometric representation of spatial entities. Vector data comes in the form of points, lines, or polygons that are geometrically and mathematically associated. Points are stored using the coordinates, for example, a two-dimensional point is stored as (x, y). Lines are stored as a series of point pairs, where each pair represents a straight-line segment, for example, (x1, y1) and (x2, y2) indicating a line from (x1, y1) to (x2, y2). Vector data models are diverse both in information content and structure (with or without attributes, with or without topology, etc.). The most complex contain thematic and temporal attributes for each geographic entity, that can be linked to a relational model, hierarchic model, etc.

-**Data processing levels.** To facilitate the discussion of data processing in practice, several processing "levels" were first defined in 1986 by NASA as part of its Earth Observing System and steadily adopted since then, both internally at NASA (*e.g.,* http://eospso.gsfc.nasa.gov/ftp_docs/2006ReferenceHandbook.pdf) and elsewhere (*e. g.,* http://www.grassaf.org/general-documents/products/grassaf_pum_v121.pdf); these definitions are Level 0, 1a, 1b, 2, 3, 4. It is important to note that not all space agencies and data distributors follow these definitions.
A Level 1 data record is the most fundamental (*i.e.,* highest reversible level) data record that has significant scientific utility, and is the foundation upon which all subsequent data sets are produced. Level 2 is the first level that is directly usable for most scientific applications. Its value is much greater than the lower levels. Level 2 data sets tend to be less voluminous than Level 1 data because they have been reduced temporally, spatially, or spectrally. Level 3 data sets are generally smaller than lower level data sets and thus can be dealt with without incurring a great deal of data handling overhead. These data tend to be generally more useful for many applications. The regular spatial and temporal organization of Level 3 datasets makes it feasible to readily combine data from different sources.

-**Radiance.** Radiance and spectral radiance are radiometric measures that describe the amount of light that passes through or is emitted from a particular area, and falls within a given solid angle in a specified direction. They are used to characterize both emission from diffuse sources and reflection from diffuse surfaces. The SI unit of radiance is watts per steradian per square metre ($W \cdot sr^{-1} \cdot m^{-2}$).

-**Reflectance.** In optics and photometry, reflectivity is the fraction of the energetic flux reflected by a surface and the incident radiation. It is a dimensionless parameter, values between 0 and 1, although commonly expressed as a percentage. It is a key parameter in remote sensing because it is dependent on the nature of the object. In general it must be treated as a directional property that is a function of the reflected direction, the incident direction, and the incident wavelength.

-**Scales.** Definitions of scales can be based on geographical, political or physiographic considerations, considerations of climate homogeneity, or considerations of model resolution. Hence, in most cases, an operational definition must be adopted. In this deliverable, the spatial scale is viewed from the extent perspective. At this stage, it is preferred to assume a general and objective concept to cope with the variety of datasets, from diverse disciplines. Thus, in WP7, so far we will attach to the following approximations:

- Global: *e.g.,* planetary
- Continental: *e.g.,* European scale
- Regional: *e.g.,* Scandinavian Peninsula, Alps.
- Local: *e.g.,* municipality.

-**Societal Benefit Area (SBA).**

The Societal Benefit Areas (SBAs) are nine environmental fields of interest around which the Global Earth Observation System of Systems (GEOSS) project is exerting its efforts. These include: Agriculture, Biodiversity, Climate, Disasters, Ecosystems, Energy, Health, Water, and Weather around which a preliminary hierarchical vocabulary has been created. Each area has some related fields (fifty-eight in total) of interest addressed as subcategories that specialise the goals of the entire project, the focus of the subcategories being mainly the relation between environmental issues and human activity and health.
One of the aims of GEOSS is to implement a proper system of earth monitoring and to render information deriving from this process available to a global range of users.
Sources:

-**Sensor**. This term refers to the device used to acquire data *i.e.,* to measure the radiation arriving to the satellite instrument **http://www.esa.int/esaMI/Education/SEMP67SJR4G_0.html**

- passive sensors. Sensors that use external energy sources to “observe” an object (*e.g.,* the sun light to observe the Earth) **http://www.esa.int/esaMI/Education/SEMP67SJR4G_0.html**
- active sensors. Sensors that rely on their own sources of radiation to “illuminate” objects so that the energy reflected and returned to the sensor may be measured. The most common active sensors used in remote sensing are radar and lidar. **http://www.esa.int/esaMI/Education/SEMJ97SJR4G_0.html**

-**User requirement**. This concept is linked to WP2 in GeoViQua. In this deliverable, the concept has been embedded in the “Role in GeoViQua” field of the pilot cases datasets description table. The relationship between specific datasets and user requirements, user stories, scientific objectives and user groups is recorded for each dataset, thus providing a first match between datasets/requirements and definition of testable requirements. This is the way of reflecting the association between WP7 and WP2 at this stage of the project. Likely, working with the selected datasets, will result in a dynamic process, and results will generate new inputs into WP2 and other work packages (*e.g.,* introduction of user requirements derived from experimental testing, that are difficult to envisage a priori).

