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 D2.1
User Requirements for GeoViQua

Version 1.0

Due date of deliverable: 31/02/2012
Actual submission date: 23/03/2012

<table>
<tr><th colspan="5">Document control page</th></tr>
<tr><td>Title</td><td colspan="4">D2.1 User Requirements for GeoViQua</td></tr>
<tr><td>Creator</td><td colspan="4">Simon Thum</td></tr>
<tr><td>Editor</td><td colspan="4">Simon Thum and Thorsten Reitz</td></tr>
<tr><td>Description</td><td colspan="4">This document summarizes the user requirements for GeoViQua components gathered in the first year of the project. It serves as a common reference for the project partners for driving the design and development of the software components as well as for upcoming discussions and decisions.</td></tr>
<tr><td>Publisher</td><td colspan="4">GeoViQua Consortium</td></tr>
<tr><td>Contributors</td><td colspan="4">GeoViQua Partners</td></tr>
<tr><td>Type</td><td colspan="4">Text</td></tr>
<tr><td>Format</td><td colspan="4">MS-Word</td></tr>
<tr><td>Language</td><td colspan="4">EN-GB</td></tr>
<tr><td>Creation date</td><td colspan="4">17/01/2012</td></tr>
<tr><td>Version number</td><td colspan="4">1.0</td></tr>
<tr><td>Version date</td><td colspan="4">20/03/2012</td></tr>
<tr><td>Last modified by</td><td colspan="4">Simon Thum</td></tr>
<tr><td>Rights</td><td colspan="4">Copyright © 2011, GeoViQua Consortium</td></tr>
<tr><td rowspan="5">Dissemination level</td><td></td><td colspan="3">CO (confidential, only for members of the consortium)</td></tr>
<tr><td>X</td><td colspan="3">PU (public)</td></tr>
<tr><td></td><td colspan="3">PP (restricted to other programme participants)</td></tr>
<tr><td></td><td colspan="3">RE (restricted to a group specified by the consortium)</td></tr>
<tr><td colspan="4">When restricted, access granted to:</td></tr>
<tr><td rowspan="4">Nature</td><td>X</td><td colspan="3">R (report)</td></tr>
<tr><td></td><td colspan="3">P (prototype)</td></tr>
<tr><td></td><td colspan="3">D (demonstrator)</td></tr>
<tr><td></td><td colspan="3">O (other)</td></tr>
<tr><td rowspan="4">Review status</td><td></td><td>Draft</td><td colspan="2">Where applicable:</td></tr>
<tr><td>X</td><td>WP leader accepted</td><td>X</td><td>Accepted by the PTB</td></tr>
<tr><td></td><td>PMB quality controlled</td><td>X</td><td>Accepted by the PTB as public document</td></tr>
<tr><td>X</td><td>Coordinator accepted</td><td colspan="2"></td></tr>
<tr><td rowspan="4">Action requested</td><td></td><td colspan="3">to be revised by all GeoViQua partners</td></tr>
<tr><td></td><td colspan="3">for approval of the WP leader</td></tr>
<tr><td></td><td colspan="3">for approval of the PMB</td></tr>
<tr><td></td><td colspan="3">for approval of the Project Coordinator</td></tr>
</table>

		for approval of the PTB
Requested deadline	31/01/2012	

Revision history			
Version	Date	Modified by	Comments
0.1	17/01/2012	FRAUN_ST	Initial Draft
0.2	03/02/2012	FRAUN_ST	Included AST_VL(3.1.2; 2.2.1), CNR_LB input(3.2.5)
0.3	10/02/2012	FRAUN_ST	Included 52N_DN (3.2.1), UREAD_JB(3.2.7) input
0.4	22/02/2012	FRAUN_ST	Included S&T_JS (3.2.2), AST_DC(3.2.4) input
0.5	23/02/2012	FRAUN_ST	Included FRAUN_EK review comments
0.6	26/02/2012	FRAUN_ST	Included CREAF_PD(3.2.6; 3.2.3) input
0.7	29/02/2012	FRAUN_ST	Updated 52N_DN contribution
0.8	02/03/2012	FRAUN_ST	Included CREAF () input, AST_VL diagram/review
0.9 PRE_FINAL	02/03/2012	FRAUN_ST	Finishing
1.0 FINAL	23/03/2012	FRAUN_ST	Review comment from FRAUN_EK, CREAF_PD

Institution	Contributors
FRAUN	ST Simon Thum
FRAUN	EK Eva Klien
AST	DC Dan Cornford
AST	VL Victoria Lush
S&T	JS Joost Smeets
CNR	LB Lorenzo Bigagli
52N	DN Daniel Nüst
UREAD	JB Jon Blower
CREAF	PD Paula Díaz, JM Joan Masó, IS Ivette Serral

Table of Contents

Abstract

With increasingly diverse uses and users for geographical information, together with growing accessibility to EO data and applications, the provision of meaningful data quality information with geographic data products is ever more important. To enable effective communication on data quality, it is equally important to thoroughly analyse the users, purposes and perspectives and to derive a set of requirements. In the GeoViQua project (7th Framework Programme; ENV.2010.4.1.2-2), two deliverables deal with the user and technical requirements that have been gathered, structured and documented during the first year of the project. Deliverable 2.1 documents the user requirements, while Deliverable 2.2 documents the technical requirements. The requirements have been collected by different means, in particular user interviews and state of the art analysis, and subsequently integrated, reviewed and validated collaboratively. They represent the essential initial step to the development of the systems in GeoViQua.

This deliverable focuses on the user requirements, i.e. requirements described from an end-user perspective, as opposed to a technical or implementers' perspective which is subject to deliverable D2.2. This document consists of two main parts: first the methodology used in acquisition, integration, review and validation of user requirements is described, followed by a documentation of the specific user requirements collected for GeoViQua.

This document provides a common ground for driving the design and the development of the software components and serves as reference for upcoming discussions and decisions in the GeoViQua project. To enhance its usefulness in this crucial area, the requirements in this document are also presented according to project-specific criteria which are not relevant to a user.

Acronyms and Abbreviations

AIP	Architecture Implementation Pilot, an OGC activity
DOW	Description of Work
ICT	Information and Communication Technology
GCI	GEOSS Common Infrastructure
GEOSS	Global Earth Observations System of Systems
OGC	Open Geospatial Consortium
RCP	Rich Client Platform
UncertML	Uncertainty Mark-up Language
WP2	Work package 2 *(or others)*

Purpose of this document

This document summarizes the user requirements for GeoViQua components gathered in the first year of the project. It serves as a common reference for the project partners for driving the design and development of the software components as well as for upcoming discussions and decisions.

1. Introduction

Despite advances in GIS field, development of better tools for capturing and analysing geospatial data, and numerous standardisation efforts to impose standards on geospatial data quality, to date, it is not fully apparent what defines quality in geospatial data to the actual data users. Quality is a crucial element in effective selection and use of spatial data, yet quality paradigms are subjective and can only be assessed by the user when performing a certain task. For these reasons, the main objective of the GeoViQua project is to define the notion of geospatial data quality from a user perspective and to provide innovative tools to support quality-enabled data search and selection based on these user defined quality elements. The aim of the research done thus far, therefore, was to elicit high level user requirements related to geospatial data quality, including data selection processes, data visualisation, data application and use, potential quality indicators, and users' initial views on the establishment of a 'seal-of-approval' for geospatial data. In order to acquire the data, information has been gathered and analysed via a series of initial interviews with actual geospatial data users and geospatial experts. From these, we established user stories and elicited high level user requirements. This document presents the practical work carried out to elicit high level user requirements that relate to geospatial data quality and quality-enabled data selection. It gives details on the methodology used to collect the data and the analysis carried out to elicit user requirements. It also describes the research findings – high level user requirements that were discovered through a series of interviews with geospatial data users. The method of data gathering and analysis is described in detail in Section 2.

The task of collecting specific user requirements is part of the work carried out in work package 2 "Requirements for data quality visualisation" (WP2). The main goal of WP2 is to provide a set of consistent high-level functional and technical requirements. To reach this goal, the collected requirements were streamlined and refined to meet certain criteria, making them useable as an analysis model which in turn was used to define the basic goals, responsibilities and collaborations of each component to be developed within GeoViQua. This baseline will evolve with future project results, such as prototype evaluations or as a consequence of the ongoing requirements management.

Additionally, the intention of the early-project activities in gathering requirements was to help refine and delineate the scope of the individual contributions and the project at large, and to clarify the junction points in between.

The process behind WP2 is based on object-oriented analysis and design, as well as established requirements analysis practices as described in books such as User-centered Requirements Engineering (Sutcliffe, 2002) and User Stories applies (Cohn, 2004). Requirements collection and modelling endeavours always need to be adapted to match the project's demands. For example, if only a minor part of the project's requirements can be adequately described in a behavioural manner, non-behavioural descriptions need to be focused. Thus at the beginning of the project, we identified our demands for requirements management (sometimes called meta-requirements) and determined appropriate adaptations to our requirements process.

In the following, the basic idea of requirements and core principles for requirements management in GeoViQua are described. We then continue to explain the methodology behind our work in Section 2 and the results in section 3.

1.1 What are requirements?

It is said that requirements cannot be "elicited" in the sense that they are there, and just need to be collected and brought into shape. According to Cohn (2004) a more suitable metaphor is that of trawling the field for requirements. The mental image of a net that will get the big ones and leave smaller fish in the water applies quite well to requirements elicitation. No-one will be able to come up with everything a system might do for him if the system does not yet exist. Most certainly, he will be able to come up with good ideas nonetheless.

This matches well with the intent of requirement engineering in the first phase of GeoViQua. But requirement do not just come from users, although that is the scope of this document. Other requirements may reflect system boundary conditions that cannot be changed or scientific goals the project intends to tackle even if users did not (yet) anticipate the usefulness of them. As a result, when this document talks about requirements without further qualification, the full gamut of a high-level user requirement that is not much more than an idea to a system implementation constraint may be meant.

Drawing from the fish net analogy a bit more, we will likely be able to elicit more user requirements when prototypes are evaluated, because ideas and expectations will become more concrete.

1.2 Guiding principles

To ensure the successful application of requirements management procedures, several key principles should be adhered to. For reference purposes, this section gives a short overview.

Traceability is the ability to trace back any requirement, even a very technical one, to its origin(s). This intends to retain the capability to reassess a requirement in light of unexpected findings or circumstances.

Unambiguousness of a requirement is given if the requirement is commonly understood in the same way by people knowledgeable in the subject(s). Proof-reading and discussion are means to ensure unambiguousness.

A requirement is **classifiable** if a pre-determined classification can be applied. The most common is functional vs. non-functional requirements, but this dichotomy is usually just the top level (also see 2.1.4).

Completeness of a requirement is given if it contains all necessary information (but no more). This obviously depends on the project context and at what level a requirement is, but should include the prerequisites and constraints under which the requirement applies. It should *not* be understood as requiring excessive detail.

A requirement should be **relevant and current**, sometimes referred to as “necessary”. This bears a bit of a tautology, but is often overlooked (in particular when requirements change) and thus allowed to bind resources which could be deployed more efficiently.

A requirement should be **implementable**. This means it should not conflict with other requirements and be technologically viable.

Verifiability is given when a means to test if a given system fulfils a requirement is described or easily deductible. A success condition should be clearly stated.

An **understandable** requirement is one which is understandable by all relevant actors. Often, excessive detail or use of jargon can prevent sufficient understandability.

A good requirement is minimal but complete according to the above (often referred to as a *canonical* requirement).

Together, these principles help forming a requirements-based analysis model. The list is not exhaustive, i.e. one could add some more desirable properties mainly depending on project characteristics, or weight them differently. In particular when aiming for a readily implementable solution, one might give more focus on completeness and verifiability, thereby implicitly discounting on understandability.

2. Method

In this section, we describe the methodology we have used to elicit and analyse high level user requirements in earth observation (EO) data quality and quality visualisation. We focus on the overall process of requirements engineering that includes data gathering (i.e. user interviews), data analysis, user stories elicitation, and requirements elicitation. Further, we give an overview on the resources and tools used to record and analyse the derived high level user requirements. It is not the intent to describe every aspect of the process in detail, but to give an overview and some notable details, which will help understanding the way how the user requirements are described, categorised and managed in this specific project.

The overall process for requirements management in GeoViQua is summarized in figure 1. The initial specification phase is the largest and most intense part of WP2. The results (in form of requirements, use cases and architectural boundary conditions) are fed into the project's development and communications processes of the remaining work packages. The work on requirement management in WP2 continues in a facilitating role for the other work packages.

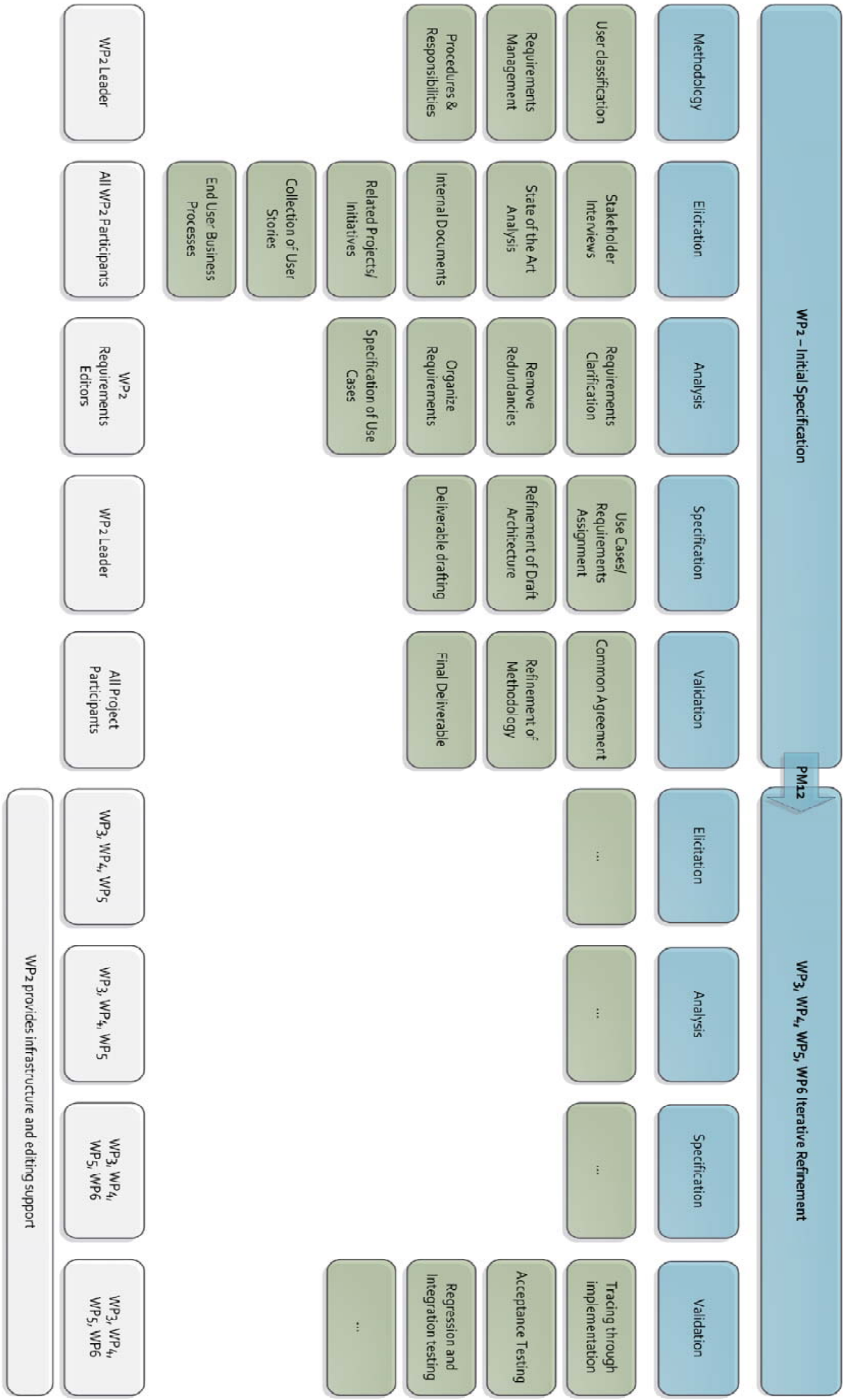

Figure 1: The requirements specification and management process for GeoViQua

Establishing the **methodology** for requirement specification in GeoViQua included the following activities:

- Clarification of the way requirements engineering will be applied in GeoViQua
- Definition of the key procedures and responsibilities (see Section 2.1.1)
- Selection of a requirements management support tool (see Section 2.1.2)
- Determination of user classification (see Section 2.1.3)
- Creation of an interview guideline (see Annex B)

The methodical baseline agreed upon so far is intended to be reviewed after we went through WP2's initial specification phase in order to prepare for the following work packages. Requirements management is to be applied to the individual work packages in an adapted manner as the nature of the actual work shifts from specification to design and implementation.

Executing the requirement **elicitation phase** included:

- structured stakeholder interviews, which are the main source of our user requirements
- state-of-the-art analysis
- documents from related projects or initiatives

During the elicitation phase, we acquired requirements from the various sources without much regard for the project's anticipated boundaries. This eased elicitation and allowed us to define the project's scope in terms of requirements. The most important sources are the structured stakeholder interviews, which are the main source of our user requirements. Other sources include a state-of-the-art analysis and documents from related projects or initiatives.

Some of the sources we envisioned to play a role did not prove to be suitable sources. Most prominently, the GEOSS User requirements registry (GEOURR, 2011), at the time of this writing, contains mostly sketchy draft requirements which provide little discernible value for someone who outlines or plans an actual implementation.

The actual sources of most requirements are thus the following:

- user interviews
- related projects and initiatives
    - QA4EO
- state-of-the-art analysis
    - GEOSS Recommendations and work plan
- internal documents
    - GeoViQua Description of Work

In the **analysis and specification phases**, the requirements were reconciled to avoid duplication, ensure quality according to the principles laid out in 1.2 and enhance the requirements usefulness for guiding or deciding implementation questions. This work has been carried out jointly by the partners involved in WP2.

In the **validation phase**, the requirements were aligned to prominent use cases - in particular those from our pilot cases – to validate their applicability and orthogonality.

This cannot be done sensibly without creating orthogonal requirements which mix up input from various sources in a more organized way; the downside is that, while user input played a vital role in the process it is hard to assess this role systematically.

However, the resulting requirements have been fit for the purpose of discussion and deciding about vital project aspects. This was a precondition for closing the common agreement as foreseen by the overarching requirements process.

2.1 Process details

This section intends to give the information required to properly read and understand GeoViQua's user requirements and how they were elicited. It goes into some detail to explain the concepts of our requirements engineering process and the rationale for the way they are applied.

2.1.1 Roles and Responsibilities

We defined a few roles and associated responsibilities to ease the distribution of work in the process, mainly in the analysis, specification and validation phases.

A **requirements editor** was named by each of the WP2 partner. This role is defined as follows: "The requirement editor will carry out review and maintenance of requirements and related items within the GeoViQua requirements management infrastructure (JIRA). This includes classification, clarification, establishing relations, groups and general editing. It should not include higher-level decisions, except if authorized by the Requirement Editors Board or another formal body."

The mentioned **Requirements Editors Board (REB)** was established in a motion to the project management board. Its aim is to provide a formal body which is able to vote on requirements-related issues in cases consensus cannot be reached. Fortunately, so far there was no need to call the REB.

A requirements editor can take over the role as the **author** or **primary reviewer** of a requirement, which is discussed in 2.1.6.

A requirements editor can also take over the role as a **cluster adopter**, which obliges him to act as a steward for a specific "cluster", a thematic grouping discussed in 2.1.8. This role was added during the GEOVIQUA PTB meeting (July 2011) in recognition of the importance of thematically focused maintenance of requirements.

2.1.2 Tool support

The methodology and elicitation phases were coordinated over the wiki and mailing lists provided by the project's coordinator and regular teleconferences.

In the analysis, specification and validation phases a ticketing system, Atlassian JIRA ("Jira"), has been used to model and review the requirements, to serve as a common reference and generally enable proper requirements management procedures. The Jira instance is provided by the WP2 lead, Fraunhofer, on behalf of the project and will be available during the whole project lifetime.

Requirements editors are allowed to modify requirements in the system, all other project members are only allowed to read them. Jira tracks modifications and enables to communicate about individual requirements.

Jira was chosen for its generic configurability, feature completeness and because similar tools were missing required functionality. As a general ticketing system, Jira needed extensive tailoring to fulfil our demands, which was expected. In some minor aspects it was not possible to adequately alter its behaviour however. In practice we could work with the system and mitigate its weaknesses, which were outweighed by the availability of a shared repository.

2.1.3 User classification

Requirements can, in general, be understood more easily if they pertain to a known user group. We pre-defined some broad user categories and decided to add more of the user's context in the individual requirements. We decided against more complex classifications available, because this additional complexity was rather hindering than helpful in our case. Our users are typically very specialized professionals, so that any conceivable taxonomy would likely lose a lot of information.

ESA and UREAD have contributed significantly to this discussion. We agreed on a **user group description,** which has the following coarse categories:

- **Scientists** understand GeoViQua's research objectives, are familiar with EO data, EO quality indicators and parameters, and the meaning of the GEO Label for the entire data processing chain/generation and the like.
- **System developers** understand the issues derived from lack of systems interoperability. They know EO data formats, typical data volume, standards, the GEOSS common infrastructure's architecture and the GEO portal
- **Generic users** are not necessarily aware about EO data processing/generation chain. For example, an authority representative or policy maker on disaster management, climate change. He may not understand data quality related indicators, but is concerned about data reliability.
- **Administrators** are actually running the systems involved, and ensure their proper operation.

More details about users can be added as informal free text to each individual requirement. The description field is used for this purpose.

2.1.4 Requirements classification

GeoViQua has opted for a strong involvement of users in the requirements gathering phase. However, as a scientific project, GeoViQua also has scientific goals which provide the framework in which the user's input is in fact considered. Also, many of the interviewees have a scientific background.

To deal with this fact, our requirements classification was enhanced with the notion of "scientific objectives" which complements the traditional (non-)functional dichotomy at the top level of the requirements classification.

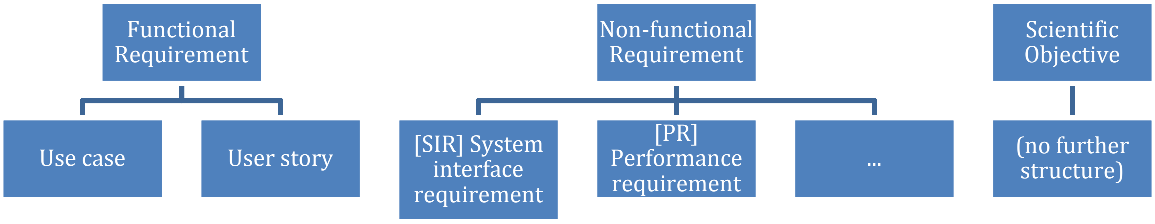

Figure 1 GeoViQua high level requirements structure

A **functional requirement** specifies observable behaviour of a system in terms of functionality. Besides a formal description which covers certain aspects such as rationale or acceptance criteria, it is often easier to bridge from users to development with **user stories**. A user story is a structured text statement about using a system from the perspective and in the language of a specific user. User stories are typically easy to derive from user input, and when applying some care they become suitable as input for a systems engineer or developer. In a similar vein, **use cases** are an again more formal but still behavioural description of observable system behaviour.

A **non-functional requirement** describes a property of a system that is not adequately described in terms of system behaviour. A better way to think of non-functional requirements is as a constraint on the system to be built. An example might be a security requirement. Specifying a user story that tells how the system conforms to a security specification is possible, but it would contain a lot of dead weight. Thus, non-functional requirements are used to describe crosscutting concerns. Commonly (and in GeoViQua), the following classes of non-functional requirements are defined:

- SIR System interface requirement
- RR Reliability requirement
- PR Performance requirement
- SR Security requirement
- STR Standard (conformance) requirement
- HMR Human-Machine interface requirement
- DR Documentation Requirement

A **scientific objective** describes some issue that affects the system(s) to be built, but not in an easily predictable manner. This property is the main reason why it somewhat falls out of the requirements perspective. We nonetheless decided that we want to allow for user requirements to influence our scientific objectives or how we approach them, so they have to be managed in the same process.

2.1.5 Traceability

After the requirements management tool (Atlassian JIRA (Jira) provided by Fraunhofer IGD, see 2.1.2) was chosen, we had to determine how to ensure traceability. Jira naturally provides versioning, but this alone is not sufficient to ensure traceability. We therefore used "links", i.e. qualified associations between selected requirements as a means to comply with the traceability principle.

We employed the following “link types”, i.e. available qualifications of relations between requirements:

- **relates to** is a generic, undirected relation.
- **is derived from / has been derived into** is used to signal that something was significantly influenced by something else.
- **references / is referenced by** is a generic reference, but implies directionality of the relation (as opposed to **“relates to”**).
- **depends on / is a dependency of** signals that a requirement cannot be implemented or even designed without another one being fulfilled or resolved.
- **blocks / is blocked by** signals that two issues cannot sensibly be implemented within one system due to their nature
- **conflicts with** is used to signal significant conflict potential or inappropriate overlap. As opposed to "Blocks", it does not imply that an ultimate decision about the underlying problem is required.
- **duplicates / is duplicated by** signals that significant overlap exists between two issues. It is a special type of conflict.

Jira is not a requirement management tool by nature (it is a software ticketing system), so there is no semantic tool support e.g. to ensure that a requirement cannot block and depend on another requirement at the same time or avoid cyclical dependencies. It is possible, therefore, that some contradictions in the link structure have been overseen.

Along those lines, we did not really ensure traceability but promoted to maintain it. This has worked fairly well using the *“is derived from/has been derived into”* links. The other link types were used during analysis and specification to signal the hidden relations we uncovered in the process, so they can be taken into account for analysis.

2.1.6 Review of requirements

Similar to e.g. the GEOSS user requirements registry (GEOURR, 2011), we have established a formal intra-project requirements review process. Each requirement bears a state from the gamut “Abandoned” to “Accepted”, beginning in “Draft”. A state diagram (as portrayed in Jira) can be seen in figure 2. The states at the bottom of the figure, “Accepted”, “Rejected” and “Abandoned”, are considered *end states*.

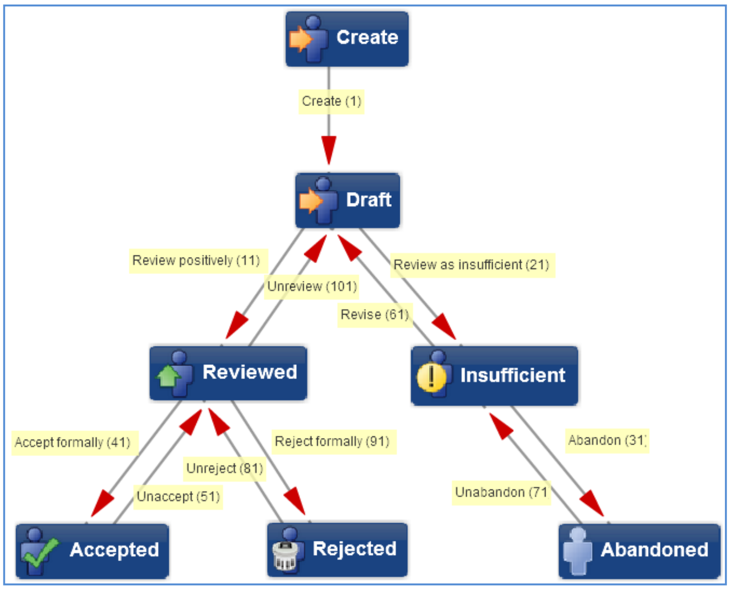

Figure 2: Possible review states and transitions

The process was driven by two disjoint roles, the author and the primary reviewer. In principle, every requirements editor could assume each role at the level of an individual requirement, but not both on the same requirement (thus disjoint). The author, naturally, is the one who initially creates the requirement in JIRA. The primary reviewer is chosen by the author, but can be changed afterwards. His responsibility is to declare a draft requirement as reviewed when the formal criteria are fulfilled.

The author's responsibility is to get his requirements into one of the end states, chiefly into "Accepted". Only the primary reviewer can declare a requirement as "Reviewed", to prevent the possibility of a "review" without even a second pair of eyes. Changes are on record to ensure accountability and to provide access to old versions.

Reviewed requirements can then be accepted; declaring that consensus about their applicability to GeoViQua has been reached. Rejected requirements have been determined to be out of scope for GeoViQua, although they are formally valid requirements.

Requirements can also be abandoned for various reasons, they might be duplicates or no longer considered worthwhile. When the reason for abandoning or rejecting a requirement is overlap with other requirements (e.g. duplication), then the requirement should be linked to the requirement(s) which shall replace the one in question. This is important to guarantee traceability because a rejected or abandoned requirement still remains in the system.

2.1.7 Functional Components

A functional component is a part of an ICT system and is specified during the analysis phase. It is used to structure the analysis model. Its definition makes no assertions about whether it represents exactly one logical or physical (deployed) component, or of what type (service, RCP, Application ...) this component is.