-**User story.** In WP2, user stories have been defined as short natural language sentences of the intended functionality of a software system, written in the language of the user or stakeholder of the system. User stories are composed of three aspects:

- a written description of the story used for planning and as a reminder
- conversations about the story that server to flesh out the details of the story
- tests that convey and document details and that can be used to determine when a story is complete

It must be noted that, in this first deliverable, focused in WP7, the user story of the datasets is a field in the datasets description table, slightly modified as a description field where to provide the story and relevant highlights of the dataset, so these aspects can be taken into account in the experimental design in WP7.
Strictly speaking, in GeoViQua, this concept is fully described and properly implemented in WP2.

8. REFERENCES

Allen T.F.H., Hoekstra T.W. (1991) “Role of Heterogeneity in Scaling of Ecological Systems under Analysis.” In Ecological Heterogeneity, edited by J. Kolasa and S.T.A. Pickett, pp. 47–68. New York: Springer.

Cao C., Lam N. (1997) “Understanding the Scale and Resolution Effects in Remote Sensing and GIS.” In Scale in Remote Sensing and GIS, edited by D. A. Quattrochi and M. F. Goodchild. New York: CRC Press.

Carroll J.M. (1999) “Five reasons for scenario-based design.” Proceedings of the 32nd Hawaii International Conference on System Sciences.

Carroll J.M., Rosson, M.B. (1990) “Human-computer interaction scenarios as a design representation.” In Proceedings of the 23rd Annual Hawaii International Conference on Systems Sciences. (Kailua-Kona, HI, January 2-5, 1990). Los Alamitos, CA: IEEE Computer society Press,1990, pages 555-561.

Ceccato P., Flasse S., Tarantola S., Jacquemond S., Gregoire J.M. (2001) Detecting vegetation water content using reflectance in the optical domain. Remote Sensing of Environment 77: 22–33.

Chuvieco E. (2002) Teledetección ambiental. Ed. Ariel. 2002.

Clark W. C. (1987) “Scale Relations in the Interactions of Climate, Ecosystems, and Societies.” In Forecasting in the Social and Natural Sciences, edited by K. C. Land and S. H. Schneider. Boston: D. Reidel Publishing Company.

Korzukhin M.D., Ter-Mikaelian M.T., Wagner R.G. (1996) “Process versus empirical models: which approach for forest ecosystem management?”Canadian Journal of Forest Research 26(5):879-887.

Fernández N., Delibes M., Palomares F. (2007) “Habitat-related heterogeneity in breeding in a metapopulation of the Iberian lynx.” Ecography 30(3): 431–439.

Haines-Young R., Chopping M. (1996) “Quantifying landscape structure: a review of landscape indices and their application to forested landscapes.” Progress in Physical Geography 20(4): 418–445.

Iampietro P.J., Kvitek R.G., Morris E. (2005) “Recent advances in automated genus-specific marine habitat mapping enabled by high-resolution multibeam bathymetry.” Marine Technology Society Journal 39(3): 83–93.

ISO-TC211(2011) ISO/DIS 19157:2011: Draft International Standard: Geographic information -- Data quality. International Organization for Standardization (ISO). Geneva, Switzerland, 2011.

ISO-TC211(2006) ISO/TS 19138:2006: International Standard: Geographic information -- Data quality measures. International Organization for Standardization (ISO). Geneva, Switzerland, 2006.

King A. W. (1998) “Hierarchy Theory: A Guide to System Structure for Wildlife Biologists.” In Wildlife and Landscape Ecology, Effects of Pattern and Scale, edited by J. A. Bissonette, pp. 3–32. New York: Springer.

Koy K., Van Wart S., Galey B., O’Connor M., Kelly M. (2011) “Cal-Adapt: Bringing global climate change data to a local application.” Photogrammetric Engineering and Remote Sensing 77(6): 547-550.

Lilesand T. M., Kiefer R.W., Chipman J.W. (2003) Remote Sensing and image interpretation. New York.

McGarigal K, Cushman S.A., Neel M.C., Ene E. (2002) "FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps." Project homepage, University of Massachusetts, Amherst. URL (cited on 24 February 2009): http://www.umass.edu/landeco/research/fragstats/fragstats.html.

McQuail D. (2000) McQuail's Mass Communication Theory (4th edition), Sage, London, pp. 16-34.

Pereira G.M. (2002) "A Typology of Spatial and Temporal Scale Relations." Geographical Analysis 34 (1): 21-33.

Pons X., Solé L. (1994) "A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data". Remote Sensing of Environment 48: 191-204.

Schils T. (2006) "The tripartite biogeographical index: a new tool for quantifying spatio-temporal differences in distribution patterns." Journal of Biogeography 33(4): 560–572.

Turner M. G., O'Neill R.V., Gardner R.H., Milne B.T. (1989) "Effects of Changing Spatial Scale on the Analysis of Landscape Pattern." Landscape Ecology 3:153–62.

Vivoni E.R., Teles V., Ivanov V.Y., Bras R.L., Entekhabi D. (2005) "Embedding landscape processes into triangulated terrain models." International Journal of Geographical Information Science 19(4): 429–457.

Baulies X., Szejwach G. (1997) "Survey of needs, gaps and priorities on data for land-use/land-cover change research." Land-Use and Land-Cover Change (LUCC) data requirements workshop. Barcelona, Spain, LUCC Report Series No. 3 1997.

Weins J. A. (1989) "Spatial Scaling in Ecology." Functional Ecology 3: 385–97.