GeoViQua's stated goal is to implement new approaches to quality visualization into GEOSS. Thus, there is a lot of pretext regarding the components in the field which cannot simply be ignored. When we needed to identify our functional components, we used the concepts present in GEOSS and used them to identify our functional components.

As a result, our functional components are not pure "functional components" as the concept outlined above, but they already have varying degrees of implementation constraints. The following figure 3 shows these functional components as they are currently established.

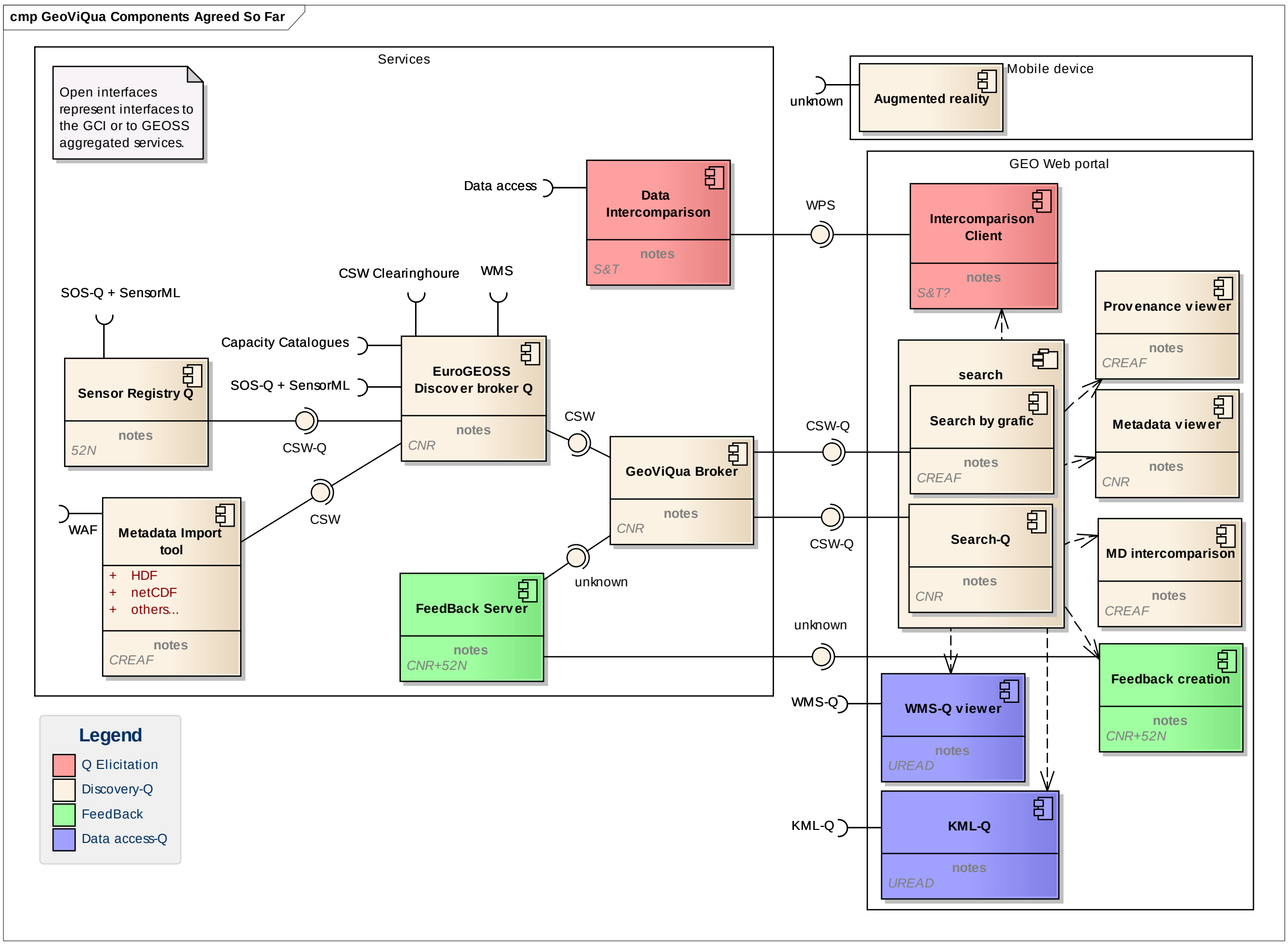

Figure 3: Functional Components in GeoViQua

It should be noted that, for the purpose of this document, functional components are largely irrelevant because users typically don't care very much about components. However they represent a common structural element in requirements engineering which we will use further on. They are explained here for completeness and will be subject to the following project deliverable D2.2 – Technical requirements for GeoViQua.

2.1.8 Clustering

In preparation of the requirements for analysis, requirements clusters have been identified to get an overview of the topics covered by the requirements. A cluster aims to represent the canonical requirements for GeoViQua pertaining to a certain subject. By canonical, it is meant that while there are more requirements pertaining to a given subject, the ones in the cluster broadly cover the topics of the subject in enough detail to delineate the scope of the requirements collected.

Initially, clusters were identified by the WP2 lead as follows:

1. Create a unordered list of requirements
2. Sort them so similarly-themed requirements are next to each other
3. Identify the emerging thematic clusters
4. Take out requirements that have thematically similar counterparts still in the cluster

Some of our requirements turned out to be borderline cases, so they ended up in multiple clusters. Further analysis showed these were thematically broad requirements, and we could derive more focused requirements where thematic classification was clear. This helped to identify and clarify the junction points between topics.

The requirements clusters have proven to be a very good unit of organization in preparation of more formal criteria such as functional components. They enabled thematically focused discussion of our requirements, which was also fostered by naming responsible persons, the *cluster adopters*, to curate them as long as is deemed necessary.

The following clusters have been identified:

Name	*Adopter*
Intercomparison	Joost Smeets (S&T)
Metadata	*(discontinued)*
Visualisation	Jon Blower (UREAD)
Feedback	Daniel Nüst (52N)
Quality-aware Search	Lorenzo Bigagli (CNR)
Standards and Communication	Joan Masó (CREAF)
Quality and Uncertainty	Dan Cornford (ASTON)
Provenance and Purpose	Joan Masó (CREAF)

The metadata cluster turned out to contain mostly requirements about the metadata users would like to be provided. We decided this is better handled by incorporating these requirements in the quality model which is an artefact of the project, discontinuing the cluster as a separate entity.

2.2 Process implementation

The following subsection provides details on the data gathering process that was carried out to elicit high level user requirements. It explains how the initial data was collected using semi-structured interviews with geospatial data users and experts. The interviews' structure is presented providing the questions addressed during interviews. The section also addresses the process of user and expert group selection.

The upcoming sections will discuss issues in the analysis and, to some extent, the specification phases. Focus will be given to applications aspects, not a fully detailed description of the process or its implementation.

2.2.1 Capturing user requirements in Interviews

During the spring 2011 the GeoViQua project carried out a number of interviews with geospatial data users and experts to elicit some user stories from potential GeoViQua users. The idea was not to go down to the level of individual requirements, but to gain initial information about different application areas and how data is found, selected and used. The interviews were intended to be relatively informal, and guided by a set of general questions. The questions were not intended to be shown to the interviewees, but were used to provide guidance to the interviewer. The set of questions used in the interviews are presented below; the full guide is in Annex B.

a. Please describe a current area of your work in which you use external data sources.
b. What data do you use in your work, and where do they come from?
c. How do you choose which datasets to use in your work? What are the reasons for your decisions?
d. Are you aware of any data certificates or seals in selecting your data? Do you look for specific certificates or meta-information in a data set you use? How do you know whether to trust the data?
e. Do the data you use come with sufficient supporting information to allow you to make an informed judgement about which one(s) to choose? How much information do you need?

Follow-up and clarification questions were asked depending on specific interview circumstances.

All GeoViQua project partners were asked to identify and contact relevant user and expert groups to participate in the initial interviews. The user and expert groups included geospatial data users, researchers, data archivists, academics and data producers. Such range of expert groups selected allowed us to elicit an interesting variety of user stories and develop a wide-ranging picture of user needs. The interviewers contacted the potential interviewees by e-mail to arrange the telephone or face-to-face interviews. A total of 16 participants were recruited for interviewing with interviews being half an hour to an hour long. The interviews were introduced by explaining that the GeoViQua is a European project, whose aim is to help build a global system for discovering and accessing earth science data. The goal in the project is to ensure that the infrastructure gives users the information they need to discover dataset and choose between them. After a short introduction, the interviewers were required to check if the interviewees is happy for his/her name to be recorded, or whether they would prefer to remain anonymous. During the conversations the interviewers were expected to capture and note down the answers to the questions listed above. If the interviewees used community-specific jargon, interviewers were required to ask the interviewees to explain their exact meaning.

After each interview the interviewers were required to complete a form (see Annex B) that captured the essential details from the interview. The output from the interviews was intended to capture enough context to allow more specific requirements to be distilled out of them if appropriate. The completed interview forms were used to derive user stories – very high-level informal statements of the requirements that capture what the users want to achieve. When creating the user stories the interviewers followed a generally accepted template: "As a <role>, I want <goal/desire> so that <benefit>". To formally record the elicited user stories, it has been decided to use Jira (see section 2.1.2) where the user stories were recorded as “issues. From each user story a number of high level user requirements were derived and also recorded in Jira, at which point they became available to other project members. Due to a large number of reporters, it has been decided to allow requirements duplication during the requirements gathering phase and carry out a requirements "clean up" when all the data is recorded and finalised.

2.2.2 Difficulties and remedies

This section describes a few instances where we encountered problems in executing the process as intended. The remedies we applied were not without drawbacks, so for completeness we are describing them here.

2.2.2.1 Traceability in a shared resource

As is mentioned under 2.1.2, Jira was not designed with requirements management in mind. After the elicitation phase, we discovered that users will talk about requirements or ideas at the level they are comfortable with, which differs among individuals. That is why there are often user requirements that look like being a consequence of other requirements. That matches the "derived from" link qualification described under 2.1.5, which originally was not intended for that use case.

So we took the pragmatic decision to favour structural integrity of our requirements over complete and untainted history. A deciding factor was the fact we are building an analysis model where the actual structure counts, not the exact way of arriving at it.

2.2.2.2 Organization of requirements

The whole notion of clusters owes itself to the fact that we did not have appropriate organizational structures in place after it emerged that assignment of functional components would be harder than anticipated. Therefore, the clustering technique was employed to help us organize the requirements without needing to make premature decisions.

In hindsight, this is one of the consequences of our functional components sometimes being close to actual components. Assigning a requirement looked like assigning work duties, which most people did not want to bear with even if they knew what had to be done.

The clusters allowed us to avoid that connotation, and it was so useful that we decided to name adopters for them, which seems to have been a very successful strategy.

(a) Clusters vs. Functional Components

Functional components (see 2.1.7) in Jira are realized as an association of requirements to functional components. Functional components act as a precursor of actual deployable components a system may contain.

Because GeoViQua's functional component are in part predetermined by GEOSS, they could not act as "pure" conceptual units. This left us with a larger conceptual gap to bridge when structuring our requirements. For example, many requirements were too broad to assign them to a single functional component. Thus, we determined that a purely topical organization is an essential organizational tool.

Obviously, topical and conceptual or functional units can be expected to be in a certain correlation, which sometimes is quite close. For example, the Quality-aware Search cluster has a major share of requirements for the GeoViQua broker.

Thus, it is expected that the clusters will help with properly assigning requirements to functional components, but when this assignment has settled the cluster's role will be less pronounced.

2.2.3 Reviews

At a point before the first Project Technical Board meeting it emerged that there was certain reluctance to reviewing requirements. It was not unwillingness to do the review work, much of that worked pretty well without interference. From comments and questions that came up we identified the cause to be in the process, which foresaw the primary reviewer to act "when a requirement has seen enough review". That point in time is poorly defined, and since all actions are on record no-one wanted to take blame. Also, since no-one at the time had declared a requirement as reviewed, the impression always was that it is too early.

We then introduced the ability to name requirements for being considered reviewed at a common point in time. Any disagreement could be aired easily, so clearly premature reviews would not slip through.

In hindsight, the obligation to act for declaring as reviewed was ill-fated and failed for predictable reasons, as it had a fuzzy definition and an uncertain outcome for the reviewer.

2.3 User vs. technical requirements

The structure of the project's deliverables makes a distinction between high-level user and technical requirements. A natural question to ask, then, is "Which user requirements have had an impact on the project?"

While there are some criteria for distinguishing the concepts, doing so it is surprisingly hard. We have interviewed a lot of users, but some of what they say can indeed be found in GeoViQua's Description of Work. In other cases, they are asking for highly specific metadata to be provided, which is mostly not GeoViQua's business.

Fortunately, a lot of their input was being accepted without intermediates (i.e. derived requirements or similar ones from other sources), so the question of the role of user input can be answered sufficiently well without considering influence through derivation and refinement.

This answer, however, is not a definite one that can be applied to later outcomes of the project. For them, a separate assessment will be required. Part of the reason is that at this stage, where lower level constraints and requirements are just beginning to emerge, a lot of our information is not better than any well-informed users', especially since they often are scientists themselves.

3. Results

This section describes the results we achieved so far. For brevity, we have opted not to include the whole requirements base which, without editing, spans 300 pages.

References to individual requirements take the form “GEOVIQUA-nnn Title”. Annex D describes how to access these resources in electronic form.

3.1 Overview of user requirements

This section intends to highlight user input from different angles such as collaboration intensity or user influence. To begin with, a few numbers are presented to continue to more in-depth topical discussions about the requirements we collected.

The following table 1 shows accepted requirements from interviews. As can be seen, they were expressed either as functional requirements or user stories.

Requirement Type	*Cluster*						
	Feedback	*Intercom-parison*	*Provenance and Purpose*	*Quality and Uncertainty*	*Quality-aware Search*	*Standards and Commu-nication*	*Sum*
Functional Require-ment	3	1	8	9	2	4	18
User Story	2	0	3	3	1	0	8
Total:							26

Table 1 Accepted requirements from interviews

When including requirements not directly from interviews, the base and diversity of requirements grows, as shown in the following Table 2 Accepted requirements by cluster and type.

Requirement Type	*Cluster*							*Sum*
	Feedback	*Intercom-parison*	*Provenance and Purpose*	*Quality and Uncertainty*	*Quality-aware Search*	*Standards and Communication*	*Visualisation*	
Functional Requirement	6	2	9	22	8	10	2	42
Non-functional Requirement	1	0	0	1	1	3	0	6
Scientific Objective	0	0	0	1	1	1	2	4
User Story	2	0	4	3	2	0	8	16
Total:	9	2	13	27	12	14	12	68

Table 2 Accepted requirements by cluster and type

This base of requirements is explored in more detail in the following sections. It should be noted that this represents a snapshot; requirements engineering is to continue over the project lifetime. At the time of this writing, there are 48 requirements in “transit states” (i.e. not in one of the defined end states) which may become accepted in the future. Moreover, completely new requirements might arise in the development process.

3.1.1 By collaboration intensity

This section is intended to focus on user requirements which require a lot of collaboration from our side or from outside groups and projects, *e.g.* because they require components from different partners to implement a difficult protocol.

3.1.1.1 By collaboration intensity within the GeoViQua project

Requirements should be testable; one of the best ways to do so is to verify them in the project pilot cases, as described in the deliverable 7.1 “Document pilot case studies in a standard format”. For that reason, ideally most of the requirements are related to the task 7.2 “Quality product parameterisation” and 7.3 “Validate quality aware visualisation and search services” which will be developed, treated and managed by a task in the project that will test this requirement in a pilot case.

Requirements that deal with the data model definition will impact several tasks in the project such as Task 3.1 “Metadata extraction quality component”; Task 3.5 “User feedback and quality assessment of data sets in GEOSS”; Task 4.1 “Quality-aware catalogue service”; Task 4.2 “Quality-aware catalogue client”; Task 5.1 “Integration of quality information with OGC visualisation services”; Task 5.3 “Implementing quality-aware visualisation components”.

Currently the list of requirements that is related with the quality model is composed by 24 requirements, as shown in the table below:

Req. Code	Requirements description
GEOVIQUA-166	The data quality model and its encoding(s) need a link mechanism to resources and parts of resources.
GEOVIQUA-165	The quality model shall define a hierarchical multilevel approach to metadata, quality indicators and traceability information
GEOVIQUA-162	The quality model and the system shall support recording, editing, and querying/searching some important metadata elements (complete list in the description)
GEOVIQUA-158	The quality model and the system shall support recording, editing, reviewing and subsequent querying of who has used the dataset
GEOVIQUA-150	Data model shall include information about redundant datasets
GEOVIQUA-148	The quality model and quality-aware search shall support sensors with quality description
GEOVIQUA-140	The quality model shall use ISO standards for metadata (19115) and quality (19157) with extensions for UncertML and O&M
GEOVIQUA-137	The system shall provide quantitative data quality estimation by comparison with reference data (e.g. Bayesian)
GEOVIQUA-136	The quality model shall support qualitative statements of belief about the data quality
GEOVIQUA-135	The quality model shall support statistical quantitative estimates of data quality (probabilistic, variance, etc.)
GEOVIQUA-125	The quality model shall document calibration and validation campaigns their statistically derived quality indicators
GEOVIQUA-120	The quality model and the system should support the recording, editing, reviewing, and querying of information about the purpose for which a dataset was originally collected.
GEOVIQUA-119	The quality model and the system shall support the recording, editing, reviewing, querying and peer review of the methodology adopted for dataset data collection.
GEOVIQUA-113	The quality model shall support links to journal articles in which dataset quality checks are reported.
GEOVIQUA-107	The quality model and the system shall be able to record, edit, review and query dataset limitations and boundaries of use provided by the creator of the dataset.
GEOVIQUA-105	The quality model and the system shall support recording, editing, reviewing and subsequent querying of dataset provenance.
GEOVIQUA-100	The system should support the recording, editing, review, and querying of a quality mask for digital elevation models.
GEOVIQUA-96	The quality model shall support point-specific quality information.
GEOVIQUA-91	The system should support the recording and subsequent review of methodology information associated with a dataset.
GEOVIQUA-89	The quality model shall support recording and searching of 3 types of quality/usability statements: syntactic; schematic; and semantic.
GEOVIQUA-61	Users want to know the build-up history of the dataset
GEOVIQUA-55	Users want to know whether the dataset has gone through certain 'validation campaign'

Req. Code	Requirements description
GEOVIQUA-23	The quality model shall support traceability for the methodology used to determine data quality indicators (QA4EO Principle)
GEOVIQUA-22	The quality model shall support quality indicators (QA4EO Principle)

One of the important findings in the requirements process is the importance the users give to the consumer's metadata and user feedback. There is no component in the GCI that currently addresses this issue. Requirements that are related with user feedback and costumer metadata will require effort from different tasks. That is because there is no previous component on the GCI that deals with user feedback, so that this part has to be developed from scratch. A data model has to be created in Task 3.5 "User feedback and quality assessment of data sets in GEOSS"; the catalogue has to be modified in Task 4.1 "Quality-aware catalogue service"; the client must be adapted in Task 4.2 "Quality-aware catalogue client"; the search results have to be integrated in visualization services in Task 5.1 "Integration of quality information with OGC visualisation services" and the implementation of the visualization components can be affected in "Task 5.3 Implementing quality-aware visualisation components". The following list reflects the requirements that affect user feedback.

Req. Code	Requirements description
GEOVIQUA-163	Consumer metadata (feedback) shall include a list of metadata records
GEOVIQUA-136	The quality model shall support qualitative statements of belief about the data quality
GEOVIQUA-118	Rating of dataset
GEOVIQUA-116	Peer review of datasets to establish a community of practice
GEOVIQUA-88	Expert reviews with value judgements about datasets
GEOVIQUA-45	Peer review of datasets
GEOVIQUA-34	Users want to know the problems identified and highlighted about this dataset
GEOVIQUA-33	Users want to know the citations of dataset
GEOVIQUA-29	Adopt user feedback comments and rates

Also, it is important that producer metadata and consumer metadata are harmonized and presented together. So, this will require that the task in charge for the producer's data model in task 6.1 has to be coordinated with the feedback task 3.5.

Req. Code	Requirements description
GEOVIQUA-134	Combine provider-generated metadata with user-generated metadata

GeoViQua components will be developed mainly for web browsers and desktop PCs, but also for low bandwidth transmission and small screen devices (*e.g.* smart phones and PDA components). A version of the visualization tool for mobile devices is expected in month 30 (milestone M 5.3). This will require collaboration on several tasks related with web clients, mainly tasks: 4.2, 5.3 and 6.2. This collaboration will respond to the following requirements:

Req. Code	Requirements description

GEOVIQUA-172	An application for a mobile device shall react to a display rotation
GEOVIQUA-171	An application for a mobile device shall be easy to install
GEOVIQUA-170	An application for a mobile device shall be as portable as possible to other operating systems
GEOVIQUA-169	The user shall be able to set manually its current location in a menu in an application for mobile devices
GEOVIQUA-168	An mobile device application shall use a location detection system (GPS, etc)
GEOVIQUA-79	Users want the system to provide a quality-aware augmented reality browser including mobile devices
GEOVIQUA-78	Visualisation of data using mobile and low-bandwidth devices should display quality information

Inside the GeoViQua project a couple of partners are interested in developing components for better integrating sensor data in GEOSS. This will require collaboration from the search (WP4) and the visualization activities (WP5). These are the requirements that deal with sensors:

Req. Code	**Requirements description**
GEOVIQUA-148	The quality model and quality-aware search shall support sensors with quality description
GEOVIQUA-44	Quality parameters support continuity of different in situ observations.

Intercomparison activities (WP3) are based in the ability to find similar or compatible datasets to compare and that is provided in WP4. Also, they will require some visualization of the results of the intercomparison in WP5.

Req. Code	**Requirements description**
GEOVIQUA-160	Users should be able to find datasets that they can intercompare with others.
GEOVIQUA-159	Capture output of intercomparison exercises in useful format.
GEOVIQUA-137	The system shall provide quantitative data quality estimation by comparison with reference data (e.g. Bayesian)
GEOVIQUA-111	The system should support visual side-by-side comparison of metadata attributes.
GEOVIQUA-109	Users of land use data should be supported in closely examining, easily comparing, and contributing to a wide range of metadata associated with land use datasets.
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.
GEOVIQUA-41	Users want the system to provide a mechanism so that they can compare this dataset with "trusted" reference dataset

3.1.1.2 By collaboration intensity with other projects or activities

One of the GeoViQua project's main objectives is the contribution to the GEO Label development. GeoViQua shall significantly contribute to the definition of a strong GEO Label, by allowing users to query, explore and contribute to data quality information within a flexible and interoperable framework. GEO Label requirements will be developed in WP3, integrated and validated in WP6, applied in pilot cases in WP7 and disseminated to the community in WP8. The activities will be completed in collaboration with the GEO task ID03 (formerly ST-09-02). A final report about this will be provided in month 24 (deliverable D 6.2). The implementation of the following requirements that affect the GEO Label will require collaboration of several partners in several tasks in the project:

Req. Code	Requirements description
GEOVIQUA-153	The GEO Label shall be applicable to a diversity of heterogeneous datasets allowing a consistent comparison among them the dataset and who created the label.
GEOVIQUA-133	GEO Label shall be based on well established trust triggers and quality perception mechanisms
GEOVIQUA-130	Quality search and visualisation to find the best dataset for users purpose
GEOVIQUA-80	The system should employ widely used trust mechanism and should be designed towards usability
GEOVIQUA-63	The GEO label shall give access to the methodology used to create the dataset and who created the label.
GEOVIQUA-31	The GEO Label shall reflect scientific relevance, quality, acceptance and societal needs criteria

As part of the WP8, GeoViQua provides funding for OWS-9 activities, mainly on a provenance activity that will produce the following OGC deliverables to the project:

- OWS-9 CCI (Cross Community Interoperability) Conflation with Provenance Engineering Report in the Cross-Community Interoperability thread
- OWS-9 OWS (Open GIS Web Service Innovation) Innovations Data Quality for Web Mapping Engineering Report in the OWS Innovations thread

Apart from GeoViQua sponsorship, it is expected that GeoViQua members and task teams contribute collaboratively to these external activities. Particularly, involvement of the WP5 team is expected. This process will help to develop provenance cluster deliverables and also some of the visualization requirements:

Req. Code	Requirements description
GEOVIQUA-167	The system shall support querying of dataset provenance
GEOVIQUA-161	A download data service shall provide data with associated uncertainty with each pixel in a standardized way
GEOVIQUA-139	Elaborate best practices for using visualisation standards that includes quality visualization
GEOVIQUA-129	The records in the catalogue shall contain access metadata to allow rich data visualization and propagation
GEOVIQUA-85	Metadata parameters shall be standardised such that their processing can be automated.
GEOVIQUA-77	Users want to visualise or access the data provenance information from each data source using workflow
GEOVIQUA-74	Users want to visualise the quality information in 3D format using

	several KML clients (e.g. Google Earth, NASA World Wind, ESA MIRAVI)
GEOVIQUA-71	The system should provide a standard means to allow the structured data quality information (e.g. error estimates, provenance) embedded or linked to WMS /WMTS and KML
GEOVIQUA-70	Users want to retrieve quality information about a particular pixel on a map using the standard WMS 'GetFeatureInfo' operation.

As part of WP8, GeoViQua will also participate in OGC AIP-5; in this case mainly on developing the quality data model and defining a unique identifier rule for GEOSS federated catalogues. As we said before, the quality data model will influence the whole development of the project, but may also impact the development of GEOSS as a whole. This is particularly the case if we are able to respond the QA4EO main requirements (called “principles” in QA4EO). Contributions from WP3 and WP6 teams are expected. These are the requirements that will affect this activity:

Req. Code	**Requirements description**
GEOVIQUA-166	The data quality model and its encoding(s) need a link mechanism to resources and parts of resources.
GEOVIQUA-140	The quality model shall use ISO standards for metadata (19115) and quality (19157) with extensions for UncertML and O&M
GEOVIQUA-85	Metadata parameters shall be standardised such that their processing can be automated.
GEOVIQUA-23	The quality model shall support traceability for the methodology used to determine data quality indicators (QA4EO Principle)
GEOVIQUA-22	The quality model shall support quality indicators (QA4EO Principle)

Some requirements directly affect the GCI and will require an integration phase characterized by communication and participation in GEOSS Design and Interoperability (former Architecture and Data Committee).

Req. Code	**Requirements description**
GEOVIQUA-93	The system should support the recording and subsequent querying information about the origin of a given metadata record (e.g. the harvesting metadata pathways)
GEOVIQUA-40:	Users could query and filter the GCI information content quality
GEOVIQUA-43	GCI component managers cannot take responsibility for the quality registered in GEOSS resources.
GEOVIQUA-14	The GCI shall allow inclusion, edition and retrieval of quality assurance information in an easy to compare in an interoperable form, and following the QA4EO guidelines.
GEOVIQUA-21	Elaborate procedures, guides and examples on how to follow QA4EO and how to record it in the GCI

Collaboration with UncertML standard and the UncertWeb project is needed to include UncertML in the quality model, in Task 6.1. These are the requirements that affect this collaboration:

Req. Code	**Requirements description**

GEOVIQUA-161	A download data service shall provide data with associated uncertainty with each pixel in a standardized way
GEOVIQUA-155	Spatially display the distribution of the uncertainty of the data values
GEOVIQUA-141	A visualization service shall provide a list of data quality indicators/measures (at the dataset or pixel level)
GEOVIQUA-140	The quality model shall use ISO standards for metadata (19115) and quality (19157) with extensions for UncertML and O&M
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.
GEOVIQUA-103	Users of climate data should be supported in being able to effectively view uncertainty information about data within a dataset and other key metadata for their climate data.
GEOVIQUA-72	Users want to visualize data uncertainties both at the map level (overall accuracy) and at the pixel level (spatially distributed accuracy)
GEOVIQUA-67	Users want to find certain error distribution pattern via visualisation of error estimates
GEOVIQUA-54	Collect the relevant, helpful and practicably extractable uncertainty.
GEOVIQUA-53	Suitable uncertainty evaluation methods obtained through user interaction.
GEOVIQUA-51	To identify relevant sources of uncertainty.
GEOVIQUA-50	To apply uncertainty as a parameter in the visualization.
GEOVIQUA-49	Graphical separation of inherently different kinds of uncertainty
GEOVIQUA-28	Propagation of uncertainty in a distributed information system
GEOVIQUA-13	The system shall use UncertML encoding for quantitative QI
GEOVIQUA-12	Statistical framework to quantify and manage uncertainty

3.1.2 By user influence

Our analysis identified that geospatial data users are exceedingly interested in good quality metadata records. The users and experts stated that complete and well documented metadata records are essential in the assessment of geospatial data quality. Core metadata defined in ISO and Dublin Core standards must, according to the majority of our interviewees, be provided with the geospatial datasets to enable data quality evaluation. In line with metadata standards, users are also interested in soft knowledge about data quality – i.e., data providers' comments on the overall quality of a dataset, any data errors, potential data use, and any other information that can help to assess fitness-for-use of the data. Another important aspect when selecting a quality dataset is peers' recommendations and reviews. Users would appear to be exceedingly interested in being able to obtain feedback from their peers and accept recommendations when selecting the right dataset.

Our study also revealed the importance of dataset provenance as well as citation and licensing information when assessing whether a data is fit for purpose. Users confirmed that provenance information is usually incomplete, citation information is hard to acquire, and licensing information is normally missing from the metadata records of datasets. Our survey results showed that, when selecting a dataset, users typically seek information about dataset providers, and in particular, valid contact details, because personal contact is often the best way to get a good overview of a dataset's quality. Our findings also

indicated that having side-by-side dataset and metadata comparison functionality would make the dataset selection process much easier for users.
The figure below represents an overall picture of most influential user requirements:

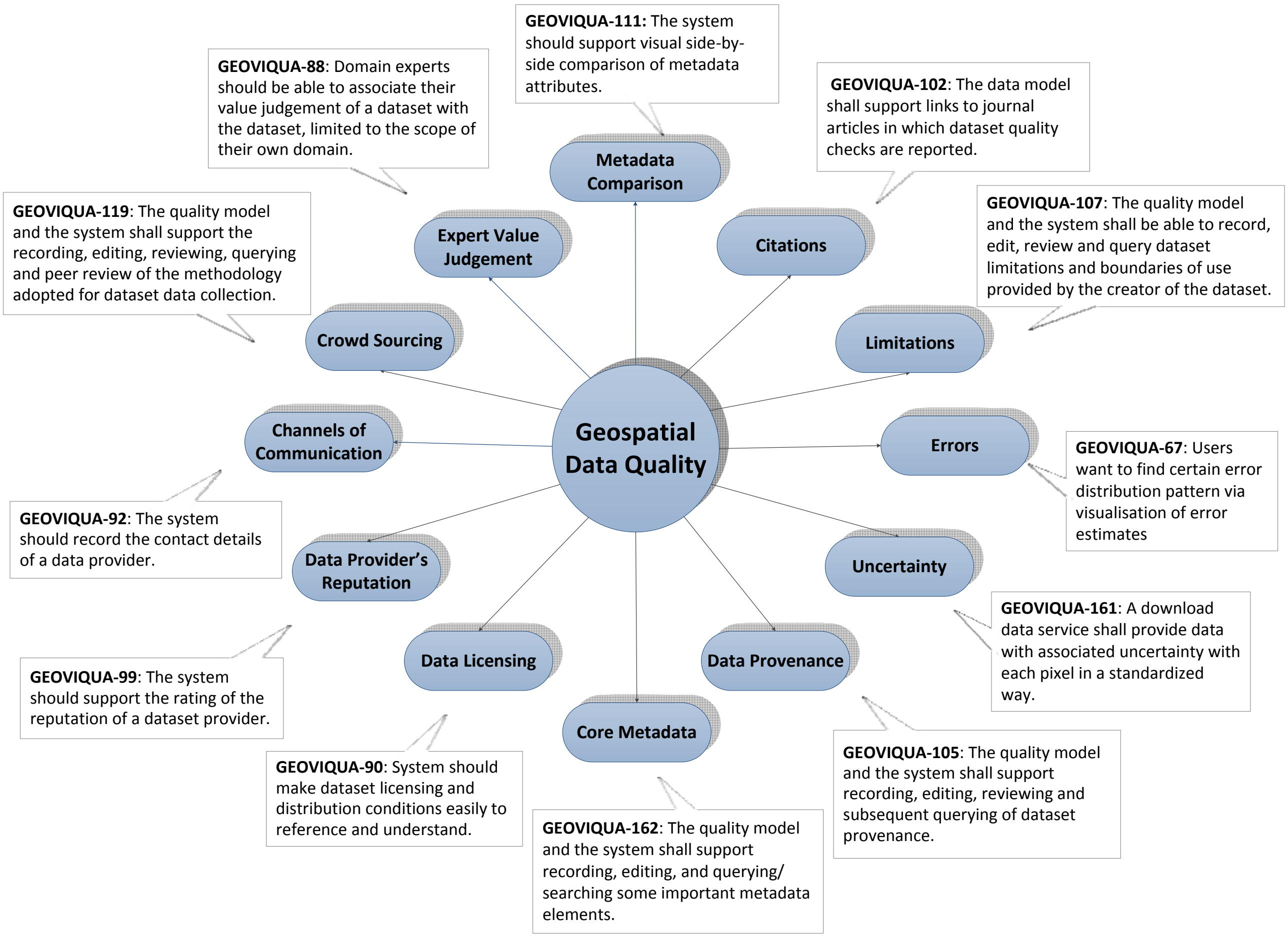

Figure 5: Most influential user requirements.

(1) Metadata Completeness, Visualisation and Comparison

Our survey revealed that, at present, users find metadata records are typically incomplete with a lot of essential data omitted. Despite the standardisation bodies' work towards establishing core metadata elements and enforcing good metadata practices, dataset providers do not always follow any standards and leave metadata records incomplete. As a result, users suffer from insufficient metadata records which make the dataset selection process more difficult. Faced with incomplete metadata records, users have to inspect data manually to assess datasets' quality and fitness-for-use. Our interviews indicated that users require at least core metadata to be provided with all the geospatial data. The interviewees stated that, when searching for geospatial data or evaluating data quality, they at least need to have coverage, licensing, methodology, uncertainty, resolution, source, acquisition date, pre-processing steps, and accuracy information provided. Our interviewees confirmed that metadata records that they come across are far from complete and are usually missing core elements. These results signify the importance of metadata standards and show that geospatial metadata still lacks metadata quality control.

Our interviews confirmed that geospatial data users and experts require more sophisticated tools for visualisation of metadata records. Users stated that, at present, metadata records are typically not easy to examine and assimilate. The non-expert users suffer the most from not being able to absorb and understand all of the information recorded in metadata. Considering the importance of metadata to geospatial data users and experts, effective visualisation methods for metadata records need to be identified to support users in the data quality evaluation and decision making process.

Another important aspect of metadata visualisation identified during the interviews is the ability to easily and systematically compare metadata records. The interviewed users indicated that side-by-side visualisation of all metadata elements would allow them to compare geospatial datasets more effectively. This feature would be most beneficial when comparing datasets that are very similar and the differences are hard to distinguish. Users stressed that selection of a suitable dataset can be a very difficult task, especially when there are several that appear to fit the purpose at first sight. Side-by-side visualisation of metadata information would support and simplify data searches, decision making processes and data quality evaluation. Metadata comparison would be particularly invaluable for less knowledgeable and non-expert users who find it hard to manually inspect the datasets to assess their fitness-for-use.

(2) Provenance

Our interviews results confirmed the importance of provenance information to geospatial data users. Users listed the following provenance elements as essential to data quality evaluation:

- original dataset provider
- party (parties) who has subsequently processed the dataset
- party (parties) who has used the dataset before
- methodology adopted for dataset data collection
- how a dataset was derived and on what it is based
- dataset harvesting pathway
- the purpose for which a dataset was originally collected
- dataset processing log

The interviewed users especially stressed the importance of information about the dataset provider, methodology and data processing steps. Our survey results revealed that the reputation of dataset providers has a significant influence on dataset selection. The methodology adopted for data collection also plays an important role in evaluation of data quality. Users are particularly interested in what equipment was used for data gathering and any problems that occurred during the gathering process. They are also interested in having a processing log provided with every dataset. A complete processing log would allow users to trace the events and transformations in the life of a dataset. Having all the information on processing steps and algorithms applied to data would allow users to measure data accuracy and revert any processing if necessary.

(3) Crowd Sourcing

Users of geospatial data stated that they rely heavily on peer recommendations when selecting a dataset to use. They contact their peers to obtain suggestions on what datasets are most suitable and are of good quality. Their peers provide valuable feedback on what datasets they have used, what these datasets were good for, potential problems with the datasets and other occasionally useful information. Having this sort of feedback available for every geospatial dataset would facilitate improved data selection and quality evaluation. The peer review functionality would allow users and experts to provide their comments on datasets they used and to indicate any limitations or issues they encountered.

Geospatial data producers stated that they are also interested in having their datasets peer reviewed. Having user feedback on their products would allow data producers to identify and resolve any issues within their datasets, and also to respond to users' comments.

One suggestion is that such peer review functionality should be facilitated via a standardised form that users can fill in to provide their feedback. Our interviews highlighted the following potential fields for the peer review form:

- general user comments about the dataset, i.e., whether user liked or disliked the dataset and why
- general user comments on the dataset metadata record – metadata completeness, usefulness of metadata record, etc.
- what a user used the dataset for and was the dataset fit for the expected use
- dataset's frequency of use, i.e., how often do users use the dataset
- user suggestions on the dataset's expected use
- the dataset's limitations and problems

Our interviewees also indicated that they would not accept recommendations from just anyone and noted that they typically rely on peers that they know and trust.

(4) Rating of Datasets and Data Providers

In line with having functionality to provide comments on datasets, users stated their interest in having an ability to rate datasets and data providers. Users mentioned that having something similar to 5-star ratings assigned to datasets would make the dataset selection process much simpler. Unfortunately, however, having just a signal numeric rating assigned to a dataset would be unlikely to provide enough information to evaluate data quality and trustworthiness. But they have the advantage of being machine interpretable and therefore can be used for sorting and filtering of search results. Rating functionality can potentially be combined with peer review information, where ratings would act as visual indicators of overall quality and user comments and reviews would provide the underlying rationale.

(5) Citation Information

Our interviews identified that the majority of users base their dataset quality evaluation on dataset citation information. That is, when making dataset selections, users are largely interested in accessing the publications where data quality checks are reported for the dataset. As part of this, journal articles that describe dataset use and evaluation are also considered to be very important in dataset quality assessment.

(6) Data Provider

The reputation of data providers was identified as one of the key factors in dataset selection. Users typically rely on data from producers that they already know or those who have a very good reputation in the community. Considering the fact that users are more likely to select data from producers that already have a good established reputation, smaller and unknown data producers will experience much lower data demand. Our interviewees also indicated the importance of documentation supplied by the data provider with the data they produce when assessing the trustworthiness of data providers. Well organised and easily accessible documentation engenders user trust in both data provider and the datasets they produce.

(7) Licensing

ISO 19115 and Dublin Core metadata standards include restrictions and licensing information as core metadata elements. Despite the fact that standardisation bodies insist on including licensing information with all geospatial data, data producers often fail to provide restrictions information with the datasets that they publish. Indeed, a number of interviewed users and experts pointed out that the licensing information is nearly always missing.

(8) Soft Knowledge

Our interviewees highlighted that there are cases when data quality measures cannot be recorded in standard metadata records. For instance, providers might be aware of problems with a satellite such as higher uncertainty in certain areas of the image it takes; in such cases, they provide soft knowledge about data quality including information which they think may be relevant to potential users. Users stressed the importance of data producers' comments and recommendations of this kind that are provided with the datasets they produce. They stated that having at least some soft knowledge about data uncertainty and error estimates would significantly help in evaluation of data quality and in more effective use of the data.

3.2 Thematic summaries of user requirements

This section contains user requirements organized by the clusters described in 2.1.8. Each section was contributed by the respective cluster adopter who curated the cluster and is most aware of its contents.

3.2.1 Feedback

This cluster aimed at identifying the requirements for users to provide feedback about data. The overall goal is to give users who search for data a mean to identify if a dataset is suitable for their particular task at hand. At the beginning of the clustering of requirements, this one was named “VGI and Feedback”, using the broad term of volunteered geographic information (Goodchild, 2007) to describe the user contributed information with respect to spatial data. However, this term commonly refers to tools for users to create and share geographic data (for example OpenStreetMap), whereas the core of the cluster boiled down to users leading some form of discussion about a geographic dataset. The cluster was therefore renamed.

Users were identified as an interesting source for metadata in several user interviews. In fact this can be seen in the broader context of the social web, or Web 2.0, where online users use the internet not only for consumption of information, but also for interaction with peers and publication. Social media, such as Facebook, LinkedIn or Twitter, are an important part of many people’s everyday life, be it private or professional. General concepts of discussion threads, online recommendations and votings permeate everyday use of the internet. This makes the features suggested in this cluster very concrete as they are close to user’s work patterns, but also well connected to today’s most powerful applications of the World Wide Web.

The feedback cluster contains requirements that transfer general concepts of the social web to GEOSS, such as discussions about items and ratings. The sources of the requirements can often be traced back to common everyday websites, such as shopping platforms, and interviewees directly mentioned sites such as Amazon or eBay. The advantage of qualitative statements is that is very well fits the most important property of quality: it is subjective. When reading qualitative statements (GEOVIQUA-136), the opinion is build up by each user himself and users can rate or apply only aspects of the discussion as they fit their need.

Discussed items could be a raster dataset, a specific time series, or only a part of a remote sensing image. Here we already identified one main requirement – feedback scope (GEOVIQUA-88). All aspects of feedback described later must be linked to a dataset, but often also only to specific parts of a dataset to preserve the usefulness.

Naturally feedback could also head to a more generic direction, from a dataset to the dataset provider. While a user mentioned this requirement, we ruled the rating of dataset providers out-of-scope for GeoViQua. Such ratings would first of all be hard to communicate to data providers, which are a crucial partner in making user feedback in spatial data infrastructures work. Secondly they would very much suffer from the aforementioned scope issues. Apart from the spatio-temporal and thematic scope of the data, data sets by independent groups of people might be published under the umbrella of one organisation, so provider ratings underlie organisational issues (GEOVIQUA-99).

The most important requirement is textual comments on datasets (GEOVIQUA-88). A free text cannot directly be aggregated/evaluated, but has the simple advantage of being able to capture any information a user wants to add. Free text also ensures the most basic feature of data quality is covered: it is inherently subjective. If a user reads comments from other users about how they used the data and how the data performed, he can form an adapted interpretation of the quality for his specific task at hand. We think this adaptation is much more powerful than automatic systems could be as of today. But of course comments could be evaluated using text-mining, too. We also identified a middle way of review fields that are so important that they should be modelled explicitly (GEOVIQUA-163) to ensure machine-readability and combine the power of automatic and manual interpretation of comments. These could all be included into a free text field with the risk of not being able to actually link to the referenced items. The requirement for a full-blown quality control process was deemed beyond the scope for the cluster (GEOVIQUA-47), as it targets rather the data provider side of quality.

These explicit review fields are as follows.

- Comparable datasets, i.e. a user would like to know a specific dataset that could be a replacement or supplement for the currently shown dataset.
- Publications and citations (GEOVIQUA-33), i.e. a user wants to know if other people published a scientific article using a dataset or about the data itself.
- Problems (GEOVIQUA-34), i.e. users are specifically interested in the problems that others had using the data. This shows the critical aspect that the quality of the data has for the whole analysis process.
- Fitness of purpose, i.e. a user wants to know if someone else used the dataset in the same way as he plans to. This is a supplementary item in response to possible problems, because the users are eager to know if someone positively identified a dataset as fit for a specific analysis.

The process of people commenting about a given item can be seen as a first variation of a peer-review process. This process could also be focused on specific communities of practice (GEOVIQUA-116), which allows users to easier identify interesting datasets and can be able to foster a discussion because of the common (research) interest. While the inclusion of a peer review architecture itself is out of scope, different levels of review that can be accomplished using discussion forums and user accounts are imaginable.
Up to now, users are not commonly identified when visiting data catalogues, such as the GEOPortal, so user accounts are considered a practical requirement. This is not a crucial requirement for user feedback to work, but can be very important to enable specific features. For example the classification of users allows expert reviews of a dataset or double-blind respectively non-public peer reviews. Experts' value judgements could be shown more prominently, compared to non-specialist and not-logged-in user's comments, and help the user in the evaluation of the collection of comments while building an own opinion. Although we see user encouragement respectively incentives for rating and active community building beyond the scope of the project (GEOVIQUA-38), user accounts could be used to ease the process of rating, for example by asking users to comment on a dataset they browsed or downloaded in a previous session in a data portal.

A totally different component is the requirement for a rating system (GEOVIQUA-118). This comprises a numerical evaluation of a dataset on a common scale, such as "one to five stars", as most users are familiar with from aforementioned websites or also hotel classifications. Such a rating system allows to influence the visualization of datasets in an infrastructure, for example for ordering of search results, and can be visualized very effectively (using colours or icons). This facilitates a user the exploration of the existing feedback landscape.

An important task is the integration of these dynamic structures into the static world of geospatial metadata. Here this cluster also faces the main challenges of the project: raising awareness about data quality while targeting existing, sometimes large, infrastructure systems. These challenges are technological as well as social and can take very different forms as described in this section. We estimate the presented requirements to be able to capture the most of the potential of user generated feedback and that implementing these requirements as features for data portals would be a strong incentive for users to contribute to the process and generate added value quickly.

3.2.2 Intercomparison

Intercomparison in the GEOVIQUA context can be seen as assessing the quality of data by comparing it to other data representing the same measured quantity, but originating from a different source. Such sources include:

1. In situ observations that typically originate from ground measurement stations or aircraft.
2. Remotely-sensed data that typically originate from space-borne instruments on satellites
3. Output from numerical simulations/models (excellent spatiotemporal coverage, variable accuracy, almost always gridded)

During user interviews and user requirements analysis, several user requirements for intercomparison were derived, such as:

Identifier	*Description*
GEOVIQUA-39	Users want to know any inter-comparison result(s) with other dataset
GEOVIQUA-41	Users want the system to provide a mechanism so that they can compare this dataset with "trusted" reference dataset
GEOVIQUA-46	Users want to know the correlations between the error estimates
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.
GEOVIQUA-111	The system should support visual side-by-side comparison of metadata attributes.
GEOVIQUA-151	The system has to be able to compare/combine different quality indicators and measures (by mapping)
GEOVIQUA-159	Capture output of intercomparison exercises in useful format.
GEOVIQUA-160	Users should be able to find datasets that they can intercompare with others.

One conclusion from the user interviews was that intercomparison is seen as a very important quality assessment mechanism for datasets, as it provides a means of verification using *independent* data.

These user requirements were further analyzed and structured. A structuring of requirements has been proposed based on envisaged user scenarios, such as:

- *GeoViQua as a source of (references to) intercomparison study results:* this applies to users that want to gain trust in a measurement dataset they are considering for use but do not want to intercompare anything by themselves. These could be supported by showing references to publications issued by scientists that have performed intercomparisons of the dataset to other datasets. Another option is to add results of such intercomparison studies as quantitative quality information to the *GeoViQua* system in order for users to perform quantitative analysis on intercomparison results. Note that given the complexity of the intercomparison, this is not trivial to do (at least not trivial on a single measurement value scale). In this case there is a link with VGI and Feedback.
- *GeoViQua as a search engine for finding intercomparison data:* this applies to users that want to gain trust in a measurement dataset and want to perform an intercomparison by themselves, for example to their own data. These users require information on which other datasets covering the same range in time and space are available for comparison (collocation). As a next step, these users then require the actual datasets for performing the numerical intercomparison on their own system, and possibly SW-tools that support the intercomparison.
- *GeoViQua as an intercomparison engine that produces and reports quality data:* This applies to users that want to gain trust in a measurement dataset and want to have the *GeoViQua* system perform an intercomparison on datasets that they selected. The intercomparison may apply to:
    - Intercomparing meta-data on datasets, such as comparing geospatial coverage, time coverage, measured quantities in the dataset etc. This is similar to the *collocation* function mentioned in the previous item.
    - Intercomparing the actual measurement data and deriving data quality statistics. This is seen as quite useful, given the number of user requirements that refer to this function, but also quite challenging to obtain within the GEOVIQUA system, as this requires on-line data-access, implementation and execution of the specific algorithms mentioned above somewhere in the GEOVIQUA system, an additional GUI for controlling the intercomparison, and standards on data representation and intercomparison.

3.2.3 Provenance and Purpose

In order to clarify the concrete objectives of this cluster, we have attached the definition of provenance. This is extracted from the dictionary GeoViQua is developing (http://twiki.geoviqua.org/twiki/bin/view/GeoViQua/GeoviQuaTerminology)

Provenance is the capacity to annotate the sources and processes that were used in the creation of a resource (e.g., datasets and operations involved in the creation of a particular dataset). It provides traceability and trust in the resource by knowing its origins and giving credit to the processors of the resource. If it is complete and has access to the same *sources and processes, it potentially allows recreating the resource. It is considered a component of the quality description of the dataset.*

The main process consisted in deriving user requirements from interviews and real user stories. Also we compiled user requirements contained in relevant GEOSS documentation. We reviewed, one by one, the following documents: The GEOSS 10-Year Implementation Plan (February 2005), GEO Report on Progress (November 2010), Strategic Guidance for Current and Potential Contributors to GEOSS (October 2007), White Paper for GEOSS Data Sharing Policies (September 2008) and the GEOSS Architecture Implementation Pilot 3 (AIP3) Summary (February 2011). Finally we extracted many requirements related with quality and provenance. After that the cleansing and refining of those user requirements was done in order to focus on the more important ones within the scope of GeoViQua.

The provenance and purpose requirements are clearly related with information the users want to access in order to know or better understand the origin and the transformation the datasets. The main requirements we have considered as both complete and relevant for the project are eight. These requirements contain the information needed to develop the GeoViQua components; these can be classified in three groups, as shown in Table 3 below. The first one is the requirement for visualization, containing a single requirement that suggest access the data provenance information from source dataset in an automatic and systematic manner. The second group covers the subject of information access. This is mostly related with the origin (producers) and transformations (methodology, standards, codes and versions). This can be considered pure requirements, directly related with the definition of provenance. In the third group we have the user requirements related with the system, that is to say the implementations that have to be done to achieve the user needs. The main needs are to record the details of data providers and methodology information.

Group	*Description*	*Identifier*
Information visualization	Data provenance information from each data source using workflow	GEOVIQUA-77
Access to information	Data provider or data author	GEOVIQUA-59
	Build-up history of the dataset	GEOVIQUA-61
	Version of code, observation data, methodology, and assimilation system that were used to generate the reanalysis product	GEOVIQUA-62
	Record the contact details of a data provider.	GEOVIQUA-92
System requirement	Support the recording and subsequent review of methodology information associated with a dataset.	GEOVIQUA-91

Table 3 Provenance user requirements classified in the three main groups of purpose

This provenance and purpose requirements are relevant in order to ease the way the users access quality information, as the data quality strongly depends on the original data and transformations. The provenance is also a main element in metadata schemas, in order to provide a complete description of the dataset, and concretize the quality information. For instance, in the ISO 19115 metadata schema, the provenance is a relevant factor in quality data, as it is one of the two entities directly depending on the quality information. Moreover, the requirements in the third group pertain information that can be collected using ISO 19115, since the schema contains it in several ways. The main problem related with the lack of provenance information in the metadata is the complexity of the schema itself. ISO 19115 allows to document provenance information in three different ways, all of them very useful for different purposes. Nevertheless, there is a potential for misunderstanding of the way metadata may be described. This fact was backed substantially in the analysis conducted in Task 7.2 of GeoViQua about the analysis of the quality data in GEOSS Clearinghouse.

The main difficulty during the process of delimiting the user requirements in this cluster was caused by the fact that most of the quality information the users want to access, is considered subject to quality-aware search and provenance clusters simultaneously. Thus, some other requirements such as “The quality model and the system shall support recording, editing, reviewing and subsequent querying of dataset provenance, GEOVIQUA-105” and “The system should support the recording and subsequent querying information about the origin of a given metadata record (e.g. the harvesting metadata pathways) GEOVIQUA-93” are considered in both clusters, and their handling had to be coordinated closely.

The next step in the delimitation of the provenance and purpose requirements was to link those requirements with the scenarios, which is an important task in the project to test the developed components (Task 7.1). The project decided to accomplish this process by relating the user requirements with user stories. We built one UML diagram (see **¡Error! No se encuentra el origen de la referencia.**) per cluster in order to better visualize and simplify the requirements. The result of this activity are useful UML diagrams in which real cases are related with our user requirements gathered in order to assess them in the future (WP7).

One important point to highlight within the GeoViQua requirement validation process is that actually we have tools to inform about all this information related to provenance (ISO 19115), but the accessible lineage and citations are often missing. That raised the question of the creation of a machine-processable provenance profile, maybe as a functional requirement, within GeoViQua, but it was agreed that archival is out of scope of the project. Instead, the tasks that GeoViQua will develop on this matter will be a manual of good practices related with standards and quality models to enhance users and producers to better describe the quality information and provenance.

3.2.4 Quality and Uncertainty

The quality and uncertainty cluster grouped together user requirements that focussed on both qualitative and quantitative descriptions of the quality of the data, typically characterised as uncertainty with respect to the real value of the quantity of interest. At the meta level of user stories, or more correctly epics, which encompass the broad scope of the user requirements we can see two main stories. On one hand we have the producer story about being able to publish their data with associated quality information and having a system that guides them in achieving that (GEOVIQUA-66), presumably to increase the usage and usability of their data. On the other hand we have the users of the data, and they are asking for information about the quality so they can find the best quality data for their application (GEOVIQUA-146) and data which they know has gone through a proper validation mechanism (GEOVIQUA-55). Associated with the user views on quality are the identification of a number of facets of quality information, through the use of quality indicators (GEOVIQUA-22):

- quantitative (GEOVIQUA-135)
  - statistical / Bayesian characterisation of relation to reality (GEOVIQUA-137, -96);
- qualitative (GEOVIQUA-136);
  - Provenance / lineage (GEOVIQUA-105);
  - Usage, both intended (GEOVIQUA-120, -107) and actual (GEOVIQUA-158);
  - Citation of calibration / validation (GEOVIQUA-113) and usage (GEOVIQUA-158);
  - User feedback (GEOVIQUA-136).

Alongside the definition of the types of quality information sought are statements that relate to aspects that these quality indicators should consider when we are implementing them:

- traceability, following QA4EO principles (GEOVIQUA-23, -119, -93, -125)
- consistency with existing standards and an extensions proposed to those (GEOVIQUA-140, -89)
- support quality descriptors across different scales and data model components – hierarchical models (GEOVIQUA-165) and sensor quality information (GEOVIQUA-148)

Finally there are also some functional issues raised by our requirements capture process, which addresses issues such as ensuring the system we develop makes it possible to create, edit and search, the quality indicators (GEOVIQUA-162). Finally there is also some mention of the GEOLabel and that this should use trust triggers to denote that quality information is available and meets the requirements of having traceable quality indicators associated with it (GEOVIQUA-133).

To summarise the findings from the requirements capture exercise users want a system that:

1. Supports both quantitative and qualitative quality indicators including:
   a. accuracy, completeness, consistency, provenance / lineage, citation, usage (both intended and actual) and user feedback.
2. Employs Quality Indicators that are:
   a. traceable, that is supported by evidence;
   b. applicable to data at various levels (single observation, coverage, data set).
3. Provides both producer and user quality indicators.

The user interviews also recognise the barriers to achieving this, and suggest we should be developing systems to support the creation, possibly in an automatic manner, of Quality Indicators associated with data.

3.2.5 Quality-aware Search

This cluster aimed at identifying the capabilities needed to support data discovery and evaluation based on quality assessment.

As by the DoW, one of GeoViQua objectives is to equip the GCI with a graphical search interface to represent statistics and metadata summaries (e.g. in the form of tables,

charts), thus enabling end users to leverage quality constraints for data discovery and evaluation.
The rationale is that current query interfaces over catalogues and processing services are poor in presenting results via a coherent and meaningful aggregation of information, namely by means of graphs and/or statistic summaries.
Topics concerned with improving the current visualisation of metadata information and the use, if any, of graphs and statistics to add value to the geospatial data discovery process, received little attention so far and should be further investigated.

Hence, we identified the development of data quality aware geo-search tools as a scientific objective for this cluster, related to the overall project success (GEOVIQUA-142).

In fact, GeoViQua has a specific Work Package (WP4) for “Enhanced geo-search tools”. WP4 started in January 2012 and its major anticipated technical innovations are search and visualization tools that communicate and exploit data quality information from the GCI.

Based on user interviews, two user stories were identified, indicating the need for the “best possible” data for the users’ purpose (GEOVIQUA-146 and GEOVIQUA-130).
This implies a high degree of subjectivity in judgment and suggests (GEOVIQUA-134) that the GeoViQua system should exploit a combination of provider-generated metadata (objective indicators such as summary statistics), system-generated metadata (contextual/tracking information such as provenance of data and metadata), and user-generated metadata (informal user comments, usage information, rating, etc.).

Other identified functional requirements pertain to query language, provenance, and support to sensor data. A requirement on the provenance of metadata (meta-provenance; GEOVIQUA-93) is also considered, possibly with a lower priority as far as GeoViQua scope is concerned.

Moreover, metadata should include sufficiently complete access information, to allow rich data visualization and propagation. This identifies a non-functional requirement on metadata completeness (GEOVIQUA-129).

To implement the above requirements, GeoViQua architecture will build on the brokering approach successfully experimented with in the EuroGEOSS project, to augment the current GCI capabilities with quality-aware components. This way, GeoViQua can effectively complement and integrate the EuroGEOSS Brokering Framework.

3.2.6 Standards and Communication

This cluster mostly contains requirements that pertain to the project’s involvement in standardization efforts and other communication-intensive goals. The interviews, the real use cases and scientific papers were the base for the extraction of user requirements related with Standard and Communication.

Once we had the cluster cleaned and finished we also made the UML diagrams explaining the origin of the final user requirements related with Standards and Communication. These diagrams are shown in **¡Error! No se encuentra el origen de la referencia.**

The scientific documents used to retrieve these user requirements were: GEOSS Citation Standard. DRAFT for Discussion by ST-09-02, (March 2011), AIP3 Data harmonization Engineering report (OGC - February 2011) and GEOSS Draft White Paper for GEOSS Data Sharing Policies (September 2008).

This cluster is the result of the importance the users grant to the standardization. In a global and free access system for Earth Observation data, as the GEOSS has becoming. It is of paramount importance to provide effort to built a robust intercomparison system in order to accomplish several of the main targets the GEOSS published in the GEOSS 10-Year Implementation Plan (February 2005): "Ensure that the shared observations and products are accessible, comparable, and understandable, by supporting common standards and adaptation to user's needs" and "encompass all areas of the world, and to cover in situ, airborne, and space-based observations". As is stated in the previous affirmations, the importance of the standards lead on the interoperability facilities it gives, the way of organizing, accessing and building data also by an automatic process (UML, XML, KML, WMS or WMTS are examples of it).

There are eleven final user requirements, contained in the cluster which are related to standards and communication. We have classified these in four groups; see Table 4 Standards and Communication requirements classified in the four main groups of purpose.Table 4. The first group is related to information access, in the sense that standardization is relevant when handling big amounts of data. The standardization is a functional mechanism to produce data and metadata always using common, so interoperable, guidelines, which consequently helps when accessing them. For instance, users could find the same information from many catalogues by requesting them using a predefined, standardized, strategy. Standards are tools to homogenize the data processing, the way the information is created, retrieved and accessed can be also automated so that users can search for multiple kinds of information in huge catalogues connected to many producers by using a single request. In our specific case, users want to be in the know about the following aspects: methodology used to create the dataset, scientific relevance, quality, acceptance and societal needs.

Two of these requirements, included in the first group, point to the GEO Label. This is a concept directly related with communication, a relevant issue also in GEOSS. The information users require is specifically related with the data provider and the societal relevance. The GEO Label is a key for communication in the GCI. That was clearly reflected in the GELACO document, (the GEO Label concept document in draft), made by the ST-09-02 Task Team on September 2010. The document title is: *Informing Users about the Quality, Relevance and Acceptance of Services, Data Sets and Products Provided by GEOSS.* It highlights the importance of user requirements related to communication. This task is trying to give response to the user needs, by developing *A Tri-Faceted Label.* This concept aims to inform users of services, data sets and products accessible through GEOSS about the societal significance and quality of the resources provided. All these targets are also reflected in the GeoViQua Standards and Communication requirements.

The second group is related to the user's ability to download data with the quality information associated, i.e. the capacity of the system to do so in a standardized way and, if possible, in different levels of detail.

The third group gathers the requirements directly related with visualization. The users mostly request visualization tools to easy compare heterogeneous datasets. They are also worried about the possibility of using the system as an integrated unit, following the same guidance, i.e. interoperability and quality directives, so that they put out the need of following the QA4EO guidelines and to have best practices documents in the GEO Portal.

The fourth group contains the requirements that directly mention the need of using specific standards. In order to let users work in an interoperable manner, with common procedures and trusted criteria regarding quality in Earth observations, it is necessary to enhance the development of a standard for EO Data Quality, QA4EO. Moreover, users need to know how this standard should be applied, including practical examples of proper use.

Group	*Description*	*Identifier*
Access to information	The data quality model and its encoding(s) need a link mechanism to resources and parts of resources.	GEOVIQUA-166
	The GEO label shall give access to the methodology used to create the dataset and who created the label.	GEOVIQUA-63
	The GEO Label shall reflect scientific relevance, quality, acceptance and societal needs criteria	GEOVIQUA-31
Download information	A download data service shall provide data with associated uncertainty with each pixel in a standardized way	GEOVIQUA-161
Visualization and comparison of the information	The GCI shall allow inclusion, edition and retrieval of quality assurance information in an easy to compare in an interoperable form, and following the QA4EO guidelines.	GEOVIQUA-14
	The GEO Label shall be applicable to a diversity of heterogeneous datasets allowing a consistent comparison among them	GEOVIQUA- 153
	A visualization service shall provide a list of data quality indicators/measures (at the dataset or pixel level)	GEOVIQUA-141
	Elaborate best practices for using visualisation standards that includes quality visualization	GEOVIQUA-139
Conformance of the information	Dataset metadata shall conform to international standards (potentially legalised) to increase conformity.	GEOVIQUA-117
	Metadata parameters shall be standardised such that their processing can be automated.	GEOVIQUA-85

	Elaborate procedures, guides and examples on how to follow QA4EO and how to record it in the GCI	GEOVIQUA-21

Table 4 Standards and Communication requirements classified in the four main groups of purpose.

All the user requirements contained in this cluster support the next GEOSS targets, also important within the GeoViQua project:

- Use Earth observation data and products (e.g., process, integrate, model) following accepted standards;
- Contribute to access and retrieve data from global data systems and networks;
- Analyze and interpret data to enable development of decision-support tools and to advance understanding in the nine societal benefit areas;
- Integrate Earth observation data and products with other data and products, for a more complete view and understanding of problems and derived solutions;
- The success of GEOSS that will depend on data and information providers accepting and implementing a set of interoperability arrangements, including technical specifications for collecting, processing, storing, and disseminating shared data, metadata, and products.
- In GEOSS, interoperability will be based on non-proprietary standards. Attention is drawn to the importance of using existing international standards organizations and institutes.

It is recommended to enhance best practices for data quality assurance by observing system data in the GCI and the clearing house. Currently the Best Practices Wiki of the GCI is poorly developed (many groups have no input or have a link to the same event: SPRING). The GEO members must be able to include quality indicators extracted from the assurance strategy associated with the GEOSS (QA4EO) when they register data in the Component and Service Registry (CSR).

Also, recognizing that there are several equivalent encoding possibilities that can make interoperability difficult, it's a good practice to provide comprehensive documentation regarding quality standards in the GEO Portal.

3.2.7 Visualization

This cluster defines the high-level requirements for data quality visualization. The requirements encompass several important and distinct aspects of quality visualization, including the visualization of uncertainty fields, the visualization of quality metadata, the interaction with international standards and the use of 3D as well as mobile devices and "augmented reality" as a means to overlay quality information.
End-user requirements were derived from user interviews and from a search of the scholarly literature. These sources were combined with the original Description of Work and provided valuable detail, refinements and context. In general, the DoW was found to be valid, requiring no major modification in order to meet the elicited user requirements.
At the time of writing of this document, work package 5 (which focuses on visualization in GeoViQua) is just beginning, and therefore these requirements will be refined and expanded into lower-level technical requirements in the next few months.
The requirements fall into two main categories: requirements that concern *what* information is to be visualized, and requirements that concern *how* the information is to be

visualized. The information to be visualized includes error and uncertainty estimates, both at the dataset and pixel level (e.g. GEOVIQUA-72), provenance information (GEOVIQUA-77) and quality indicators (GEOVIQUA-68). This information is to be visualized in both desktop and web-based tools (GEOVIQUA-76) and mobile devices (GEOVIQUA-79).
There is also a requirement to link quality information to the OGC visualization technologies WMS and KML (GEOVIQUA-71).
Given the variety of different visualization methods and technologies that are anticipated to be relevant to the GEOSS, there is a requirement for the quality visualization to adapt to the user's context in order to provide the most helpful and comprehensible information (GEOVIQUA-126).

4. Summary

In this document we presented the user requirements we found valid for GeoViQua and looked at them from different angles. We presented the way we elicited and analysed the input we got from users and other sources, enabling us to deepen our understanding of how the GeoViQua project can make a lasting contribution while staying connected to the demands of relevant user communities and building incrementally on established practice and technology.

In the upcoming project deliverable "Technical requirements for GeoViQua" we will further explore the technical side of the requirements.

5. References

Cohn, M. (2004). *User Stories Applied.* Addison-Wesley.

GEOURR. (2011). Retrieved 02 2012, from http://www.scgcorp.com/urr: GEOSS User Requirements Registry

Goodchild, M. (2007). Citizens as Sensors: the world of volunteered geography. *GeoJournal 69 (4)* , pp. 211-221.

Sutcliffe, A. (2002). *User-centered Requirements Engineering.* London: Springer.

Annex

Annex A Use case diagrams

One of the difficulties when working with requirements is ensuring that all requirements are present, their redundancies and implicit relationships are really discovered. In order to mitigate this problem, the group elaborated use case diagrams where atomic use cases were identified for the integrated system we are trying to build. These use cases were related with the requirements. In a use case diagram physical users and automatic systems (agents) are presented by simplified human figures that are interacting with the system (that is represented by boxes) by means of individual actions that are represented by ellipses.

Individual actions in the project we extended a classical use case diagram by adding requirements boxes and considering the use case a possible application of the requirement. During this process many relations between requirements were identified and documented in the Jira system as links between requirements. Also, the process helped to identify the targets of the requirements and to clean up the description of them.

In the following, you can find four examples of the use case diagrams for Standards and communication, Provenance and Visualization cluster and also a specific one for the GEO Label.

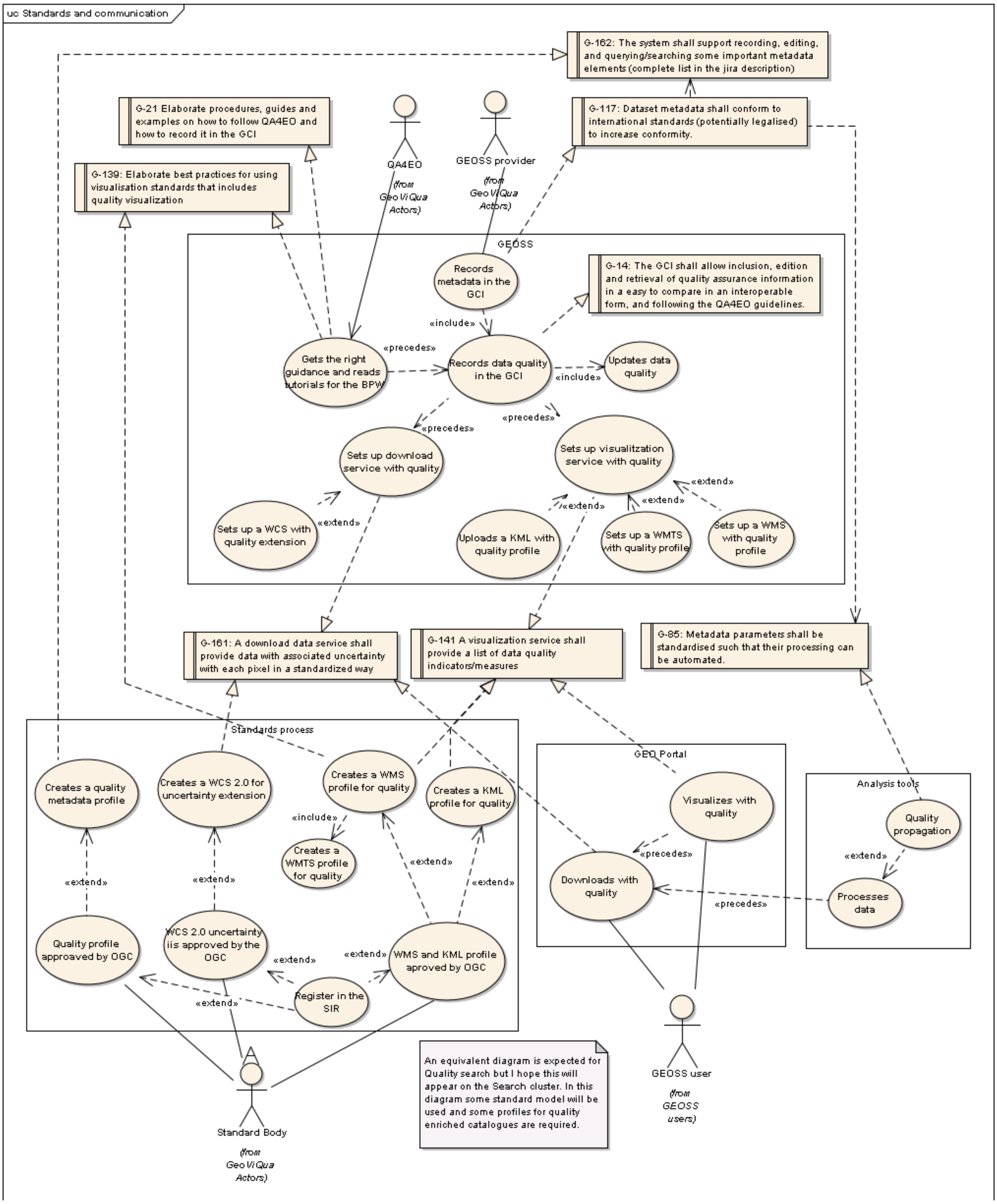

Figure 4 Standard and communication use cases

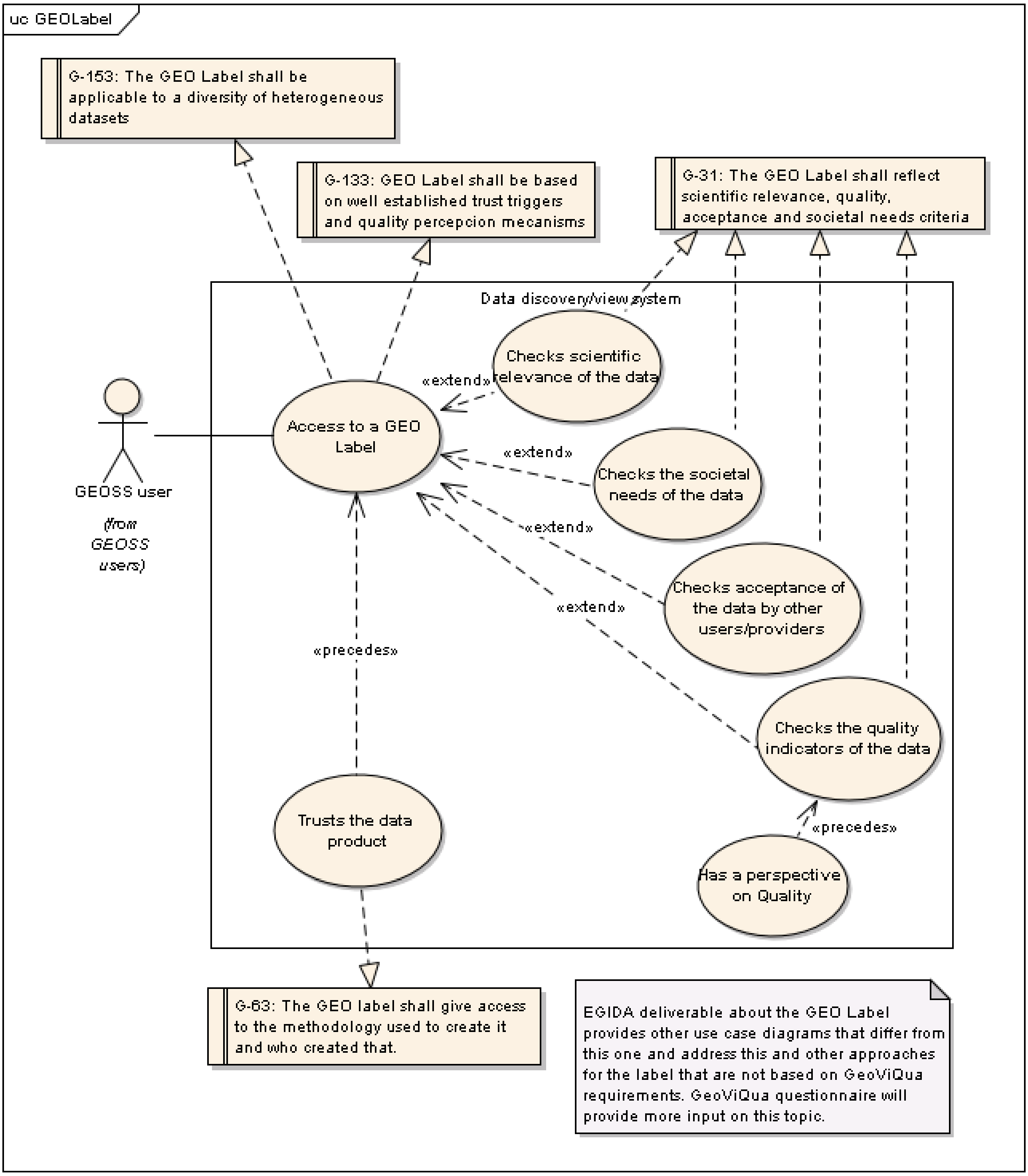

Figure 5 GEO Label UML Use Case diagram

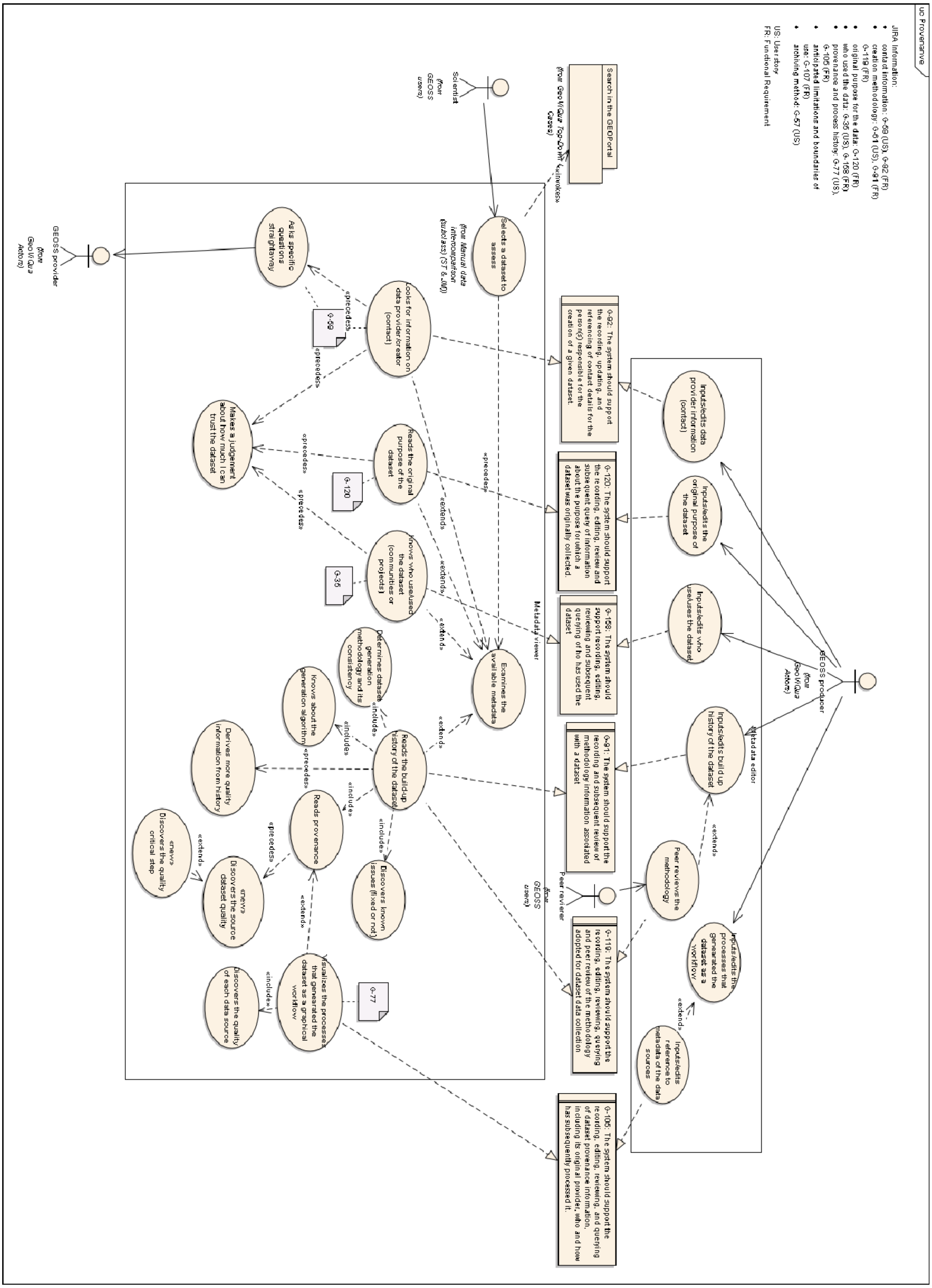

Figure 6: Provencance use case diagram (turn page please)

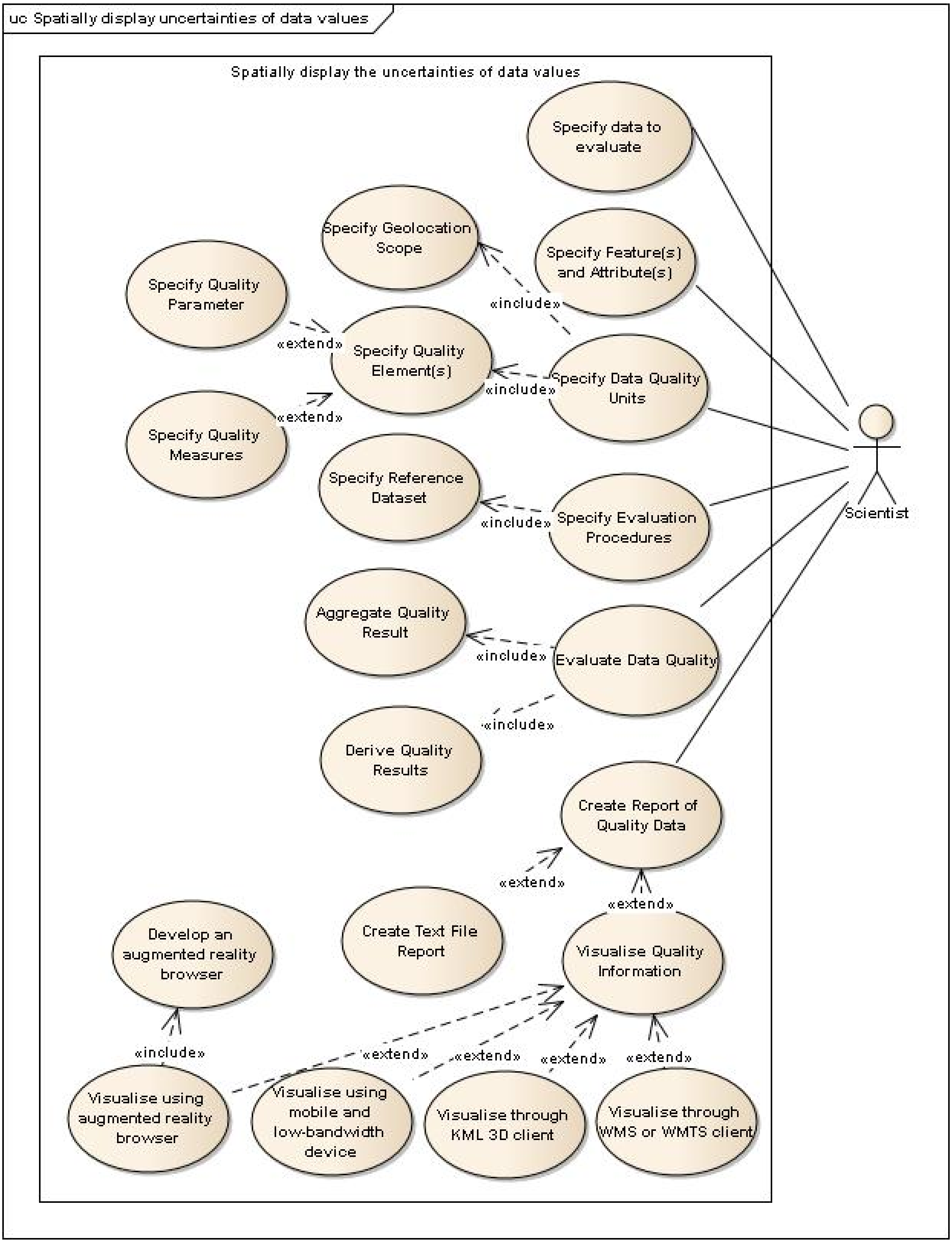

Figure 7 Quality visualization use case diagram

Annex B Interview Guide

This section contains the interview guide as it was used internally, i.e. things the interviewee wouldn't be asked directly are included.

(1) Introducing the interview

The interview should be introduced by explaining that we are working on a European project, whose aim is to help build a global system for discovering accessing earth science data. Our goal in the project is to ensure that the infrastructure gives users the information they need to discover dataset and choose between them. Check to see that the interviewee is happy for his/her name to be recorded, or whether they would prefer to remain anonymous.

Note: we could go into more details of GEOSS here if appropriate to the interview in question.

(2) Questions

(The notes below the questions are not intended to be shown to the interviewee, but are used to provide more guidance to the interviewer.)

1. Please describe a current area of your work in which you use external data sources

Notes: A specific project might be more helpful than a general research theme. The work must involve the use of external data sources or it won't be very relevant to us.

2. What data do you use in your work, and where do they come from?

Notes: Where appropriate, we should capture relevant details that are specific to the community in question. This may include:

- *Processing levels for remotely-sensed data in the (semi-)standard "L0 to L4" terminology.*
- *Data type (raster, vector)*
- *Scale:*
  - *Spatial and temporal extent and resolution*
  - *Radiometric(bits/pixel) | Spectral (spectral regions, number of bands, band width)*

3. How do you choose which datasets to use in your work? What are the reasons for your decisions?

Notes: The question doesn't mention "quality" explicitly because this can mean different things to different people. However, the user's perspective of quality, or "suitability for the job" should drop naturally out of this. (It may be that there is only one dataset that provides the information they need!) If the user responds "I choose the best-quality datasets" we must be sure to explore this in more detail.

If the user has a choice of similar datasets, we must be sure to explore why they choose a particular dataset over the "competitors".

Issues of data visualization might also come from this (if the user likes to see a picture of the data before downloading it for instance).

4. Are you aware of any data certificates or seals in selecting your data? Do you look for specific certificates or meta-information in a data set you use? How do you know whether to trust the data?

5. Do the data you use come with sufficient supporting information to allow you to make an informed judgement about which one(s) to choose? How much information do you need?

Notes: "Supporting information", perhaps tautologously, includes any information that helps a user make a decision. This might include quantitative auxiliary information (errors, biases) or provenance information (history, traceability). We should ask direct questions about these categories if the user doesn't volunteer the information.

If the answer is "no" then we can follow up with questions about how the users elicit quality/suitability for themselves. They might compare datasets with each other, or they might perform some statistical analysis.

Annex C Example Requirement

This is an example requirement extracted from an interview.

[GEOVIQUA-23] The quality model shall support traceability for the methodology used to determine data quality indicators (QA4EO Principle) **Created: 01/Jun/11 Updated: 18/Nov/11**	
Status:	Accepted
Project:	GeoViQua

Type:	Functional Requirement		
Reporter:	Joan Maso	**Assignee:**	Joanna Lumsden
Resolution:	Unresolved	**Votes:**	0
Labels:	fast-track, quality_model		

Issue Links:	**Derivation**			
	is derived from	GEOVIQUA-21	Elaborate procedures, guides and exam...	Accepted
	Relates			
	relates to	GEOVIQUA-151	The system has to be able to compare/...	Reviewed
	relates to	GEOVIQUA-61	Users want to know the build-up	Accepted

	histo... relates to ~~GEOVIQUA-42~~ Users want to know the Rejected methodology th...
Impact:	Many users
Related Initiatives:	GEOSS, QA4EO
Requirement:	A QI shall be based on a documented and quantifiable assessment of evidence demonstrating the level of traceability to internationally agreed (where possible SI) reference standards
Rationale:	Comparison between quality measures is needed
Acceptance Criteria:	QI is also acompanied by a way to know how the QI is calculated (reference to a document, methodology or standard)
Cluster:	Quality and Uncertainty

Description

Traceability information of the QI determination methods need to be available. It is important to know what methodology is used to determine the error estimates, and the peer-reviewed comments or publications for the methodology, so that I can understand better about how much uncertainty is contained in its error estimates.

Generic standardized procedures can be referenced adding the specifics of this particular measure.

Extracted from a set of interviews we carried out. See http://twiki.geoviqua.org/twiki/bin/view/GeoViQuaIntranet/WP2_Initial_User_Interviews_Results

Annex D Accessing the requirements

Every requirement has a unique identifier, such as GEOVIQUA-23. This requirement can be located in the web-based requirements management under

https://projects-ext.igd.fraunhofer.de/jira/browse/GEOVIQUA-23

Authentication is required and can be obtained from Fraunhofer or the project lead.